■ Nonlinear regression models

20185293 Joo, Jiwon

· Sometimes, we have to consider non-linear regression models when a regression model is intrinsically non-linear in their parameters.

cf. 2 types of regression models

① Linear regression model

- Includes models which appear to be nonlinear, but it can be linearized after transformations (e.g., taking logarithms).

② (Intrinsically) non-linear regression model

-What is the intrinsically non-linear model?

: If we cannot transform a non-linear model into linearlized model, the model is called 'intrinsically' non-linear.

For example, the Cobb-Douglas production function that relates output (Y) to labor (L) and capital (K) can be written as $Y=\alpha L^{\beta}K^{\gamma}$.

Taking logarithms yields $\log(Y)=\delta+\beta\log(L)+\gamma\log(K)$, where $\delta=\log(\alpha)$.

This function is non-linear in the variables Y, L, and K, but it is linear in the parameters δ, β, and γ. Models of this kind can be estimated using the least squares method.

On the other hand, there are other types of non-linear models which are non-linear in the parameters and which cannot be made linear in the parameters after a transformation.

For instance, the CES Production Function is one of well-known intrinsically non-linear model.

$Y_i = A[\delta K_i^{-\beta}+(1-\delta)L_i^{-\beta}]^{-1/\beta}$, where $\boldsymbol{\theta}=\{A,\delta,\beta\}$.

Aside from having an interest in well-known non-linear models such as above, we shall consider non-linear models when estimation results with linear regression shows fatal problems. For example, it seems there are some violations of the classical assumptions such as heteroskedasticity problem.

Especially, when the response variable has binary rather than continuous in nature, we have to consider Logit or Probit model.
Logistic regression is useful for situations in which you want to be able to predict the presence or absence of a characteristic or outcome based on values of a set of predictor variables.
From now on, I'd like to focus on these two regression models.

● **Logit/Probit model**

· Two of the most common methods to solve for Binary Classification.
· Frequently used as non-linear probability models.

-When do we consider Logit/Probit (non-linear) model rather than linear regression?

1) If the type of dependent variable is not a continuous (but a nominal), we cannot use the existing linear regression model.
Note that a nominal variable has no numerical meaning (e.g., 1 (male), 0 (female)).

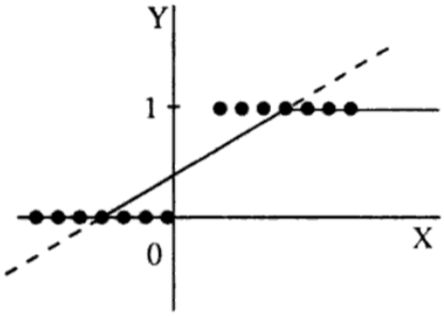
2) There is no linear relationship between a dependent variable and independent variables.

For example, consider a simple linear regression model.

$Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i$
, where $Y_i = 1$ (if the ith household own a car), 0 (o.w.) and $X_i =$ the ith household's income).

This model assumes that the probability of Y increases linearly with X (Unrealistic).

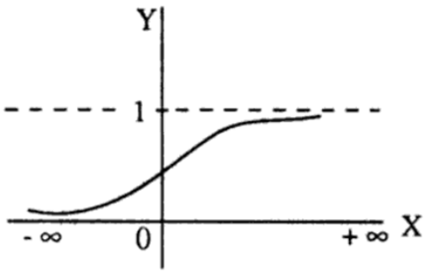
Below scatter plot shows that the linear model cannot fit such data well, resulting in low R^2 value.



To overcome overall problems, we have to seek an alternative (non-linear) model which satisfies some characteristics as shown below.

① As the regressor increases, the probability increases, but never gets outside the interval (0,1).
② The relationship between X_i and Y_i is non-linear.
③ Increases in X are associated with increases in Y for all values of X.
④ As $X_i \rightarrow -\infty, Y \rightarrow 0$ at slower and slower rates.
⑤ As $X_i \rightarrow \infty, Y \rightarrow 1$ at slower and slower rates.

Above five properties can be satisfied when we utilize the cumulative distribution function (CDF).



Non-linear S-shaped Function

Generally, we choose the logistic (Logit model) or normal (Probit model) distribution as the CDF.

3) Use when we are interested in binary outcome as a dependent variable (e.g., Yes/No, Success/Failure, Bankrupt/Non-bankrupt) and when we want to predict discrete outcomes based on related predictors.

cf. Of course, response variables with three or more categorical outcome can be analyzed using Multinomial Logit/Probit model.

Overall, below **[Table 1]** summarizes main differences between linear regression and logistic regression.

[Table 1]

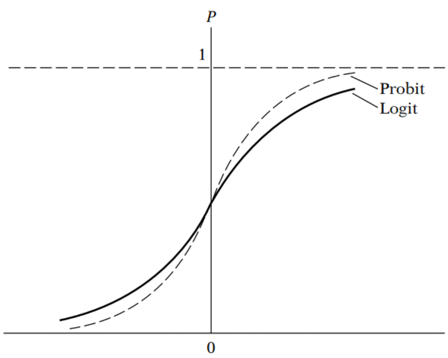
Linear regression	Logistic regression
A linear approach that models the relationship between a dependent variable and one or more independent variables.	A statistical model that predicts the probability of an outcome that can only have two values.
Used to solve regression problems.	Used to solve classification problems (binary classification).
Estimates the dependent variable when there is a change in the independent variable.	Calculates the possibility of an event occurring.
Output value is continuous.	Output value is discrete.
Uses a straight line.	Uses an S-curve or sigmold function.

◆ **Choice between Logit and Probit**

· Logistic distribution has fat tails than Probit.

· The logit model performs well in heterogeneous data, moderately balanced data as well as data with outliers.

· Logit and Probit estimates are approximately related by the following rule: $\widehat{\beta_{\text{logit}}} = 1.6\widehat{\beta_{probit}}$.



· When choosing between Logit and Probit in the dichotomous case, there is no basis in statistical theory for preferring one over the other. In most applications, it makes no difference which one uses. If we have a small sample, the two distributions can differ significantly in their results, however, they are quite similar in large samples.

· Probit models are popular in social sciences such as economics.

· While the outcomes tend to be similar, the underlying distributions are different.

· In order to use MLE (Maximum Likelihood Estimation: Find parameter values that maximize the probability of the observed data.) method instead of LSE, we need to make some assumption about the distribution of the errors. The difference between Logistic and Probit models lies in this assumption about the distribution of the errors.

-The OLS model (by vector notation)

: $\boldsymbol{y}=\boldsymbol{x}'\boldsymbol{\beta}+\boldsymbol{\varepsilon}$, $y_i \in (-\infty, \infty)$

-Binary outcome models

: Estimate the probability that Y=1 as a function of the regressors.

$p = P(y=1|\boldsymbol{x}) = F(\boldsymbol{x}'\boldsymbol{\beta})$

By assuming a specific distribution function of $F(\boldsymbol{x}'\boldsymbol{\beta})$, we can call the name of a binary outcome model. The popular two models are based on logistic and normal distribution.

◆ **Logit model**

-For the logit model, $F(\boldsymbol{x}'\boldsymbol{\beta})$ is the cdf of the logistic distribution.

$$F(\boldsymbol{x}'\boldsymbol{\beta}) = \Lambda(\boldsymbol{x}'\boldsymbol{\beta}) = \frac{\exp(\boldsymbol{x}'\boldsymbol{\beta})}{1+\exp(\boldsymbol{x}'\boldsymbol{\beta})} = \boldsymbol{p}$$

The predicted probabilities ($\hat{\boldsymbol{p}}$) are limited between 0 and 1.

Then, do logit transformation (i.e., $g(\Lambda)$).

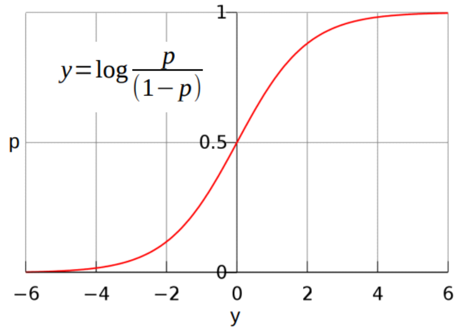
$$g(\Lambda) = \log\left[\frac{\Lambda}{1-\Lambda}\right] = \boldsymbol{x}'\boldsymbol{\beta}$$

· **Odd ratio**: $\log\left[\frac{\boldsymbol{p}}{1-\boldsymbol{p}}\right]$

(**Odds** = Prob. of presence of characteristic (i.e., success)/ Prob. of absence of characteristic (i.e., failure))

· **Logit model**: $Y(=\log[\frac{p}{1-p}])=x'\beta+\varepsilon$, where $Y_i\in(-\infty,\infty)$.

Therefore, the odd ratio is linear in X_i (and is also linear in β_i in terms of estimation).



Next, to estimate parameters, we have to know the values of odd ratios (i.e., $\log[\frac{p_i}{1-p_i}]$).

Based on the raw data, we can get estimated values of p_i by calculating **relative frequency**.

$\hat{p}_i=\frac{n_i}{N_i}$

(e.g., Corresponding to each income level X_i, there are N_i families, n_i among whom are car owners ($n_i \leq N_i$).)

So, we get $\hat{Y}(=\log[\frac{\hat{p}}{1-\hat{p}}])$ and use as data of the dependent variable when estimating the logit model.

In addition, to eliminate the problem of heteroscedasticity, multiply by the reciprocal of standard deviation of $\hat{\varepsilon}_i$ (= $\sqrt{N_i\hat{P}_i(1-\hat{P}_i)}$) both sides of the logit model ($\hat{\varepsilon}_i\sim N[0,\frac{1}{N_i\hat{P}_i(1-\hat{P}_i)}]$).

Finally, estimate parameters by OLS method.

· **Note**

1) We have to estimate the logit model without intercept because of above multiplication.

2) Statistical inferences are vaild only in the case of large sample size.

-Interpreting parameter estimates in Logit model

: Interpreting the estimated values literally has little meaning.

First, we can do **the odds interpretation**.

By taking the antilog of the estimated logit model, we can obtain:

$$[\frac{p}{1-p}] = \exp(\boldsymbol{x}'\hat{\boldsymbol{\beta}})$$

For example, the estimated model: $\widetilde{Y} = -1.59w_i + 0.08\widetilde{X}_i$

, where $w_i = \sqrt{N_i\hat{P}_i(1-\hat{P}_i)}$ and $\widetilde{X}_i = X_i w_i$, $\widetilde{Y}_i = \log[\frac{\hat{p}}{1-\hat{p}}]w_i$.

$$\log[\frac{\hat{p}}{1-\hat{p}}] = \frac{1}{w_i}(-1.59w_i + 0.08\widetilde{X}_i)$$

Then, taking antilog leads to $[\frac{\hat{P}_i}{1-\hat{P}_i}] = \exp(-1.59 + 0.08X_i)$

$= \exp(-1.59)\exp(0.08X_i)$

For the related term (i.e., $\exp(0.08X_i)$), let $X_i = 1$.

Then $\exp(0.08 \times 1) = 1.083$, which means that for a unit increase in X_i (income), the odds in favor of $Y_i = 1$ (owning a car) increases by 1.083 or about 8.3%.

In general, if you take the antilog of the jth slope coefficient (in case there is more than one regressor in the model), subtract 1 from it, and multiply the result by 100, you will get the percent change in the odds for a unit increase in the jth regressor.

Overall, types of interpretations can be categorized into 3 cases:
① if $0 < \exp(\hat{\beta}_j) < 1$, then Y and X_j are negatively correlated.
(i.e., The event is less likely to happen when X_j increases.)
② if $1 < \exp(\hat{\beta}_j) < \infty$, then Y and X_j are positively correlated.
(i.e., The event is more likely to happen when X_j increases.)
③ if $\exp(\hat{\beta}_j) = 1$, then the event is exactly as likely to occur regardless of the level of X_j.

Second, we can compute **probabilities** with respect to the binary outcome.

That is, compute the probability of owning a car at a certain level of income.

Suppose we want to compute the probability at X=20 (i.e., $20,000).
Plugging this value in the estimated logit model, then we obtain $\hat{Y}$.

To be specific,
$$\log[\frac{\hat{p}}{1-\hat{p}}] = \frac{1}{w_i}(-1.59w_i + 0.08\widetilde{X}_i) = -1.59 + 0.08X_i = -1.59 + 0.08(20) = 0.01$$
In other words, at the income level of $20,000, we have $0.01 = \log[\frac{\hat{p}}{1-\hat{p}}]$.
Then, solving the equation regarding $\hat{p}$ results in:
$$\hat{p} = (\frac{1.01}{1+1.01}) = 0.5025.$$

That is, given the income of $20,000, the probability of a family owing a car (i.e., $p = P(y=1|\boldsymbol{x})$) is about 50.25%.

Third, compute **the rate of change of probability**.

It is clear that the probability of owning a car depends on the income level. The rate of change of probability depends not only on the estimated slope coefficient ($\hat{\beta}_j$), but also on the level of the probability from which the change is measured.

For example, suppose we want to measure the change in the probability of

owning a car at the income level of $20,000. Then, the change in probability for a unit increase in income from the level 20 (thousand) is:
$\hat{\beta}_j(1-\hat{P})\hat{P} = 0.08(1-0.5025)0.5025 = 0.0199995 \approx 0.02$.

◆ **Probit model** (suggested by McFadden)
· Sometimes, the normal CDF would be more desirable than the logistic CDF.

-For the probit model, $F(\boldsymbol{x}'\boldsymbol{\beta})$ is the cdf of the standard normal distribution:
$F(\boldsymbol{x}'\boldsymbol{\beta}) = \Phi(\boldsymbol{x}'\boldsymbol{\beta}) = \int_{-\infty}^{\boldsymbol{x}'\boldsymbol{\beta}} \phi(z)dz$

The predicted probabilities are limited between 0 and 1.

To induce the probit model (for the previous example), assume that whether the decision of the ith family to own a car or not depends on **an unobservable utility index** I_i, a latent variable, that is determined by one or more regressors, say income X_i, in such a way that the larger the value of the index I_i, the greater the probability of a family owning a car.

We express the index I_i as $I_i = \beta_1 + \beta_2 X_i$ ⋯ **(1)**
, where X_i is the income of the ith family.

Assume that there is **a critical or threshold level of the index** (I_i^*). If $I_i \geq I_i^*$, then the family (i) will own a car, otherwise it will not. The threshold is not observable, but if we assume that it is normally distributed with the same mean and variance, it is possible not only to estimate the parameters of the index given in $I_i = \beta_1 + \beta_2 X_i$, but also to get some information about the unobservable index itself.

Given **normality assumption**, the probability that $I_i \geq I_i^*$ can be computed from the standardized normal CDF as:
$P_i = P(Y=1|X) = P(I_i^* \leq I_i) = P(Z_i \leq \beta_1 + \beta_2 X_i) = F(\beta_1 + \beta_2 X_i)$ ⋯ **(2)**
, where $P(Y=1|X)$ means the probability that an event occurs given the values of the X, and Z_i is the standard normal variable (i.e., $Z_i \sim N(0,\sigma^2)$).

F is the standard normal CDF.

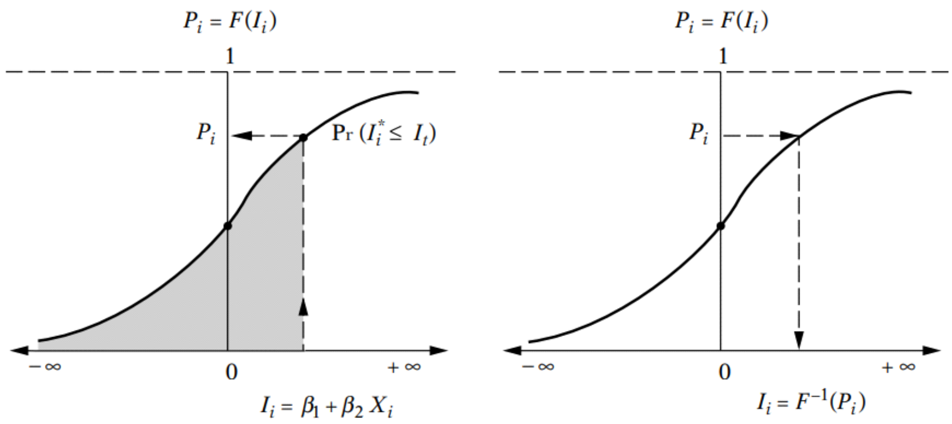
$F(I_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{I_i(=\beta_1+\beta_2 X_i)} \exp(-z^2/2)dz$ ⋯ **(3)**

Since P indicates the probability that an event will occur, here the probability of owning a car, it is measured by the area of the standard normal curve from $-\infty$ to I_i.

Now to obtain information on I_i (the utility index), as well as on β_1 and β_2, we take the inverse of **equation (2)** to obtain:

$I_i = F^{-1}(P_i) = \beta_1 + \beta_2 X_i$ ⋯ **(4)**

, where F^{-1} is the inverse of the normal CDF.



· For the left graph, the shaded area is equal to $P(I_i \geq I_i^*)$ (i.e., the cumulative probability of owning a car in case of $I_i \geq I_i^*$).

· For the right graph, it just shows the reverse flow of the left graph.

Next, to estimate parameters, we can estimate I_i with the information of relative frequency.

Calculate $\hat{p_i} = \frac{n_i}{N_i}$ ($\hat{p_i}$: the vertical axis of the standard normal CDF).

Then, estimate the index I_i from the standard normal CDF:

$I_i = F^{-1}(\hat{p_i})$ (I_i: the horizontal axis of the standard normal CDF).

(5 used to be added to avoid negative values in hand calculation.)

-Interpreting parameter estimates in Probit model
: Suppose we want to know the effect of a unit change in X on the probability that Y=1 (i.e., a family buy a car.).

For **equation (2)**, take the derivative with respect to X (i.e., the (instantaneous) rate of change of the probability (= marginal effect) with respect to income.).

$$\frac{dP_i}{dX_i} = f(\beta_1 + \beta_2 X_i)\beta_2 \quad \cdots \quad \textbf{(5)}$$

, where $f(\beta_1 + \beta_2 X_i)$ is the standard normal p.d.f. evaluated at $(\beta_1 + \beta_2 X_i)$.

So, let $X = 6$ (thousand dollars).
Find the normal density function at $f[-1.0166 + 0.04846(6)] = f(-0.72548)$.

Related Z-tables show that for $Z = -0.72548$, the normal density is about 0.3066. Then, multiply 0.3066 by 0.04846 (the estimated slope coefficient).

It gives 0.01485, which means that starting with an income level of $6,000, if the income goes up by $1,000 (one unit), then the probability of a family purchasing a car goes up by about 1.4%.

Besides, we can predict the odds of being a case based on the certain values of the predictors.

For example, Ohlson (1980) examines some empirical results of a study predicting corporate failure as evidenced by the event of bankruptcy. He uses logit model and classify the company if and only if $P(X_i, \hat{\beta}) > 0.5$ (a cutoff point, which should be determined at the discretion of the researcher.).

To be specific, he selects four factors (or predictors) (i.e., the size of the company, a measure of the financial structure, a measure of performance, a measure of current liquidity). Using the logit model, he shows that the statistic "Percent Correctly Predicted" equals 96.12%.

◆ **Practice example**

-Examine the factors influencing the purchase of health insurance by estimating Logit and Probit models.

· **Y**: '**INS**'

(1 if a person has health insurance (39%: Actual), 0 o.w. (61%: Actual)).

· **X**: '**RETIRE**' (1=Yes, 0=No), '**AGE**', '**HSTATUSG**' (1 if a person has a good health status, 0 o.w.), '**HHINCOME**' (household income, ten thousand USD/unit), '**EDUCYEAR**' (education years), '**MARRIED**' (1 if a person got married, 0 o.w.), '**HISP**' (1 if a person is a hispanic, 0 o.w.)

[Raw dataset (with variables of interest)]

	ins	retire	age	hstatusg	hhincome	educyear	married	hisp
1	0	0	62	0	0	12	0	0
2	0	0	59	0	0	12	0	0
3	0	1	60	1	0	13	0	0
4	0	0	62	0	0	10	0	0
5	0	0	54	0	0	9	0	0
6	0	1	62	1	0	12	1	0
7	0	0	59	0	0	5	1	0
8	0	0	59	0	0	11	0	0
9	0	0	65	0	0	14	0	0
10	0	0	58	0	0.101	12	0	0
11	0	0	60	0	0.12	11	0	0
12	1	1	61	1	0.124	12	0	0
13	0	0	64	0	0.2	10	1	0
14	0	0	59	0	0.2	10	1	0
15	0	0	58	0	0.24	8	0	0
16	0	0	62	0	0.396	14	0	0
17	1	0	65	1	0.4	14	0	0
18	1	0	63	1	0.5	16	0	0
19	0	0	67	0	0.528	8	0	1
20	1	0	69	1	0.6	9	0	0

[Table 2]

Model coefficients	Multiple linear regression	Logit	Probit
Constant	0.12	-1.71*	-1.06*
Retired	0.04*	0.19*	0.11*
Age	-0.002	-0.01	-0.008*
Good health status	0.06*	0.31*	0.19*
Household income	0.0004*	0.002*	0.001*
Education years	0.02*	0.11*	0.07*
Married	0.12*	0.57*	0.36*

Hispanic	-0.12*	-0.81*	-0.46*

* Indicates significance at the 5% level.

-Coefficient interpretation based on the above **[Table 2]**

1) We can only interpret significant estimates in terms of the sign.

Retired individuals (in comparison to non-retired individuals), individuals with good health status, higher household income, higher education, married are more likely to have health insurance, and Hispanic are less likely to have health insurance.

2) Each model's coefficients differ by a scale factor.

So, we cannot compare the magnitudes among different models.

[Table 3]

Have health insurance	Regression marginal effects	Logit marginal effects at the mean	Probit marginal effects at the mean
Retired	0.04*	0.04*	0.04*
Age	-0.002	-0.003	-0.003
Good health status	0.06*	0.07*	0.07*
Household income	0.0004*	0.0005*	0.0004*
Education years	0.02*	0.02*	0.02*
Married	0.12*	0.12*	0.13*
Hispanic	-0.12*	-0.16*	-0.16*

-Marginal effects based on the above **[Table 3]**

1) The marginal effect is a measure of the instantaneous effect that a change in a particular regressor has on the predicted probability of Y_i, when the other covariates are kept fixed. The effects are nonlinear functions of the parameter estimates and levels of the regressors.

Marginal effects for distributions such as logit and probit can be computed with PROC QLIM by using MARGINAL option. The output dataset (“mfx”) after executing PROC QLIM with MARGINAL option evaluates marginal effects for each observation. Among related added columns, we only take ‘**Meff_P2_**(covatiate)’ columns since it indicates the marginal effect of a specific regressor on the probability of (the dependent variable=1).

[Probit model (“mfx” dataset)]

	agewhi	Marginal effect of retire on the probability of ins=1	Marginal effect of retire on the probability of ins=2	Marginal effect of age on the probability of ins=1	Marginal effect of age on the probability of ins=2	Marginal effect of hstatusg on the probability of ins=1	Marginal effect of hstatusg on the probability of ins=2
1	0	-0,035096847	0,0350968471	0,0026301511	-0,002630151	-0,058635472	0,0586354722
2	59	-0,035810929	0,0358109288	0,0026836642	-0,002683664	-0,059828472	0,0598284717
3	0	-0,044163632	0,0441636317	0,0033096142	-0,003309614	-0,073783135	0,0737831348
4	62	-0,031159278	0,0311592778	0,0023350704	-0,00233507	-0,052057068	0,0520570683
5	54	-0,031165222	0,0311652225	0,0023355159	-0,002335516	-0,052067	0,0520669999
6	62	-0,047018753	0,0470187534	0,0035235764	-0,003523576	-0,078553119	0,078553119
7	0	-0,032155571	0,032155571	0,0024097324	-0,002409732	-0,053721552	0,0537215518
8	59	-0,033890632	0,033890632	0,0025397575	-0,002539758	-0,056620277	0,0566202771
9	0	-0,038092018	0,0380920183	0,0028546086	-0,002854609	-0,063639434	0,0636394339
10	58	-0,036049797	0,0360497966	0,0027015649	-0,002701565	-0,060227542	0,0602275424

	Marginal effect of hhincome on the probability of ins=1	Marginal effect of hhincome on the probability of ins=2	Marginal effect of educyear on the probability of ins=1	Marginal effect of educyear on the probability of ins=2	Marginal effect of married on the probability of ins=1	Marginal effect of married on the probability of ins=2	Marginal effect of hisp on the probability of ins=1	Marginal effect of hisp on the probability of ins=2
1	-0,000365638	0,0003656379	-0,020979158	0,020979158	-0,107443116	0,107443116	0,1402934847	-0,140293485
2	-0,000373077	0,0003730772	-0,021406001	0,0214060007	-0,109629157	0,1096291575	0,143147901	-0,143147901
3	-0,000460095	0,0004600954	-0,026398833	0,0263988331	-0,135199558	0,1351995576	0,1765363644	-0,176536364
4	-0,000324616	0,0003246165	-0,018625474	0,018625474	-0,095388907	0,0953889073	0,1245537425	-0,124553743
5	-0,000324678	0,0003246784	-0,018629027	0,0186290274	-0,095407106	0,0954071059	0,1245775053	-0,124577505
6	-0,00048984	0,00048984	-0,028105484	0,0281054835	-0,143940034	0,1439400341	0,1879492119	-0,187949212
7	-0,000334996	0,0003349958	-0,019221009	0,0192210088	-0,098438892	0,0984388921	0,1285362498	-0,12853625
8	-0,000353072	0,0003530716	-0,020258142	0,0202581423	-0,10375049	0,1037504903	0,1354718513	-0,135471851
9	-0,000396842	0,0003968415	-0,022769523	0,0227695231	-0,116612331	0,1166123307	0,1522661558	-0,152266156
10	-0,000375566	0,0003755657	-0,021548784	0,0215487841	-0,110360411	0,1103604111	0,1441027329	-0,144102733

2) Interpretations

: The values exist for each observations, however, calculate mean value of each column (‘**Meff_P2_**(covatiate)’). The mean values for Logit and Probit models are summarized in **[Table 3]**.

Retired individuals are 4% more likely to have insurance (in comparison with those that are not retired).
For each additional year in education, individuals are 2% more likely to have insurance. Hispanics are 16% less likely to have insurance than non-Hispanics.

· Note that unlike the coefficients which are different, the marginal effects are almost identical in the models.
· The marginal effects at the mean and the average marginal effects are almost identical (especially for large sample sizes). However, for smaller samples, averaging the individual marginal effects is preferred.
· The signs of the coefficients and marginal effects are the same for the logit and probit models.

[Logit model ("lpred" dataset)]

	ins	retire	age	hstatusg	hhincome	educyear	married	hisp	Estimated Probability
1	0	0	62	0	0	12	0	0	0.2228155839
2	0	0	59	0	0	12	0	0	0.2304894588
3	0	1	60	1	0	13	0	0	0.3550969545
4	0	0	62	0	0	10	0	0	0.1857533277
5	0	0	54	0	0	9	0	0	0.1861319353
6	0	1	62	1	0	12	1	0	0.4597053522
7	0	0	59	0	0	5	1	0	0.1936115422
8	0	0	59	0	0	11	0	0	0.2108507724
9	0	0	65	0	0	14	0	0	0.2564301574
10	0	0	58	0	0.101	12	0	0	0.23312978

: For $i=1$, $\widehat{P}_1 = 0.2228$ (: the estimated probability of the first man having insurance is about 22.28%.)

[Probit model ("ppred" dataset)]

	ins	retire	age	hstatusg	hhincome	educyear	married	hisp	Estimated Probability
1	0	0	62	0	0	12	0	0	0.2205777209
2	0	0	59	0	0	12	0	0	0.2285479237
3	0	1	60	1	0	13	0	0	0.3573073711
4	0	0	62	0	0	10	0	0	0.1809517025
5	0	0	54	0	0	9	0	0	0.1810041726
6	0	1	62	1	0	12	1	0	0.4634244884
7	0	0	59	0	0	5	1	0	0.1903664953
8	0	0	59	0	0	11	0	0	0.2077114256
9	0	0	65	0	0	14	0	0	0.2561215381
10	0	0	58	0	0.101	12	0	0	0.2312780813

: For $i=1$, $\widehat{P}_1 = 0.2206$ (: the estimated probability of the first man having insurance is about 22.06%.)

Finally, if you want to predict the odds of being a case based on the values of the predictors, examine outputs of the **[Classification table]** from the CTABLE option and choose the cutoff point which shows the highest correct rate (the lowest error rate).

Below **[Classification table]** suggest that the prob. level of 0.48 or 0.54 would be desirable to build a superior prediction model.

[Classification table (of logit model)]

Classification Table									
	Correct		Incorrect		Percentages				
Prob Level	Event	Non-Event	Event	Non-Event	Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.020	1241	0	1965	0	38.7	100.0	0.0	61.3	.
0.040	1241	6	1959	0	38.9	100.0	0.3	61.2	0.0
0.060	1240	24	1941	1	39.4	99.9	1.2	61.0	4.0
0.080	1240	51	1914	1	40.3	99.9	2.6	60.7	1.9
0.100	1239	90	1875	2	41.5	99.8	4.6	60.2	2.2
0.120	1235	122	1843	6	42.3	99.5	6.2	59.9	4.7
0.140	1232	149	1816	9	43.1	99.3	7.6	59.6	5.7
0.160	1226	191	1774	15	44.2	98.8	9.7	59.1	7.3
0.180	1214	238	1727	27	45.3	97.8	12.1	58.7	10.2
0.200	1206	294	1671	35	46.8	97.2	15.0	58.1	10.6
0.220	1195	365	1600	46	48.7	96.3	18.6	57.2	11.2
0.240	1172	447	1518	69	50.5	94.4	22.7	56.4	13.4
0.260	1150	530	1435	91	52.4	92.7	27.0	55.5	14.7
0.280	1105	622	1343	136	53.9	89.0	31.7	54.9	17.9
0.300	1074	715	1250	167	55.8	86.5	36.4	53.8	18.9
0.320	1026	828	1137	215	57.8	82.7	42.1	52.6	20.6
0.340	984	936	1029	257	59.9	79.3	47.6	51.1	21.5
0.360	933	1019	946	308	60.9	75.2	51.9	50.3	23.2
0.380	865	1092	873	376	61.0	69.7	55.6	50.2	25.6
0.400	800	1164	801	441	61.3	64.5	59.2	50.0	27.5
0.420	729	1262	703	512	62.1	58.7	64.2	49.1	28.9
0.440	644	1346	619	597	62.1	51.9	68.5	49.0	30.7
0.460	523	1472	493	718	62.2	42.1	74.9	48.5	32.8
0.480	404	1594	371	837	62.3	32.6	81.1	47.9	34.4
0.500	337	1657	308	904	62.2	27.2	84.3	47.8	35.3
0.520	288	1703	262	953	62.1	23.2	86.7	47.6	35.9
0.540	247	1749	216	994	62.3	19.9	89.0	46.7	36.2

[Appendix. Interpretation with log-transformed variables]

① $\log(Y) = \beta_0 + \beta_1 X + \epsilon$ (**Log-Level Regression**)

: A change in X by 1 unit is associated with a $(e^{\widehat{\beta_1}} - 1)100\%$ change in Y.

② $Y = \beta_0 + \beta_1 \log(X) + \epsilon$ (**Level-Log Regression**)

: A 1% change in X is associated with a $\frac{\widehat{\beta_1}}{100}$ units of Y.

③ $\log(Y) = \beta_0 + \beta_1 \log(X) + \epsilon$ (**Log-Log Regression**)

: A 1% change in X is associated with a $\widehat{\beta_1}\%$ change in Y, so $\widehat{\beta_1}$ is the elasticity of Y with respect to X.

[SAS Code (with nonlinear_insurance.csv file)]

```
proc import out= work.data
datafile= "C:\Users\user\Downloads\nonlinear_insurance.csv"
dbms=csv replace; getnames=yes; datarow=2;
run;

proc means data=data;
var ins retire age hstatusg hhincome educyear married hisp;
run;

proc freq data=data;
tables ins;
run;

*OLS;
proc reg data=data;
model ins = retire age hstatusg hhincome educyear married hisp;
run;

*Logit;
proc logistic data=data descending;
model ins = retire age hstatusg hhincome educyear married hisp/ ctable pprob=0.5;
output out=lpred predicted=plogit;
run;

*Logit;
proc qlim data=data;
model ins = retire age hstatusg hhincome educyear married hisp / discrete (dist=logit);
output out=mfx marginal;
run;

*Logit marginal effects;
proc means data=mfx mean std;
var Meff_P2_retire Meff_P2_age Meff_P2_hstatusg Meff_P2_hhincome Meff_P2_educyear
Meff_P2_married Meff_P2_hisp;
run;

*Logit predicted probabilities;
proc means data=lpred;
var ins plogit;
run;
```

```
*Probit;
proc logistic data=data descending;
model ins = retire age hstatusg hhincome educyear married hisp / link=probit ctable
pprob=0.5;
output out=ppred predicted=pprobit;
run;

*Probit;
proc qlim data=data;
model ins = retire age hstatusg hhincome educyear married hisp / discrete (dist=normal);
output out=mfx marginal;
run;

*Probit marginal effects;
proc means data=mfx mean std;
var Meff_P2_retire Meff_P2_age Meff_P2_hstatusg Meff_P2_hhincome Meff_P2_educyear
Meff_P2_married Meff_P2_hisp;
run;

*Probit predicted probabilities;
proc means data=ppred;
var ins pprobit;
run;
```

[Stata Code (with nonlinear_insurance.dta file)]

```
clear all
set more off

use C:\Users\user\Downloads\nonlinear_insurance

global ylist ins
global xlist retire age hstatusg hhincome educyear married hisp

describe $ylist $xlist
summarize $ylist $xlist

tabulate $ylist

* Regression
reg $ylist $xlist

* Probit model
probit $ylist $xlist

* Logit model
logit $ylist $xlist


* Marginal effects (at the mean and average marginal effect)
quietly reg $ylist $xlist
margins, dydx(*) atmeans
margins, dydx(*)

quietly logit $ylist $xlist
margins, dydx(*) atmeans
margins, dydx(*)

quietly probit $ylist $xlist
margins, dydx(*) atmeans
margins, dydx(*)


*Logistic model gives odds ratio
logistic $ylist $xlist
```

```
* Predicted probabilities
quietly logit $ylist $xlist
predict plogit, pr

quietly probit $ylist $xlist
predict pprobit, pr

quietly regress $ylist $xlist
predict pols, xb

summarize $ylist plogit pprobit pols


* Percent correctly predicted values
quietly logit $ylist $xlist
estat classification

quietly probit $ylist $xlist
estat classification

* Probit and Logit Models;
* Copyright 2013 by Ani Katchova;
```