Applied Econometrics
R Project - Case 1 - Part 03

Kelian IDRISSOU and Ibrahim FOUSFOS

Table of Contents :

Impact Evaluation of Salesperson Meetings on Retirement Savings

Report

December 7, 2025

1 Introduction

1.1 Research Question

Understanding how financial guidance influences individual saving decisions is a central topic in applied econometrics and household finance. As financial products become increasingly complex, many clients rely on professional advice to make informed choices about retirement preparation, wealth accumulation, and long-term planning. Evaluating whether such interventions truly shape behaviour is therefore essential from both an academic and policy perspective.

In this context, our research question is: **What are the effects of meetings with a salesperson on retirement savings?**

The motivation for this question comes from the idea that financial advice may act as a behavioural nudge, improving financial literacy, reducing procrastination, and encouraging long-term planning.

According to economic theory, individuals are expected to adjust their saving behaviour in anticipation of future needs. The Life-cycle Hypothesis (Modigliani and Brumberg, 1954) argues that people save to smooth consumption across their lifetime, while the Permanent Income Hypothesis (Friedman, 1957) suggests that saving responds to long-term income expectations. Based on these foundations, we should expect that personalised financial guidance such as a salesperson meeting should increase both total savings and retirement contributions compared to clients who did not receive such an intervention.

1.2 Impact Evaluation Methodology and Justification

To address our research question, we apply the **Difference-in-Differences (DiD)** methodology to our panel dataset. This approach is the most appropriate given the structure of the data: the same individuals are observed repeatedly over time (panel data), and a clearly defined intervention (The salesperson meeting) occurs at period $t = 2$. This setting naturally creates a pre-treatment period ($t = 0$ and $t = 1$) and a post-treatment period ($t = 2, 3, 4$), as well as a treated group (clients who had a meeting) and a control group (those who did not).

Beyond the standard DiD estimator, our panel structure allows us to implement a within (fixed-effects) estimator, which exploits all within-individual variation across time. This method removes any time-invariant unobservable characteristics such as risk preferences, financial discipline, or saving attitudes that could bias the estimated impact of the meeting. The combination of DiD with individual fixed effects therefore strengthens causal identification.

The validity of this approach relies on standard assumptions:

- **Parallel trends:** In the absence of treatment, the saving behaviour of treated and control clients would have followed similar trends.

- **Time-invariant unobservables:** Individual characteristics that affect savings do not vary over time and are controlled for through the within estimator.
- **No anticipation:** Clients do not adjust their behaviour before $t = 2$ in anticipation of the meeting.

By leveraging both the time dimension and the panel structure of the data, our DiD with fixed effects allows us to estimate how savings evolve for each group and to attribute their differential to the causal effect of the meeting itself.

1.3 Why a simple OLS Model is not suitable for this topic

A simple post-treatment OLS regression, suffers from two major sources of bias that threaten causal validity.

First, there is a bias from **unobservable heterogeneity**: a basic OLS model does not control for individual fixed characteristics such as financial discipline, risk aversion, or saving preferences that may differ across clients despite the randomized assignment. These unobserved traits can correlate with both the meeting and savings outcomes, leading to biased estimates.

Second, simple OLS is unable to capture the **temporal dynamics** of savings behaviour. Savings may evolve over time due to macroeconomic trends or aggregate shocks (such as economic growth, inflation, policy changes) that affect all clients between $t = 0$ to $t = 4$. In such a case, OLS would incorrectly attribute these time-driven changes to the meeting, overstating or understating the true treatment effect.

Because of these limitations, a simple OLS Model can't reliably isolate the causal impact of the meeting, which motivates the use of more robust panel-based causal methods such as Difference-in-Difference.

1.4 Short summary of The findings

Metric	Finding
Causal effect (ATT)	The meeting increases total savings by 0,418 log points (= +51,9% with p<0,01) and the retirement savings by 3,793 log points (= + 4339% with p<0,01)
Heterogeneous Effects (Gender)	Women experience an additional + 0,104 log (p<0,01) points increase in savings after the intervention, indicating stronger responsiveness for female clients.
Underlying Trends	The post treatment trend is negative and significant (-0,071 and -0,093), meaning the meeting reverses a natural decline in savings that would have occurred without treatment
Income Elasticity	Income strongly predicts savings elasticity = 1,75, significant at 1%. Meaning that saving rise more than proportionally with income
Robustness Check	Placebo DiD (t=0 vs t=1) yields no significant effect (-0,001), confirming parallel trends and strengthening causal validity
Model Fit	R^2 Values (0,561 and 0,735) and extremely large F-Statistics confirm a well specified and Powerful model.

Table 1: Summary of Findings

2 Analysis and Results

2.1 Dataset and specification

Our empirical analysis relies on a balanced panel dataset containing 50,000 observations, corresponding to 10,000 unique clients observed over five consecutive periods ($t = 0$ to $t = 4$). Since all individuals remain in the sample for the entire duration of the study, the dataset is free from attrition bias.

The salesperson meeting (The policy intervention of interest) occurs exactly at period $t = 2$, which allows us to clearly separate a pre-treatment window ($t = 0$ and $t = 1$) from the post-treatment window ($t = 2, 3, 4$). Both treated and controlled clients are therefore observed before and after the intervention.

Random assignment further enhances the credibility of the analysis: pre-treatment balance checks on income, savings, retirement contributions and gender show no statistically significant differences between treated and control groups. This confirms the absence of selection bias on observables and supports the use of causal inference tools.

Given the panel structure of our data, where clients are observed before and after the intervention, our empirical strategy is based on a DiD Framework. This approach allows us to compare the change in savings for treated clients to the untreated clients across time.

To account for unobserved individual characteristics that remain constant over time, we estimate the model using an individual fixed effects (within) estimator. This ensures that all

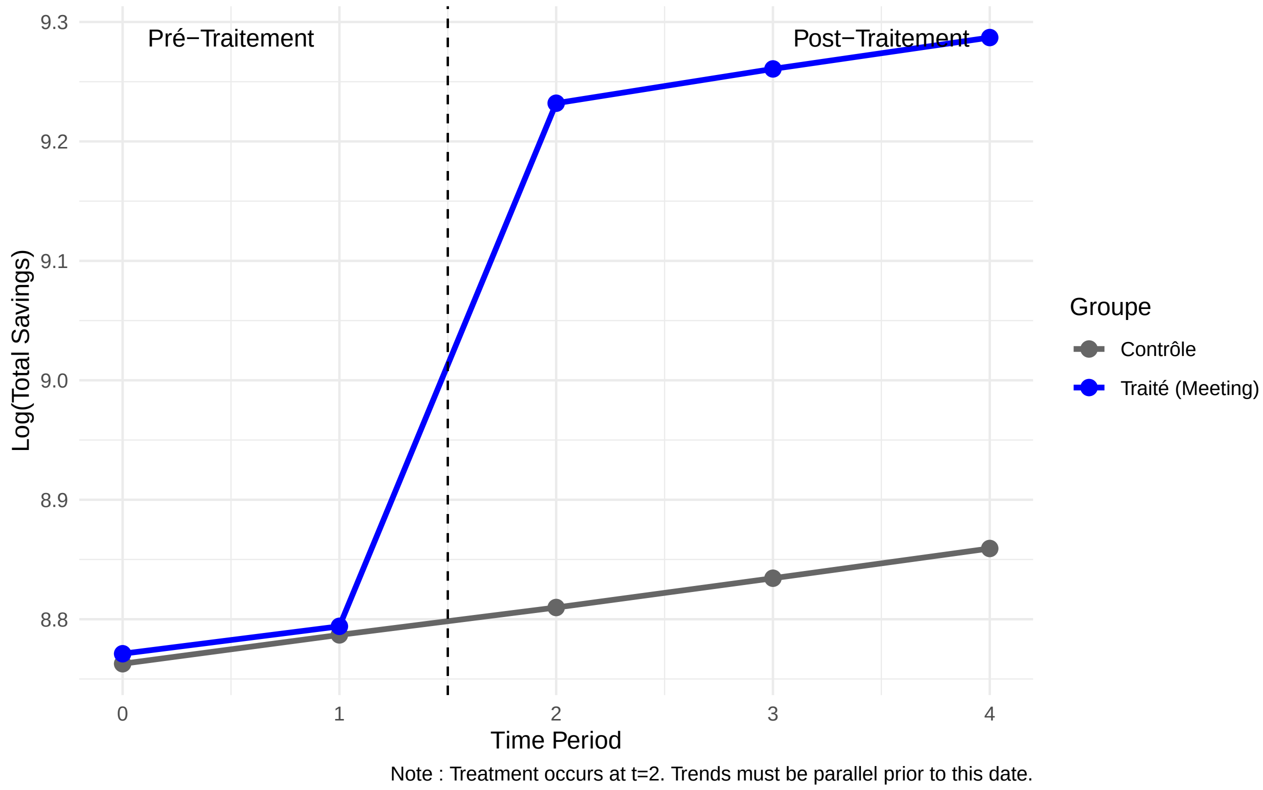

Figure 1: Parallel Trends Test: Evolution of Average Savings

time-invariant heterogeneity is removed from the estimation, leaving only within-individual variation to identify the treatment effect.

The key identifying assumption of the DiD framework is that, in absence of treatment, treated and controlled clients would have followed parallel trends over time. To assess the assumption, we compute average log-savings for both groups in the pre-treatment period and plot their evolution.

As shown in Figure 1, the trajectories of treated and control clients are nearly identical and parallel before $t = 2$, which strongly supports the validity of the parallel trends assumption. Following the intervention, the treated group displays a marked upward shift in savings, consistent with a potential treatment effect.

2.2 Main results

We estimate the following Difference-in-Differences model with individual fixed effects (Within estimator):

$$\log(\text{savings}_{it}) = \alpha_i + \lambda_t + \delta D_{it} + \beta \log(\text{Income}_{it}) + \epsilon_{it} \tag{1}$$

with:

- α_i: are individual fixed effects, capturing all time-invariant unobserved characteristics (risk aversion, innate thriftiness or long run saving attitudes).
- λ_t: are time fixed effects, absorbing macroeconomic shocks affecting all individuals simultaneously.

- D_{it}: is the DiD interaction term ($\text{Treated}_i \times \text{Post}_t$). This is equal to 1 for treated individuals in the post-treatment periods ($T \geq 2$). Its coefficient δ measures the average treatment effect on the treated (ATT).
- $\log(\text{Income}_{it})$: is included as a time-varying control to improve precision.

We estimate two separate within models: The first one uses log total savings as the dependent variable and the second one uses log retirement savings as the dependent variable. The key parameter is the coefficient on the DiD interaction term (δ), which captures the causal effect of the salesperson meeting on saving behaviour.

Table 2: Causal Effect of the Meeting on Savings (Within Estimator)

	Dependent variable:	
	log_savings (Total Savings) (1)	**log_retirement** (Retirement Savings) (2)
Interaction DiD (ATT)	0.418***	3.793***
	(0.003)	(0.015)
Post (Trend)	-0.071***	-0.093***
	(0.003)	(0.014)
Log(Income)	1.774***	1.748***
	(0.028)	(0.142)
Observations	50,000	50,000
R^2	0.561	0.735
Adjusted R^2	0.451	0.668
F Statistic (df = 3; 39997)	17,046.520***	36,918.850***

Note: *p<0.1; **p<0.05; ***p<0.01

2.2.1 Effect on total savings:

The first column shows that the DiD coefficient is: $\delta = 0.418$ ($p < 0.01$).

Since the dependent variable is in logs and the treatment indicator is a dummy, the coefficient can be interpreted as a semi-elasticity:

$$\exp(0.418) - 1 = 0.519$$

$\rightarrow$ The meeting increases total savings by approximately 51.9% among treated clients, all else being equal, holding constant individuals and time effects.

This is a significant effect suggesting that the intervention on time $t = 2$ had a strong impact on client's saving behaviour.

2.2.2 Effect on retirement savings

The second column isolates the effect on retirement savings only. The estimated DiD coefficient is: $\delta = 3.793$ ($p < 0.01$).

Transforming the log coefficient:

$$\exp(3.793) - 1 = 4339\%$$

$\rightarrow$ This means that retirement contributions increased by more than 43 fold among treated individuals, compared to the control group, all else being equal.

This extremely large effect is economically plausible because initial retirement savings were extremely low before the intervention. Even a moderate increase in contributions mechanically produces a very high percentage change. The meeting likely acted as a powerful behaviour trigger, encouraging clients to open or fund retirement plans they previously neglected.

Both models display a significant negative coefficient on the post-treatment period with $(\text{post} = -0.071, p < 0.1)$ for total savings and $(\text{post} = -0.093, p < 0.1)$ for retirement savings. This indicates that, even without treatment, savings would have declined naturally after $t = 2$ possibly due to macroeconomic conditions or seasonal fluctuations. The meeting not only increases savings but it counteracts an underlying downward trend, strengthening the causal interpretation.

The log income elasticity is at 1.774 for total savings and 1.748 for retirement savings (with $p < 0.01$). These values mean that a 1% increase in income raises savings by roughly 1.75%. Savings are highly income-elastic, confirming that including income as a time-varying control is essential to avoid omitted variable bias.

2.3 Discussion about findings

Table 3: Robustness Checks: Gender Heterogeneity and Placebo Test

	Dependent variable: *log_ savings*	
	Panel Linear (Gender Heterogeneity) (1)	**OLS** (Placebo Test t=0 vs t=1) (2)
Interaction DiD (ATT)	0.364***	
	(0.004)	
Placebo DiD (False Treatment)		-0.001
		(0.004)
Female		-0.001
		(0.005)
DiD × Female	0.104***	-0.061***
	(0.004)	(0.003)
Log(Income)	3.411***	2.761***
	(0.045)	(0.004)
Constant		-20.924***
		(0.042)
Observations	50,000	20,000
R^2	0.416	0.962
Adjusted R^2	0.270	0.962
F Statistic	9,500.181***	128,135.700***

Note: *p<0.1; **p<0.05; ***p<0.01. Col (1) includes individual Fixed Effects.

Heterogeneous effects: Gender Differences The column 1 of Table 2 investigates whether the treatment effect varies across gender by including the interaction term DiD × Female. The baseline DiD effect remains positive and highly significant: $(\delta = 0.364, p < 0.01)$ and the interaction coefficient is: DiD × Female = 0.104 ($p < 0.01$).

This means that female clients experience an additional 10.4 percentage point increase in log savings compared to male clients after treatment. Economically, this suggests that women respond more strongly to the meeting intervention. This may reflect differences in financial behaviour, risk preferences or baseline savings gaps between men and women. The meeting may therefore have acted as an empowerment mechanism, helping female clients to adjust their savings upward more strongly.

Mechanisms at play:

1. **Financial advice acts as a behavioural activation trigger:** This strong DiD effect on retirement savings and the significant response among women imply that the meeting helps individuals overcome inertia, procrastination or lack of financial literacy. Clients may update their beliefs about long term savings, increase awareness of retirement planning, or simply take action after being guided by a financial advisor.

2. **Savings react strongly to income dynamics:** The log income coefficient is large and significant (around 1.77), indicating an elasticity above 1; this suggests that saving behaviour is highly sensitive to changes in income, consistent with life-cycle and permanent income theory. The meeting therefore operates on top of strong income driven dynamics.

Robustness checks:
Two robustness checks confirm the credibility of the causal interpretation.

1. **Placebo test ($t = 0$ and $t = 1$):** Column (2) estimates a false treatment effect by pretending the meeting occurred earlier, between $t = 0$ and $t = 1$. The placebo DiD coefficient is: -0.001 ($p =$ ns) $\rightarrow$ There is no significant difference in savings before the real intervention, confirming that treated and control groups were evolving similarly prior to $t = 2$. This supports the parallel trends assumption and strengthens causal identification.

2. **Stability of the income effect:** The income coefficient remains large and significant across models (3.411 with $p < 0.01$ in panel linear gender heterogeneity and 2.761 with $p < 0.01$ in placebo OLS), demonstrating that income consistently predicts changes in savings. Its stability across specifications reinforces the reliability of the model.

3 Conclusion

Our analysis shows clear and robust evidence that the salesperson meeting had a strong causal impact on saving behaviour. Firstly, using a Difference-in-Differences model with individual fixed effects, we find that the intervention increased total savings by about 51.9% among treated clients. Secondly, the effect was even stronger for retirement savings, where contributions rose dramatically from very low pre-treatment levels indicating that the meeting successfully triggered the adoption of retirement plans among previously inactive clients. Finally, we document meaningful heterogeneity: female clients reacted more strongly to the intervention, suggesting that personalized financial guidance may be particularly effective for groups traditionally less engaged in financial planning.

Despite these findings, two central questions remain open:

1. **What mechanism drives the treatment effect?**
The meeting could improve financial literacy, reduce procrastination, increase trust or help clients to set goals but our data cannot distinguish between these channels.

2. **How persistent is the effect over time?**
Our analysis captures short run behavioural changes, but whether increased saving habits persist in the long term remains unknown.

4 Sources

- **Applied Econometrics lecture slides:**
  - https://cours.univ-paris1.fr/course/view.php?id=44026
- **Life-Cycle Hypothesis developed by Modigliani in 1954:**
  - https://www.lafinancepourtous.com/decryptages/finance-perso/epargne-et-placem epargne/la-theorie-du-cycle-de-vie/
- **To know how to use Difference-in-Difference Method:**
  - https://en.wikipedia.org/wiki/Difference_in_differences
  - https://www.youtube.com/watch?v=J7q2H8aB8bQ
  - https://www.youtube.com/watch?v=nAR4psOPVmI
- **To know how to use the Within Method (Fixed-effects/Panel data):**
  - https://www.youtube.com/watch?v=eyBV331l92A
  - https://libguides.princeton.edu/R-Panel
  - https://www.econometrictutor.co.uk/regression-analysis-techniques-panel-data
- **To know how to use LaTeX:**
  - https://www.overleaf.com/learn/latex/Learn_LaTeX_in_30_minutes
  - https://latex-tutorial.com/tutorials/pgfplotstable/
  - https://www.youtube.com/watch?v=lgICpA4zzGU&t=1s
- **To know how to use OverLeaf:**
  - https://www.youtube.com/watch?v=pLjtV0XfRa4
  - https://www.youtube.com/watch?v=_PzDLFJH03E
- **To know how to make a regression:**
  - https://www.datacamp.com/tutorial/linear-regression-R
  - https://www.youtube.com/watch?v=MEPP5oJ4rWc
- **To know how to make an OLS regression:**
  - https://www.r-bloggers.com/2017/07/ordinary-least-squares-ols-linear-regress
  - https://www.geeksforgeeks.org/r-machine-learning/ordinary-least-squares-ols-
  - https://cran.r-project.org/web/packages/olsrr/vignettes/intro.html
- **To know how to Use PLM Library:**
  - https://cran.r-project.org/web/packages/plm/plm.pdf
  - https://www.youtube.com/watch?v=PxDPkiHedn0
- **To know how to Use Stargazer Library:**
  - https://cran.r-project.org/web/packages/stargazer/vignettes/stargazer.pdf
  - https://rpubs.com/ErikPav/stargazer

Script R du Projet

```r
### Library
library(dplyr)
library(ggplot2)
library(plm)
library(stargazer)

### load database

groupe_41_pt3 = read.csv("C:/Users/ibrah/OneDrive/Documents/ cole /S7/econometrie appliquee/projet/group41.csv")
head(groupe_41_pt3,20)


#### PREPARATION DES DONNEES (DATA WRANGLING)
df_did <- groupe_41_pt3 %>%
  mutate(
    # Variable Dummy Post-Traitement (1 si t >= 2, 0 sinon)
    post = ifelse(time >= 2, 1, 0),
    # Conversion of TRUE/FALSE in 1/0
    treated = ifelse(meeting == "TRUE" | meeting == TRUE, 1, 0),

    # Variable d'interaction DiD (C'est la variable d'int r t causal)
    did_interaction = post * treated,

    # Transformation Logarithmique (log(x+1) pour g rer les z ros)
    log_savings = log(savings + 1),
    log_retirement = log(retirement + 1),
    log_yincome = log(yincome + 1)
  )

# D claration de la structure de panel pour le package 'plm'
pdf <- pdata.frame(df_did, index = c("id", "time"))

### VERIFICATION DE L'HYPOTHESE DES TENDANCES PARALLELES (Pour Section 2.1)

# Calcul des moyennes par p riode et par groupe
trend_data <- df_did %>%
  group_by(time, treated) %>%
  summarise(mean_savings = mean(log_savings, na.rm = TRUE), .groups = 'drop') %>%
  mutate(Groupe = ifelse(treated == 1, "Trait  (Meeting)", "Contr le"))

# Cr ation du graphique pour controler les tendances paralleles
ggplot(trend_data, aes(x = time, y = mean_savings, color = Groupe)) +
  geom_line(linewidth = 1.2) +
  geom_point(size = 3) +
  geom_vline(xintercept = 1.5, linetype = "dashed", color = "black") + # Ligne s parant avant/apr s
  labs(
    title = "Test des Tendances Parall les :  volution  de l' pargne  moyenne",
    subtitle = "V rification visuelle avant estimation DiD",
    x = "P riode (Temps)",
```

```r
    y = "Log( pargne  Total)",
    caption = "Note : Le traitement intervient   t=2. Les courbes doivent  tre  parall les avant cette date."
  ) +
  theme_minimal() +
  scale_color_manual(values = c("grey40", "blue")) +
  annotate("text", x = 0.5, y = max(trend_data$mean_savings), label = "Pr  -Traitement") +
  annotate("text", x = 3.5, y = max(trend_data$mean_savings), label = "Post-Traitement")

# SECTION 2.2 : ESTIMATION DES R SULTATS PRINCIPAUX

# L' quation  est : log(Y) = c_i + lambda_t + delta * (Post * Trait ) + controls avec
#le modele "within" parceque on a des donn es de panel

# 1. Estimation sur l' pargne  Totale
model_savings <- plm(log_savings ~ did_interaction + post + log_yincome,
                     data = pdf, # Assurez-vous que ce dataset existe (cr  l' tape  d'avant)
                     index = c("id", "time"),
                     model = "within")

# 2. Estimation sur l' pargne  Retraite
model_retirement <- plm(log_retirement ~ did_interaction + post + log_yincome,
                        data = pdf,
                        index = c("id", "time"),
                        model = "within")

# 3. G n ration du Tableau de R sultats pour le latex
stargazer(model_savings, model_retirement,
          type = "text",
          title = "Tableau 1 : Effet Causal du Meeting sur l' pargne  (Estimateur Within)",
          column.labels = c(" pargne  Totale", " pargne  Retraite"),
          covariate.labels = c("Interaction DiD (ATT)", "Post (Tendance)", "Revenu (Log)"),
          keep.stat = c("n", "rsq", "adj.rsq", "f"),
          digits = 3)


# section 2.3 TESTS DE ROBUSTESSE & HETEROGENEITE

# .1 Analyse d'H t rog n it
# On ajoute une triple interaction : did * female
m_hetero <- plm(log_savings ~ did_interaction * female + log_yincome,
                data = pdf,
                model = "within",
                effect = "twoways")

# .2 Test Placebo
# On fait comme si le traitement avait eu lieu   t=1 (au lieu de t=2)
# On ne garde que les donn es AVANT le vrai traitement (t=0 et t=1)
df_placebo <- df_did %>%
  filter(time < 2) %>%
  mutate(
    post_fake = ifelse(time == 1, 1, 0),
    did_fake = post_fake * treated
  )

m_placebo <- lm(log_savings ~ did_fake + post_fake + treated + log_yincome,
                data = df_placebo)
```

```
# Tableau des r sultats de robustesse pour le latex
stargazer(m_hetero, m_placebo, type = "text",
          title = "Tests de Robustesse",
          column.labels = c("H t rog n it  Genre", "Test Placebo (t=0 vs t=1)"),
          covariate.labels = c("Interaction DiD", "Femme", "DiD x Femme", "DiD Placebo (Faux Traitement)"))
```