The Treatment Effect, the Cross Difference, and the Interaction Term in Nonlinear "Difference-in-Differences" Models

Patrick A. Puhani

DISCUSSION PAPER SERIES

The Treatment Effect, the Cross Difference, and the Interaction Term in Nonlinear "Difference-in-Differences" Models

Patrick A. Puhani
*Leibniz University of Hannover, SIAW, University of St. Gallen,
ERMES, Université Paris II and IZA*

ABSTRACT

The Treatment Effect, the Cross Difference, and the Interaction Term in Nonlinear "Difference-in-Differences" Models[*]

I demonstrate that Ai and Norton's (2003) point about cross differences is not relevant for the estimation of the treatment effect in nonlinear "difference-in-differences" models such as probit, logit or tobit, because the cross difference is not equal to the treatment effect, which is the parameter of interest. In a nonlinear "difference-in-differences" model, the treatment effect is the cross difference of the conditional expectation of the *observed* outcome minus the cross difference of the conditional expectation of the *potential* outcome without treatment. Unlike in the linear model, the latter cross difference is not zero in the nonlinear model. It follows that the sign of the treatment effect in a nonlinear "difference-in-differences" model with a strictly monotonic transformation function is equal to the sign of the coefficient of the interaction term of the time and treatment group indicators. The treatment effect is simply the incremental effect of the coefficient of the interaction term.

Corresponding author:

Patrick A. Puhani
Leibniz University of Hannover
Institut für Arbeitsökonomik
Königsworther Platz 1
D-30167 Hannover
Germany
E-mail: puhani@aoek.uni-hannover.de

1   Introduction

Difference-in-differences estimation is one of the most important identification strategies in applied economics (Meyer, 1995; Angrist and Krueger, 1999; Heckman, LaLonde and Smith, 1999; Bertrand, Duflo and Mullainathan, 2004; Athey and Imbens, 2006). Especially when using micro data, the dependent variable may be binary or censored. One possibility to address this issue is to transfer the difference-in-differences identification strategy to the latent variable in nonlinear models like probit, logit or tobit. In a widely cited paper, Ai and Norton (2003) suggest that "most applied researchers misinterpret the coefficient of the interaction term in nonlinear models", for example in the context of difference-in-difference estimation.[1] In this note, I show that while the remark made by Ai and Norton (2003) on the cross difference (or derivative) in a nonlinear model such as probit is correct, the cross difference does not represent the treatment effect and is thus not an interesting parameter in a nonlinear "difference-in-differences" model. Instead, it is correct to focus on the coefficient of the interaction term.

2   Difference-in-Differences in the Linear and the Nonlinear Model

2.1   *Difference-in-Differences in the Linear Model*

I start by presenting the commonly known linear difference-in-differences model for a continuous and uncensored outcome. Define the treatment effect on the treated at the time of treatment by

$$\tau\left(T=1,G=1\right)=E\left[Y^1\middle|T=1,G=1,X\right]-E\left[Y^0\middle|T=1,G=1,X\right], \qquad (1)$$

where Y^1 and Y^0 are potential outcomes with and without treatment, respectively. T and G are binary time period and group indicators, which are coded as either 0 or 1. X is a vector

---

[1] 272 citations in "Google Scholar" as of April 26th, 2008.

of control variables. $E[\cdot]$ is the expectation operator. Note that the treatment effect is assumed to be constant over X. Participation in treatment is indicated by

$$I = 1[T = 1, G = 1] = T \times G,$$ (2)

where $1[\cdot]$ is the indicator function. Equation (2) implies that group $G = 1$ is treated in period $T = 1$. There is no treatment for group $G = 0$ in either period and no treatment for either group in period $T = 0$. The observational rule for the outcome variable Y is

$$Y = I \times Y^1 + (1 - I) \times Y^0.$$ (3)

In the linear model, the expected potential outcome Y^0 conditional on T, G, and X is specified as

$$E\left[Y^0 \middle| T, G, X\right] = \alpha T + \beta G + X\theta.$$ (4)

This restriction implies that the time effect α is constant for both groups and that the group difference β is constant across time. If Y^0 is not bounded, this restriction, which is the identifying assumption of the standard difference-in-differences model, may hold.

As specified, treatment occurs for group $G = 1$ in time period $T = 1$. Hence for the treated group the potential outcome under treatment

$$E\left[Y^1 \middle| T = 1, G = 1, X\right]$$

is observed during the treatment period. By equations (1) and (4), it can be written as

$$E\left[Y^1 \middle| T = 1, G = 1, X\right] = \alpha + \beta + \tau(T = 1, G = 1) + X\theta.^2$$ (5)

Note that the counterfactual outcome

$$E\left[Y^0 \middle| T = 1, G = 1, X\right]$$

is unobserved but modeled by equation (4):

$$E\left[Y^0 \middle| T=1, G=1, X\right] = \alpha + \beta + X\theta. \tag{4*}$$

Because (1) implies that

$$E\left[Y^1 \middle| T=1, G=1, X\right] = \tau(T=1, G=1) + E\left[Y^0 \middle| T=1, G=1, X\right] \tag{1*}$$

and because of the observational rule (3), the treatment allocation rule (2), and the functional

form restriction (4), the treatment effect on the treated at the time of treatment

$\tau = \tau(T=1, G=1)$ is identified by an interaction effect in a linear model:

$$\begin{aligned} E\left[Y \middle| T, G, X\right] &= TG \times \left[\tau + \alpha T + \beta G + X\theta\right] + (1 - TG) \times \left[\alpha T + \beta G + X\theta\right] \\ &= \alpha T + \beta G + \tau TG + X\theta. \end{aligned} \tag{6}$$

2.2 *"Difference-in-Differences" in the Nonlinear Model*

Now consider a simple nonlinear model, such as probit. Model the linear index as before, so

that the conditional expectation of the binary potential outcome Y^0 becomes

$$E\left[Y^0 \middle| T, G, X\right] = \Phi(\alpha T + \beta G + X\theta), \tag{7}$$

where $\Phi(\cdot)$ is the conditional distribution function of the standard normal distribution.

In addition, let equations (2) and (3) hold. Equation (1) does not hold, because the treatment

effect *cannot* be constant across the treated population if the expectation of the outcome

variable is bounded, as it is on probit, logit or tobit models (cf. Athey and Imbens, 2006,

p.451ff.). Instead, we implicitly define the treatment effect on the treated at the time of

treatment by specifying

$$E\left[Y^1 \middle| T=1, G=1, X\right] = \Phi(\alpha + \beta + \gamma + X\theta).^3 \tag{8}$$

---

[2] The treatment effect of the untreated, as well as treatment effects at time $T = 0$ would also be identified if we
made the additional assumption $E\left[Y^1 \middle| T, G, X\right] = \tau + \alpha T + \beta G + X\theta$.

[3] Analogously to the linear model, the treatment effect of the untreated, as well as treatment effects at time
$T = 0$ would also be identified if we made the more restrictive assumption
$E\left[Y^1 \middle| T, G, X\right] = \Phi(\alpha T + \beta G + \gamma + X\theta)$.

Combining (2), (3), (7), and (8), we obtain the standard probit "difference-in-differences model":

$$E\left[Y\middle|T,G,X\right]=\Phi\left(\alpha T+\beta G+\gamma TG+X\theta\right). \tag{9}$$

The treatment effect on the treated at the time of treatment conditional on X in this "difference-in-differences" probit model is

$$\begin{aligned}\tau\left(T=1,G=1,X\right)&=E\left[Y^{1}\middle|T=1,G=1,X\right]-E\left[Y^{0}\middle|T=1,G=1,X\right]\\&=\Phi\left(\alpha+\beta+\gamma+X\theta\right)-\Phi\left(\alpha+\beta+X\theta\right).\end{aligned} \tag{10}$$

Hence, the treatment effect is zero if and only if the coefficient γ of the interaction term TG is zero. In addition, because $\Phi(\cdot)$ is a strictly monotonic function, the sign of γ is equal to the sign of the treatment effect. The treatment effect is simply the incremental effect of the coefficient of the interaction term γ. The cross difference, as discussed in Ai and Norton (2003), is not relevant for estimating the treatment effect in the probit "difference-in-differences" model or in any other nonlinear "difference-in-differences" model with a structure like (9). This is because the treatment effect is not the cross difference, but the difference between two cross differences: it is the cross difference of the conditional expectation of the *observed* outcome Y minus the cross difference of the conditional expectation of the *counterfactual* outcome Y^{0} (see equation (10)* below). In the rest of this section, I will illustrate this point.

Unlike in the linear model, α in equation (7) is not a time effect that is constant across groups and β is not a group difference that is constant across time. It is still true, however, that α and β – together with the nonlinear transformation rule $\Phi(\cdot)$ – implicitly define the time and group effects, respectively. These effects are:

$$\frac{\Delta E\left[Y^{0}\middle|T,G,X\right]}{\Delta T}=\Phi\left(\alpha+\beta G+X\theta\right)-\Phi\left(\beta G+X\theta\right), \tag{11}$$

$$\frac{\Delta E\left[Y^0 \middle| T,G,X\right]}{\Delta G} = \Phi(\alpha T + \beta + X\theta) - \Phi(\alpha T + X\theta). \tag{12}$$

Hence, the time effect (11) varies by group and by X (both times in a parametrically restricted way). However, it is driven by α in the sense that there would be no time effect if α were zero. An analogous result holds for the difference between the two groups (12), which is driven by β.

What is the difference-in-differences in the expected value of the potential outcome Y^0 between groups and time? In the linear model, this difference-in-differences in the potential outcome Y^0 (i.e. the cross difference of equation (4)) is zero. Indeed, this is the identifying assumption that guarantees that the cross difference in the conditional expectation of the *observed* outcome Y is equal to the treatment effect in the linear model. However, neither of the last two statements holds in the nonlinear "difference-in-differences" model. As Ai and Norton (2003) point out, the cross difference in the probit model *without* interaction term is generally nonzero:[4]

$$\frac{\Delta^2 E\left[Y^0 \middle| T,G,X\right]}{\Delta T \Delta G} = \left[\Phi(\alpha + \beta + X\theta) - \Phi(\beta + X\theta)\right] - \left[\Phi(\alpha + X\theta) - \Phi(X\theta)\right] \neq 0. \tag{13}$$

The fact that time and group differences (and the cross differences) in the conditional expectation of the potential outcome Y^0 are not constant is a natural implication of the nonlinearity of the probit model. Unlike in the linear model, identification is not provided by the assumption that the cross difference is zero for the expected potential outcome Y^0, but by a nonlinear parametric restriction on that cross difference. Although this makes the identification assumption in the nonlinear "difference-in-differences" models less intuitive,

---

[4] The authors provide an example with continuous regressors in their paper, but the idea is analogous.

this assumption guarantees that all expected outcomes (factual or counterfactual) are bounded as required (in the example of the probit or logit model, bounded by zero and one).

To see this, reconsider equation (4)*, where time and group effects are constant. Nothing restricts the counterfactual in equation (4)* to be bounded by zero and one. However, this requirement is satisfied in equation (7).

Next, consider the cross difference of equation (9), which is

$$\frac{\Delta^2 E\left[Y|T,G,X\right]}{\Delta T \Delta G} = \left[\Phi(\alpha+\beta+\gamma+X\theta) - \Phi(\beta+X\theta)\right] - \left[\Phi(\alpha+X\theta) - \Phi(X\theta)\right].$$

(14)

This cross difference is different from equation (10) and thus does not represent a treatment effect. Why? Because the treatment effect in the probit "difference-in-differences" model is equal to the difference in two cross differences:

$$\tau(T=1, G=1, X)$$

$$= \frac{\Delta^2 E\left[Y|T,G,X\right]}{\Delta T \Delta G} - \frac{\Delta^2 E\left[Y^0|T,G,X\right]}{\Delta T \Delta G}$$

$$= \left[\Phi(\alpha+\beta+\gamma+X\theta) - \Phi(\beta+X\theta)\right] - \left[\Phi(\alpha+X\theta) - \Phi(X\theta)\right]$$
$$\quad - \left[\Phi(\alpha+\beta+X\theta) - \Phi(\beta+X\theta)\right] + \left[\Phi(\alpha+X\theta) - \Phi(X\theta)\right]$$

$$= \Phi(\alpha+\beta+\gamma+X\theta) - \Phi(\alpha+\beta+X\theta).$$

(10)*

This result applies to all nonlinear models of this type of parametric structure, as $\Phi(\cdot)$ could be any nonlinear (strictly monotonic) transformation function.

3   Conclusions

A widely cited paper by Ai and Norton (2003) provides a correct derivation for the cross difference (and derivative, in the continuous case) in a nonlinear model like probit. In this

note, I show that this cross difference is not equal to the treatment effect and thus not an interesting parameter in a nonlinear "difference-in-differences" model. I further demonstrate that the sign of the treatment effect in a nonlinear "difference-in-differences" model with a strictly monotonic transformation function of a linear index (like probit, logit or tobit) is equal to the sign of the coefficient of the interaction term. The sign of the cross-difference is irrelevant. Hence, researchers carrying out "difference-in-differences" estimates in nonlinear models like probit, logit or tobit are correct to focus their attention on the coefficient of the interaction term of the group and time dummy.

To gauge the statistical significance of an estimate of the treatment effect in a finite sample, applying the delta method to the nonlinear transformation of the estimated coefficients is appropriate. However, this question of inference is different from the question of identification. Only the latter is the focus of this note.

References

Ai, D. and E.C. Norton (2003): Interaction Terms in Logit and Probit Models, *Economics Letters* 80: 123-129.

Angrist, J.D. and A. Krueger (1999): Empirical Strategies in Labor Economics, in O. Ashenfelter and D. Card (eds.): *Handbook of Labor Economics*, *Vol. 3A*, Amsterdam: Elsevier, 1277-1366.

Athey, S. and G. Imbens (2006): Identification and Inference in Nonlinear Difference-in-Differences Models, *Econometrica* 74: 431-497.

Bertrand, M., E. Duflo und S. Mullainathan (2004): How Much Should We Trust Differences-in-Differences Estimates?, *Quarterly Journal of Economics* 119: 249-275.

Heckman, J.J. R.J. LaLonde and J.A. Smith (1999): The Economics and Econometrics of Active Labor Market Programs, in O. Ashenfelter and D. Card (eds.): *Handbook of Labor Economics*, *Vol. 3A*, Amsterdam: Elsevier, 1865-2097.

Meyer, B. (1995): Natural and Quasi-Experiments in Economics, *Journal of Business and Economic Statistics* 13: 151-161.