

# La méthode de la double différence

Introduction aux méthodes d'évaluation d'impact  
des politiques publiques

Philippe De Vreyer

Quand va-t-on employer cette  
méthode ?

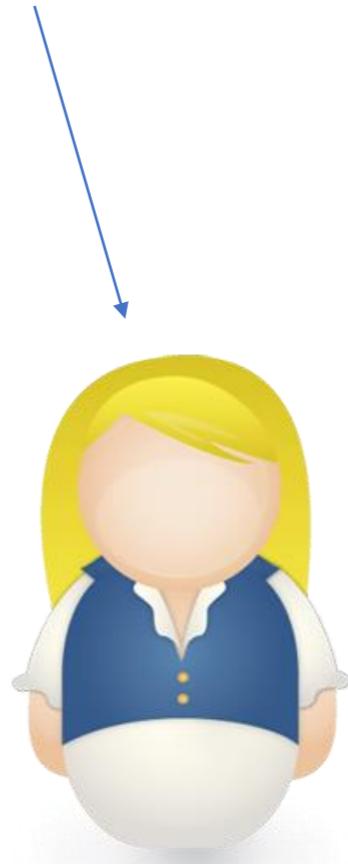
# Conditions d'emploi de la méthode

- Quand on suspecte la présence de caractères inobservés des individus qui expliquent à la fois la participation au programme et la valeur prise par la variable d'intérêt (indépendamment de la participation). On parle d'hétérogénéité inobservée.
- Quand cette hétérogénéité est *fixe* et *additive*.
- Quand on dispose *d'observations répétées* sur les participants et les non participants.

# Exemple: traitement destiné à favoriser la croissance des enfants

1<sup>ère</sup> méthode: comparaison après le traitement entre enfants traités et enfants non traités

Traité

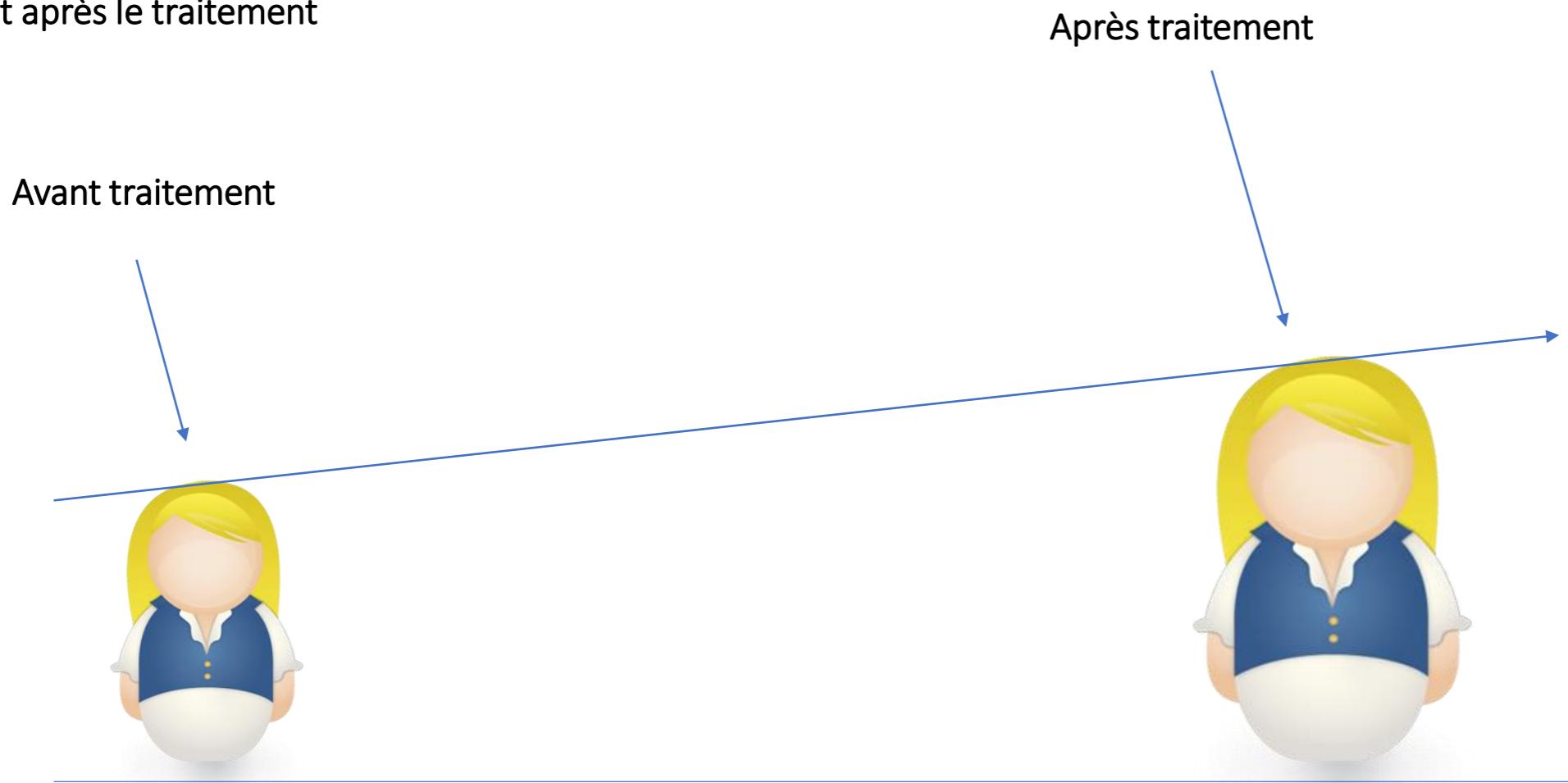


Non traité



# Exemple: traitement destiné à favoriser la croissance des enfants

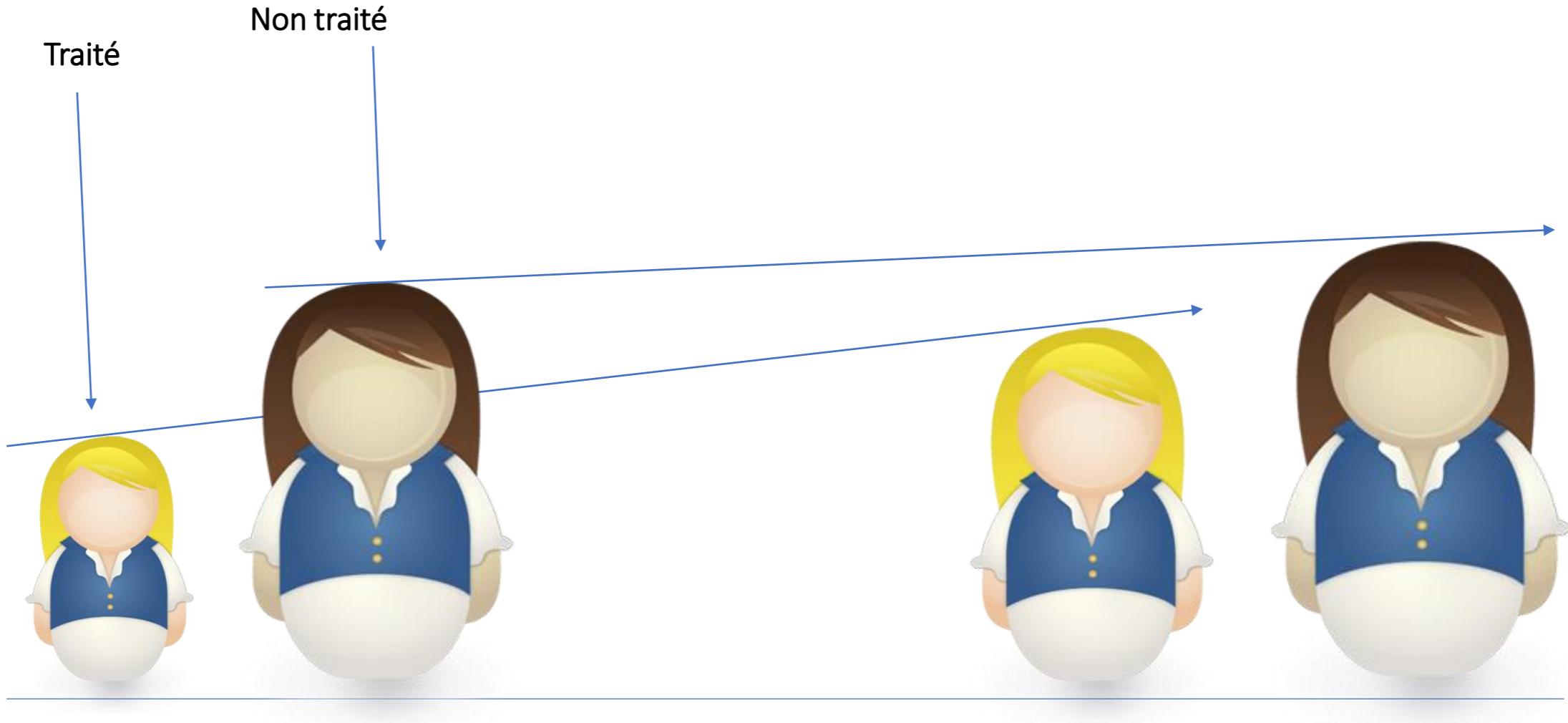
2<sup>ème</sup> méthode: comparaison des enfants traités avant et après le traitement



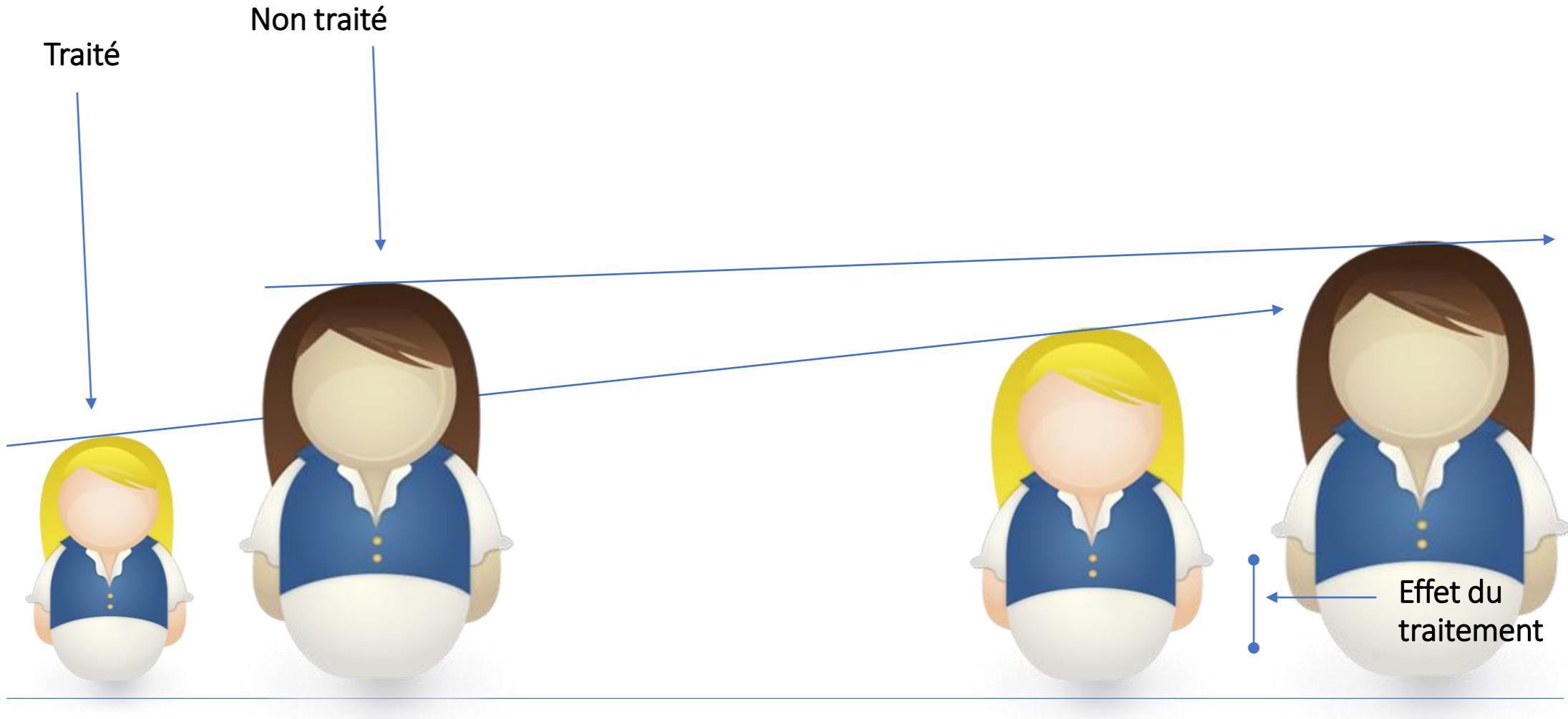
# Exemple: traitement destiné à favoriser la croissance des enfants

- Les deux méthodes conduisent à des conclusions opposées:
  - La comparaison des enfants traités avec les enfants non traités après le traitement néglige la possibilité que des différences existent entre enfants en l'absence du traitement
  - La comparaison des enfants traités avant et après le traitement néglige la possibilité que la variable d'intérêt (ici la taille) varie au cours du temps indépendamment du traitement
  - La bonne solution est de marier les deux méthodes: c'est ce que fait la double différence.

# Exemple: traitement destiné à favoriser la croissance des enfants



# Exemple: traitement destiné à favoriser la croissance des enfants



# Théorie

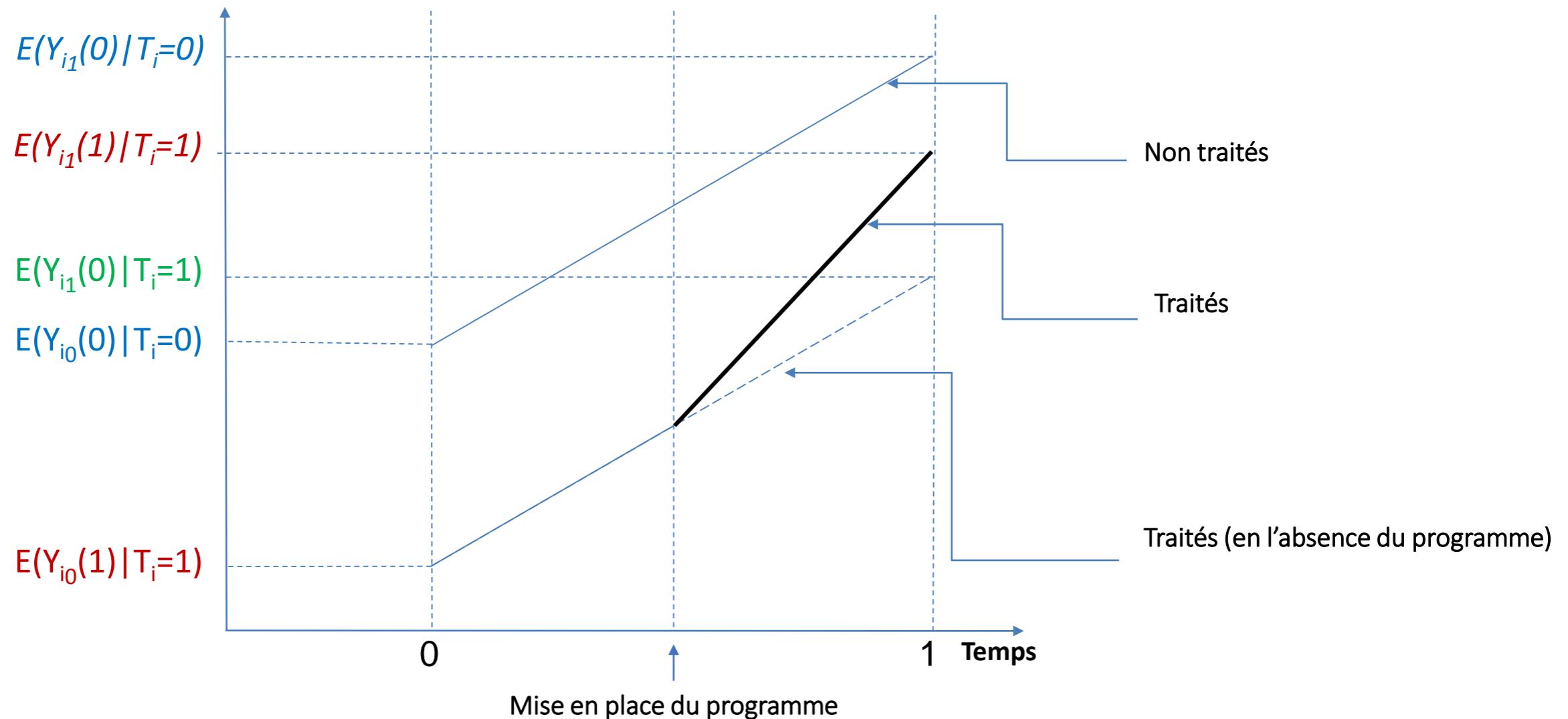
# Hypothèse principale

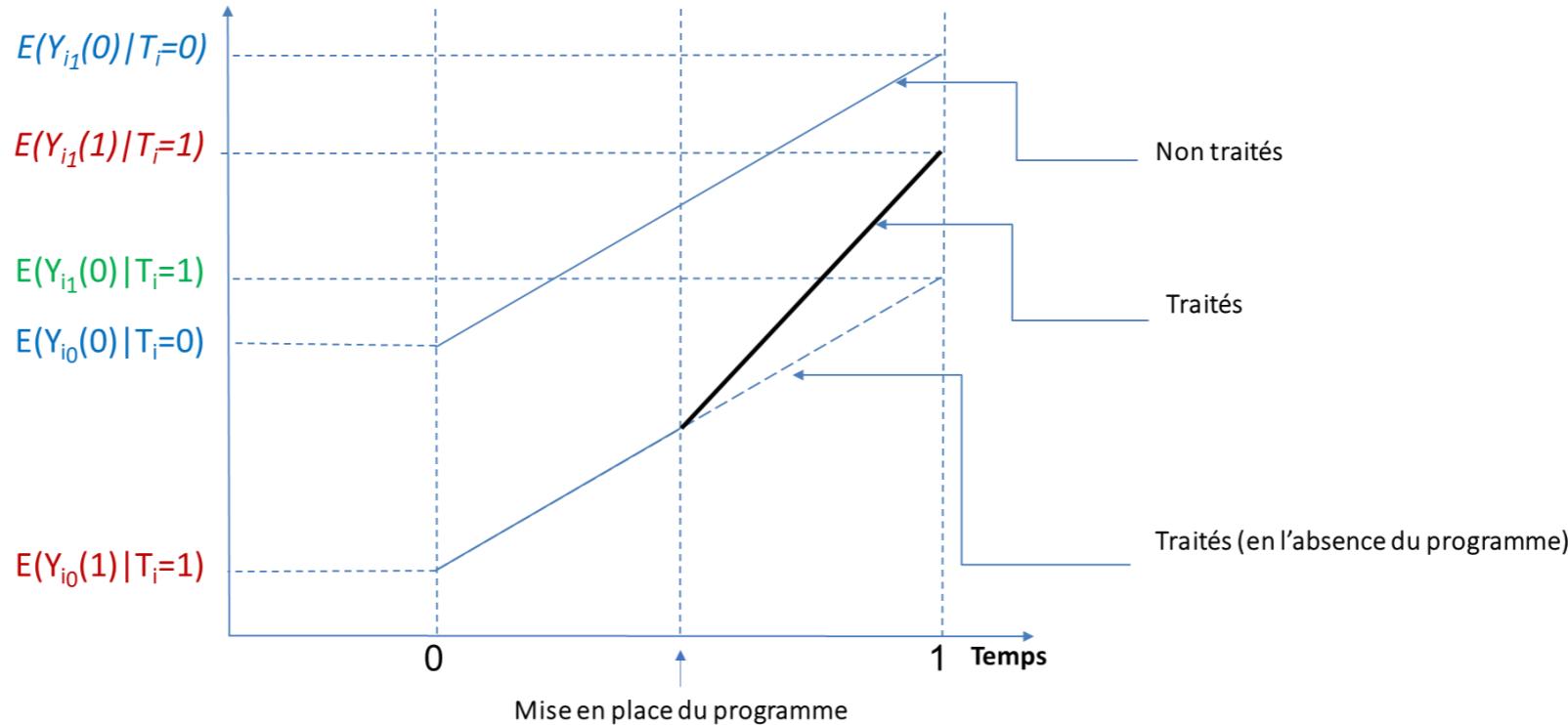
- La méthode repose sur l'hypothèse que les différences observées avant le traitement entre traités et non traités seraient restées identiques si les traités ne l'avaient pas été (hypothèse de « trends » parallèles).

# Formalisation

- Deux dates d'observation: 0 et 1.
- En  $t=0$  le programme n'existe pas.
- En  $t=1$  le programme est en place et a produit ses effets.
- On note  $Y_{it}(0)$  la valeur de la variable d'intérêt pour l'individu  $i$  s'il n'est pas traité à la date  $t$  et  $Y_{it}(1)$  la valeur si cet individu est traité.
- L'estimateur de la double différence s'écrit:

$$DD = E[(Y_{i1}(1) - Y_{i0}(1))|T_i = 1] - E[(Y_{i1}(0) - Y_{i0}(0))|T_i = 0]$$





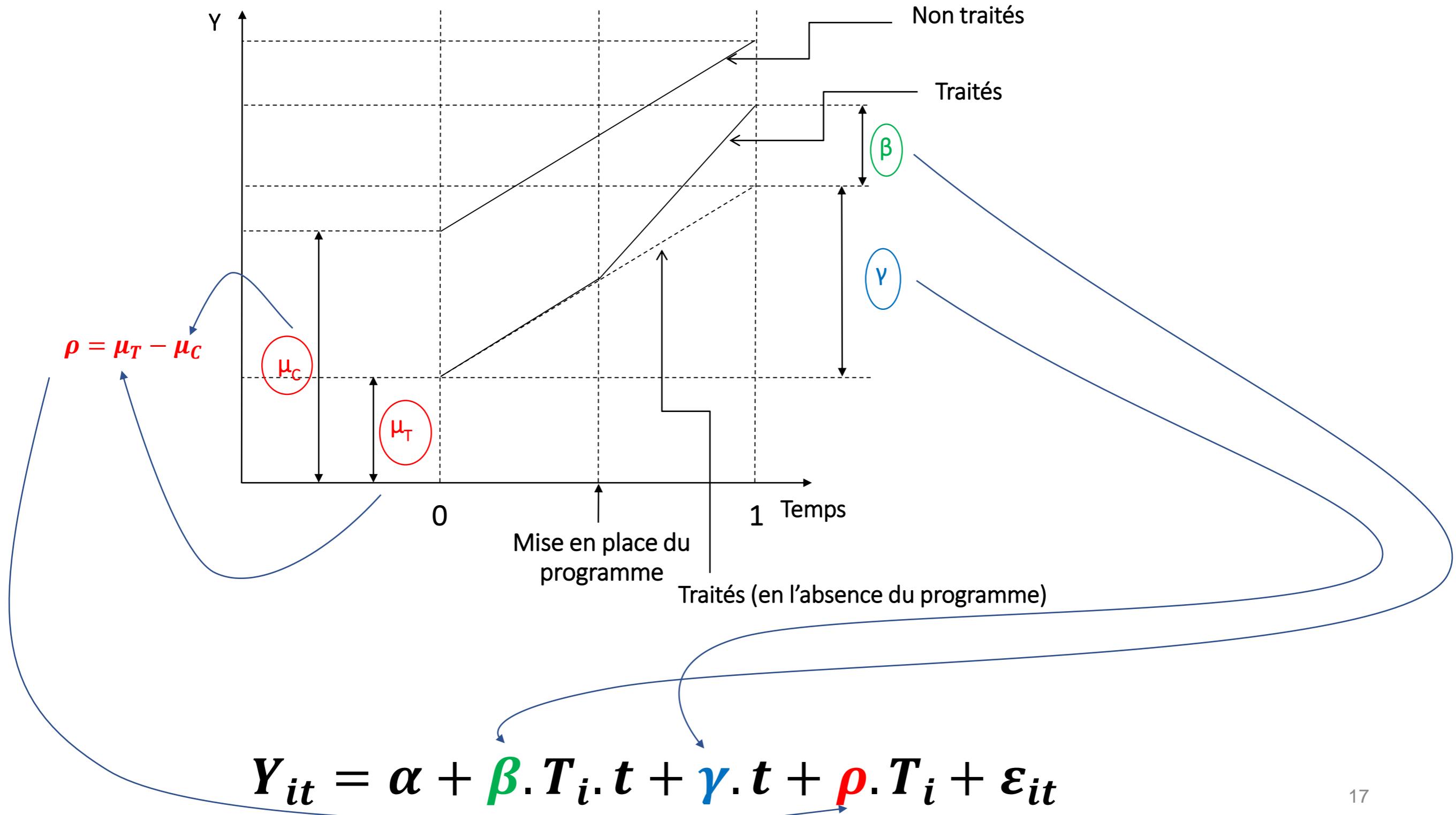
- Traités à la date 0 :  $E(Y_{i_0}(1) | T_i=1)$ , non traités à la date 0:  $E(Y_{i_0}(0) | T_i=0)$
  - Traités à la date 1 :  $E(Y_{i_1}(1) | T_i=1)$ , non traités à la date 1:  $E(Y_{i_1}(0) | T_i=0)$
  - Traités à la date 1, *s'ils n'avaient pas été traités* :  $E(Y_{i_1}(0) | T_i=1)$
  - Impact du programme :  $E(Y_{i_1}(1) | T_i=1) - E(Y_{i_1}(0) | T_i=1)$
  - Estimateur de l'impact:  $DD = E(Y_{i_1}(1) | T_i=1) - E(Y_{i_1}(0) | T_i=0) - (E(Y_{i_1}(0) | T_i=1) - E(Y_{i_1}(0) | T_i=0))$
  - Sous l'hypothèse de trends parallèles ceci s'écrit :
- $$DD = E(Y_{i_1}(1) | T_i=1) - E(Y_{i_1}(0) | T_i=0) - (E(Y_{i_0}(1) | T_i=1) - E(Y_{i_0}(0) | T_i=0))$$

# Mise en œuvre pratique

- **Méthode 1:** Emploi des moyennes calculées sur les échantillons des individus traités et non traités aux deux dates d'observation.
- **Méthode 2:** Emploi d'une régression linéaire. Le modèle à estimer est alors de la forme:

$$Y_{it} = \alpha + \beta \cdot T_i \cdot t + \gamma \cdot t + \rho \cdot T_i + \varepsilon_{it}$$

- Le coefficient du terme d'interaction entre le temps,  $t$ , et la variable de traitement  $T_i$ ,  $\beta$ , est l'estimateur de la double différence DD.
- Les variables  $t$  et  $T_i$  sont incluses séparément pour tenir compte d'un effet potentiel du temps qui passe (trend) et d'un effet provenant du fait d'être, ou non, inclus dans l'échantillon test (a priori non nul si le tirage des échantillons n'est pas aléatoire).



- Il est facile de montrer que les estimations des coefficients de cette régression conduisent au résultat attendu:

$$Y_{it} = \alpha + \beta \cdot T_i \cdot t + \gamma \cdot t + \rho \cdot T_i + \varepsilon_{it}$$

$$DD = [E(Y_{i1} - Y_{i0} | T_i = 1)] - [E(Y_{i1} - Y_{i0} | T_i = 0)]$$

$$\begin{aligned} &= [\alpha + \beta * 1 * 1 + \gamma * 1 + \rho * 1 - \alpha - \beta * 1 * 0 - \gamma * 0 - \rho * 1] \\ &- [\alpha + \beta * 0 * 1 + \gamma * 1 + \rho * 0 - \alpha - \beta * 0 * 0 - \gamma * 0 - \rho * 0] \end{aligned}$$

$$= [\alpha + \beta + \gamma + \rho - \alpha - \rho] - [\alpha + \gamma - \alpha]$$

$$= [\beta + \gamma] - [\gamma]$$

$$= \beta$$

$$Y_{it} = \alpha + \beta \cdot T_i \cdot t + \gamma \cdot t + \rho \cdot T_i + \varepsilon_{it}$$

- Simple différence entre groupes:

$$\begin{aligned} E(Y_{i1}|T_i = 1) - E(Y_{i1}|T_i = 0) &= \alpha + \beta * 1 * 1 + \gamma * 1 + \rho * 1 - (\alpha + \beta * 0 * 1 + \gamma * 1 - \rho * 0) \\ &= \beta + \gamma + \rho - \gamma \\ &= \beta + \rho \end{aligned}$$

$\rho$  = biais = différence fixe entre les valeurs moyennes de Y pour les groupes traités et non traités en l'absence de tout traitement.

$$Y_{it} = \alpha + \beta \cdot T_i \cdot t + \gamma \cdot t + \rho \cdot T_i + \varepsilon_{it}$$

- Simple différence intertemporelle:

$$E(Y_{i1} - Y_{i0} | T_i = 1) = \alpha + \beta + \gamma + \rho - \alpha - \rho = \beta + \gamma$$

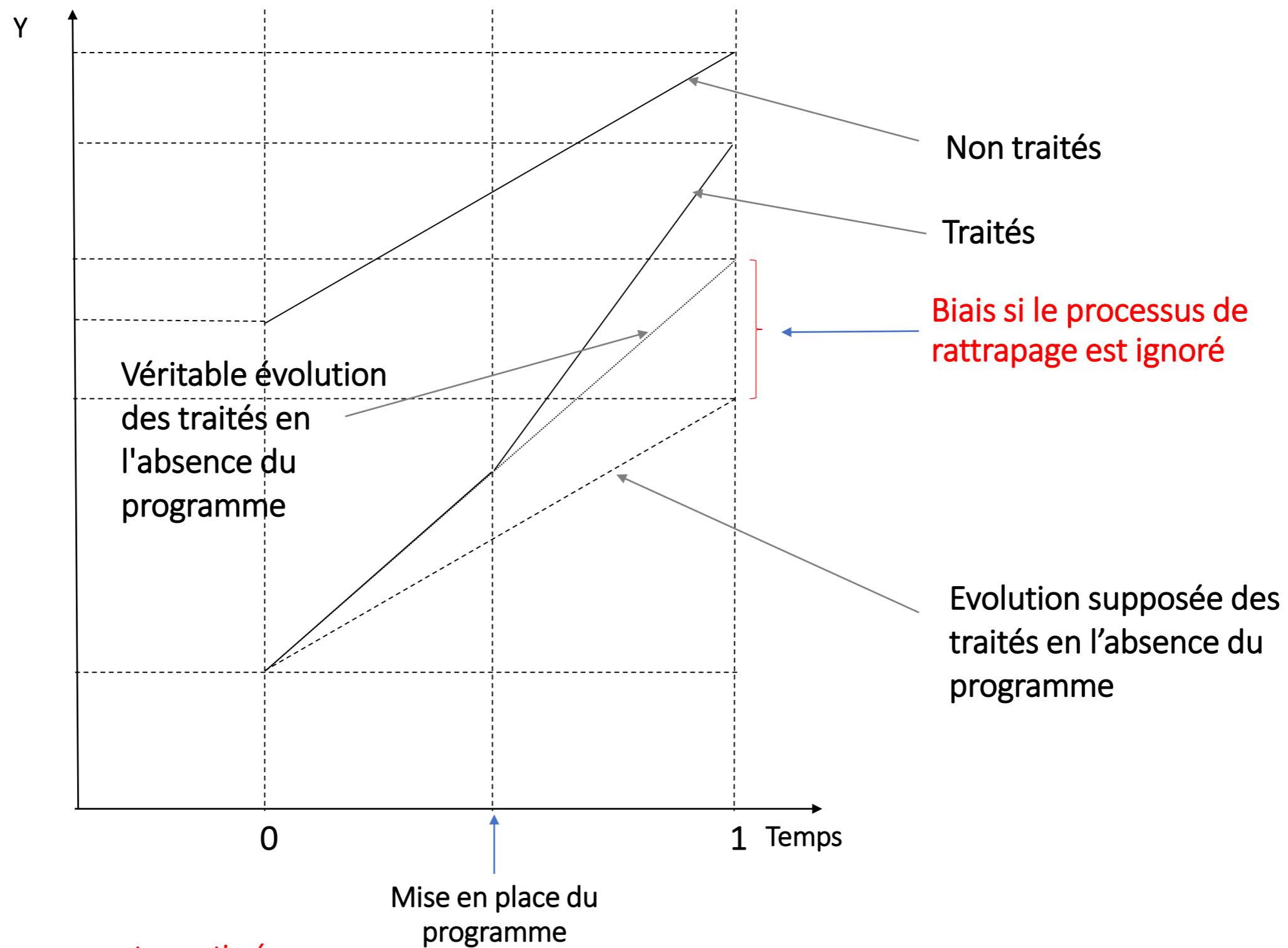
$\gamma$  = biais = mesure de l'évolution de la valeur moyenne de Y en l'absence de tout traitement.

# Limites

Validité de l'hypothèse des trends parallèles ?

- Evaluation d'un programme ciblé sur certaines régions dont l'objectif est de favoriser la croissance des revenus de leurs habitants.
- Les régions ciblées peuvent l'avoir été (ou s'être autosélectionnées) parce qu'elles ont connu une baisse de leurs revenus (suite à un choc comme par exemple une catastrophe naturelle) juste avant l'introduction du programme.

- Ces régions, comparées à d'autres n'ayant pas connu le choc, peuvent alors connaître un processus de rattrapage, qui va entraîner une hausse de leurs revenus plus importante et indépendante de leur participation au programme.
- Si cette hausse n'est pas prise en compte, l'estimateur DD conduira à une sur-estimation de l'efficacité de la politique (« Ashenfelter Dip »)



- Ainsi, l'emploi de l'estimateur DD ne dispense pas de s'assurer que les échantillons test et de contrôle présentent des caractéristiques identiques avant l'application du traitement.

# Comment faire mieux ?

# Combiner Matching et Double Différence pour réduire les sources de biais

- L'idée est d'employer la méthode du score de propension pour renforcer la comparabilité des échantillons test et de contrôle avant d'appliquer la double différence.
- On dispose d'au moins deux années d'observation.
- Sur l'année de base, on emploie la méthode du score de propension pour appariер les traités avec les non traités qui ont un score de propension proche.
- Ensuite on applique la méthode DD en calculant la différence de variation moyenne de la variable d'intérêt entre l'échantillon des traités et celui des individus qui leurs sont appariés.

L'estimateur de l'impact du programme a alors pour expression:

$$DD_{PSM} = \frac{1}{N_T} \sum_{i=1}^{N_T} \left[ Y_{i1}(1) - Y_{i0}(1) - \sum_{j \in M(i)} \omega_j \cdot (Y_{i1}(0) - Y_{i0}(0)) \right]$$

Traité

Non traité

Nombre d'individus traités

Date d'observation

Ensemble des non traités appariés à i

Pondération issue de la méthode PSM

# L'estimateur de la triple différence

- Supposons que l'on dispose de plus d'un groupe de contrôle.
- Deux possibilités:
  - On peut les utiliser successivement pour vérifier que l'on obtient le même résultat quelque soit le groupe de contrôle utilisé.
  - On peut combiner ces différents groupes pour obtenir un seul estimateur, tel que l'estimateur de la triple différence.

# Exemple

Evaluation d'un programme destiné à réinsérer les chômeurs âgés de plus de 50 ans

- Le programme est testé dans une région.
- Deux groupes de contrôle peuvent être envisagés :
  - Celui des chômeurs de plus de 50 ans habitant une autre région.
  - Celui des chômeurs de la même région, âgés de moins de 50 ans.
- Si les données le permettent, l'estimateur de la double différence peut être calculé en employant chacun de ces deux groupes de contrôle.

- Des estimations porteuses de biais spécifiques:
  - L'emploi du groupe des travailleurs âgés de + de 50 ans habitant une autre région peut conduire à un biais si les deux régions connaissent des dynamiques différentes.
  - L'emploi du groupe des travailleurs âgés de – de 50 ans habitant la même région peut conduire à un biais si les deux populations connaissent des dynamiques d'insertion différentes.
  - Dans les deux cas, l'hypothèse des trends parallèles sous-jacente à l'emploi de la double différence est invalidée.

- L'estimateur de la triple différence consiste à employer les deux groupes de contrôle pour éliminer les deux sources de biais, moyennant une hypothèse moins forte que celle qui sous-tend la double différence.
- V: chômeurs âgés de 50 ans ou plus
- J: chômeurs âgés de moins de 50 ans
- T: région où le programme est expérimenté
- C: région où il ne l'est pas.

$$\begin{aligned} DDD &= \left[ E \left( Y_{i1}^{V,T}(1) - Y_{i0}^{V,T}(1) \right) - E \left( Y_{i1}^{J,T}(0) - Y_{i0}^{J,T}(0) \right) \right] \\ &\quad - \left[ E \left( Y_{i1}^{V,C}(0) - Y_{i0}^{V,C}(0) \right) - E \left( Y_{i1}^{J,C}(0) - Y_{i0}^{J,C}(0) \right) \right] \end{aligned}$$

$$\begin{aligned}
DDD &= \left[ E(Y_{i1}^{V,T}(1) - Y_{i0}^{V,T}(1)) - E(Y_{i1}^{J,T}(0) - Y_{i0}^{J,T}(0)) \right] \\
&\quad - \left[ E(Y_{i1}^{V,C}(0) - Y_{i0}^{V,C}(0)) - E(Y_{i1}^{J,C}(0) - Y_{i0}^{J,C}(0)) \right]
\end{aligned}$$

Les **deux premiers termes** correspondent à l'estimateur de la double différence lorsque le groupe de chômeurs de moins de 50 ans est pris comme groupe de contrôle.

Il est possible que ce groupe connaisse une dynamique d'insertion différente de celle des plus de 50 ans. Sous l'hypothèse que cette différence de dynamique est identique d'une région à l'autre, on peut utiliser les données observées dans la région C pour évaluer cette différence et la retirer : c'est le rôle des **deux derniers termes** de l'équation.

$$\begin{aligned}
DDD &= \left[ E \left( Y_{i1}^{V,T}(1) - Y_{i0}^{V,T}(1) \right) \right. \\
&\quad \left. - E \left( Y_{i1}^{J,C}(0) - Y_{i0}^{J,C}(0) \right) \right]
\end{aligned}$$

$$\begin{aligned}
DDD &= \left[ E(Y_{i1}^{V,T}(1) - Y_{i0}^{V,T}(1)) - E(Y_{i1}^{V,C}(0) - Y_{i0}^{V,C}(0)) \right] \\
&\quad - \left[ E(Y_{i1}^{J,T}(0) - Y_{i0}^{J,T}(0)) - E(Y_{i1}^{J,C}(0) - Y_{i0}^{J,C}(0)) \right]
\end{aligned}$$

- Dans cette écriture, les deux premiers termes correspondent maintenant à *l'estimateur de la double différence lorsque le groupe de chômeurs de plus de 50 ans habitant l'autre région est pris comme groupe de contrôle.*
- Il est possible que cette région groupe connaisse une dynamique différente de celle de la région traitée. *Sous l'hypothèse que cette différence de dynamique est identique pour les jeunes chômeurs,* on peut utiliser les données observées sur le groupe des moins de 50 ans pour évaluer cette différence et la retirer : c'est le rôle des deux derniers termes de l'équation.