Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.

This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

Economics Letters 107 (2010) 291–296

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Testing hypotheses about interaction terms in nonlinear models

William Greene *

Stern School of Business, New York University, United States

ARTICLE INFO

Article history:
Received 19 May 2009
Received in revised form 17 February 2010
Accepted 25 February 2010
Available online 4 March 2010

JEL classification:
C12
C25
C51

Keywords:
Interaction effect
Interaction term
Partial effect
Probit
Nonlinear models

ABSTRACT

We examine the interaction effect in nonlinear models discussed by Ai and Norton (2003). Tests about partial effects and interaction terms are not necessarily informative in the context of the model. We suggest ways to examine the effects that do not involve statistical testing.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

A widely discussed contribution to econometric practice by Ai and Norton (2003) has proposed an approach to analyzing *interaction effects* in nonlinear single index models. The main result applies to nonlinear models such as

$$E[y|x_1,x_2,z] \;= F(\beta_1x_1 + \beta_2x_2 + \beta_{12}x_1x_2 + \delta z). \tag{1}$$

The authors argue that the common computation of the *partial effect of the interaction term*, $\gamma_{12}=\beta_{12}F'(.)$ provides no information about the *interaction effect* in the model $\Delta_{12}=\partial^2F(.)/\partial x_1\partial x_2$. They provide results for examining the magnitude and statistical significance of estimates of Δ_{12}. This note argues that Δ_{12} is also difficult to interpret in terms of the relationships among the variables in the model. The difficulty is a missing element in the measure of partial effects — the 'unit change' in the relevant variable may itself be unreasonable. An example given below illustrates the point. We argue that graphical devices can be much more informative than the test statistics suggested by the authors.

As a corollary to this argument, we suggest that the common practice of testing hypotheses about partial effects is less informative than one might hope, and could usefully be omitted from empirical analyses. The paper proceeds to a summary of the Ai and Norton (2003) results in Section 2, some discussion of the results in Section 3 and an application in Section 4. Conclusions are drawn in Section 5.

2. Estimation and inference for interaction effects

Consider the model in Eq. (1), where $F(.)$ is a nonlinear conditional mean function such as the normal or logistic cdf in a binary choice model, x_1 and x_2 are variables of interest, either or both of which may be binary or continuous, and z is a related variable or set of variables, including the constant term. For convenience, we will specialize the discussion to the a probit model,

$$\begin{aligned} E[y|x_1,x_2,z] &= Prob\,(y=1|x_1,x_2,z) \\ &= \Phi(\beta_1x_1+\beta_2x_2+\beta_{12}x_1x_2+\delta z) = \Phi(A), \end{aligned} \tag{2}$$

where $\Phi(A)$ is the standard normal cdf. The results will generalize to other models with only minor modification.[1] Partial effects in the model are

$$\begin{aligned} \partial E[y|x_1,x_2,z]\,/\,\partial x_1 &= \Phi'(A)\times\partial A\,/\,\partial x_1 \\ &= \phi(A)\times(\beta_1+\beta_{12}x_2), \end{aligned} \tag{3}$$

* Department of Economics, Stern School of Business, New York University, 44 West 4th St., 7-90, New York, NY, 10012, United States. Tel.: +1 212 998 0876; fax: +1 212 995 4218.
E-mail address: wgreene@stern.nyu.edu.
URL: http://pages.stern.nyu.edu/~wgreene.

[1] Ai and Norton (2003) analyze a logit model as a convenient application. The results for the probit model differ only by a minor change in notation. Their results apply in other models as well.

0165-1765/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.econlet.2010.02.014

Author's personal copy

for a continuous variable, where $\phi(A)$ is the standard normal pdf, or

$$\Delta E[y|x_1,x_2,z] / \Delta x_1 = E[y|x_1 = 1, x_2, z] – E[y|x_1 = 0, x_2, z] = \Phi(\beta_1 + \beta_2 x_2 + \beta_{12} x_2 + \delta z) - \Phi(\beta_2 x_2 + \delta z) \quad (4)$$

for a binary variable.

The interaction effect is the effect of a change in one of the variables on the partial effect of the other variable; for two continuous variables,

$$\frac{\partial^2 E[y|x_1,x_2,z]}{\partial x_1 \partial x_2} = \beta_{12}\Phi^{'}(A) + (\beta_1 + \beta_{12}x_2)(\beta_2 + \beta_{12}x_1)\Phi^{''}(A) = \beta_{12}\phi(A) + (\beta_1 + \beta_{12}x_2)(\beta_2 + \beta_{12}x_1)[-A\phi(A)]. \quad (5a)$$

Differentiation is replaced with differencing when the variables are binary;

$$\frac{\partial(\Delta E[y|x_1,x_2,z] / \Delta x_1)}{\partial x_2} = (\beta_2 + \beta_{12})\phi(\beta_1 + \beta_2 x_2 + \beta_{12} x_2 + \delta z) - \beta_2 \phi(\beta_2 x_2 + \delta z) \quad (5b)$$

or,

$$\frac{\Delta^2 E[y|x_1,x_2,z]}{\Delta x_1 \Delta x_2} = [\Phi(\beta_1 + \beta_2 + \beta_{12} + \delta z) - \Phi(\beta_2 + \delta z)] - [\Phi(\beta_1 + \delta z) - \Phi(\delta z)]. \quad (5c)$$

The notation $\Delta_{12}(A) = \Delta_{12}$ will be used for all three cases.

The coefficient on the interaction term, β_{12}, does not provide the change in the partial effect of either variable on the conditional mean function if the function is nonlinear. Even after scaling by $\phi(A)$ as in Eq. (3), the mismeasured interaction effect, $\phi(A)\beta_{12}$, which is what is likely to be reported by software that reports partial effects in the form of scaled coefficients, does not provide a useful measure of any interesting quantity. Ai and Norton (2003) note, "[However,] most applied economists instead compute the marginal effect of the interaction term, which is $\partial\Phi(.)/\partial(x_1 x_2) = \beta_{12}\Phi'(.)$." The discussion is motivated by the fact that statistical software (in their case, Stata® 7) typically mechanically computes a separate "partial effect" for each variable that appears in the model. The product variable would naturally appear as a separate variable in the model specification, and the software would have no way of discerning that it is a product of two variables that appear elsewhere in the model.

3. Inference about the interaction effect

Asymptotic standard errors for the partial and interaction effects in any of these cases may be computed using the delta method, as suggested by the authors in their Eqs. (4) and (5a)–(5c). They argue that inference about interaction effects should be based on Eqs. (5a), (5b), and (5c), not $\beta_{12}\Phi'(A)$. In an application, they demonstrate with plots of observation specific results from Eq. (5a) and associated t ratios against the predicted probabilities for models without, then with second order terms, $\beta_{12}x_1x_2$. It seems unclear what inference should be drawn from a plot of the "t statistics" associated with individual observations on Eq. (5a) against the fitted probabilities in Eq. (1), as is done in their Fig. 1b and our Fig. 3.

Consider, first, a case in which x_1 and x_2 are both continuous variables. Even if β_{12} in Eq. (1) equals zero, the interaction effect in Eq. (5a) is likely to be nonzero;

$$\left\{\left[\partial^2\Phi(A) / \partial x_1 \partial x_2\right] | \beta_{12} = 0\right\} = \beta_1\beta_2\Phi^{''}(A) = \beta_1\beta_2 \times [-A\phi(A)]. \quad (6)$$

For a probit (or logit) model, if β_1 and β_2 are nonzero, since the density must be positive, the result in Eq. (6) can be zero if and only if the index function, $(\beta_1 x_1 + \beta_2 x_2 + \delta z)$ is zero. Thus, the *interaction effect* is zero when and only when the index function is zero, or the probability equals one half. That the effect is nonzero when the probability deviates from one half makes sense. The conditional mean function, $\Phi(A)$, is essentially linear at $A = 0$ and the linear function would exhibit no interaction effects. Where the function departs from linearity, the effect in Eq. (6) becomes nonzero, even in a model in which there are no 'interaction effects.' This raises the complication that the interaction effect in the model (any model) is at least partly an artifact of the functional form chosen. In practice, the fitted probability will tend to mimic the sample proportion of ones. If the sample is unbalanced, A will generally not be close to zero. Insurance purchases in a population are very common; labor force participation rates vary by group, but could be around half; incidence of a disease might be quite rare.

In a model in which there is a true second order term, the result is less transparent. Then, the desired effect is that in Eq. (5a). As suggested in the article, one cannot assess the statistical significance of this interaction effect with a simple t test on the coefficient on the interaction term, β_{12}, or even $\beta_{12}F'(A)$. All coefficients and data could be positive; but $\Phi''(A) = -A\phi(A)$ might still tip the result toward zero. Even the sign of the interaction term depends on the data and, therefore, it could be close to zero numerically and insignificantly different from zero statistically. In this case, $\Phi(A) = .5$ is neither necessary nor sufficient for the interaction effect to equal zero. On the other hand, if $\Phi(A) = .5$, the interaction effect cannot be nonzero unless β_{12} is also nonzero. Once again, it is unclear how one is to interpret this result in economic terms.

The fault in this methodology is in the *statistical testing about these partial effects*. The computations are correctly prescribed. But, the *hypothesis* can be meaningless in the context of the model. In the probit model with no second order term, from Eq. (3),

$$\Delta_1 = \partial E[y|x_1,x_2,z] / \partial x_1 = \beta_1 \times \phi(A). \quad (7)$$

The density is positive if the index is finite. Suppose one has (as in our example below), fit the model and found the estimate of β_1 to be highly significantly different from zero. In testing the hypothesis that Δ_1 equals zero, one is in the contradictory position of testing the hypothesis that the product of two nonzero terms is zero. In practice, it rarely occurs that an estimated coefficient is statistically nonzero while the corresponding partial effect is not. But it could be that the sampling variance of the estimate of a highly nonlinear function such as $\phi(A)$ could be very large, particularly if the index, A, is large so that $\phi(A)$ is very small. If it should occur, the two outcomes are directly contradictory. Intuition would suggest that the test based on Δ_1 is the one that should be ignored. If the hypothesis that β_1 equals zero has already been rejected, it is pointless to test the hypothesis that Δ_1 equals zero.

An interesting case is that in which x_1 is a dummy variable and x_2 is continuous. In this instance, the *interaction effect* shows how the partial effect of the continuous x_2 varies with a regime switch in x_1;

$$\frac{\Delta(\partial E[y|x_1,x_2,z] / \partial x_2)}{\Delta x_1} = (\beta_2 + \beta_{12})\phi(\beta_1 + \beta_2 x_2 + \beta_{12} x_2 + \delta z) - \beta_2\phi(\beta_2 x_2 + \delta z). \quad (8)$$

As the authors note, the interaction effect depends on all variables in the model. A test of zero interaction effect could now be carried out based simply on $\beta_2 = \beta_{12} = 0$, which is sufficient, but not necessary. But, this is likely to be stronger than desired, since one might not have in mind to eliminate the interaction effect by eliminating the second variable entirely. The hypothesis that the combination of terms in Eq. (8) equals zero at a specific data point could be tested using the methods suggested earlier, as there are configurations of the data that will equate Eq. (8) to zero (statistically) without imposing

Author's personal copy

Table 1
Probit estimates for doctor. (Absolute asymptotic *t* ratios in parentheses).

Variable	Mean	Std. Dev. (range)	Model 0	Model 1	Model 2	Model 3	Model 4
Constant			−.1243 (2.138)	−.1422 (2.44)	−.2510 (4.03)	−.3058 (3.56)	−.4664 (5.18)
Female[a]	0.4788	.4996	.3559 (22.22)	.4552 (13.94)	.7082 (11.16)	.3453 (22.11)	.7647 (11.58)
Age	43.53	11.33 (25–64)	.01189 (14.95)	.01137 (14.05)	.01559 (15.20)	.01589 (9.89)	.01963 (10.96)
Income	.3521	.1769 (0.0–3.1)	−.1324 (2.84)	−.1197 (2.56)	−.1371 (2.94)	.4060 (2.20)	.4885 (2.51)
Married[a]	.7586	.4279	.07352 (3.56)	.1387 (4.99)	.06241 (3.01)	.07877 (3.80)	.1168 (4.11)
Young kids[a]	.4027	.4905	−.1521 (8.30)	−.1613 (8.71)	−.1588 (8.64)	−.1525 (8.32)	−.1658 (8.94)
Education	11.32	2.325 (7–18)	−.01497 (4.19)	−.01587 (4.43)	−.01641 (4.58)	−.01452 (4.06)	−.0165 (4.59)
Female × married				−.131 (3.49)			−.09607 (2.52)
Female × age					−.00820 (5.74)		−.00787 (5.40)
Income × age						−.01241 (2.86)	−.01418 (3.27)

[a] Binary variable.

Table 2
Estimated partial effects for age, female and income based on Model 2.

Variable	Coefficient	*t* ratio	Average partial effect[a]	Minimum effect[b]	Maximum effect[b]	Minimum *t* ratio[b]	Maximum *t* ratio[b]
Age	0.01559	15.20	0.00433	0.00205	0.00622	6.07	18.87
Female	0.7072	11.16	0.313	0.225	0.387	17.71	65.42
Income	−0.1371	−2.94	−0.0499	−0.0547	−0.0381	−3.23	−2.56

[a] Averaged across all observations.
[b] Based on values computed for each observation.

$\beta_2 = \beta_{12} = 0$. When the model contains numerous variables however, interpreting the outcome would be difficult at best. In our application, the model contains, in addition to x_1 = gender and x_2 = age, z = (constant, income, education, marital status, and presence of children in the household). Any number of different combinations of these variables could interact with age and gender to equate (8) to zero statistically. It is unclear what meaning one should attach to this.

The preceding suggests that one *can* test the hypothesis that the interaction effect is zero for a particular individual, or for the average individual in the data set. It is unclear, however, what the hypothesis means. Moreover, in any case, at least some of what we find is an artifact of the functional form, not necessarily an economically meaningful result.

4. Graphical analysis of partial effects

Riphahn, Wambach and Million (RWM) (2003) constructed count data models for physician and hospital visits by individuals in the German Socioeconomic Panel (GSOEP). The data are an unbalanced panel of 7293 families, with group sizes ranging from one to seven, for a total of 27,326 family-year observations.[2] We fit pooled probit models for *Doctor*, defined to equal one if the individual reports at least one physician visit in the family-year observation and zero otherwise.[3] In the full sample, 62.9% of the individuals reported at least one visit. The probit models include binary variables for marital status, gender and presence of young children in the household and continuous variables income, age and education. Estimates for several specifications are given in Table 1. All estimated coefficients in all models are statistically significant at the 99% level save for that on *Income* in Models 1, 3 and 4, which is significant at the 95% level.

Table 2 reports the estimated partial effects for *Age*, *Female* and *Income* based on Model 2. The highly significant relationship between the binary gender variable and the probabilities is evident in the tabulated partial effects. On average, the probability that an individual reports at least one visit to the doctor is .146 larger for women than men. Since the overall proportion is .629, this effect is extremely large. The sample splits roughly equally between men and women, so these values suggest that the average probabilities are about .629–.146/2 = .556 for men and about .702 for women, for a difference of roughly 26%. The partial effect of *Age* on the probability appears to be about +.004 per year. This is statistically highly significant for the average individual and for every individual in the sample. Over a 40 year observation period, if everything else were held fixed, this would translate to an increase in the probability of about 0.16. By age 65, the probability of a doctor visit in any given year would increase from .702 at age 45 for women to .782, and from .556 to .636 for men.

Based on the regression results in Table 1, conclusions based on Table 2 about statistical significance of the partial effects of *Age*, *Female* and *Income* seem foregone. The economic content of the results is shown in Fig. 1 which traces the impacts of *Age* and *Income* for a particular demographic group, married women aged 46 with children, average income and 16 years of education. The rows of teeth at the top and bottom of the figure show the range of variation of *Age* and *Income* (times 10), respectively. The trios of vertical lines show the means and plus and minus one sample standard deviation of *Income* at the left and *Age* at the right.

Among the problems of partial effects for continuous variables such as *Age* and *Income* is accommodating the units of measurement. For *Age* in particular, it is useful to examine the impact graphically as in Fig. 1. The partial effect per year as well as the range of variation is evident in the figure. To underscore the point, consider *Income*, which ranges from 0.0 to about 3.1 in the sample and has a measured partial effect in Model 2 of about −0.0499 with *t* ratios that range from −3.23 to −2.56. A change of one *unit* in *Income* is larger than five sample standard deviations, so the partial effect (per unit change) could be quite misleading. Once again, a graphical device that accommodates this scaling issue, such as Fig. 1, is likely to be more informative than a simple report of the partial derivative, even if averaged over the sample observations. If one presents partial effects numerically, it will be useful to provide as well a meaningful unit of change of the relevant variable. For example, for our income variable, a change of one standard deviation equal to 0.1769 is associated with a change in the probability of a doctor visit of 0.0499 × 0.1769 = 0.0088. Notwithstanding the persistently statistically significant effect of income in the model, the economic magnitude does

[2] The data are described in detail in RWM (2003) and in Greene (2008, p. 1088). Descriptive statistics for variables used in the models are given in Table 1.
[3] All computations were done with *NLOGIT 4.0* (http://www.nlogit.com). The raw data may be downloaded from the data archive of the *Journal of Applied Econometrics* at http://qed.econ.queensu.ca/jae/2003-v18.4/riphahn-wambach-million/. The data in the form of an *NLOGIT* project file may be downloaded from http://pages.stern.nyu.edu/~wgreene/healthcare.lpj. Program commands for replicating the results may be downloaded from http://pages.stern.nyu.edu/~wgreene/InteractionEffects.lim.

Author's personal copy

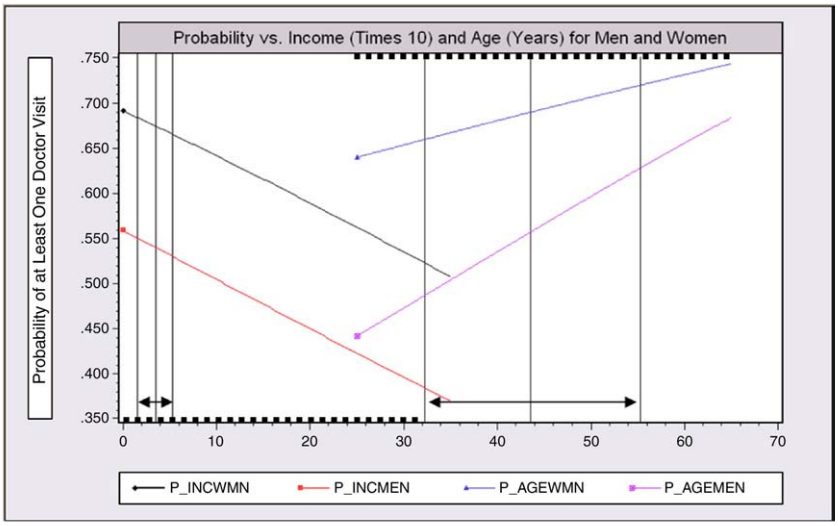

Fig. 1. Relationship between age and income and probability of doctor visit for men and women.

seem quite small. Fig. 1 also reveals the interaction between the gender and age effects, where we find that the gap between women and men appears to diminish as they age.

Fig. 1 is based on Model 2. The left side plots the predicted probabilities for men and women as a function of income. The predictors are essentially parallel. (There is a very small, indiscernible difference in the slopes.) The right hand side of the figure shows the probabilities plotted against age. We could interpret *an* "interaction effect" in this model as the change in the distance between the two sets of predicted probabilities. Perhaps confirming expectations, the impact of the interaction is to narrow the gap between predicted probabilities for men and women as age increases. The numerical results in Table 2 would not reveal this feature of the model.

We could examine the interaction effect more directly, as is done in Fig. 2. Fig. 2 shows a plot of Eq. (3) with *Income*, *Kids*, *Married* and *Education* held constant. The figure shows some of the interaction effects in Model 2, specifically for *Age* and *Female*. In the figure, the partial effect of changes in *Income*, not the probability is plotted on the vertical axis, as a function of *Age*. The lower curve is for men; the upper is for women. The interaction effect is the change in the partial effect of *Income* with respect to change in gender, which is the distance between the two curves. The distance is given by Eq. (5b), $\Delta\{\partial(E[y|Female, Income, Age, z]/\partial Income)\}/\Delta Female$. (We have multiplied the ordinate scale by 100 to enhance the visibility.)

Table 3 displays the estimated probabilities and interaction effects based on Eq. (5a) in Model 3, where there is a second order, term involving *Age* and *Income*. The interaction effect is statistically significant for every observation in the sample. Fig. 3 is the counterpart to Ai and Norton's 2(b). In these data, we find no observations for which the interaction effect is even close to zero, at least statistically. The numerical values average between −0.005 and −0.001. This seems economically small, though there is no obvious metric on which to base an evaluation. The measured value is a second derivative of the probability. The second derivative must be multiplied by meaningful changes of the variables in question to produce an appropriate measure of the partial effect. The case of a unit change in income discussed above suggests an example.

Fig. 3 does not suggest obviously how the probability of interest varies with *Age* or *Income*. We do note, based on Model 0 in Table 1, that *Age* and *Income* appear to act in opposite directions. Thus, the

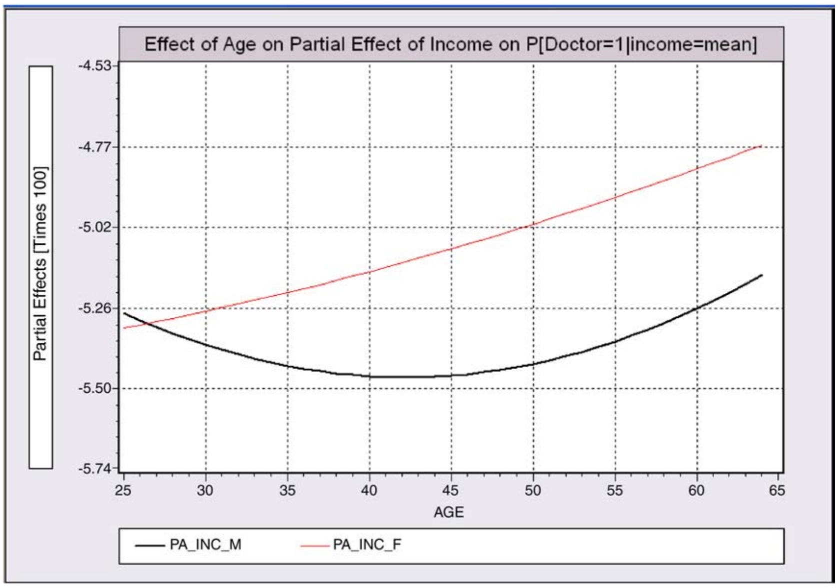

Fig. 2. Effect of gender on the partial effect of age.

Author's personal copy

Table 3
Estimated interaction effects between age and income in Model 3.

	Mean	Standard Deviation	Minimum	Maximum
Probability	0.6291	0.1002	0.4015	0.8366
Interaction effect	−0.004244	0.0007766	−0.005085	−0.001669
t ratio	−2.73	0.12	−3.18	−2.17

striking result in Fig. 4 might not be unexpected. Fig. 4 displays the relationship between *Income* and the fitted probability for four ages, 25, 35, 45 and 55 based on the estimates of Model 3. The interaction effect between *Age* and *Income* acts to reverse the sign of the partial effect of *Income* at about age 34.

5. Conclusions

The preceding does not fault Ai and Norton's (2003) suggested calculations. Rather, we argue that the process of statistical testing about partial effects, and interaction terms in particular, produces generally uninformative and sometimes contradictory and misleading results. The mechanical reliance on statistical measures of significance obscures the economic, numerical content of the estimated model. We conclude, on the basis of the preceding, in the words of the authors, that "to improve best practice by applied econometricians," a useful way to proceed in the analysis is a two step approach:

1. Build the model based on appropriate statistical procedures and principles. Statistical testing about the model specification is done at this step Hypothesis tests are about model coefficients and about the structural aspects of the model specifications. Partial effects are neither coefficients nor elements of the specification of the model. They are *implications* of the specified and estimated model.
2. Once the model is in place, inform the reader with analysis of model implications such as coefficient values, predictions, partial effects and interactions. We find that graphical presentations are a very informative adjunct to numerical statistical results for this

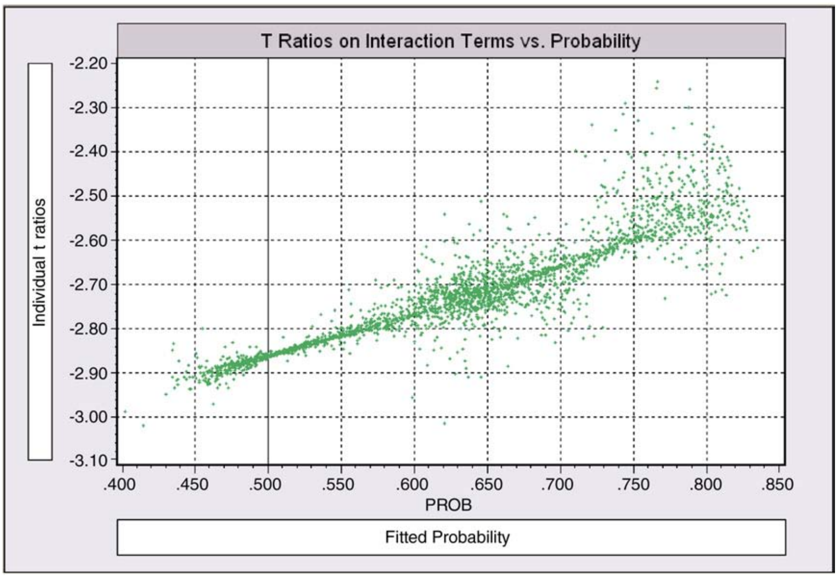

Fig. 3. Relationship between *t* statistics and probabilities in a model with interaction between age and income.

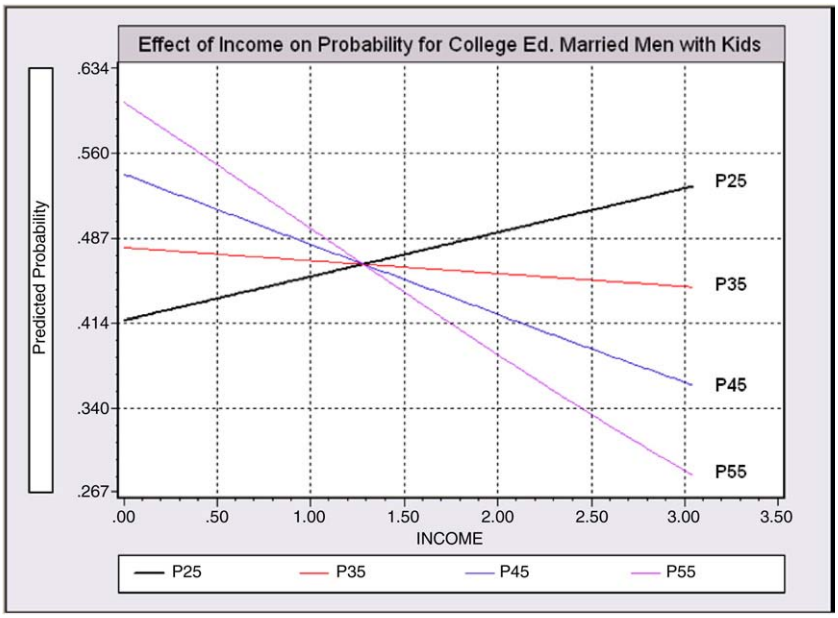

Fig. 4. Relationships between age and income and probability of doctor visit.

Author's personal copy

purpose. Hypothesis testing need not be done at this point. Even where the partial effects are the ultimate target of estimation, it seems it would be rare for a model builder to build a structural model by hypothesizing (statistically) about partial effects and/or predictions that would be made by that model.

This prescription conflicts with common practice. Widely used software packages such as *Stata*, *NLOGIT*, *EViews*, and so on all produce standard errors and t statistics for estimated partial effects, and it has become commonplace to report them among statistical results. The computations detailed in the example above were also simple to apply. See footnote 3. Ai and Norton have also made generally available a set of *Stata* code, *INTEFF*, for this purpose. In this note, we suggest that in spite of the availability of off the shelf software which facilitates the computations, the most informative point in the analysis at which to do hypothesis testing is at the model building step, not at the analysis step.

References

Ai, C., Norton, E., 2003. Interaction terms in logit and probit models. Economics Letters 80, 123–129.

Greene, W., 2008. Econometric Analysis, 6th Ed. Prentice Hall, Englewood Cliffs, NJ.

Riphahn, R., Wambach, A., Million, A., 2003. Incentive effects in the demand for health care: a bivariate panel count data estimation. Journal of Applied Econometrics 18 (4), 387–405.