

# **AULA 17**

# **MÉTODO DE DIFERENÇA**

# **EM DIFERENÇAS**

**Ernesto F. L. Amaral**

**09 de maio de 2013**

**Técnicas Avançadas de Avaliação de Políticas Públicas (DCP 098)**

**Fonte:**

**Curso “Técnicas Econométricas para Avaliação de Impacto” do “International Policy Centre for Inclusive Growth” (IPC-IG) da “United Nations Development Programme” (UNDP) (<http://www.ipc-undp.org/evaluation>).**

## EXPERIMENTO NATURAL

- Em economia, muitas pesquisas são feitas analisando os chamados experimentos naturais (quase-experimentos).
- Os experimentos naturais ocorrem quando algum evento exógeno, como, por exemplo, uma mudança de política do governo muda o ambiente no qual indivíduos, famílias, instituições ou municípios operam.
- Para analisarmos um experimento natural sempre temos que ter: (1) um grupo de controle que não foi afetado pela mudança; e (2) um grupo de tratamento que foi afetado pelo evento.
- Estes grupos têm características semelhantes, com exceção do recebimento da política.

## DADOS NECESSÁRIOS

- Ao contrário de um experimento real (ou, simplesmente, experimento) em que os grupos de tratamento e controle são escolhidos aleatoriamente para impedir viés nas estimativas, os grupos em um experimento natural (quase-experimento) emergem da forma com que a mudança é efetuada.
- Para estudarmos as diferenças entre os dois grupos precisamos de dados de antes e depois do evento para os dois grupos.
- Assim, nossa amostra é dividida em quatro grupos: (1) grupo de controle antes da mudança; (2) grupo de controle depois da mudança; (3) grupo de tratamento antes da mudança; e (4) grupo de tratamento depois da mudança.

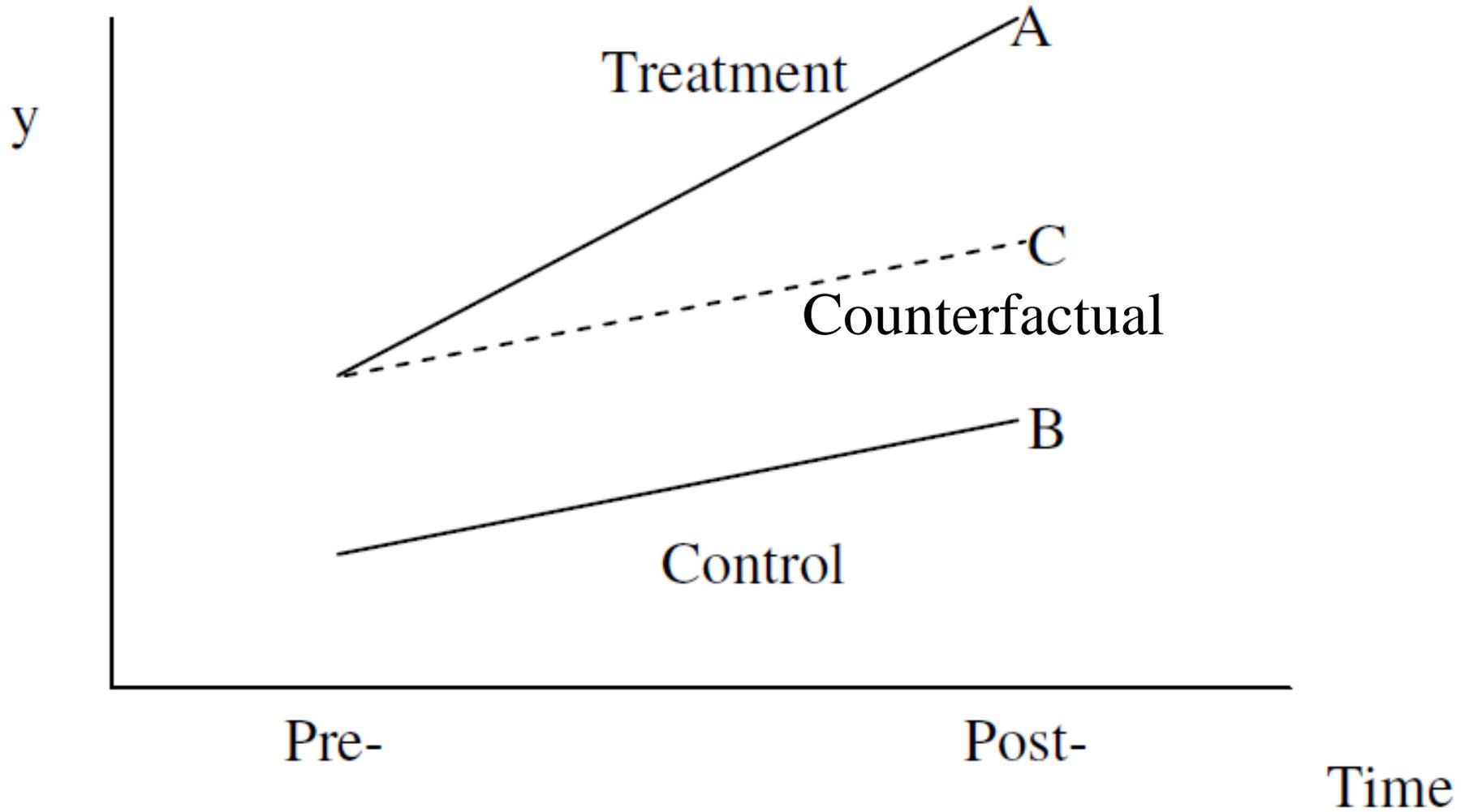
# AVALIAÇÃO DE IMPACTO

- “Diferença em diferenças” ou “dupla diferença” (DD) estima:
- 1) Diferença dentro de cada grupo (tratamento e controle).
  - 2) Diferença dessas duas médias.

<b>GRUPO</b>	<b>ANTES</b>	<b>POLÍTICA</b>	<b>DEPOIS</b>	<b>DIFERENÇAS</b>
<b>Tratamento</b>	$T_0$	<b>X</b>	$T_1$	$T_1 - T_0$
<b>Controle</b>	$C_0$		$C_1$	$C_1 - C_0$
<b>Diferenças</b>	$T_0 - C_0$		$T_1 - C_1$	$(T_1 - T_0) - (C_1 - C_0)$ ou $(T_1 - C_1) - (T_0 - C_0)$

- Testamos hipóteses com base na literatura de pesquisa.

# DIFERENÇA ENTRE GRUPOS NO TEMPO



## ESTIMADOR ANTES X DEPOIS

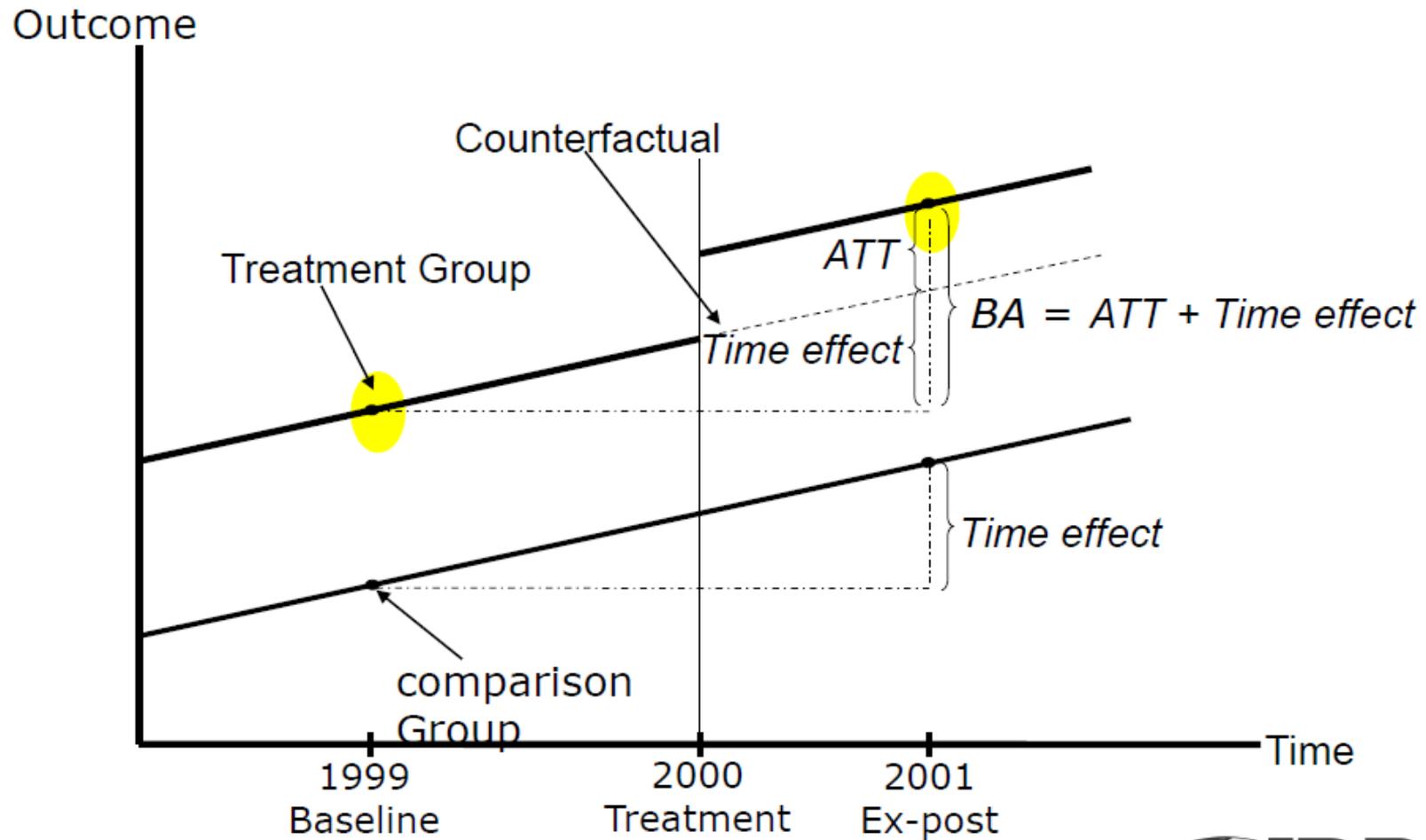
- $T_1 - T_0$  e  $C_1 - C_0$  representam em que medida o grupo de tratamento e o de controle se alteraram, respectivamente, entre o período anterior e posterior ao evento que está sendo examinado.
- Como o grupo de controle não sofreu impacto do evento, essas mudanças ocorreram por fatores que também devem ter influenciado o grupo de tratamento.
- Equação de regressão MQO para estimador antes-depois, considerando somente grupo de tratamento:

$$Y_i = \alpha_1 + \delta_1 t_i + \varepsilon_i$$

- Este estimador não será correto se existir uma tendência no tempo, independente do efeito da política (seria preciso retirar o efeito de tempo de  $\delta_1$ ).

# BEFORE-AFTER (ANTES X DEPOIS) = IMPACTO + TEMPO <sup>7</sup>

## Graphical intuition: BA estimator

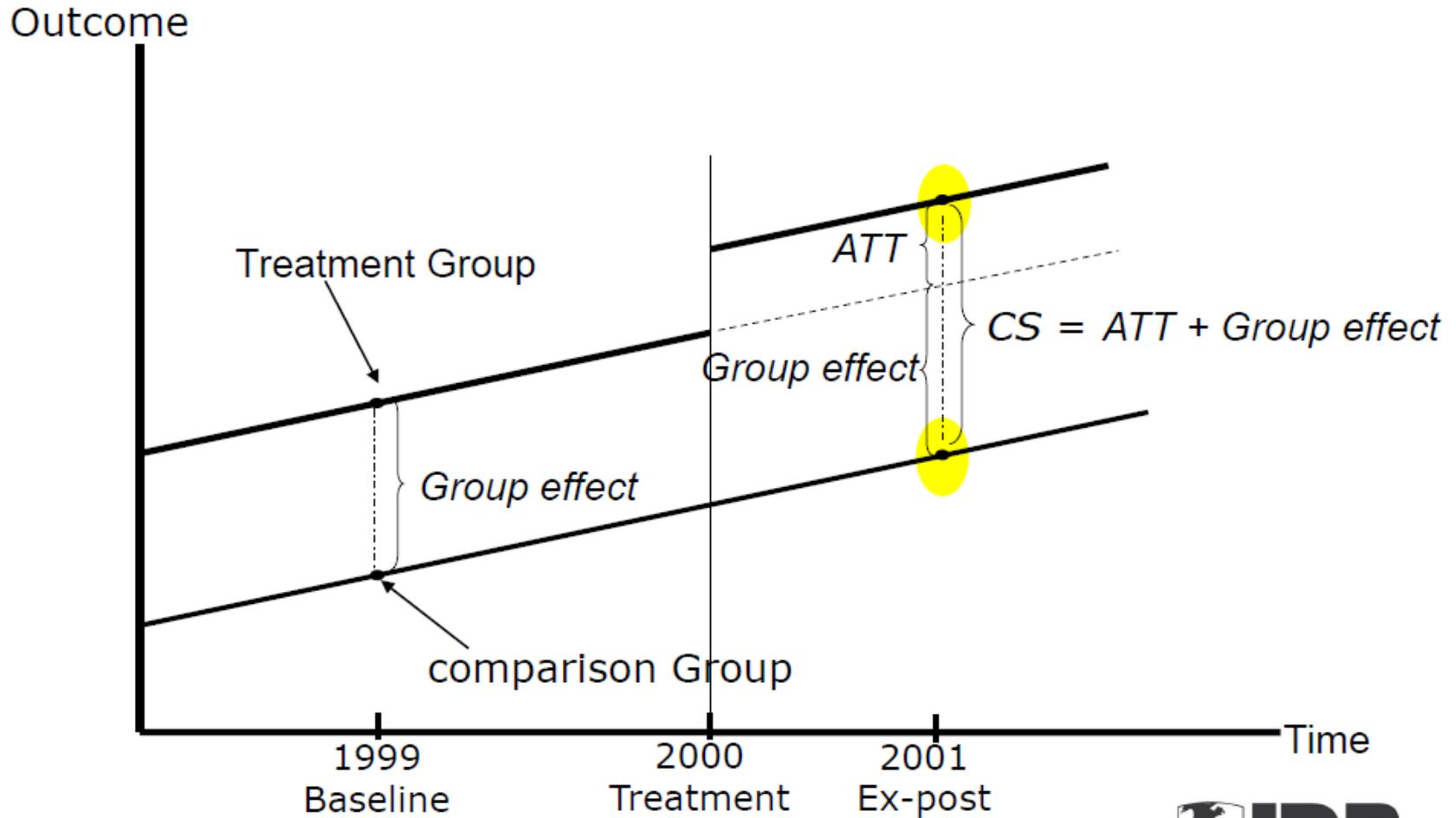


## ESTIMADOR CONTROLE X TRATAMENTO

- $T_0 - C_0$  e  $T_1 - C_1$  representam as diferenças entre os grupos de tratamento e de controle antes e depois do evento, respectivamente.
- Equação de regressão MQO para estimador controle-tratamento, considerando somente momento após a política:
$$Y_i = \alpha_2 + \delta_2 T_i + \varepsilon_i$$
- Este estimador não será correto se existir uma diferença permanente entre grupos, independente do efeito da política (seria preciso retirar o efeito de grupo de  $\delta_2$ ).

# CROSS-SECTION (C X T) = IMPACTO + GRUPO

## Graphical intuition: CS estimator

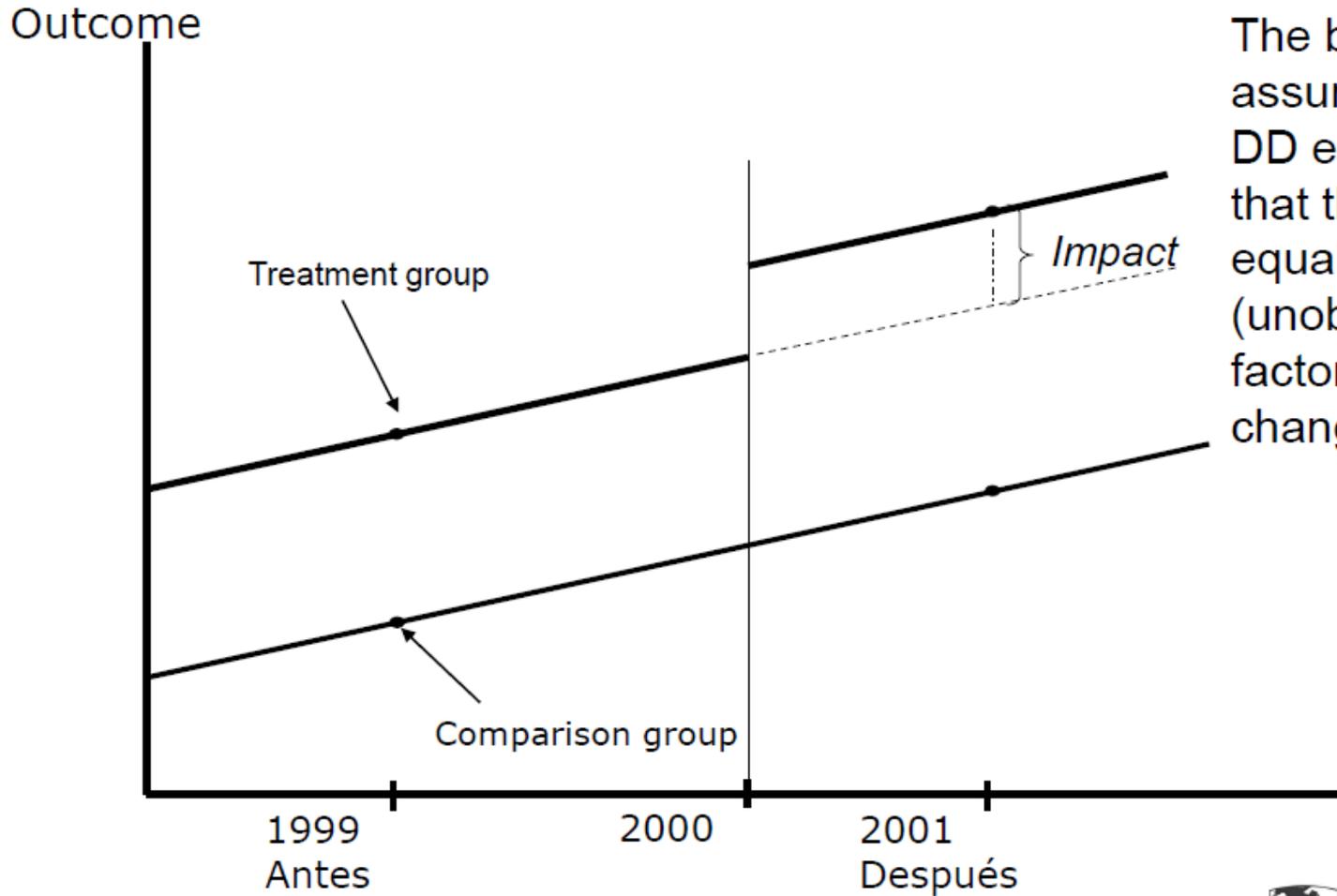


## ESTIMADOR DE DIFERENÇA EM DIFERENÇAS

- Calculando  $(T_1 - T_0) - (C_1 - C_0)$ , temos a diferença verificada entre as diferenças entre os dois períodos em cada um dos grupos.
- Isto é o mesmo que  $(T_1 - C_1) - (T_0 - C_0)$ , o que é a diferença verificada entre as diferenças entre os dois grupos em cada um dos períodos.
- Esta é a origem do nome diferença em diferenças (*diff-in-diff*).

# ESTIMADOR DD = IMPACTO

## Graphical intuition: DD estimator



The basic assumption of the DD estimator is that the trends are equals (unobservable factors do not change over time)

# EXPRESSÃO MATEMÁTICA

- Matematicamente, podemos representar o método de diferenças em diferenças com a seguinte equação:

$$DD = (y_{T1} - y_{C1}) - (y_{T0} - y_{C0})$$

- Cada “y” representa a média da variável estudada para cada ano e grupo.
- A letra representa o grupo ao qual o dado pertence (C para o grupo de controle; e T para o grupo de tratamento).
- O número representa o período da amostra (0 para antes da mudança; e 1 para depois da mudança).
- “DD” é a estimativa a partir da diferença em diferenças, a qual determina o impacto do experimento natural sobre a variável que gostaríamos de explicar.

# EQUAÇÃO DE REGRESSÃO MQO PARA ESTIMADOR DD

- Representando o método através de uma regressão e criando as variáveis indicadoras (ou *dummies*).
- $T_i$  igual a um para os indivíduos do grupo de tratamento e zero para os indivíduos do grupo de controle.
- $t_i$  igual a um quando os dados se referem ao momento pós-política e zero para o período pré-política.
- Temos:

$$Y_i = \alpha + \beta T_i + \gamma t_i + \delta(T_i * t_i) + \varepsilon_i$$

## COMPONENTES DA REGRESSÃO

$$Y_i = \alpha + \beta T_i + \gamma t_i + \delta(T_i * t_i) + \varepsilon_i$$

- $Y$ : representa a variável estudada.
- $\alpha$ : valor esperado da variável estudada quando se analisa o grupo de controle antes da mudança (termo constante).
- $\beta$ : impacto do grupo de tratamento na variável estudada (diferença permanente entre controle e tratamento).
- $\gamma$ : impacto do segundo período sobre a variável estudada (tendência no tempo comum entre controle e tratamento).
- $\delta$ : impacto pós-evento do grupo de tratamento, em relação ao grupo de controle, sobre a variável estudada (efeito verdadeiro do tratamento).
- $\varepsilon_i$ : erro aleatório não observado (variáveis independentes não incluídas).

## VALORES ESPERADOS

- Os valores esperados (preditos) médios da variável dependente, para cada grupo e tempo, são dados por:

$$Y_i = \alpha + \beta T_i + \gamma t_i + \delta(T_i * t_i) + \varepsilon_i$$

$$E[Y_{T0}] = \alpha + \beta$$

$$E[Y_{T1}] = \alpha + \beta + \gamma + \delta$$

$$E[Y_{C0}] = \alpha$$

$$E[Y_{C1}] = \alpha + \gamma$$

## É PRECISO CONTROLAR POR OUTROS FATORES

- É preciso controlar por outros fatores relevantes na regressão.
- Antes de dizer que  $\delta$  nos dará o impacto da política exógena, temos que descobrir e isolar o efeito de todas as outras variáveis que podem estar causando mudanças na variável estudada.
- Isso é feito inserindo variáveis de controle relevantes na regressão, evitando-se assim que efeitos de outras variáveis produzam viés na nossa estimação.
- Com esse procedimento, determinamos o efeito puro do experimento natural sobre a variável que gostaríamos de explicar, mantendo as demais variáveis constantes (efeito *ceteris paribus*).

## PRESSUPOSTOS PARA ESTIMADOR DD NÃO VIESADO

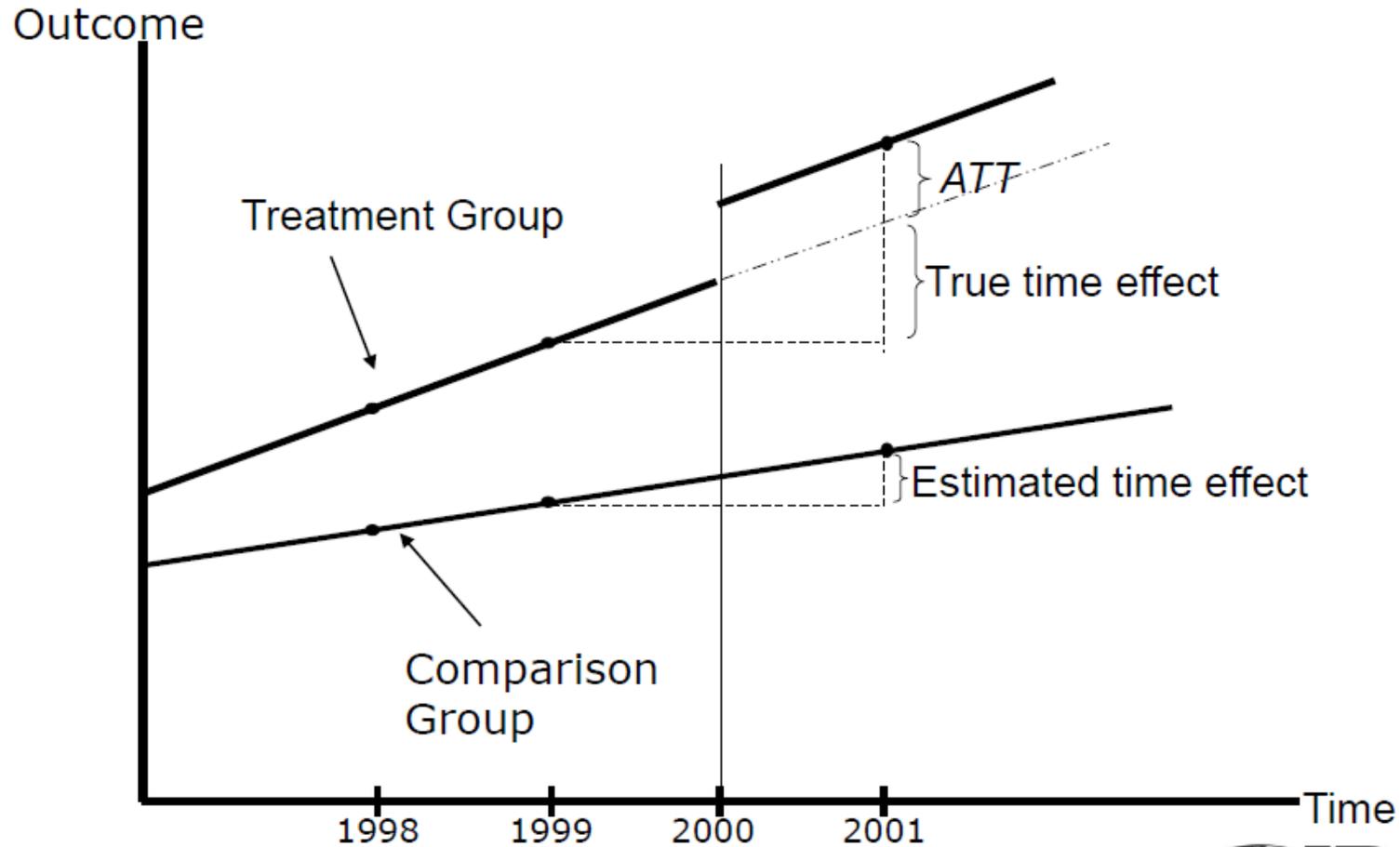
- Sob quais condições o estimador DD estima efetivamente o parâmetro de interesse? Quando o grupo C será um bom contrafactual para o grupo T?
  
- Os pressupostos para obter um estimador correto são:
  - 1) O **modelo de regressão** foi corretamente especificado, com inclusão de outras variáveis independentes de controle.
  - 2) O **termo de erro é em média igual a zero**:  $E[\varepsilon_i]=0$ . Esse não é um pressuposto forte, já que temos a inclusão do termo constante ( $\alpha$ ).
  - 3) O termo de erro não é correlacionado com outras variáveis na equação. Esse pressuposto, conhecido como pressuposto de **tendência paralela**, é o mais difícil de ser alcançado...

## PRESSUPOSTO DA TENDÊNCIA PARALELA

- O pressuposto-chave de tendências comuns é chamado de *parallel trend assumption*.
- Tempo deve ser igual entre grupos de tratamento e controle.
- Fatores não-observáveis específicos dos grupos devem ser constantes no tempo.
- Se grupos possuírem tendências diferentes no tempo ( $\gamma_C = \gamma$  e  $\gamma_T = \gamma + \Delta$ , por exemplo), o estimador DD será enviesado.
- Esse pressuposto não é testável, porém algumas evidências de sua validade podem ser observadas.
- Por exemplo, podemos observar a tendência da variável dependente em momentos diferentes, antes da política, nos grupos de controle e tratamento.

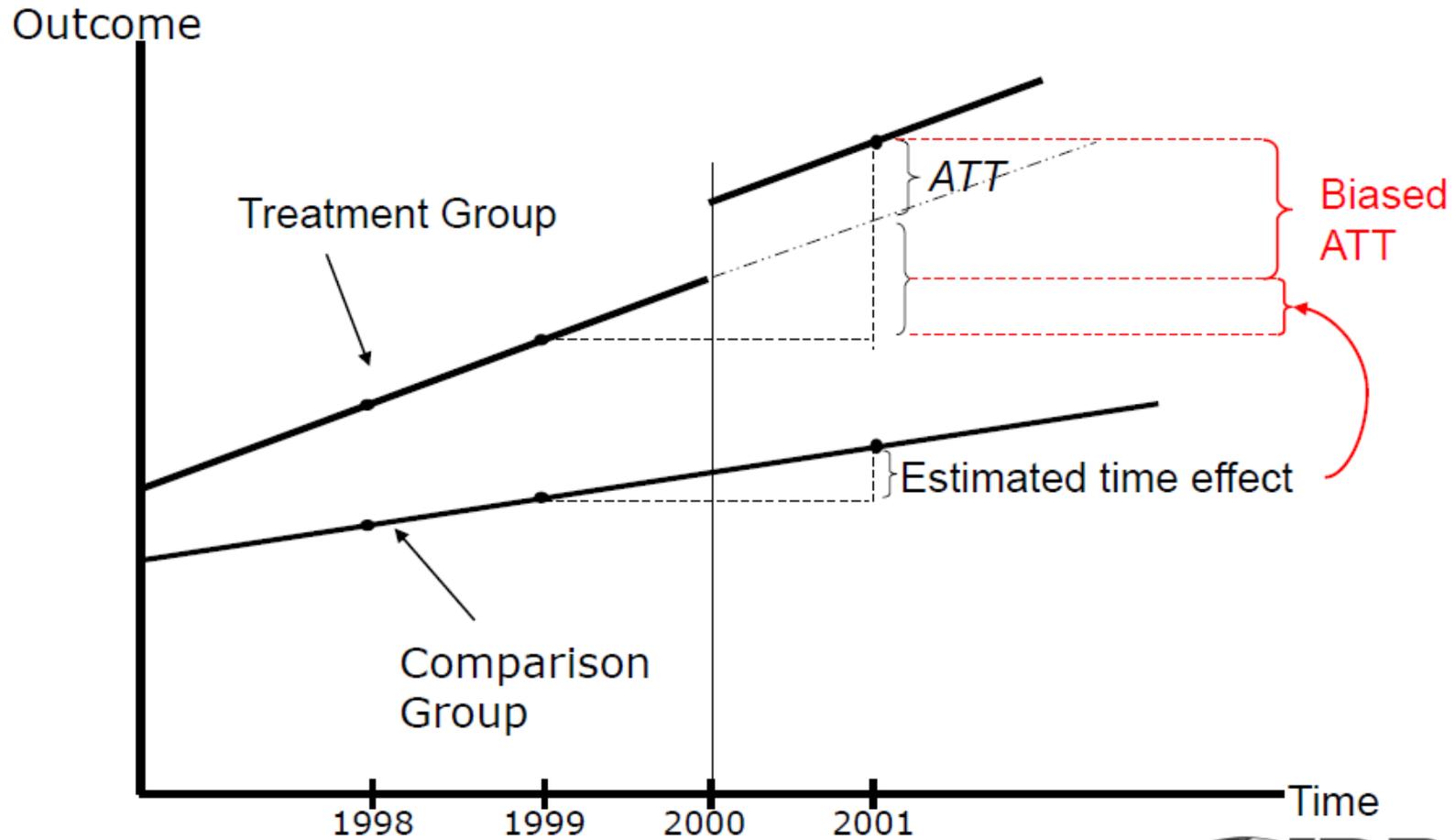
# FALHA DO PRESSUPOSTO

## Example: failure of assumption



# FALHA DO PRESSUPOSTO

## Example: failure of assumption



# **FORMATO DO BANCO DE DADOS**

# FORMATO DO BANCO DE DADOS NO STATA

- O comando “reshape” muda o formato dos bancos de dados, segundo a necessidade do estudo.

```

wide form (amplo)
      i          ..... x_ij .....
id  sex  inc80  inc81  inc82
-----
  1   0   5000   5500   6000
  2   1   2000   2200   3300

```

long form (dados em painel): usado pelo método de diferença em diferenças

```

      i      j          x_ij
id  year  sex  inc
-----
  1    80    0  5000
  1    81    0  5500
  1    82    0  6000
  2    80    1  2000
  2    81    1  2200
  2    82    1  3300

```

- Mudar de “wide” (amplo) para “long” (dados em painel):  
`reshape long inc, i(id) j(year)`
- Mudar de “long” (dados em painel) para “wide” (amplo):  
`reshape wide inc, i(id) j(year)`

## PRINCIPAIS REFERÊNCIAS

- Albouy, David. s.d. “Program Evaluation and the Difference in Difference Estimator”. Section Notes Economics 131.  
([http://emlab.berkeley.edu/users/webfac/saez/e131\\_s04/diff.pdf](http://emlab.berkeley.edu/users/webfac/saez/e131_s04/diff.pdf))
- FGV. s.d. “Metodologia de Diferenças-em-Diferenças”.  
([http://www.fgv.br/cps/pesquisas/Políticas\\_sociais\\_alunos/2009/BES\\_Diferen%C3%A7asemDiferen%C3%A7as.pdf](http://www.fgv.br/cps/pesquisas/Políticas_sociais_alunos/2009/BES_Diferen%C3%A7asemDiferen%C3%A7as.pdf))
- Maffioli, Alessandro. 2011. “Non-Experimental Methods: Propensity Score Matching and Difference in Difference”.  
IDB: Cuernavaca.  
(<http://www.impactevaluation2011.org/forum/wp-content/uploads/2011/06/C5-commissioners-of-evaluations-file-1.pdf>)