Maestría en Economía
Fundamentos Microeconómicos Del Capital Humano
Universidad de Montevideo

Smart Tools, Smarter Students?: Technology's Influence on Global Education Metrics
The case of Uruguay

Abstract

Access to the internet and digital devices is becoming increasingly widespread. The impact of this phenomenon on education remains inconclusive, largely depending on how these devices are used. This study leverages the roll-out strategy of Fiber to the Home (FTTH) in Uruguay to estimate the impact of enhanced internet quality on learning outcomes. We employ a staggered difference-in-differences model at artificial neighborhoods geographical level to analyze this impact. Additionally, we propose an analysis of heterogeneous effects and explore the possible mechanisms behind the observed results.

Keywords: Difference-in-difference, ICT, PISA, staggered

Horacio Castellanos Muñoa
Sofía Spektor Steiner

October 16, 2024

1 Introduction

Access to the internet and digital devices is becoming easier and more widespread every day. According to the World Bank, the percentage of individuals worldwide who use the internet has risen from 29% in 2010 to 63% in 2021, and in 2023, the number of smartphones surpassed the world population.

Given this reality, the time teenagers spend on the internet is increasing, and evidence suggests that it is detrimental to rest, physical endurance, and social well-being. However, evidence regarding the impact on learning has been inconclusive and mostly depends on how the internet and digital devices are used.

For example, (Navarro-Martinez and Peña-Acuña, 2022) analyzed the effects of Information and Communication Technology (ICT) usage on the PISA test scores of 2018 for Spanish students. They found that excessive use of technology and social networks during the week and weekends was associated with impaired academic performance. This effect was particularly pronounced among male students, who started using the internet at an earlier age and were more likely to engage in online gaming, leading to lower academic achievement. The study also acknowledged that when social networks were used appropriately, they could positively impact academic performance and contribute to students' development. Proper utilization of social networks for communication among peers was associated with better academic outcomes.

(Haleem et al., 2022) addresses the role of digital technologies in education. They state that the use of digital devices can enhance access to educational resources, help students complete the syllabus via online platforms, transform learning through more engaging resources, foster self-learning abilities by browsing online, and allow exploration of self-interests via video-based instructional learning. However, these implementations can only succeed if there are qualified teachers and students who have and know how to use the resources.

Regarding the psychology literature, (Bickham, 2021) explores the current research and viewpoints on problematic interactive media usage (PIMU) in adolescents, focusing on aspects such as nomenclature, prevalence rates, potential determinants, co-morbid disorders, and treatment strategies. They highlight that Internet Gaming Disorder (IGD) was added to the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) in 2013 by the American Psychiatric Association, with symptoms such as an increased need to spend more time gaming, inability to reduce game time, lying to others about the amount of gaming, and using gaming to reduce a negative mood. As potential causes of PIMU, (Bickham, 2021) highlights a poor parent-child relationship and personality traits such as impulsivity, low self-control, and problems when dealing with anxiety.

According to (Bickham, 2021), Problematic Interactive Media Usage (PIMU) is co-morbid with other mental health problems such as depression, ADHD, and autism. (Vural et al., 2015) also explores the complex relationship between ADHD and internet usage. The study surveyed 1,389 secondary school students in grades six to eight residing in or near the center of the city of Bursa, Turkey. They concluded that ADHD is associated with riskier internet behavior, such as meeting with strangers. Therefore, we consider it important to highlight the close relationship between disruptive behavior and problematic internet use, as (Carrell et al., 2018) previously studied that adding disruptive peers to a classroom has a lasting negative impact on academic achievement.

In the last PISA test, conducted in 2022, students across the OECD reported being distracted by using digital devices in at least some maths lessons. Argentina, Brazil, Canada, Chile, Finland, Latvia, Mongolia, New Zealand, and Uruguay topped this question, with 80% of students reporting distraction. The 2022 PISA report also highlights a negative trend in scores in almost all the countries surveyed.

In light of the above, this research aims to leverage the roll-out strategy of fiber optic installation in Uruguay to understand how enhanced access to the Internet has affected students' PISA test scores.

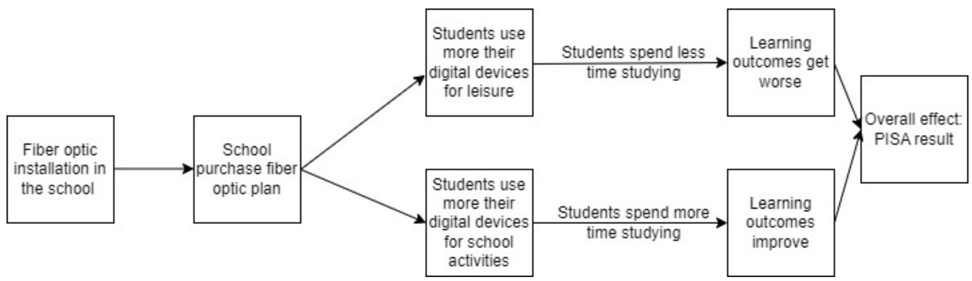
2 Conceptual Framework

Taking advantage of the staggered nature of public policies is a common natural experiment that scientists tend to exploit. The installation of fiber optic in some countries is one such example, and it has been used before to study the effect of technology on education. For example, (Grimes and Townsend, 2018) investigates the impact of ultra-fast broadband (UFB) on the academic performance of students in New Zealand schools using a difference-in-difference (DiD) approach. The article provides evidence that the introduction of fibre broadband in New Zealand schools had a positive effect on student performance in standardised assessments, with some evidence suggesting larger benefits for students in low-socioeconomic schools.

Another example of this literature is (Boeri, 2020), which examines the short-run effects of high-speed internet access on educational disparities before and after the COVID-19 pandemic. The study focuses on Italy and follows 3 million students across six different cohorts over the period 2012-2022, leveraging the roll-out of broadband infrastructure under the National Ultra-Broadband Plan (NUBP). The key findings were that the introduction of 30 Mbit/s broadband showed an average null effect on 8th grade student performance in numeracy and literacy. However, the effect of access to high-speed internet was not homogeneous. For example, students who were low performers in grade 5 and those with a better socioeconomic background experienced positive and significant gains from increased internet speed. The broadband infrastructure might have amplified the gap between students from different socioeconomic backgrounds, with wealthier students benefiting more from the enhanced internet access. In line with this idea, (Boeri, 2020) makes a policy recommendation that policymakers consider these heterogeneous effects when designing and implementing broadband infrastructure projects, ensuring that interventions are tailored to support the most disadvantaged students effectively.

The theory of change motivating this study posits that improved internet access leads to increased online activity among students. While this additional time may be spent on activities such as playing video games or using social media, it is also likely to be utilized for homework and online learning. In our study, the available data includes whether a student attended a school with fiber optic internet and their PISA test scores. This data limits our ability to distinguish between positive effects from increased learning and potential negative effects from more leisure time usage. However, we can still assess the overall impact. To understand the underlying reasons behind the results, we will conduct a mechanism analysis in a subsequent section. We will leverage findings from (Bickham, 2021), which indicate that the peak of problematic interactive media usage (PIMU) occurs at 15.9 years, while the PISA test is administered at 15 years. Additionally, we will explore leisure usage data from the PISA database as a potential mechanism, connecting it with the psychology literature previously cited. Figure 1 illustrates the theory of change.

Figure 1: Theory of Change of Fiber Optic Implementation



Source: Own Elaboration

3 Background and Data

In Uruguay, the National Telecommunications Administration (ANTEL for its acronym in Spanish) is the state telecommunications company for mobile, fixed, and cable internet connections. In 2010, ANTEL started a project to install fiber optic infrastructure to deliver fast internet connections to everyone in Uruguay. According to (Failache-Mirza, 2022), who compiled and analyzed data regarding this project, the installation was done gradually by geographical areas, eventually reaching all households. (Failache-Mirza, 2022) states that by the end of 2018, 83% of households with fixed telephone lines had fiber optic accessibility.

The installation did not guarantee access to fiber optic; for that, people had to switch to fiber optic services with ANTEL. Therefore, this research focuses on the intention-to-treat effects.

Internet data

For the internet data, we will rely on the meticulous work previously conducted by (Failache-Mirza, 2022), which involved compiling all data regarding FTTH implementation and estimating values for missing years. The research team gathered administrative data from two different sources. From online information, they collected data on the proportion of fixed telephone lines with FTTH connections by year and department for the period 2012-2018, and constructed granular data on the 2012 installation at the block level from publications on ANTEL's webpage available in the Internet Archive. Additionally, ANTEL provided detailed block-level data on the 2020 installations.

To estimate the installation data for the missing years (2013-2018), the team used the proportion of fixed telephone lines with FTTH connections for the period 2012-2018 to impute an adjusted probability of FTTH installation to artificial neighbourhoods. These were defined based on the information of the years 2012 and 2020. They grouped the neighbourhoods into three categories. The three sets of areas, A_1, A_2, and A_3, can be defined as follows: d identifies each geographical area a_d, y refers to the year, and $FTTH_t$ is an indicator variable for fiber optic accessibility.

- $a_{d,y} \in A_1 : FTTH_{2012} = 1 \quad \rightarrow FTTH_t = 1, t \in [2012, 2018]$
- $a_{d,y} \in A_2 : FTTH_{2020} = 0 \rightarrow FTTH_t = 0, t \in [2012, 2018]$
- $a_{d,y} \in A_3 : \max\left(FTTH_t = 1\right), t \in [2012, 2018]$

For more details about this procedure, see Appendix A.3 on (Failache-Mirza, 2022).

Once we have the data, we need to ensure that the implementation criteria were exogenous to our study. ANTEL publicly stated that the implementation would occur nationwide, regardless of potential profit losses. However, this does not mean the order of implementation was entirely without criteria. (Failache-Mirza, 2022) conducted a Principal Component Analysis (PCA), which suggested that sanitation levels and family income are correlated with the timing of fiber optic implementation. Based on this information, we propose doing a regression analysis to analyse the correlation between the average presence of fiber optic for each geographical area over the specified time period (e.g., 2010-2020) and the variables of interest at the start of the policy (family income and sanitation[1] levels).

To calculate the average, we will define the variable D_{it} as an indicator of presence of fiber optic in the geographical area for each year. (1 if fiber is present in year t, 0 otherwise)

$$\text{Average Fiber Optic Presence}_i = \frac{1}{T}\sum_{t=1}^{T} D_{it} \tag{1}$$

This average gives the proportion of years fiber optics were present in each geographical area i.

Next, we will tansform our panel data to a cross-sectional data set, where each observation represents a geographical area with its average presence of fiber optics and the initial variables (e.g., income and sanitation). For example:

Area	**Average Fiber Optic Presence**	**Income 2010**	**Sanitation 2010**
A	0.7	50000	0.9
B	0.4	45000	0.8
...	...	...	...

Table 1: Transformed Cross-Sectional Data

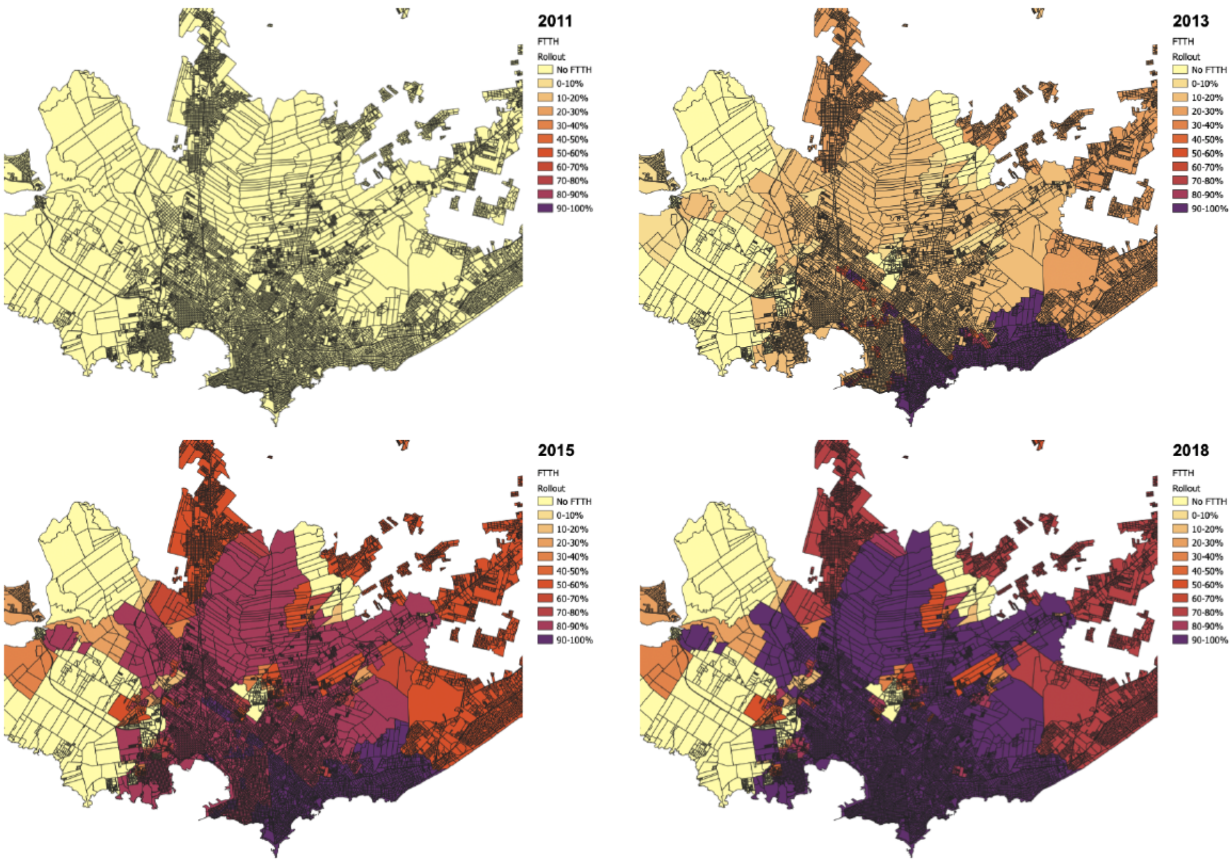
[1] The variable for sanitation is not available in our Internet or PISA datasets. To acquire this variable, we will rely on data from the Uruguayan Household Survey. This survey is conducted annually and provides nationwide information on various dimensions of the population. We will extract information on access to sanitation, along with geographic references, to match this data with our original dataset.

With this transformed data we will run the regression as follows:

$$\text{Average Fiber Optic Presence}_i = \alpha + \beta_1 \text{Income}_{i,2010} + \beta_2 \text{Sanitation}_{i,2010} + \epsilon_i \tag{2}$$

Figure 2, taken from (Failache-Mirza, 2022), shows the roll-out of fiber optic implementation, clearly illustrating the different times when the periphery and coastal areas of Montevideo received the installation suggesting that our regression might find the same results as (Failache-Mirza, 2022). We will address this issue further ahead in this article.

Figure 2: FTTH Roll-out in Montevideo by Neighborhood



Source: (Failache-Mirza, 2022)

Learning data

To measure learning outcomes, we plan to use data from the last five periods (2009, 2012, 2015, 2018, 2022) of the PISA test. Since PISA does not evaluate the same schools over time and employs different sampling methods each year, our focus will be on analyzing the impact by geographic zones (the artificial neighbourhoods mentioned in the Internet Data section), which align with the FTTH installation process. By grouping the PISA results according to these zones, we can accurately estimate the impact of the internet policy on learning outcomes.

Access to the geographic dimension of the PISA results is not publicly available, but we plan to partner with ANEP to obtain this data.

4 Our Model

We will use the following notation, N denotes the number of units i, T denotes the total of time periods t. D_{it} corresponds to a dummy indicating if unit i was treated at period t. In our case of interest, the treatment corresponds to the implementation of optic fiber internet and our units will be the provinces (departamentos) of Uruguay to a neigborhood level, as in (Failache-Mirza, 2022).

4.1 List of Variables

Our starting point for the control variables is based on (Hanushek et al., 2013), with some minor adjustments. We exclude parent income because, due to the nature of our treatment—the implementation of fiber optic internet—it might be endogenous. Instead, income will be explored as a mechanism further ahead in this article. However, as stated by (Failache-Mirza, 2022), both neighborhood and income might influence the implementation of optic fiber. To introduce the desired exogeneity into our model, we will use parents' education as a proxy. Neighborhood characteristics will be accounted for through the regional fixed effect included in our model.

Student and Family Characteristics

- Female
- Age (years)
- Immigration background
  - Native student
  - First generation student
  - Non-native student
- Other language than test language or national dialect spoken at home
- Parents' education
  - None
  - Primary
  - Lower secondary
  - Upper secondary I
  - Upper secondary II
  - University

- Books at home
  - 0–10 books
  - 11–100 books
  - 101–500 books
  - More than 500 books

School's Charcteristics

- Number of students
- Privately operated
- Share of government funding
- Share of fully certified teachers at school
- Shortage of math teachers

School's Community Location

- Village or rural area (b3000)
- Town (3000–15,000)
- Large town (15,000–100,000)
- City (100,000-1,000,000)
- Large city (>1,000,000)

4.2 Double LASSO for Control Selection

According to (Urminsky et al., 2016), applying standard LASSO regression in treatment analysis can lead to artificially weak or artificially strong regressions due to Lasso mistakenly excluding variables with non-zero coefficients, which leads to omitted variable bias. (Belloni et al., 2014) showed that Double LASSO solves this problem and that the estimates are heteroscedasticity robust.

To implement the Double LASSO method, we will follow the steps outlined in (Sice et al., 2023):

4.2.1 Step 1: Run LASSO Regressions for Control Selection

- **First LASSO (for outcome selection)**: Run a LASSO regression to identify control variables that predict the outcome Y_{it}:

$$Y_{it} = \alpha_i^1 + X_{it}'\theta_1 + v_{it} \tag{3}$$

- **Second LASSO (for treatment selection)**: Run another LASSO regression to identify control variables that predict the treatment D_{it}:

$$D_{it} = \alpha_i^2 + X_{it}'\theta_2 + u_{it} \tag{4}$$

Citing (Urminsky et al., 2016), the second LASSO regression is crucial because excluding a covariate that is a moderate predictor of the dependent variable but a strong predictor of the independent variable can result in significant omitted variable bias.

4.2.2 Step 2: Combine Selected Controls

Combine the sets of control variables selected from both LASSO regressions. Denote this combined set as X_{it}^*.

4.2.3 Step 3: Run the Final LASSO Regression

Run a LASSO regression of the outcome Y_{it} on the treatment D_{it} and the combined set of selected control variables X_{it}^*:

$$Y_{it} = \alpha_i + \beta D_{it} + X_{it}^{*\prime}\theta + \epsilon_{it} \tag{5}$$

4.3 Balance Analysis for Treatment Effects

Given that our study leverages the random staggering of the policy implementation, we need to define the control group ex post. To achieve this, we propose employing Propensity Score Matching (PSM) at each time period.

To conduct propensity score matching in staggered settings, where the treatment is implemented at different times across units, we will follow these use a logit model to estimate the propensity score, incorporating time-fixed effects to account for variations over time. The model can be specified as:

$$P(D_{it} = 1 \mid \hat{X}_{it}) = \text{logit}(\hat{X}_{it}'\theta) \tag{6}$$

where $\hat{X}_{it}$ includes covariates obtained previously by the Double LASSO procedure and γ_t represents time-fixed effects.

After the model is estimated, we will match treated units with control units based on their propensity scores. Ensure that matched units are from the same or similar time periods to account for the timing of the treatment.

Once we obtain the propensity scores for all observations, we will retain the treatment and control observations that are successfully matched, ensuring that both groups are balanced.

4.4 Identification Strategy

Considering the nature of our experiment, following (Sun and Abraham, 2018) we propose using a staggered difference-in-differences model with two-way fixed effects as follows:

$$Y_{it} = \sum_{\ell=0}^{T} \beta_\ell \cdot 1\{F_i = t - \ell\} + \lambda_i + \gamma_t + \hat{X}'_{it}\theta + \epsilon_{it} \tag{7}$$

- γ_t: **Time Fixed Effect**
  - This term captures the influence of time-specific factors that affect all individuals in the same period. It accounts for common shocks or trends over time that might influence the dependent variable Y_{it}.
- λ_i: **Regional Fixed Effect**
  - This term represents individual-specific effects that are constant over time. It accounts for unobserved heterogeneity across individuals (or regions) that might influence Y_{it}.
- $\hat{X}'_{it}$: **Controls**
  - These are the control variables, selected by the Double-Lasso procedure, included in the model to account for other factors that might affect Y_{it}.
- F_i: **Time Individual i Received the Treatment**
  - F_i denotes the time period in which individual i received the treatment. The indicator function $1\{F_i = t - \ell\}$ is equal to 1 if individual i received the treatment ℓ periods ago, and 0 otherwise. The sum $\sum_{\ell=0}^{T} \beta_\ell \cdot 1\{F_i = t - \ell\}$ captures the dynamic effects of the treatment over time. Following (de Chaisemartin and D'Haultfoeuille, 2022) we set $\ell \geq 0$ as we are interested in the cumulative effect of the treatment.

If the regression estimated in Equation 2 yields coefficients significantly different from zero, we will need to incorporate an interaction term between a trend variable that will take values equal to 1 at year one, 2 at year 2,.. etc. and the regional fixed effect. It aims to correct for different trends across the various departments for the outcome variable or for other variables correlated with the outcome variable, which were already present before the deployment of fiber optic technology.

$$Y_{it} = \sum_{\ell=0}^{T} \beta_\ell \cdot 1\{F_i = t - \ell\} + \lambda_i + \gamma_t + \hat{X}'_{it}\theta + \lambda_i \cdot \text{Trend}_t + \epsilon_{it} \tag{8}$$

4.5 Assumptions

Following (Sun and Abraham, 2018) we will verify if 3 assumptions are met, *parallel trends*, *no anticipation* and *treatment effect homogeneity*.

4.5.1 No Anticipation

Individuals or units are not expected to change their behavior in anticipation of the treatment. In other words, the treatment effect should only occur after the treatment is implemented, not before. Given the gradual and often non-publicized roll-out of fiber optic internet, residents and businesses are unlikely to alter their behavior in advance since they are not aware of the specific timing of the treatment.

4.5.2 Treatment Effect Homogeneity

The assumption here is that the effect of the treatment is consistent across all treated units. This implies that once treated, the impact on the outcome variable is uniform, making the estimated treatment effect representative of the true effect. The standardized technology and installation process for fiber optic internet ensure that all treated regions experience similar improvements in internet speed and reliability, thus supporting this assumption.

4.5.3 Parallel Trends

It is assumed that in the absence of treatment, the difference in outcomes between the treated and control groups would have remained constant over time. This ensures that the observed treatment effect is not confounded by pre-existing differences in trends. In order to verify the parallel trends assumption, we will use the model proposed in (Wooldridge, 2021):

$$\begin{aligned} Y_{it} =& \eta + \lambda_q D_{iq} + \cdots + \lambda_T D_{iT} \\ &+ \theta_2 f_{2t} + \cdots + \theta_T f_{Tt} \\ &+ \sum_{r=q}^{T} \sum_{s=2}^{r-1} \omega_{rs} (D_{ir} \cdot f_{st}) \\ &+ \sum_{r=q}^{T} \sum_{s=r}^{T} \delta_{rs} (D_{ir} \cdot f_{st}) + u_{it} \end{aligned} \tag{9}$$

Where:

- Y_{it}: The outcome variable for individual i at time t.
- η: The intercept term.
- $\lambda_q, \ldots, \lambda_T$: Coefficients for treatment dummy variables.
- $D_{iq}, \ldots, D_{iT}$: Dummy variables indicating that individual i received treatment in period q to T.
- $\theta_2, \ldots, \theta_T$: Coefficients for time dummy variables.
- $f_{2t}, \ldots, f_{Tt}$: Time dummy variables.
- ω_{rs}: Coefficients for the interaction terms between treatment dummies and time dummies (for testing parallel trends).

- δ_{rs}: Coefficients for other interaction terms between treatment dummies and time dummies.

- u_{it}: The error term.

The null hypothesis to test the parallel trends assumption is:

$$H_0 : \omega_{rs} = 0 \text{ for all } r = q, \dots, T \text{ and } s = 2, \dots, r - 1$$

The previous null hypothesis states that the coefficients ω_{rs} on the interaction terms between the treatment dummies and the time dummies are zero. If the null hypothesis is not rejected, it suggests that there are no pre-treatment differences in trends between the treated and control groups, thereby supporting the parallel trends assumption. If the assumption of parallel trends is not met, a model of the form Equation 8 might address this issue.

5 Robustness Check

In order to perform a robustness check of our model we will follow (Goodman-Bacon, 2021), in particular the Goodman-Bacon decomposition. The Goodman-Bacon decomposition provides a way to understand the $\hat{\beta}^{DD}$ estimates by breaking them into components that represent the weighted average of DiD estimates from different groups and times. Specifically, it considers comparisons between:

- Treated (T) versus Never Treated (U) units.

- Early Treated (k) versus Late Control(C^l) units.

- Late Treated (l) versus Early Control (C^e) units.

To perform the decomposition, first we need to estimate the following regression without controls:

$$Y_{it} = \lambda_i + \gamma_t + \beta^{DD} D_{it} + e_{it} \tag{10}$$

Where

$$\hat{\beta}^{DD} = \frac{\frac{1}{NT}\sum_i \sum_t Y_{it}\tilde{D}_{it}}{\frac{1}{NT}\sum_i \sum_t \tilde{D}_{it}^2}$$

And

$$\tilde{D}_{it} = (D_{it} - D_i) - \left(D_t - \overline{\overline{D}}\right)$$
$$\overline{\overline{D}} = \frac{\sum_i \sum_t D_{it}}{NT}$$

From (Goodman-Bacon, 2021) we take the following notation: Each group's sample share is n_k and the share of time it spends treated is $\bar{D}_k$. The main contribution of (Goodman-Bacon, 2021) is a theorem that states that the estimator $\hat{\beta}^{DD}$ can be decomposed as follows:

$$\widehat{\beta}^{DD} = \sum_{k \neq U} s_{kU} \widehat{\beta}_{kU}^{2\times2} + \sum_{k \neq U} \sum_{l > k} \left[s_{kl}^{k} \widehat{\beta}_{kl}^{2\times2,k} + s_{kl}^{l} \widehat{\beta}_{kl}^{2\times2,l} \right] \tag{11}$$

Were each term of the form s_{kU} corresponds to the weight of each Beta in the estimator $\widehat{\beta}^{DD}$, for a complete derivation of the weights the reader can consult (Goodman-Bacon, 2021). Problems arise with the term

$$\widehat{\beta}_{k\ell}^{2x2,\ell} \equiv \left(\bar{y}_{\ell}^{POST(\ell)} - \bar{y}_{\ell}^{MID(k,\ell)} \right) - \left(\bar{y}_{k}^{POST(\ell)} - \bar{y}_{k}^{MID(k,\ell)} \right)$$

As we are comparing in both cases treated units. This can be better understood if we recall the asymptotic distribution

$$\text{plim}_{N \to \infty} \hat{\beta}^{DD} = \beta^{DD} = VWATT + VWCT - \triangle ATT$$

Following (Goodman-Bacon, 2021) the term $-\triangle ATT$ corresponds to the change in treatment effects within each unit's post-period with respect to another unit's treatment timing. So, we can conclude that if the treatment effect is not homogeneous, the coefficient can be negative even though there is an overall positive effect. By performing the Goodman-Bacon decomposition we can evaluate how each individual 2×2 DiD is affecting the overall estimate of $\widehat{\beta}^{DD}$ and we can double check the treatment effect homogeneity assumption.

6 Heterogeneous Effects

We believe it is crucial to assess the presence of heterogeneous effects across key variables. Firstly, we hypothesize that there may be gender differences, potentially favoring girls. (Navarro-Martinez and Peña-Acuña, 2022) found that boys are more likely to use social networks to play online games. Consequently, we anticipate that girls may benefit more from increased internet access.

Another important variable to consider is parental education levels. Students whose parents have higher levels of education might experience less pronounced effects compared to those with lower parental education. Well-educated parents are more likely to assist with homework, providing support outside of school. In contrast, students with less-educated parents may gain more from the educational resources available online.

Lastly, examining the impact on migrant families is also essential. Migrant families often face more challenging conditions than native ones, making it harder for students to succeed academically. Improved internet access might provide these students with additional support and learning opportunities, helping to bridge the educational gap.

7 Mechanisms

Once we estimate the treatment effect, we propose examining various variables to understand the underlying channels behind the overall effect. This analysis strategy involves regressing the mechanism variable on the original model, as shown in the following equation:

$$I_{it} = \sum_{\ell=0}^{T} \beta_\ell \cdot 1\left\{F_i = t - \ell\right\} + \lambda_i + \gamma_t + \hat{X}_{it}'\theta + \epsilon_{it} \tag{12}$$

Where I_{it} represents the output of the mechanism of interest. This approach will allow us to identify and quantify the specific pathways through which the enhanced internet quality impacts learning outcomes.

The first analysis focuses on the double-edged nature of access to the internet and digital devices. We aim to understand the reasons behind the likely increase in digital device usage. Utilizing the PISA database, we will differentiate between the time spent on devices for leisure and the time spent on educational activities. For leisure time, we will analyze data on time spent during weekdays and weekends. For educational activities, we will use a variable related to the use of ICT for school activities outside the classroom. We expect both coefficients to be positive, with our primary analysis concentrating on their significance and magnitude.

Another mechanism involves school infrastructure. Improved internet access might incentivize enhancements to digital laboratories. PISA provides data on the availability of computers in schools, which we can use as a proxy for general infrastructure improvements to analyze this mechanism. Lastly, we plan to explore family dynamics that might have changed due to better internet access. Specifically, we will examine the type of work parents do. If the treatment improved their employment conditions, it might lead to a better disposition for their children to study and improve their learning outcomes.

We propose expanding the mechanism analysis by integrating databases from Ceibal, the national center for educational innovation and digital technologies. Ceibal began by providing a laptop to every student in public schools and has since developed a platform with valuable educational resources and apps for teachers and students. By incorporating data on computer usage from students across different schools, including hours spent on computers and the utilization of various platforms, we can gain deeper insights into the impact of digital resources on educational outcomes.

8 Limitations

The integration of Fiber to the Home (FTTH) was implemented by geographic zones, making it challenging to isolate the treatment effect from other neighborhood-related factors. For instance, access to FTTH might have enabled parents of students to secure remote jobs with better conditions, such as higher salaries or more free time due to the elimination of commuting. As a result, parents could afford to pay for additional tuition or spend more time assisting their children with homework, potentially enhancing their learning outcomes.

To address this, we propose estimating the magnitude of this effect by incorporating wage data from the Household Survey as a dependent variable in our model. However, it is important to note that the Household Survey data is based on a different sampling methodology, which may introduce complexities in drawing straightforward conclusions.

9 Conclusion

Access to the internet and digital devices is becoming easier and more widespread every day. Evidence regarding the impact on learning has been inconclusive and mostly depends on how the internet and digital devices are used.

This research aims to leverage the roll-out strategy of fiber optic installation in Uruguay to understand how enhanced access to the Internet has affected students' PISA test scores. Based on our theory of change, we expect that FTTH installation has two contrary effects: positive effects from increased learning and potential negative effects from increased leisure time usage. Although we cannot distinguish between these effects directly and can only estimate the overall impact, we plan to gain insights by including variables for internet usage for school activities and for leisure in the mechanism analysis.

Additionally, we aim to determine whether the effect on learning is due to internet access at school or if it is also influenced by parents gaining access to better job conditions as a result of the policy.

Finally, in line with existing literature, we expect to observe heterogeneous effects by gender and migration status.

Bibliography

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, 81:608–650.

Bickham, D. S. (2021). Current research and viewpoints on internet addiction in adolescents. *Current Pediatrics Reports*, 9:1–10.

Boeri, F. (2020). High-speed broadband and educational achievements. *SSRN Electronic Journal*.

Carrell, S. E., Hoekstra, M., and Kuka, E. (2018). The long-run effects of disruptive peers. *American Economic Review*, 108:3377–3415.

de Chaisemartin, C. and D'Haultfoeuille, X. (2022). Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey.

Failache-Mirza, E. (2022). *Three Essays on How New Technologies Are Changing Our Lives: A Contribution from Applied Economics*. Phd thesis, Universitat Autònoma de Barcelona. Defended September 2022.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225:254–277.

Grimes, A. and Townsend, W. (2018). Effects of (ultra-fast) fibre broadband on student achievement. *Information Economics and Policy*, 44:8–15.

Haleem, A., Javaid, M., Qadri, M. A., and Suman, R. (2022). Understanding the role of digital technologies in education: A review. *Sustainable Operations and Computers*, 3:275–285.

Hanushek, E. A., Link, S., and Woessmann, L. (2013). Does school autonomy make sense everywhere? panel estimates from pisa. *Journal of Development Economics*, 104:212–232.

Navarro-Martinez, O. and Peña-Acuña, B. (2022). Technology usage and academic performance in the pisa 2018 report. *Journal of New Approaches in Educational Research*, 11:130.

Sice, J. F., Lattimore, F., Robinson, T., and Zhu, A. (2023). Practical insights into applying double lasso. *SSRN Electronic Journal*.

Sun, L. and Abraham, S. (2018). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects.

Urminsky, O., Hansen, C., and Chernozhukov, V. (2016). Using double-lasso regression for principled variable selection. *SSRN Electronic Journal*.

Vural, P., Uncu, Y., and Kilic, E. Z. (2015). Relationship between symptoms of disruptive behavior disorders and unsafe internet usage in early adolescence. *Noro Psikiyatri Arsivi*, 52:240–246.

Wooldridge, J. M. (2021). Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators. *SSRN Electronic Journal*.