中山大學嶺南學院
LINGNAN COLLEGE SUN YAT-SEN UNIVERSITY

Estimating Interaction Effects in Probit Model with Endogenous Regressors

Xianbo Zhou, Yujun Lian, Hujie Bai
(Lingnan College, Sun Yat-sen University)

作育英才 服务社会

中山大學嶺南學院
LINGNAN COLLEGE SUN YAT-SEN UNIVERSITY

CONTENTS

Part Ⅰ

Background

Background

Linear Model: $y = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_1 x_2 + x'\beta + \varepsilon$

The Interaction Effect $=$ α_3

And

Probit Model: $y = 1\{\alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_1 x_2 + x'\beta + \varepsilon > 0\}$

$$P(y = 1 \mid x_1, x_2, x) = \Phi(\alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_1 x_2 + x'\beta)$$

The Interaction Effect $\neq$ α_3

Background

Method

Available online at www.sciencedirect.com
SCIENCE DIRECT
Economics Letters 80 (2003) 123–129
economics letters
www.elsevier.com/locate/econbase
ELSEVIER

Interaction terms in logit and probit models

Chunrong Ai[a], Edward C. Norton[b,*]
[a]University of Florida, Gainesville, FL, USA
[b]Department of Health Policy and Administration, University of North Carolina, CB#7411 McGarvan-Greenberg Building, Chapel Hill, NC 27599-7411, USA

Received 4 December 2002; accepted 14 January 2003

Abstract

The magnitude of the interaction effect in nonlinear models does not equal the marginal effect of the interaction term, can be of opposite sign, and its statistical significance is not calculated by standard software. We present the correct way to estimate the magnitude and standard errors of the interaction effect in nonlinear models.

Keywords: Interaction effect; Interaction term; Logit; Probit; Nonlinear models

JEL classification: C12; C25; C51

Command

The Stata Journal (2004)
4, Number 2, pp. 154–167

Computing interaction effects and standard errors in logit and probit models

Edward C. Norton
Department of Health Policy and Administration
University of North Carolina at Chapel Hill
Hua Wang
Department of Health Policy and Administration
University of North Carolina at Chapel Hill
Chunrong Ai
Department of Economics
University of Florida and Tsinghua University, China

Abstract. This paper explains why computing the marginal effect of a change in two variables is more complicated in nonlinear models than in linear models. The command `inteff` computes the correct marginal effect of a change in two interacted variables for a logit or probit model, as well as the correct standard errors. The `inteff` command graphs the interaction effect and saves the results to allow further investigation.

Keywords: st0063, inteff, interaction terms, logit, probit, nonlinear models

作育英才 服务社会

中山大學嶺南學院
LINGNAN COLLEGE SUN YAT-SEN UNIVERSITY

Method

$$y = 1\{\alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_1 x_2 + \alpha_4 x_3 + \alpha_5 x_3^2 + x'\beta + \varepsilon > 0\}$$

Economics Letters 198 (2021) 109695

Contents lists available at ScienceDirect

Economics Letters

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

Interaction and quadratic effects in probit model with endogenous regressors

Xianbo Zhou [*,1], Heyang Li

Lingnan College, Sun Yat-sen University, Guangzhou, China

ARTICLE INFO

Article history:
Received 29 May 2020
Received in revised form 25 September 2020
Accepted 8 December 2020
Available online 11 December 2020

JEL classification:
C12
C25
C51

Keywords:
Endogenous probit model
Interaction effect
Quadratic effect
Bootstrap standard errors

ABSTRACT

This paper proposes a method to estimate the interaction and quadratic effects in probit model with endogenous regressors, which generalizes Ai and Norton (2003)'s study on the interaction effect in nonlinear model without endogenous regressors. The method is applied to estimate the interaction effect of the new-typed information tool usage and social network and the quadratic effect of age on the household risky assets investment, where the bootstrap standard errors of the two effects are provided.

Endogenous Probit Models

The Control Function Method

$x_1 x_2$

The Interaction Effects

x_3^2

The Quadraic Effects

Give more details on ME and Interaction effect and Develop a Stata Command to implement Zhou-Li's method (eivprobit)

$$y = 1\{\alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_1 x_2 + x'\beta + \varepsilon > 0\}$$

*Case*1:
X_1: Continuous and Endogenous
X_2: Continuous

*Case*2:
X_1: Continuous and Endogenous
X_2: Dummy but Exogenous

In addition:
eivprobit can scatter the estimated marginal, interaction and quadratic effects

Part Ⅱ

Model

Continuous and Endogenous

$$Example: y = 1\{\alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_1 x_2 + x'\beta + \varepsilon > 0\}$$

Continuous but Exogenous

$Cov(\varepsilon, x_1) \neq 0$

Control Function Method

$x_1 = \gamma_2 x_2 + x'\gamma + z'\delta + v_1$, $\varepsilon = \theta_1 v_1 + e$

$$y = 1\{\alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_1 x_2 + x'\beta + \theta_1 v_1 + e > 0\}$$

Normalizing the Error Te $e \sim N(0, \sigma_e^2)$

$$y = 1\{a_1 x_1 + a_2 x_2 + a_3 x_1 x_2 + x'b + \gamma_1 v_1 + u > 0\}$$

where $a_1 = \frac{\alpha_1}{\sigma_e}, a_2 = \frac{\alpha_2}{\sigma_e}, a_3 = \frac{\alpha_3}{\sigma_e}, b = \frac{\beta}{\sigma_e},$

$\gamma_1 = \frac{\theta_1}{\sigma_e}$, and $u = \frac{e}{\sigma_e} \sim N(0, 1)$

Marginal probability and interaction effects

中山大學嶺南學院
LINGNAN COLLEGE SUN YAT-SEN UNIVERSITY

$y = 1\{a_1x_1+a_2x_2+a_3x_1x_2+x'b+\gamma_1v_1+u>0\}$, where

$$a_1 = \frac{\alpha_1}{\sigma_e}, a_2 = \frac{\alpha_2}{\sigma_e}, a_3 = \frac{\alpha_3}{\sigma_e}, b = \frac{\beta}{\sigma_e}, \gamma_1 = \frac{\theta_1}{\sigma_e}, \text{ and } u = \frac{e}{\sigma_e} \sim N(0,1)$$

$$P(y=1 \mid x_1, x_2, x, v_1) = \Phi(a_1x_1 + a_2x_2 + a_3x_1x_2 + x'b + \gamma_1v_1)$$

$$P(y=1 \mid x_1, x_2, x) = E_{v_1}[\Phi(a_1x_1 + a_2x_2 + a_3x_1x_2 + x'b + \gamma_1v_1)]$$

$$APE_1(x_1,x_2,x;\lambda) = (a_1 + a_3x_2)E_{v_1}[\phi(\tau)]$$

$$APE_2(x_1,x_2,x;\lambda) = (a_2 + a_3x_1)E_{v_1}[\phi(\tau)]$$

$$inteff(x_1,x_2,x;\lambda) = a_3E_{v_1}[\phi(\tau)] - (a_1 + a_3x_2)(a_2 + a_3x_1)E_{v_1}[\tau\phi(\tau)]$$

where $\tau = a_1x_1 + a_2x_2 + a_3x_1x_2 + x'b + \gamma_1v_1$.

Probability estimate

$$y = 1\{a_1x_1 + a_2x_2 + a_3x_1x_2 + x'b + \gamma_1 v_1 + u > 0\}$$

$$x_1 = \gamma_2 x_2 + x'\gamma + z'\delta + v_1$$

(i) regress x_1 on x_2, x, z to obtain the residuals $\hat{v}_1$;

(ii) probit y on $x_1, x_2, x_1x_2, x, \hat{v}_1$ to get coefficient estimates $\hat{\lambda} \equiv (\hat{a}_1, \hat{a}_2, \hat{a}_3, \hat{b}', \hat{\gamma}_1)'$.

Generated regressor

$\hat{\lambda} \xrightarrow{p} \lambda$.

Given x_1, x_2, x, the probability estimate is

$$P\left(x_1, x_2, x; \hat{\lambda}\right) \equiv \frac{1}{n}\sum_{i=1}^{n} \Phi\left(\tau_i\right) \xrightarrow{p} P\left(y=1 \mid x_1, x_2, x\right)$$

where $\tau_i \equiv \hat{a}_1 x_1 + \hat{a}_2 x_2 + \hat{a}_3 x_1 x_2 + x'\hat{b} + \hat{\gamma}_1 \hat{v}_{1i}$.

Asymptotic properties

$$\hat{\lambda} \xrightarrow{p} \lambda, \quad \sqrt{n}(\hat{\lambda} - \lambda) \xrightarrow{d} N(0, \Omega)$$

$$\Omega \equiv Asy.var(\hat{\lambda}) = -\frac{1}{n}\left[H_{22}^{(2)}\right]^{-1} H_{21}^{(2)} + \left[H_{22}^{(2)}\right]^{-1} \Sigma_{21} \left[H_{11}^{(1)}\right]^{-1} H_{12}^{(2)} \left[H_{22}^{(2)}\right]^{-1}$$

$$-\frac{1}{n}\left[H_{22}^{(2)}\right]^{-1} H_{21}^{(2)} \left[H_{11}^{(1)}\right]^{-1} \Sigma_{12} \left[H_{22}^{(2)}\right]^{-1} - \frac{1}{n}\left[H_{22}^{(2)}\right]^{-1}.$$

X_2: Continuous and exogeneous

The Marginal Probability Effects (MPE): x_1&x_2

$$\mathrm{MPE}_1\left(x_1, x_2, x; \hat{\lambda}\right) \equiv \frac{\partial P\left(x_1, x_2, x; \hat{\lambda}\right)}{\partial x_1} = \left(\hat{a}_1 + \hat{a}_3 x_2\right) \frac{1}{n} \sum_{i=1}^{n} \phi\left(\tau_i\right) \xrightarrow{p} APE_1(x_1, x_2, x; \lambda)$$

$$\mathrm{MPE}_2\left(x_1, x_2, x; \hat{\lambda}\right) \equiv \frac{\partial P\left(x_1, x_2, x; \hat{\lambda}\right)}{\partial x_2} = \left(\hat{a}_2 + \hat{a}_3 x_1\right) \frac{1}{n} \sum_{i=1}^{n} \phi\left(\tau_i\right) \xrightarrow{p} APE_2(x_1, x_2, x; \lambda)$$

The Interaction Effects (inteff): x_1x_2

$$\operatorname{inteff}\left(x_1, x_2, x; \hat{\lambda}\right) \equiv \frac{\partial^2 P\left(x_1, x_2, x; \hat{\lambda}\right)}{\partial x_1 \partial x_2} = \frac{1}{n} \sum_{i=1}^{n} \left[\hat{a}_3 - \left(\hat{a}_1 + \hat{a}_3 x_2\right)\left(\hat{a}_2 + \hat{a}_3 x_1\right) \tau_i\right] \phi\left(\tau_i\right)$$

$$\xrightarrow{p} inteff(x_1, x_2, x; \lambda)$$

X_2: Dummy and exogenous

The Marginal Probability Effects (MPE): x_2

$$\text{MPE}_2\left(x_1, x_2, x; \hat{\lambda}\right) \equiv P\left(x_1, 1, x; \hat{\lambda}\right) - P\left(x_1, 0, x; \hat{\lambda}\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left[\Phi\left(\hat{a}_1 x_1 + \hat{a}_2 + \hat{a}_3 x_1 + x'b + \hat{\gamma}_1 \hat{v}_{1i}\right) - \Phi\left(\hat{a}_1 x_1 + x'b + \hat{\gamma}_1 \hat{v}_{1i}\right)\right]$$

The Interaction Effects (inteff): x_1x_2

$$\text{inteff}\left(x_1, x_2, x; \hat{\lambda}\right) \equiv MPE_1\left(x_1, 1, x; \hat{\lambda}\right) - MPE_1\left(x_1, 0, x; \hat{\lambda}\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left[\left(\hat{a}_1 + \hat{a}_3\right)\phi\left(\hat{a}_1 x_1 + \hat{a}_2 + \hat{a}_3 x_1 + x'b + \hat{\gamma}_1 \hat{v}_{1i}\right) - \hat{a}_1\phi\left(\hat{a}_1 x_1 + x'b + \hat{\gamma}_1 \hat{v}_{1i}\right)\right]$$

Estimate Quadratic Effects

quadratic forms $\hat{a}_4 x_3 + \hat{a}_5 x_3^2$,

The Marginal Probability Effects (MPE): x_3

$$\text{MPE}_3\left(x_1, x_2, x; \hat{\lambda}\right) \equiv \frac{\partial P\left(x_1, x_2, x; \hat{\lambda}\right)}{\partial x_3} = \left(\hat{a}_4 + 2\hat{a}_5 x_3\right) \frac{1}{n} \sum_{i=1}^{n} \phi\left(\tau_i\right)$$

The Quadratic Effects (quaeff): $x_1 x_2$

$$\text{quaeff}\left(x_1, x_2, x; \hat{\lambda}\right) \equiv \frac{\partial^2 P\left(x_1, x_2, x; \hat{\lambda}\right)}{\partial x_3^2} = \frac{1}{n} \sum_{i=1}^{n} \left[2\hat{a}_5 - \left(\hat{a}_4 + 2\hat{a}_5 x_3\right)^2 \tau_i\right] \phi\left(\tau_i\right)$$

Test Statistic

$$\sqrt{n}\left(inteff(\hat{\lambda}) - inteff(\lambda)\right) = \frac{\partial inteff}{\partial \hat{\lambda}'}\sqrt{n}(\hat{\lambda}-\lambda) + O(\sqrt{n}\,|\,\hat{\lambda}-\lambda|^2)$$

$$\xrightarrow{d} \frac{\partial inteff}{\partial \lambda'} N(0,\Omega) = N\left(0, \frac{\partial inteff}{\partial \lambda'}\Omega\frac{\partial inteff}{\partial \lambda}\right)$$

H_0: No Interaction Effects i.e. $inteff(x_1, x_2, x; \lambda) = 0$

$$\frac{\text{inteff}\left(x_1, x_2, x; \widehat{\lambda}\right) - 0}{\sqrt{\frac{\partial \text{inteff}\left(x_1, x_2, x; \widehat{\lambda}\right)}{\partial \lambda'} \widehat{\Omega}_\lambda \frac{\partial \text{inteff}\left(x_1, x_2, x; \widehat{\lambda}\right)}{\partial \lambda'}}} \rightarrow N(0,1)$$

$\widehat{\Omega}_\lambda$ is the var-covariance estimator of $\widehat{\lambda}$

Nonparametric Bootstrap Method

Estimate Average Effects

If estimate the effects through loop for observations i, j, it will be potentially infeasible, for large datasets.

So $\left\{\hat{v}_1^{(j)}\right\}_{j=1}^{N_r} \xrightarrow{\text{Approximation}} \left\{\hat{v}_{1i}\right\}_{i=1}^{n}$ e.g. $N_r = 1000$

$$\text{MPE}_1 = \frac{1}{nN_r}\sum_{i=1}^{n}\sum_{j=1}^{N_r}\left(\hat{a}_1 + \hat{a}_3 x_{2i}\right)\phi\left(\tau_{i,j}\right),$$
$$\text{MPE}_2 = \frac{1}{nN_r}\sum_{i=1}^{n}\sum_{j=1}^{N_r}\left(\hat{a}_2 + \hat{a}_3 x_{1i}\right)\phi\left(\tau_{i,j}\right),$$
$$\text{inteff} = \frac{1}{nN_r}\sum_{i=1}^{n}\sum_{j=1}^{N_r}\left[\hat{a}_3 - \left(\hat{a}_1 + \hat{a}_3 x_{2i}\right)\left(\hat{a}_2 + \hat{a}_3 x_{1i}\right)\tau_{i,j}\right]\phi\left(\tau_{i,j}\right)$$

$$\text{where } \tau_{i,j} = \hat{a}_1 x_{1i} + \hat{a}_2 x_{2i} + \hat{a}_3 x_{1i} x_{2i} + x_i'\hat{b} + \hat{\theta}_1 \hat{v}_{1j}$$

MATA in Stata can handle averaging across columns of an $n\times N_r$ matrix by using matrix block operation

Part Ⅲ

Monte Carlo Simulations

Monte Carlo Simulations

Continuous and Endogenous

Model: $y = 1\{a_0 + a_1 x_1 + a_2 x_2 + a_3 x_1 x_2 + \beta x + \varepsilon > 0\}$

$$\rho = \frac{\mathrm{Cov}(\varepsilon, v_1)}{\sigma_\varepsilon \sigma_{v_1}} = \frac{\theta_1}{\sqrt{1+\theta_1^2}}$$

Continuous but Exogenous

DGP: $$y = 1\{a_0 + a_1 x_1 + a_2 x_2 + a_3 x_1 x_2 + \beta x + \theta_1 v_1 + e > 0\} \quad (17)$$

where $a_0 = 0$, $a_1 = 1$, $a_2 = 2$, $a_3 = -1$, $\beta = 1$, $\theta_1 = 1$, the variables $x \sim N(-1, 1)$, $e \sim N(0, 1)$, $v_1 \sim N(0, 1)$, $z_1 \sim N(0, 1)$ and $z_2 \sim N(0, 2)$ are generated independently. Then we generate $x_1 = 1 + x + z_1 + v_1$, $x_2 = 1 + 2x + 2z_2$, and y as shown in (17).

Only Simulate the interaction effect:

Type Ⅰ: *Different Interaction Effects* at different points (x_1, x_2, x)

Type Ⅱ: *Different Endogeneity*

Type Ⅰ: at different points inteff(x_1, x_2, x)

中山大學嶺南學院
LINGNAN COLLEGE SUN YAT-SEN UNIVERSITY

$$y = 1\{a_0 + a_1x_1 + a_2x_2 + a_3x_1x_2 + \beta x + \varepsilon > 0\}$$

Table 2: Simulation of interaction effect: $n = 1600$

	(x_1, x_2, x)	$(-1,-1,1)$	$(-0.5,-1,1)$	$(0,-1,1)$	$(0.5,-1,1)$	$(1,-1,1)$
	inteff	0.2340	0.4152	0.2140	-0.2867	-0.4384
Ai-Norton Method	mean	0.0103	0.1217	0.4719	0.2483	-0.5472
	bias	-0.2238	-0.2935	0.2579	0.5350	-0.1087
	std	0.0111	0.0679	0.0902	0.1727	0.0931
	rmse	0.2242	0.3015	0.2735	0.5626	0.1432
$\hat{\alpha}_3$ in IV-probit	mean	-0.0985	-0.0985	-0.0985	-0.0985	-0.0985
	bias	-0.3326	-0.5138	-0.3125	0.1882	0.3399
	std	0.0123	0.0123	0.0123	0.0123	0.0123
	rmse	0.3331	0.5143	0.3130	0.1887	0.3404
Our Method ✓	mean	0.2277	0.3907	0.2264	-0.2659	-0.4414
	bias	-0.0063	-0.0245	0.0125	0.0208	-0.0030
	std	0.0876	0.0602	0.1478	0.1141	0.0520
	rmse	0.0878	0.0650	0.1483	0.1160	0.0521

作育英才 服务社会

Type II: Different endogeneity

$$y = 1\{a_0 + a_1x_1 + a_2x_2 + a_3x_1x_2 + \beta x + \varepsilon > 0\}$$

$$y = 1\{a_0 + a_1x_1 + a_2x_2 + a_3x_1x_2 + \beta x + \theta_1 v_1 + e > 0\}$$

Groups

	θ_1	inteff
I	-2	0.1400
II	-1	0.0544
III	0	0.0031
IV	1	0.0587
V	2	0.1392

$\varepsilon = \theta_1 v_1 + e$

Endogenous

Exogenous

Endogenous

$$\rho = \frac{Cov(\varepsilon, v_1)}{\sigma_\varepsilon \sigma_{v_1}} = \frac{\theta_1}{\sqrt{1+\theta_1^2}}$$

Table 3: Simulation of interaction effect at $(-1, -1, 0)$ for different θ_1

	θ_1	-2	-1	0	1	2
	inteff	0.1400	0.0544	0.0031	0.0587	0.1392
Ai-Norton Method	mean	0.1153	0.1287	0.0039	0.0022	0.0103
	bias	-0.0247	0.0744	0.0009	-0.0565	-0.1289
	std	0.0397	0.0450	0.0043	0.0023	0.0072
	rmse	0.0468	0.0870	0.0044	0.0566	0.1292
$\hat{\alpha}_3$ in IV-probit	mean	-0.0938	-0.0994	-0.1008	-0.0985	-0.0961
	bias	-0.2339	-0.1538	-0.1039	-0.1573	-0.2353
	std	0.0122	0.0122	0.0122	0.0123	0.0118
	rmse	0.2344	0.1544	0.1047	0.1579	0.2358
Our Method √	mean	0.1360	0.0573	0.0042	0.0574	0.1327
	bias	-0.0041	0.0029	0.0011	-0.0013	-0.0065
	std	0.0150	0.0214	0.0049	0.0315	0.0160
	rmse	0.0156	0.0216	0.0050	0.0315	0.0173

$n = 1600$

Part Ⅳ

Command and Applications

The eivprobit Command

$$y = 1\{\alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_1 x_2 + x'\beta + \varepsilon > 0\}$$

eivprobit *depvar* (y) [*indepvars* (x)] (*variable1*=*varlist_iv*) (x_1, z) [*if*] [*in*], interact(*variable2*) (x_2) [options]

Options

`maeffect(`*varname*`)` specifies the model without a squared term of an interested variable, which is one of the variable in the *indepvars*.

`quaeffect(`*varname*`)` specifies the model with a squared term of an interested variable in the *indepvars*, the squared effect of which can be estimated.

`seed(`#`)` specifies the seed when bootstrapping the standard errors of the parametric estimates.

`bootstrap(`#`)` specifies the replicates of the bootstrap, default of which is `bootstrap(50)`.

`endog2(1)` specifies that *variable2* is also endogeneous, the instrumental variables of which are set in *varlist_iv*.

help for eivprobit

eivprobit — Calculate the average effect (Interaction effect or Quadratic effect) in Probit Model with Endogenous Regressor.

Syntax

eivprobit *depvar* [*varlist1*] (*variable1* = *varlist_iv*) [*if*] [*in*] [*weight*] , **interact**(*variable2*) [*options*]

variable1 is the endogenous variable.

variable2 is the interact variable (It shouldn't be involved in the varlist1 and it could be endogenous).

varlist1 is the list of control variables (exclude *variable2*).

varlist_iv is the list of instruments variables.

Options	Description
maeffect(*variable3*)	Calculate the marginal effect of variable3 which is control variable in the model ().
quaffect(*variable3*)	Calculate the marginal and quadratic effect of variable3 which is control variable in the model ().
seed(#)	Set random-number seed to #
bootstrap	Perform # bootstrap replications; default is bootstrap(50)
endog2(*string*)	endog2(1) specifies that variable2 is also endogeneous, the instrumental variables of which are also included in varlist iv

Description

eivprobit fits models for binary dependent variables where one or more of the covariates are endogenous and errors are normally distributed and estimate the interaction effect which is consistent. **eivprobit** estimation is based on the control function approach and the standard errors of the estimated effects are obtained by nonparametric bootstrapping. And **eivprobit** allows both *variable1* and *variable2* be endogenous.

Examples

```
*When x1 is endogenous, x2 is exogenous
. use Zhou2021_EL, clear
. eivprobit y $control (x1=$ivs), interact(x2) bootstrap(500)

*When x1 is endogenous, x2 is exogenous, and the result shows the marginal effect of control variable 'age'
. use Zhou2021_EL, clear
. eivprobit y $control (x1=$ivs), interact(x2) maeffect(age) bootstrap(500)

*When x1 is endogenous, x2 is exogenous, and the result shows the effect of control variable 'age' and 'age*age'
. use Zhou2021_EL, clear
. eivprobit y $control (x1=$ivs), interact(x2) quaeffect(age) bootstrap(500)

*When both x1 and x2 are endogenous
. use Zhou2021_EL, clear
. eivprobit y $control (x1=$ivs), interact(x2) quaeffect(age) bootstrap(500) tt(1)

*Draw the graph on relationship between possibility and the marginal effect for variable1
. scatter mex1 Phat,msize(vsmall)ytitle() xtitle()
. graph export, as(png) replace
```

Model: $y = 1\{\alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_1 x_2 + \alpha_4 x_3 + \alpha_5 x_3^2 + x'\beta + \varepsilon > 0\}$

Signal	Meaning	State	Effects
y	A binary variable of whether the household participates in the risky investment.	Exogenous	\
x1	Social Network	Endogenous	MPE+Interaction
x2	The degree of the APP-Internet usage	Exogenous	
x3	Age	Exogenous	MPE+Quadratic
z	*iphone, onlineshop, cell* and *fee*	Exogenous	\
x	the household head's demographic characteristics such as gender, education, marital status, risk preference, hukou and job category, and the household's characteristics such as wealth, income and family size.	Exogenous	\

Example-Codes-Main: IV-Probit estimation

```
. use Zhou2021_EL.dta,clear all

. gen x1=lnsocial      //potential endogeneity
. gen x2=inf           //APP-internet usage
. gen x1x2=x1*x2      //interaction term
. global control lnwealth lnincome age age2 edu gender marriage ///
> risk_lover risk_averter size job rural east west
. global ivs iphone onlineshop cell fee   // IVs
. foreach var of global ivs {
  gen `var´_x2=`var´*x2
. }

. global ivs_x2  iphone_x2 onlineshop_x2 cell_x2 fee_x2
. ivprobit y (x1 x1x2 = $ivs $ivs_x2) x2 $control

Probit model with endogenous regressors          Number of obs     =     37,794
                                                 Wald chi2(17)     =   21404.46
Log likelihood = -150059.19                      Prob > chi2       =     0.0000

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .3434453   .0083155    41.30   0.000     .3271473    .3597433
        x1x2 |  -.9350269   .0934335   -10.01   0.000    -1.118153   -.7519006
          x2 |   4.306428   .4848708     8.88   0.000     3.356099    5.256757
         age |   .1404128    .042971     3.27   0.001     .0561911    .2246344
        age2 |  -.0119476   .0039139    -3.05   0.002    -.0196188   -.0042764
     (output ommited)
       _cons |  -2.811661   .1681245   -16.72   0.000    -3.141179   -2.482144
-------------+----------------------------------------------------------------
```

Example-Codes-Main: *eivprobit*


```
. global control lnwealth lnincome age edu gender marriage ///
> risk_lover risk_averter size job rural east west
. qui eivprobit y $control (x1=$ivs), interact(x2)quaeffect(age)  ///
> bootstrap(500) seed(123454321)
. run=1
...
. run=500
***Result
              mean          se            z           P>|z|        95% conf.interval
 mex1:      0.0554      0.0070       7.9732       0.0000      0.0418       0.0690
 mex2:      0.1928      0.0574       3.3602       0.0008      0.0803       0.3052
 inteff:  0.0480      0.0066       7.2769       0.0000      0.0351       0.0610
 meage:    0.0068      0.0016       4.2422       0.0000      0.0036       0.0099
 qeage:  -0.0077      0.0023      -3.4004       0.0007     -0.0121      -0.0033
```

Example-Codes-Scatters

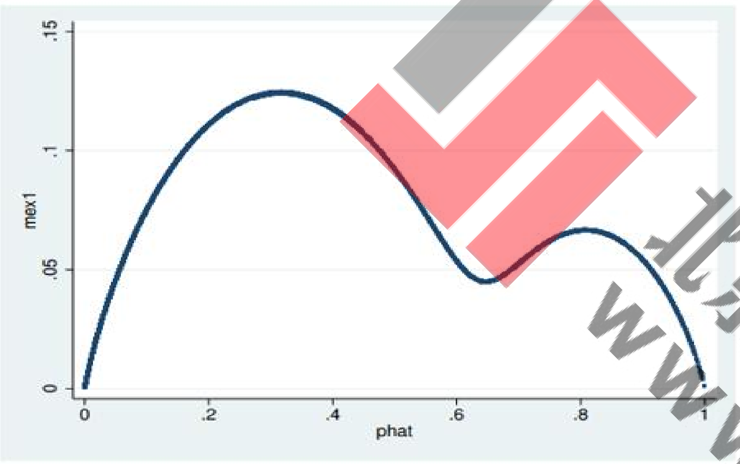

Figure 1: scatter of mex1 and phat

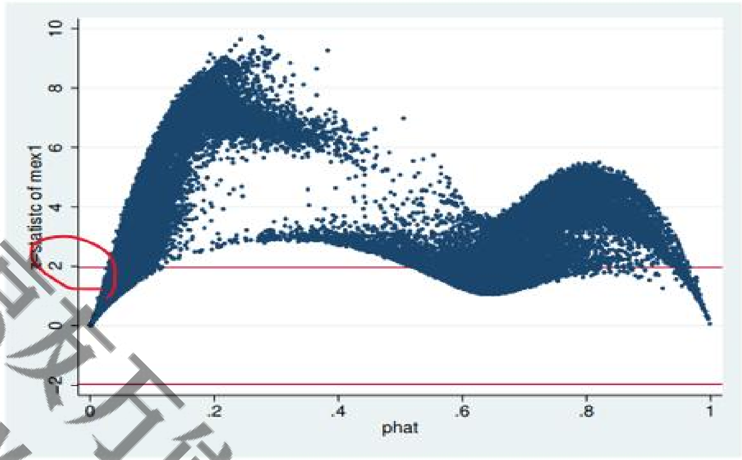

Figure 2: scatter of z-statistic and phat

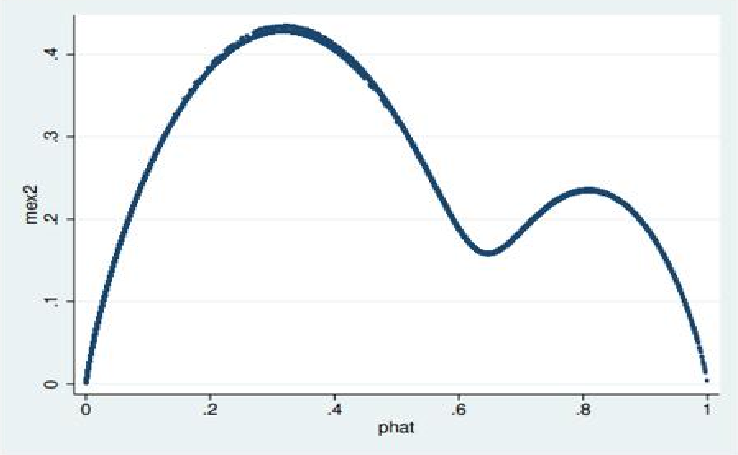

Figure 3: scatter of mex2 and phat

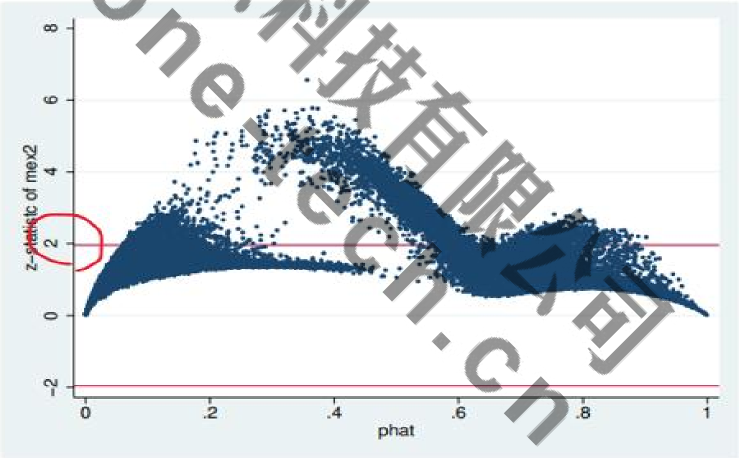

Figure 4: scatter of z-statistic and phat

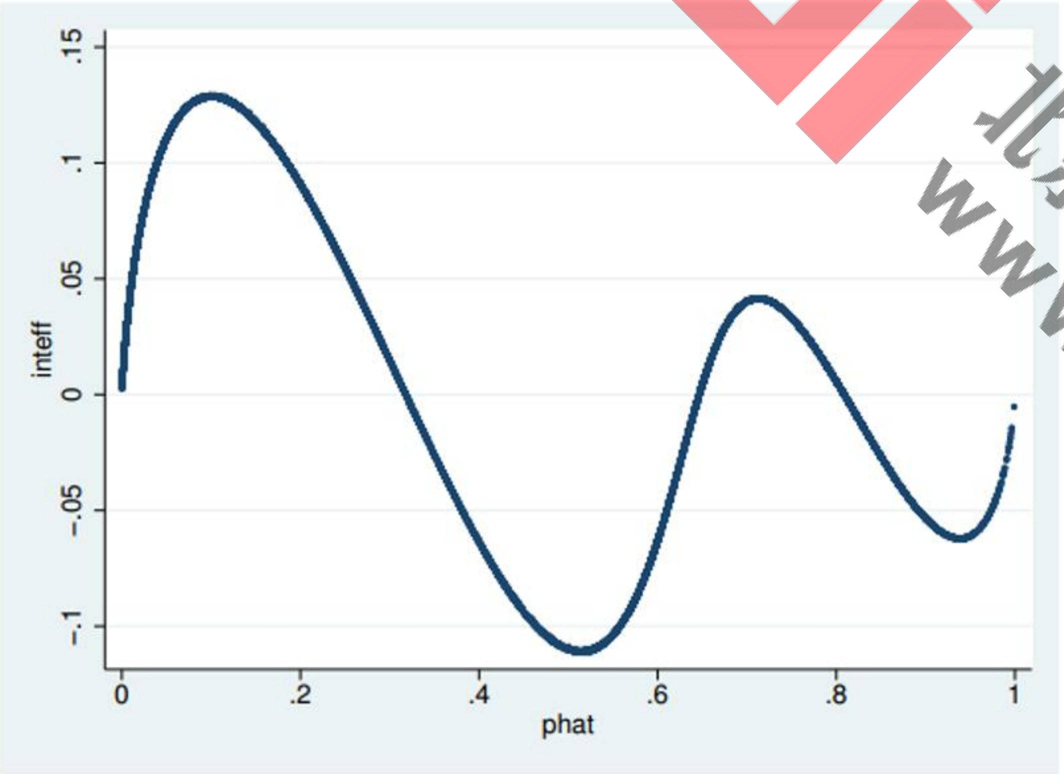

Figure 5: scatter of inteff and phat

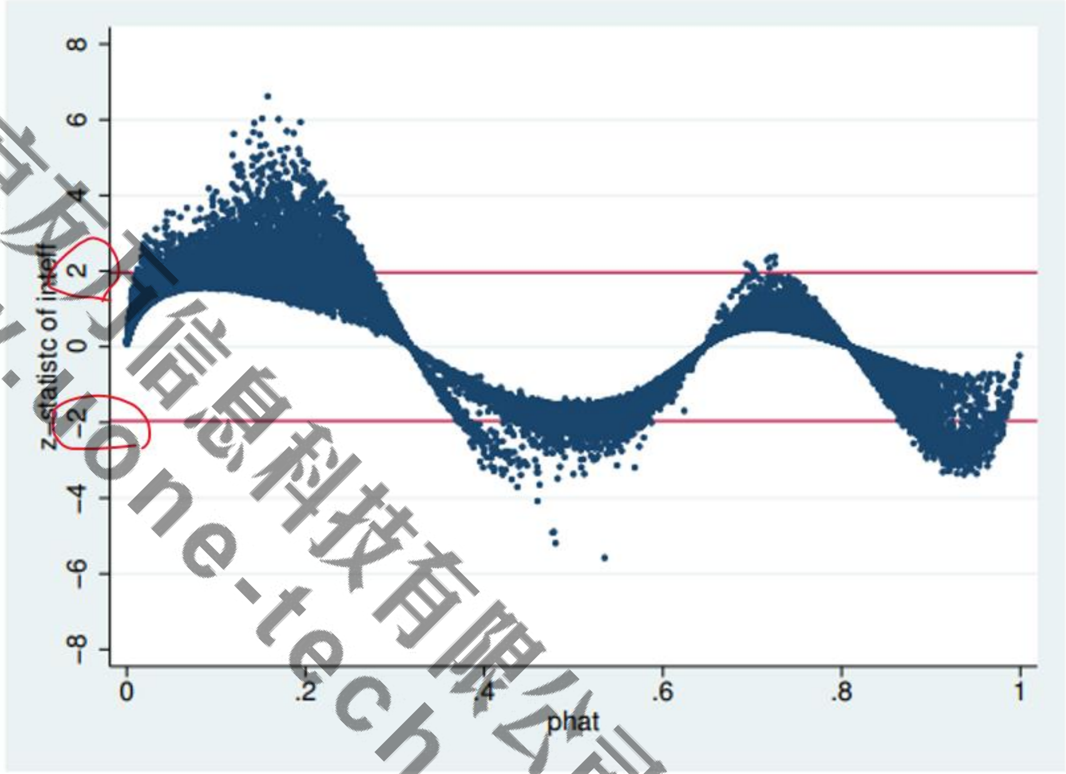

Figure 6: scatter of z-statistic and phat

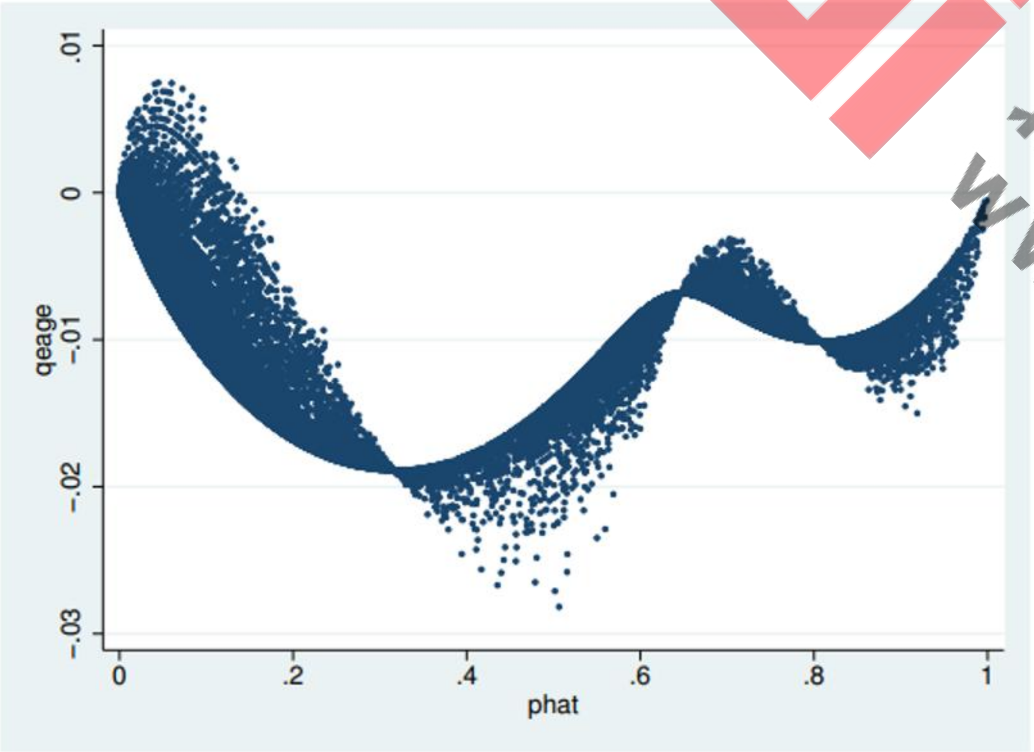

Figure 7: scatter of qeage and phat

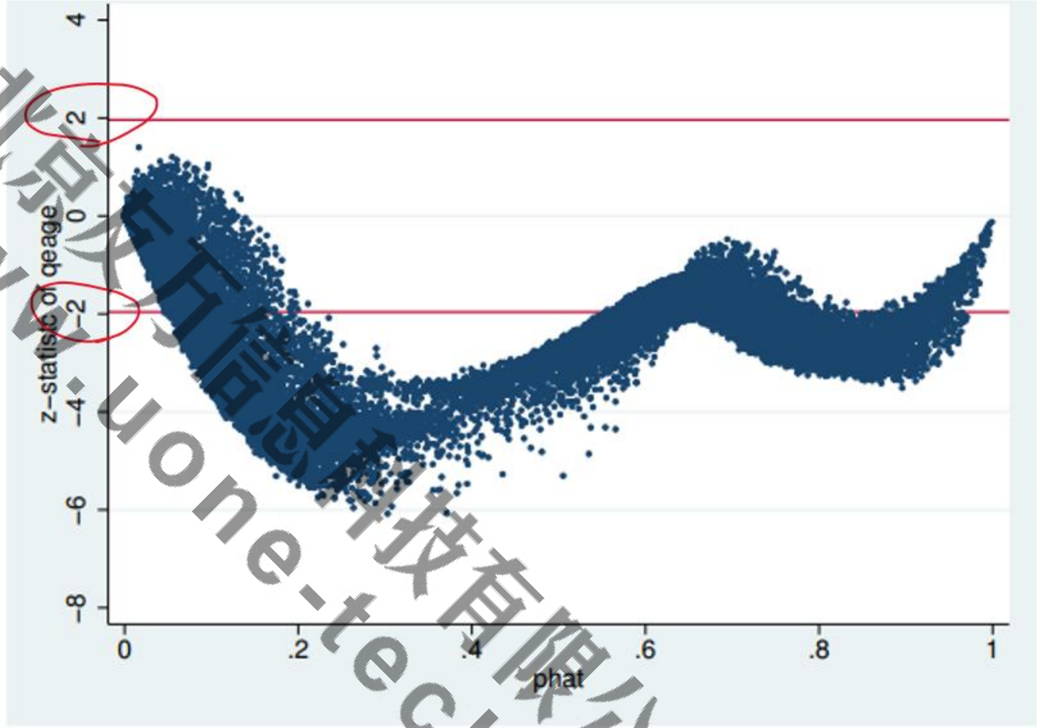

Figure 8: scatter of z-statistic of qeage and phat

Conclusion

- Provide a method to consistently estimate marginal, interaction and quadratic effects in endogenous probit models with interactive or quadratic terms.
- Our estimator performs well and better than Ai-Norton(2003)'s method and the interaction coefficient estimator in IV-probit.
- Develop a new Stata command, ***eivprobit***, to implement our method with much less time, especially for large dataset.
- An application shows the usefulness of the command.

中山大學嶺南學院
LINGNAN COLLEGE SUN YAT-SEN UNIVERSITY

Thanks!

Xianbo Zhou

作育英才 服务社会