Economics Bulletin

Volume 36, Issue 4

Comparing marginal effects between different models and/or samples

Jörg Schwiebert
Leuphana University Lüneburg

Abstract

This paper proposes a simple approach to comparing marginal effects between different models and/or samples. A Generalized Method of Moments estimation framework is set up in which the equality of marginal effects between models/samples can be tested quite easily. Three relevant examples of potential application are provided. Stata code in the appendix shows that an implementation of the proposed approach in practice is quite simple.

Citation: Jörg Schwiebert, (2016) "Comparing marginal effects between different models and/or samples", *Economics Bulletin*, Volume 36, Issue 4, pages 2324-2332
Contact: Jörg Schwiebert - schwiebert@leuphana.de.
Submitted: February 15, 2016. **Published:** December 02, 2016.

1. Introduction

In this paper, I propose a simple approach to comparing marginal effects between different models and/or samples. The suggested method is useful when one wants to statistically compare marginal effects between

1. (non-)nested models, e.g., does the marginal effect change when another set of explanatory variables is added?
2. different classes of models, e.g., are the marginal effects generated from a probit model equal to those generated from a linear probability model?
3. different samples, e.g., are the marginal effects for women equal to those for men?

While a statistical comparison of regression coefficients has already been addressed in the literature (e.g., Clogg et al. 1995) and been implemented in statistical software such as Stata (StataCorp 2015, "suest" command), there has not been presented a unified approach for a statistical comparison of marginal effects. However, such a method is practically relevant, since empirical researchers using nonlinear models typically calculate and wish to compare marginal effects rather than coefficients; see, e.g., Greene (2012, p. 729) or Wooldridge (2010, p. 575, in the context of binary response models). In previous empirical practice, researchers usually report differences between the point estimates of the marginal effects, without taking into account the statistical significance of these differences. The following examples from the literature illustrate this issue:

- Altonji et al. (2005) analyze different methods to identify and estimate the effect of Catholic schooling on high school graduation (among other academic outcome variables) and compare the estimated marginal effects obtained from a univariate probit model, a bivariate probit model, an ordinary least squares (OLS) model and a two stage least squares (2SLS) model. However, only differences in point estimates are acknowledged without stating whether these differences are significant from a statistical point of view. For example, Altonji et al. (2005) state "the bivariate probit estimate of 0.170 (0.055) is also well above the univariate probit estimate of 0.094 (0.022)" (p. 796) or "[...], the probit estimate of the effect of CH_i on college attendance is 0.068 (0.016), which [is] reasonably close to the corresponding NELS:88 coefficient of 0.094" (p. 797).

- Chiswick et al. (2004) investigate the determinants of English language proficiency among immigrants and compare probit marginal effects between models based on different outcome variables. They also only compare the point estimates without making statements about the statistical significance of the differences. For example, they conclude that "[...], the impact of the birthplace concentration variable is much weaker in the study of English reading and writing skills than it is in the study of English-speaking skills" (pp. 121/124), without stating whether this is a statistically significant difference.

- Goerke and Pannenberg (2012) study the relationship between risk aversion and trade-union membership and compare probit marginal effects between subsamples of males and females and the pooled sample of males and females. They state that "relative to the pooled sample [...], the [average marginal effects] are slightly larger [in the subsample of males]" (p. 287), without stating whether this slight difference is significant from a statistical point of view.

- Using probit models, Fougere and Safi (2008) analyze the impact of naturalization on the probability of being employed and report that "on average, gaining French nationality increases the probability of being employed [...] by 2.7 points for men and 8.2 points for women" (p. 17). They conclude that "the 'naturalization premium' is much higher for women than for men" (p. 17), also without stating whether this difference is statistically significant.

This literature review is of course not exhaustive, but it clearly demonstrates that previous empirical practice merely relies on reporting and comparing differences in point estimates of marginal effects without making statements on the statistical significance of these differences. Therefore, there seems to be a need for a method to compare marginal effects between models and/or samples on a statistical basis. This paper provides a simple approach to statistically comparing marginal effects that might prove useful in empirical practice.

2. Econometric Framework

Suppose there are two different models or samples characterized by parameter vectors θ_1 and θ_2, respectively. Assume that these parameter vectors are the solution to the following moment conditions:

$$E[g_1(Y_1, X_1; \theta_1)] = 0 \tag{1}$$
$$E[g_2(Y_2, X_2; \theta_2)] = 0, \tag{2}$$

where Y_j is the dependent variable of model/sample j, X_j the associated vector of explanatory variables, and g_j a known function. The variables Y_j, X_j, and the functions $g_j, j = 1, 2$, are allowed to be identical across models/samples. Most linear and nonlinear regression models imply moment conditions as given by (1) and (2), hence assuming that these conditions are satisfied is not restrictive.

The marginal effects considered in this paper will be average marginal effects, as opposed to marginal effects at the average. In general, marginal effects are defined as the change in the conditional expected value of the dependent variable given a one-unit change in explanatory variables. Formally, they are defined as

$$m_j(X_j; \theta_j) \equiv \frac{\partial E[Y_j|X_j; \theta_j]}{\partial X_j}, \quad j = 1, 2. \tag{3}$$

Note that $m_j, j = 1, 2$, is a vector, where each element contains the marginal effect of a specific explanatory variable included in X_j. Also note that the marginal effects depend on the explanatory variables X_j. The average marginal effects for model/sample j are defined as

$$\mu_j \equiv E[m_j(X_j; \theta_j)] = E\left[\frac{\partial E[Y_j|X_j; \theta_j]}{\partial X_j}\right], \quad j = 1, 2, \tag{4}$$

where the outer expectation averages over the distribution of X_j. Note that $\mu_j, j = 1, 2$, is also a vector, containing the average marginal effect of each explanatory variable included in X_j.

The average marginal effects are the marginal effects averaged over the distribution of explanatory variables. Hence, they represent the average response of an individual to a one-unit change in explanatory variables. By contrast, the marginal effects at the average are defined as $\tilde{\mu}_j \equiv m_j(E[X_j];\theta_j)$, hence they are marginal effects evaluated at the mean of explanatory variables. A conceptual drawback of the marginal effects at the average is that it is unclear whether the mean of explanatory variables actually represents the "average individual" in the population. In other words, the marginal effects at the average might not have a sound interpretation; see Wooldridge (2010, p. 575). For this reason, this paper focuses on average marginal effects and not on the marginal effects at the average.

For simplicity, it is assumed here that the explanatory variables are continuous. However, the framework can be easily extended to discrete explanatory variables. In that case, the derivative in Eq. (4) has to be replaced by the discrete change:

$$\mu_j \equiv E[m_j(X_j;\theta_j)] \equiv E\left[E[Y_j|X_j \oplus 1;\theta_j] - E[Y_j|X_j;\theta_j]\right], \quad j = 1, 2, \tag{5}$$

where $X_j \oplus 1$ means that each element of the vector X_j is increased by one. A mixture of continuous and discrete explanatory variables is also possible. This might be the most relevant case in practice, since most empirical applications include both continuous and discrete explanatory variables.

The functions m_1 and m_2 are typically known to the researcher. Given that θ_1 and θ_2 were also known, the average marginal effects could be derived from the following moment conditions:

$$E[m_1(X_1;\theta_1) - \mu_1] = 0 \tag{6}$$
$$E[m_2(X_2;\theta_2) - \mu_2] = 0. \tag{7}$$

These conditions follow directly from the definition of average marginal effects in Eq. (4).

However, as θ_1 and θ_2 are not known in practice, a natural suggestion would be to replace them by their estimates $\hat{\theta}_1$ and $\hat{\theta}_2$ obtained from the empirical counterparts to Eqs. (1) and (2) and then to estimate Eqs. (6) and (7) conditional on $\hat{\theta}_1$ and $\hat{\theta}_2$. A disadvantage of this procedure is that the standard errors of estimated marginal effects would have to be corrected for the fact that θ_1 and θ_2 have been estimated. In principle, such a standard error correction is possible within a two-step estimation approach. However, a simpler way to proceed in practice is to perform a one-step approach by estimating Eqs. (1), (2), (6) and (7) jointly. This gives the following set of moment conditions:

$$E\begin{bmatrix} g_1(Y_1, X_1;\theta_1) \\ g_2(Y_2, X_2;\theta_2) \\ m_1(X_1;\theta_1) - \mu_1 \\ m_2(X_2;\theta_2) - \mu_2 \end{bmatrix} = 0 \tag{8}$$

This system of moment conditions can be estimated conveniently by applying the Generalized Method of Moments (GMM) estimation approach; see, e.g., Wooldridge (2010, ch. 14) for a description of this estimation approach. Since θ_1, θ_2, μ_1 and μ_2 are estimated jointly, the standard errors of the estimated marginal effects $\hat{\mu}_1$ and $\hat{\mu}_2$ will be correct immediately.

In order to statistically compare the marginal effects from models/samples 1 and 2, a simple Wald test can be carried out. The null hypothesis is $H_0 : \mu_{1k} = \mu_{2k}$, where the k indicates the marginal effect of a specific explanatory variable X_{jk}. It is also possible to test joint hypotheses, for instance $H_0 : \mu_{1k} = \mu_{2k}\ \forall k$. For the validity of the Wald test it is important that the researcher specifies the correct standard error computation procedure. For example, heteroskedasticity is often an issue in cross-sectional data sets, hence heteroskedasticity-robust standard errors should be computed.

Statistical software like Stata allows the user to perform GMM estimation with user-specified moment conditions, and also to test hypotheses within this estimation framework. Appendix 1 of this paper contains Stata code for the examples discussed in the next section. The code shows that implementing the proposed approach in empirical practice is quite simple.

3. Examples of Application

3.1 (Non-)Nested Models

Suppose that a probit model shall be estimated and that the marginal effect of a specific variable is the object of interest. The researcher wants to compare the marginal effect from a base model with the marginal effect from the same model augmented with additional explanatory variables. Put differently, the researcher wants to test whether the marginal effect remains unaffected when additional (control) variables are added.

The moment conditions (1) and (2) can be written as

$$E[(Y - \Phi(X_1'\theta_1))X_1] = 0 \quad (9)$$

$$E[(Y - \Phi(X_2'\theta_2))X_2] = 0, \quad (10)$$

where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function. Here, X_1 is a subset of X_2, X_2 includes the additional control variables.

The moment conditions say that the prediction error of the probit model, $Y - \Phi(X_j'\theta_j)$, $j = 1, 2$, should be orthogonal to the explanatory variables (instruments). Alternative moment conditions could also be exploited; for example, one could use the first order conditions from maximum likelihood estimation as moment conditions, which might result in efficiency gains.

The moment conditions for the marginal effects are given by

$$E[\phi(X_1'\theta_1)\theta_1 - \mu_1] = 0 \quad (11)$$

$$E[\phi(X_2'\theta_2)\theta_2 - \mu_2] = 0, \quad (12)$$

where $\phi(\cdot)$ denotes the standard normal probability density function. To test whether the marginal effect of a specific variable remains constant when additional control variables are added, a Wald test can be carried out, as described in the last section. The procedure can also be used for non-nested models, i.e., when X_1 is not a subset of X_2.

3.2 Different Classes of Models

Suppose that the marginal effects from a probit model shall be compared to the marginal effects from an analogous linear probability model. As Wooldridge (2010, p. 579) points out, the linear probability model will consistently estimate the average marginal effects when the explanatory variables are jointly normally distributed. Under joint normality, one would thus not expect the marginal effects generated from the linear probability model to be different from the marginal effects of a correctly specified probit model. However, when the explanatory variables are not jointly normally distributed and/or the probit specification is incorrect, average marginal effects from both models might not be equal. To test statistically whether the marginal effects generated from the two different models are equal, the researcher can use the approach presented in this paper.

The moment conditions (1) and (2) can be written as

$$E[(Y - \Phi(X'\theta_1))X] = 0 \tag{13}$$
$$E[(Y - X'\theta_2)X] = 0, \tag{14}$$

assuming that the first model is the probit model and the second model is the linear probability model. It is also assumed that each model includes the same dependent variable and the same set of explanatory variables, hence the index on Y and X has been omitted. The only difference between the models is thus that the first model is a probit model and the second a linear probability model.

The moment conditions for the marginal effects are given by

$$E[\phi(X'\theta_1)\theta_1 - \mu_1] = 0 \tag{15}$$
$$E[\theta_2 - \mu_2] = 0, \tag{16}$$

where the second condition trivially follows from the fact that the marginal effects in the linear probability model are identical to the coefficients. To test whether the marginal effects from the probit model are equal to the marginal effect from the linear probability model, a Wald test can be carried out, as described above.

If the null hypothesis of equality of marginal effects cannot be rejected, this might indicate that the marginal effects are robust against the specification of the underlying binary response model, at least among the models under consideration.[1] However, the question might arise which model we should trust when the null hypothesis is rejected, i.e., when the marginal effects are statistically different. A way to proceed in practice is to find the "best" model among the models under consideration, for example by applying the Vuong (1989) test for non-nested models. Given such a "best" model, one may then perform model specification tests to test the validity of the current model specification. Such specification tests may address issues like heteroskedasticity, omitted variables (including nonlinear terms) and functional form (see, e.g., Wooldridge 2010 or Greene 2012). If these tests indicate that the current model specification is valid, then the marginal effects based on this model should be reported in empirical work.

3.3 Different Samples

Suppose that two probit models with the same dependent and explanatory variables are estimated, but one for a sample of males and the other for a sample of females. The researcher wants to test whether the marginal effects for females are identical to the marginal effects for males. Let D denote a dummy variable equal to one if an individual is male. Then, the parameters characterizing the male and female sample can be derived from the following moment conditions, which are analogous to Eqs. (1) and (2):

$$E[D(Y - \Phi(X'\theta_1))X] = 0 \tag{17}$$
$$E[(1 - D)(Y - \Phi(X'\theta_2))X] = 0. \tag{18}$$

[1] Formally, failing to reject the null hypothesis does not imply that one can accept the null hypothesis.

The moment conditions for the marginal effects are given by

$$E[D(\phi(X'\theta_1)\theta_1 - \mu_1)] = 0 \tag{19}$$

$$E[(1-D)(\phi(X'\theta_2)\theta_2 - \mu_2)] = 0. \tag{20}$$

Again, a Wald test can be carried out to compare the marginal effects between males and females.

4. Empirical Example

In this section, an illustrative example is provided to show the usefulness of the proposed approach in empirical practice. This example uses data from Mroz (1987) on the labor force participation of married women. The data set encompasses a sample of 753 married women and can be obtained from the data archive associated with Wooldridge's (2010) textbook "Econometric Analysis of Cross Section and Panel Data". The dependent variable of interest (*inlf*) is a binary variable equal to one if a woman belongs to the labor force and zero otherwise. Concerning the choice of explanatory variables I follow Wooldridge (2010, p. 580) and select the following set of explanatory variables: years of education (*educ*), age (*age*), the number of children aged 0-5 (*kidslt6*), the number of children aged 6-18 (*kidsge6*), labor market experience (*exper*), labor market experience squared (*expersq*) and non-wife income (*nwifeinc*).

Suppose we are interested in the impact of education on the probability of labor force participation of married women. Our goal is therefore to estimate the marginal effect of education on the probability of labor force participation. Given the explanatory variables listed above, we estimate three different binary response models: the probit model, the linear probability model (LPM) and the logit model. Estimating each model separately and using Stata's *margins* command to calculate marginal effects, we obtain the following estimates of the quantity of interest:[2]

	Probit	LPM	Logit
Estimated marginal effect of education	0.0394	0.0380	0.0395
Standard error	0.0074	0.0073	0.0075

By visual inspection, the estimates are very similar. The estimates from the probit and logit models are almost identical, while the estimate from the linear probability model is a bit smaller. Given these small discrepancies, we might wonder if the marginal effects generated from the three models are different from a statistical point of view.

In this situation, the approach proposed in this paper proves useful. In Sec. 3.2 it was shown how to compare marginal effects between a probit model and a linear probability model. It is straightforward to extend this approach to include a third model, which is

[2]The parameter estimates generated from the three models are reported in Wooldridge (2010), p. 580. In contrast to Wooldridge (2010), I computed robust (sandwich-type) standard errors to account for possible misspecification of the underlying binary response models. The standard errors of the marginal effects are based on these robust standard errors.

the logit model in this example. The moment conditions are:

$$E \begin{bmatrix} (Y - \Phi(X'\theta_1))X \\ (Y - X'\theta_2)X \\ (Y - \Lambda(X'\theta_3))X \\ \phi(X'\theta_1)\theta_1 - \mu_1 \\ \theta_2 - \mu_2 \\ \lambda(X'\theta_3)\theta_3 - \mu_3 \end{bmatrix} = 0, \tag{21}$$

where $\Lambda(\cdot)$ and $\lambda(\cdot)$ denote the standard logistic cumulative distribution and probability density functions, respectively. I applied the GMM command available in Stata 14 to obtain estimates of the parameters $(\theta_1', \theta_2', \theta_3')'$ and marginal effects $(\mu_1', \mu_2', \mu_3')'$. I used a two-step GMM approach with an initial weighting matrix given by the identity matrix in the first step and a "robust" weighting matrix in the second step. The second-step weighting matrix accounts for the potential correlation of the moment conditions as well as for potential heteroskedasticity. The exact specification of the second-step weighting matrix is given in Appendix 2 of this paper.

After estimation, I tested the null hypothesis that the marginal effects of education are identical across the models under consideration. The corresponding Wald test yields a p-value of 0.2358, which clearly exceeds the conventionally used significance levels. Hence, from a statistical point of view the marginal effects of education are not different across the models under consideration.

References

Altonji, J.G., T.E. Elder and C.R. Taber (2005) "An Evaluation of Instrumental Variable Strategies for Estimating the Effects of Catholic Schooling" *Journal of Human Resources* **40**, 791-821.

Chiswick, B.R., Y.L. Lee and P.W. Miller (2004) "Immigrants' Language Skills: the Immigrant Experience in a Longitudinal Survey" *Annals of Economics and Statistics* **71-72**, 97-139.

Clogg, C.C., E. Petkova and A. Haritou (1995) "Statistical Methods for Comparing Regression Coefficients Between Models" *American Journal of Sociology* **100**, 1261-1293.

Fougere, D. and M. Safi (2008) "The Effects of Naturalization on Immigrants' Employment Probability (France, 1968-1999)" IZA discussion paper No. 3372.

Goerke, L. and M. Pannenberg (2012) "Risk Aversion and Trade Union Membership" *Scandinavian Journal of Economics* **114**, 275-295.

Greene, W.H. (2012) *Econometric Analysis*, seventh edition, international edition, Pearson: Boston, Mass.

Mroz, T.A. (1987) "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions" *Econometrica* **55**, 765-799.

StataCorp (2015) *Stata 14 Base Reference Manual*, Stata Press: College Station, TX.

Vuong, Q.H. (1989) "Likelihood Ratio Tests for Model Selection and non-nested Hypotheses" *Econometrica* **57**, 307-333.

Wooldridge, J.M. (2010) *Econometric Analysis of Cross Section and Panel Data*, second edition, MIT Press: Cambridge, Mass.

Appendix 1: Stata Code for Examples of Application

(Non-)Nested Models

Let y denote the dependent variable and x1 and x2 the explanatory variables. To test if the marginal effect of x1 in a probit model using x1 as the only explanatory variable is equal to the marginal effect of x1 in a probit model using x1 and x2 as explanatory variables, the following Stata code must be executed:

```
gmm (y-normal({theta10}+{theta11}*x1)) (y-normal({theta20}+{
    theta21}*x1+{theta22}*x2)) (normalden({theta10}+{theta11}*x1)
    *{theta11}-{mu1}) (normalden({theta20}+{theta21}*x1+{theta22
    }*x2)*{theta21}-{mu2}), instruments(1:x1) instruments(2:x1 x2
    ) instruments(3:) instruments(4:) winitial(identity)
test _b[/mu1]=_b[/mu2]
```

Note: Each model includes a constant term.

Different Classes of Models

Let y denote the dependent variable and x1 and x2 the explanatory variables. To test if the marginal effect of x1 in a probit model using x1 and x2 as the explanatory variables is equal to the marginal effect of x1 in an analogous linear probability model, the following Stata code must be executed:

```
gmm (y-normal({theta10}+{theta11}*x1+{theta12}*x2)) (y-({theta20
    }+{theta21}*x1+{theta22}*x2)) (normalden({theta10}+{theta11}*
    x1+{theta12}*x2)*{theta11}-{mu1}) ({theta21}-{mu2}),
    instruments(1:x1 x2) instruments(2:x1 x2) instruments(3:)
    instruments(4:) winitial(identity)
test _b[/mu1]=_b[/mu2]
```

Note: Each model includes a constant term.

Different Samples

Let y denote the dependent variable and x1 and x2 the explanatory variables. Moreover, let d be a sample indicator being equal to one if the observation belongs to the first sample (e.g., males) and zero if it belongs to the second sample (e.g., females). To test if the marginal effect of x1 in a probit model using x1 and x2 as the explanatory variables is the same in both samples, the following Stata code must be executed:

```
gmm (d*(y-normal({theta10}+{theta11}*x1+{theta12}*x2))) ((1-d)*(
    y-normal({theta20}+{theta21}*x1+{theta22}*x2))) (d*(normalden
    ({theta10}+{theta11}*x1+{theta12}*x2)*{theta11}-{mu1})) ((1-d
    )*(normalden({theta20}+{theta21}*x1+{theta22}*x2)*{theta21}-{
```

```
    mu2}))), instruments(1:x1 x2) instruments(2:x1 x2) instruments
    (3:) instruments(4:) winitial(identity)
test _b[/mu1]=_b[/mu2]
```

Note: Each model includes a constant term.

Appendix 2: Specification of the Second-step Weighting Matrix in the Empirical Application

Let $\{(x_i, y_i)\}_{i=1}^n$ denote a random sample of realizations of the random variables (X, Y). Furthermore, define

$$Z_i \equiv \begin{pmatrix} x_i' & 0 & 0 & 0 & 0 & 0 \\ 0 & x_i' & 0 & 0 & 0 & 0 \\ 0 & 0 & x_i' & 0 & 0 & 0 \\ 0 & 0 & 0 & I_K & 0 & 0 \\ 0 & 0 & 0 & 0 & I_K & 0 \\ 0 & 0 & 0 & 0 & 0 & I_K \end{pmatrix} \quad \text{and} \quad u_i \equiv \begin{pmatrix} y_i - \Phi(x_i'\theta_1) \\ y_i - x_i'\theta_2 \\ y_i - \Lambda(x_i'\theta_3) \\ \phi(x_i'\theta_1)\theta_1 - \mu_1 \\ \theta_2 - \mu_2 \\ \lambda(x_i'\theta_3)\theta_3 - \mu_3 \end{pmatrix},$$

where I_K denotes an identity matrix of dimension K, with K being the dimension of μ_1, μ_2 and μ_3.

Then, the empirical counterpart to Eq. (21) can be written as $n^{-1}\sum_{i=1}^n Z_i' u_i$, and the GMM estimator minimizes

$$Q(\theta_1, \theta_2, \theta_3, \mu_1, \mu_2, \mu_3) \equiv \left(n^{-1}\sum_{i=1}^n Z_i' u_i\right)' W \left(n^{-1}\sum_{i=1}^n Z_i' u_i\right),$$

where W denotes the weighting matrix. As mentioned above, the initial (first-step) weighting matrix is the identity matrix. The second-step weighting matrix is given by $W = S^{-1}$, with

$$S \equiv n^{-1}\sum_{i=1}^n Z_i' \hat{u}_i \hat{u}_i' Z_i,$$

where $\hat{u}_i$ denotes u_i evaluated at the estimated parameters from the first step; also see StataCorp (2015, "gmm" command).