

Causal inference in educational research: Causal analysis in cross-sectional observational studies

Inferencia causal en investigación educativa: Análisis de la causalidad en estudios observacionales de carácter transversal

Inferência causal em investigação educativa: Análise da causalidade em estudos observacionais de carácter transversal

教育研究中的因果推断：对观察性横断面研究的因果关系进行分析

الاستدلال السببي في البحوث التربوية: تحليل السببية في الدراسات الرصدية المستعرضة

Martínez-Abad, Fernando ⁽¹⁾ ; León, Jaime ⁽²⁾ 

⁽¹⁾ University of Salamanca, Spain.

⁽²⁾ University of Las Palmas de Gran Canaria, Spain.

Abstract

The assumption of cause-effect relationships in ex post facto research is a widely known issue in the field of research methods in social sciences. To address this important limitation, the use of causal inference techniques has become widespread in recent years. Causal inference establishes a set of statistical procedures for drawing causal conclusions in non-experimental research. Despite its wide popularity and diffusion in the social and health sciences, its use in educational research is still marginal. Thus, this paper introduces the main causal inference techniques available to the educational researcher when observational panel data are available. After addressing the key features and potential of propensity score matching, instrumental variables, and regression discontinuity design, we present an example application of each of these techniques. We used the available databases from the PISA 2018 assessments. We included the mathematical competence as the dependent variable in all the three models implemented. Given the different characteristics of each of these techniques, the independent variable used is different in the three models applied: attendance to early childhood education in propensity score matching; student academic expectations in instrumental variables; and size of the community in which the school is located in regression discontinuity design. The article concludes by discussing the potential of this set of techniques, taking into account the needs and methodological procedures most commonly applied in educational research.

Keywords: causal analysis, statistical methods, evaluation, data analysis.

Resumen

La suposición de relaciones causa-efecto en la investigación *ex post facto* es un problema ampliamente conocido en el ámbito de la metodología de investigación en ciencias sociales. Para abordar esta importante limitación, en los últimos años se ha extendido el empleo de técnicas de inferencia causal, un conjunto de procedimientos estadísticos establecidos para poder extraer conclusiones causales en investigaciones no experimentales. A pesar de su amplia popularidad y difusión en el ámbito de las ciencias sociales y de la salud, su uso en investigación educativa es todavía marginal. Así, este trabajo introduce las principales técnicas de inferencia causal disponibles para el investigador educativo cuando dispone de datos observacionales de panel. Tras abordar las características clave y el potencial de las técnicas de emparejamiento por puntuación de propensión, variables instrumentales y diseño de regresión discontinua, se presenta un ejemplo de aplicación de cada una de ellas empleando las bases de datos obtenidas en la evaluación PISA 2018. Se incluye la competencia matemática como variable dependiente en todos los modelos propuestos. Dada las diferentes características de cada una de estas técnicas, la variable independiente empleada varía en los tres modelos aplicados: asistencia a educación infantil en el emparejamiento por puntuación de propensión, expectativas académicas del estudiante en variables instrumentales y tamaño del municipio en el que se encuentra la escuela en diseño de regresión discontinua. Se concluye el artículo discutiendo el potencial de este conjunto de técnicas, teniendo en cuenta las necesidades y procedimientos metodológicos más habitualmente aplicados en la investigación educativa.

Palabras clave: Análisis causal, metodología estadística, evaluación, análisis de datos.

Received/Recibido

Dec 13, 2023

Approved /Aprobado

Jul 12, 2023

Published/Publicado

Dec 11, 2023

Corresponding author / Autor de contacto: Fernando Martínez Abad. Instituto Universitario de Ciencias de la Educación, Paseo Canalejas, 169, 37008, Salamanca, España; correo-e: fma@usal.es

Resumo

A assunção de relações de causa-efeito na investigação *ex post facto* é um problema amplamente conhecido no domínio da metodologia de investigação em ciências sociais. Para fazer face a esta importante limitação, a utilização de técnicas de inferência causal, um conjunto de procedimentos estatísticos estabelecidos para poder tirar conclusões causais em investigações não experimentais, tem vindo a generalizar-se nos últimos anos. Apesar da sua grande popularidade e disseminação no âmbito das ciências sociais e da saúde, a sua utilização em investigação educativa é ainda marginal. Assim, este documento introduz as principais técnicas de inferência causal disponíveis para o investigador educativo quando existem dados observacionais de painel. Depois de discutir as principais características e o potencial das técnicas de correspondência por pontuação de propensão, variáveis instrumentais e conceção de regressão descontínua, apresenta-se um exemplo da aplicação de cada uma delas utilizando as bases de dados obtidas na avaliação PISA 2018. A competência matemática é incluída como variável dependente em todos os modelos propostos. Dadas as diferentes características de cada uma destas técnicas, a variável independente utilizada varia nos três modelos aplicados: assistência ao ensino infantil na correspondência por pontuação de propensão, expectativas académicas do estudante em variáveis instrumentais e dimensão do município em que a escola se localiza em conceção de regressão descontínua. O artigo conclui discutindo o potencial deste conjunto de técnicas, tendo em conta as necessidades e os procedimentos metodológicos mais comumente aplicados na investigação educativa.

Palavras-chave: Análise causal, metodologia estatística, avaliação, análise de dados.

摘要

事后回溯研究中的因果关系假设是社会科学研究方法领域普遍公认的问题。为了了解这一局限性，最近几年开始广泛地使用因果推断技术。这项技术是指通过一系列已建立的统计学过程从非实验研究中提取因果结论的技术。虽然该技术在社会科学和健康科学领域有着广泛的认知度和使用度，但在教育领域它还处在边缘位置。因此，该研究导入可用的主要因果推断技术，帮助教育研究者分析观察性面板数据。研究首先详述了倾向性得分匹配、工具变量、断点回归设计的主要特点和潜力，然后分别展示了这些技术在 2018 年国际学生评估项目 (PISA2018) 测试数据上的应用示例。在所有的建议模型中，数学能力都作为因变量出现。考虑到每项技术的特殊性，三种模型使用不同的自变量：在倾向性得分匹配模型中的幼儿教育出勤率；在工具变量模型中的学生学业期待以及断点回归设计中学校所在的城区规模。在考虑到教学研究需求和常规方法论应用过程的基础上，该研究还对一系列技术的潜力进行了讨论。

关键词: 技能、教师培养、学生、学习计划

ملخص

نطاق واسع في مجال منهجية البحث في العلوم يحد افتراض العلاقات بين السبب والنتيجة في الأبحاث بأثر رجعي مشكلة معروفة على الاجتماعية. ولمعالجة هذا القيد المهم، انتشر في السنوات الأخيرة استخدام تقنيات الاستدلال السببي، وهي مجموعة من الإجراءات الإحصائية التي تم وضعها لتكون قادرة على استخلاص استنتاجات سببية في البحوث غير التجريبية. وعلى الرغم من شعبيتها وانتشارها الواسع في مجال العلوم الاجتماعية والصحية، إلا أن استخدامها في البحوث التربوية لا يزال هامشياً. وبالتالي، يقدم هذا العمل تقنيات الاستدلال السببي الرئيسية، المتاحة للباحث التربوي عند توفر بيانات لوحة المراقبة بعد معالجة الخصائص الرئيسية وإمكانات مطابقة درجات الميل، والمتغيرات الآلية، تم. PISA2018 وتقنيات تصميم الانحدار المتقطع، يتم تقديم مثال لتطبيق كل منها باستخدام قواعد البيانات التي تم الحصول عليها في تقييم تضمنين الكفاءة الرياضية كمتغير تابع في جميع النماذج المقترحة. وبالنظر إلى الخصائص المختلفة لكل من هذه التقنيات، فإن المتغير المستقل المستخدم يختلف في النماذج الثلاثة المطبقة: الحضور في التعليم في مرحلة الطفولة المبكرة في مطابقة درجات الميل، والتوقعات الأكاديمية للطلاب في المتغيرات المفيدة والحجم البلدية التي تقع فيها المدرسة في تصميم تراجمي منقطع. ويختتم المقال بمناقشة إمكانات هذه المجموعة من التقنيات، مع الأخذ في الاعتبار الاحتياجات والإجراءات المنهجية الأكثر شيوعاً في البحث التربوي.

الكلمات الدالة: التحليل السببي، المنهجية الإحصائية، التقييم تحليل البيانات

Introduction

Causal inference in non-experimental research:

Experimental designs are used to verify research hypotheses and attribute causal relationships between two or more variables. Although some designs that would allow causality to be attributed in less-controlled environments are conceptualised in social sciences (Campbell & Stanley, 1963), experimental research fundamentally aims to strictly control research design: random group assignment, total manipulation of the independent variable, objective dependent variable measurement, and maximum control of biases associated with extraneous variable or covariates identified. However, strictly experimental designs are often not feasible in socio-educational research due to various operational, ethical or economic grounds, etc.

When randomisation and thorough variable control are not possible, observational studies are commonly conducted, also known as non-experimental or *ex post facto*. Literature identifies the essential limitation of this type of studies in their *ex post facto* nature (Altman, 2020; Kerlinger & Lee, 1999): given that groups in these studies are predetermined by prior sample conditions, the different characteristics of the subjects of each group lead to bias when comparing them.

The use of causal inference statistical techniques has become popular in recent years. These aim to overcome the correlational scope associated with non-experimental designs, proposing different procedures to control this comparison bias (Antonakis et al., 2010; Imai et al., 2011; Imbens & Rubin, 2015; Pearl & Mackenzie, 2018). The most widespread techniques in social research for causal inferences with cross-sectional observational data are:

- Propensity Score Matching (PSM)
- Instrumental Variables (IV)
- Regression Discontinuity Design (RDD)

As these techniques have been developed from clearly differentiated approaches, it is important to select the most suitable based on research needs, the nature of variables available and causal model to be validated. So given their emerging nature and potential in educational research, this paper aims to analyse the characteristics and possibilities of PSM, IV and RDD techniques for causal inference. Following a conceptual review of the three techniques, their use will be exemplified with three models proposed based on data from the Spanish PISA 2018 sample of students and schools.

Dissemination of causal inference in educational research

The use of causal inference techniques in educational research is currently limited with low levels of dissemination compared to other similar areas of social and health sciences (figure 1).

In fact, the evolution of causal inference papers published in educational journals has been very discreet during this first quarter of the century (figure 2) with a clear downward trend in recent years. Studies using Spanish samples for large-scale assessment are few and far between: most apply instrumental variables (Castro Aristizabal et al., 2017; Choi et al., 2012; Cordero & Gil-Izquierdo, 2018; Lopez-Agudo et al., 2021) and use of propensity score matching is marginal (Crespo-Cebada et al., 2014; Ignacio García-Pérez & Hidalgo-Hidalgo, 2017), while no studies use regression discontinuity designs.

Figure 1. Areas with more than 50 publications on “causal inference” (Web of Science, 6/12/2022)

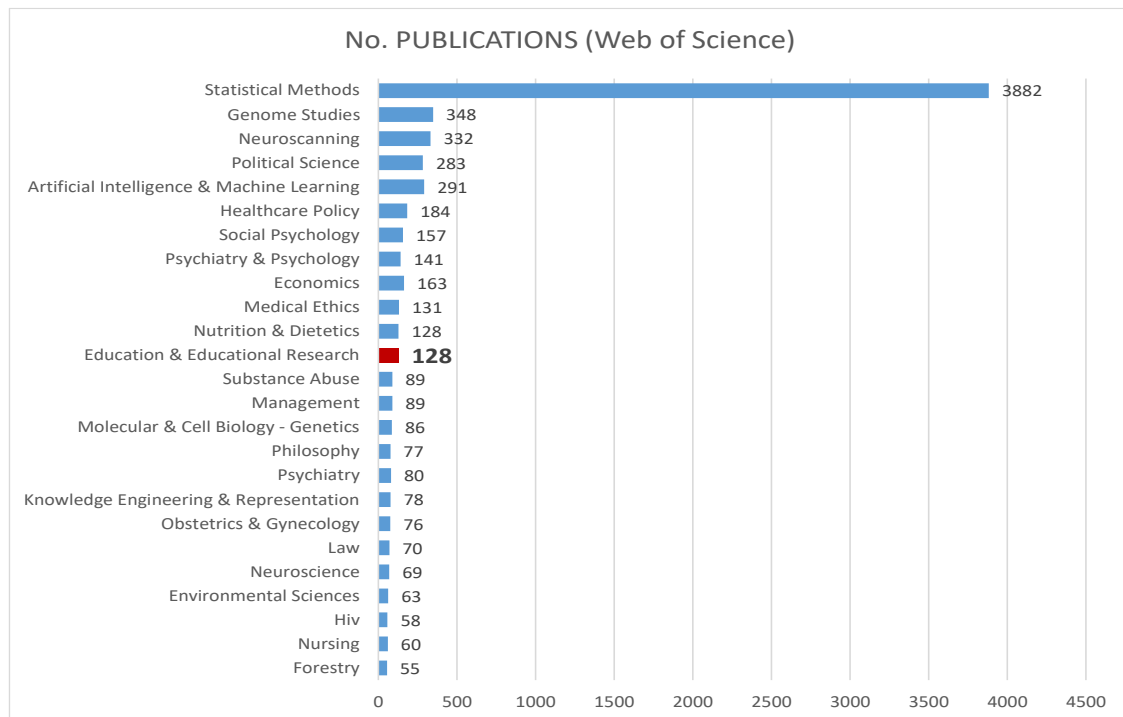
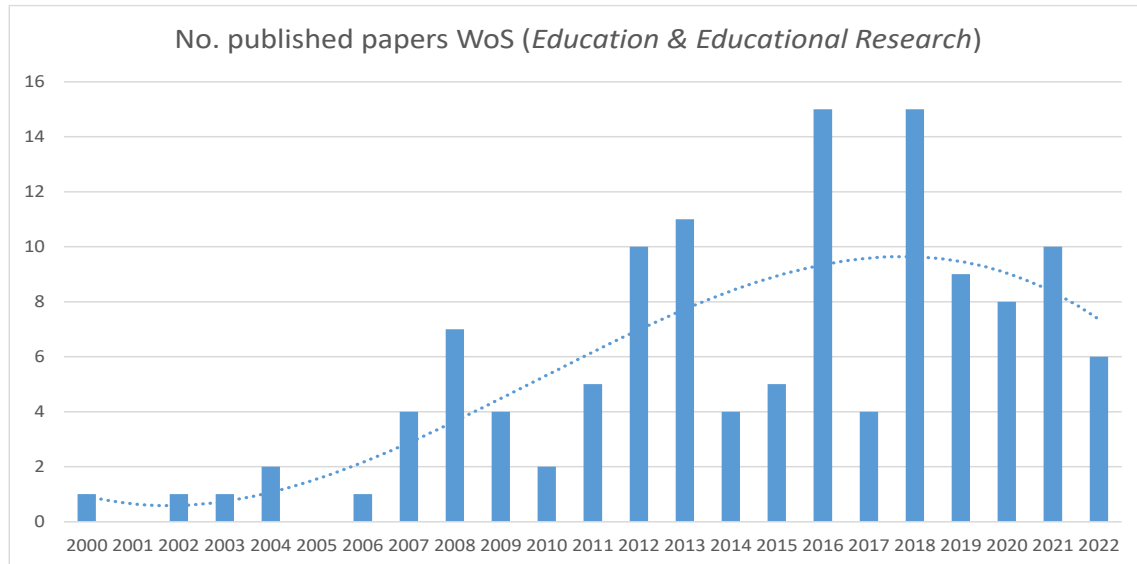


Figure 2. Publications on “causal inference” in the field of ‘Education & Educational Research’ (Web of Science, 6/12/2022)



Propensity Score Matching: grouping of ‘twin’ pairs

Designs to estimate causal effects of treatment in relation to a control, considering *treatment* and *control* as two different levels of the independent variable, focus on all treatment activities that precede the trait measured in the dependent variable or result

(Rosenbaum & Rubin, 2022). Given that we cannot observe the results of treated and controlled individuals, these designs are applied under the assumption that the dependent variable would remain stable if all treatment activities were not carried out. This logic can be defined as counterfactual reasoning (Rubin, 1974). A fundamental

problem is that this potential effect on the dependent variable is not directly observable if the *control* level were applied to the group instead of *treatment* level. Therefore, the control group is included to guarantee correct estimation of causal effects in experimental research, seeking maximum homogeneity between the experimental and control groups (Kaplan, 2016). Both groups must have similar distributions in covariates that could affect the proposed causal relationship; if any of these variables cannot be controlled, random distribution of subjects to control and experimental conditions is trusted.

However, there are practical and ethical constraints in social science research that hinder or even prevent the definition of *treatment* and *control* conditions in field work. This research is forced to use observational studies with non-experimental designs or, in the best case scenario, pre-experimental designs.

PSM is a result of this logic; it proposes matching subjects from two different groups based on their homogeneity in terms of a set of covariates. The two groups theoretically establish the control and experimental condition (e.g., having repeated a year or not, studying in a private or public school, receiving early childhood education or not, etc.), and covariates are understood as the variables controlled in the experiment. After matching, the subject pairs in the control and experimental groups can be considered as ‘twins’ and conceptually it is possible to verify the causal effect of treatment on the dependent variable.

Subjects are matched based on the propensity score (PS) of each subject in both groups, defined in Rosenbaum & Rubin (1986) as the conditional probability of assignment to treatment, given a set of covariates:

$$PS = p(Z = 1|X_i)$$

where variable *Z* refers to the subject belonging to the experimental (*Z*=1) or control group (*Z*=0), and *X_i* are the set of covariates defined.

Two fundamental aspects are assumed in applying PSM (Rosenbaum & Rubin, 2022; Rutkowski & Delandshere, 2016):

1. **Ignorability:** all key covariates affecting the causal relationship studied are identified and controlled. Any key covariates not observed may lead to significant biases in group comparability (Kaplan, 2016). Thorough planning is essential in this type of research given that compliance with this assumption cannot be directly tested.
2. **Overlap:** propensity score values (probability of the subject receiving treatment) for each pair are balanced. The most widely accepted statistic to validate this is the standardised mean difference (SMD), which indicates the difference in means between the two groups generated in PSM (Ali et al., 2014).

After complying with these assumptions, the effects of treatment on the dependent variable can be estimated by applying the appropriate hypothesis test. Literature available reveals debate on the use of inferential techniques for independent groups or for related groups (Austin, 2011); this decision depends on the researcher.

Instrumental Variables: independent variable endogeneity control

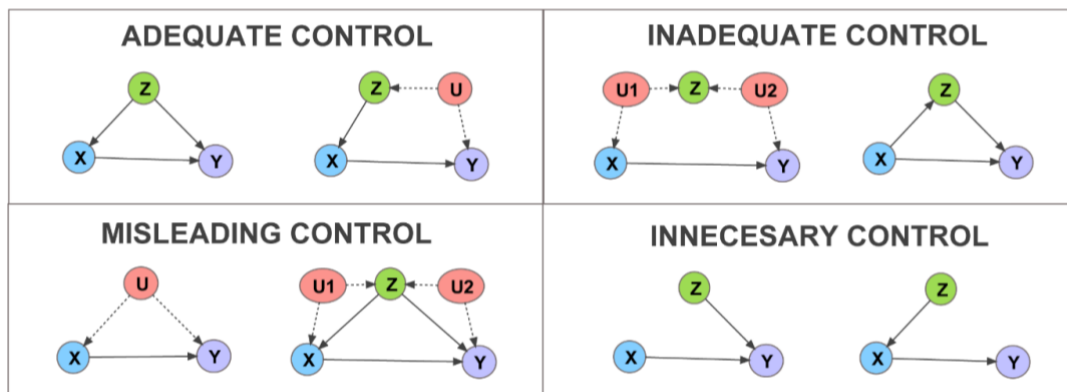
While PSM is suitable with a dichotomous variable that establishes an experimental situation and a control situation, the IV technique was developed with the attribution of causality in regression models of one or more independent variables (*X*) on a dependent variable (*Y*) in mind. Even though regression normally uses observational panel data, and is therefore correlational, it is often assumed as a *causal model* that leads to incorrect causal inference assumptions. This is due to the problem of independent variable endogeneity caused when *X* has a significant correlation with the difference between the real score and the score predicted by the regression line (*Y* error), leading to biased estimates of model parameters. Endogeneity bias has five

main causes (Maydeu-Olivares et al., 2020; Wooldridge, 2010):

1. **Omitted variable.** Variables related to X and Y are omitted from the model.
2. **Error measuring X.** Variable X* is obtained due to this bias, estimating model $X^* \rightarrow Y$ instead of the desired model $X \rightarrow Y$.
3. **Reverse causality.** The actual causal effect direction is $Y \rightarrow X$.
4. **Reciprocal causality.** There are mutual causal effects, both $X \rightarrow Y$ and $Y \rightarrow X$, but the latter is omitted in the empirical model.
5. **Selection.** Sample or treatment selection is not random, which affects the $X \rightarrow Y$ estimate.

The IV procedure aims to control endogeneity by using exogenous variables Z (Cinelli et al., 2022). Thus, as shown in figure 3, the conditions are established under which a set of Z variables can be considered to establish adequate, inadequate or misleading control of the causal relationship between a predictor X and result Y, considering the possible existence of unobserved variables U (Cinelli et al., 2022; Huenermund et al., 2022). These same authors also establish certain conditions in which control is unnecessary. Generally, the model will establish adequate control of causality: (1) if Z is a cause variable of both X and Y, with no other unobserved variables that are in turn cause of Z and Y or Z and X; and (2) if Z is a cause variable of X but not Y, and there are U variables that are in turn cause of Z and Y.

Figure 3. Control of endogeneity using instrumental variables



These Z variables are what are known as instrumental variables. They are exogenous to the $X \rightarrow Y$ regression model with the function of only selecting joint variability between X and Z ($\sigma^2_{X|Z}$). Therefore, if adequate control has been established, it is possible to control the correlation between X and dependent variable residuals. IV regression is carried out in two stages for this reason (Angrist et al., 1996; Pokropek, 2016) and the most common method for estimation is the two-stage least squares or 2SLS method (Maydeu-Olivares et al., 2020).

1. Perform the regression of X on instrumental variables Z:

$$X_i = \pi_0 + \pi_1 Z_i + U_i$$

where U_i refers to the regression error term.

2. Regression of Y on \hat{X} , which, thanks to the previous estimation, includes the common variance of X with Z:

$$Y_i = \beta_0 + \beta_1 \hat{X}_i + E_i$$

where \hat{X} refers to the score predicted in the first stage and E_i is the model error term.

In addition to testing a model with the appropriate conditions for Z to establish adequate control, as shown in figure 2, other

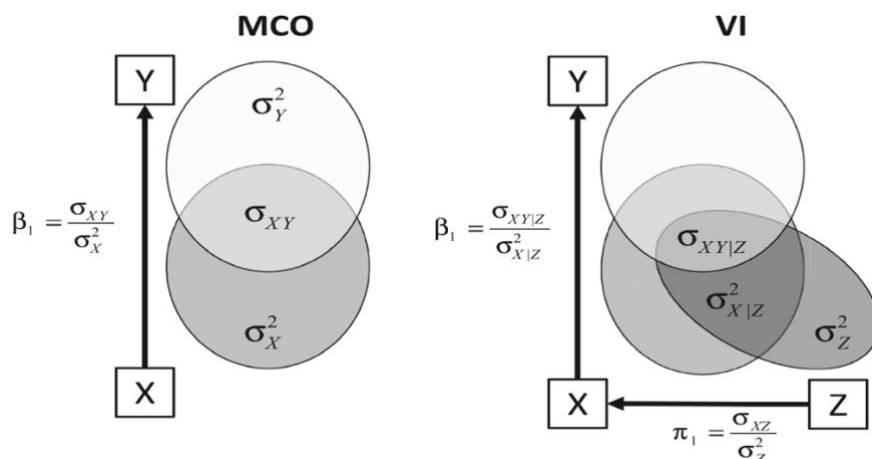
fundamental assumptions are made in IV regression:

- **Endogeneity** ($\sigma_{XE}^2 \neq 0$): 2SLS estimation makes sense as an alternative to ordinary least squares (OLC) estimation when there is endogeneity between X and residuals. If X is exogenous, both estimators are consistent and OLC is more efficient, so using OLC is recommended.
- **Relevance** ($\sigma_{XZ}^2 \neq 0$): there must be a strong correlation between the instruments and independent variable.
- **Exogeneity** ($\sigma_{ZE}^2 = 0$): instruments cannot be correlated with variable Y, i.e.,

there can be no correlation between them and the error. Exogeneity can be tested by applying an overidentification contrast such as Sargan's test (Jin, 2022), although for this there must be more instrumental variables (IV) than independent variables (X).

Figure 4 illustrates the conceptual differences between the IV and OLS models. While OLS estimates parameter of interest β (which theoretically indicates the $X \rightarrow Y$ causal effect) with the full variance of X, IV estimates β based on X variability conditional on instrument Z.

Figure 4. OLS and IV conceptual models (Pokropek, 2016).



In conclusion, when complying with the proposed conditions, in IV models it is possible to consider parameter β_1 as the causal effect X exercises over Y.

Regression Discontinuity Design: control and experiment assigned based on X

RDD is applicable when there is a continuous observed variable X that determines the assignment of subjects to treatment T, or at least influences this assignment. According to this logic, RDD aims to estimate the effects of a treatment by separating participants into two groups - treatment and control- according to a critical value X (see Figure 4), called k (Imbens &

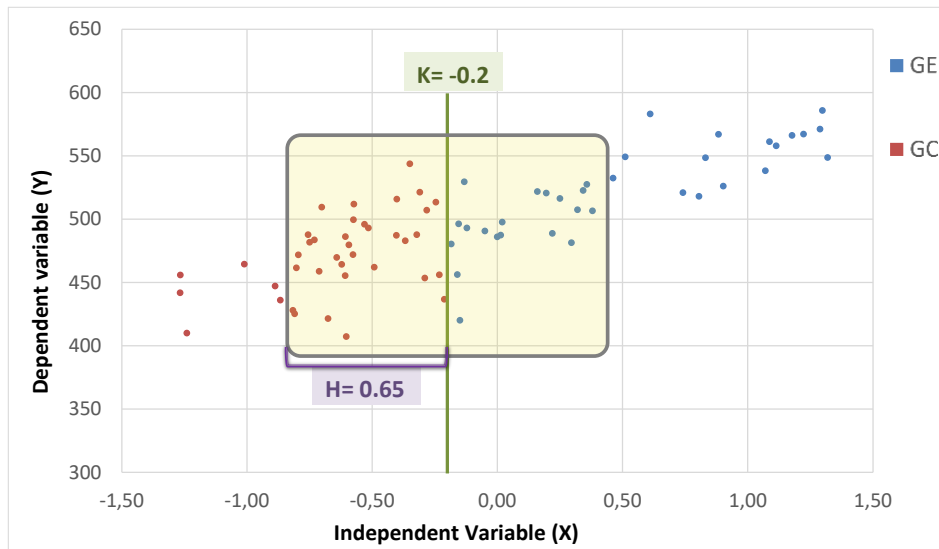
Lemieux, 2008; Lee & Lemieux, 2010). In other words, the effects of treatment T will be estimated based on a subsample of similar subjects in X, with scores around k . In addition to value k , the interval width of scores around k included in the analysis will have to be estimated. This interval width will be called h (Imbens & Kalyanaraman, 2012).

RDD therefore analyses the jump or discontinuity in the distribution of dependent variable Y between X scores immediately below and above k . The RDD model obtains the average effect of this discontinuity, in other words, the effect of treatment, estimating the weight of T in a regression model that includes T, X and $T \cdot X$ as independent variables. This

model only applies to subjects with X score within the range $(k - h, k + h)$. Figure 5 shows the visual interpretation of h and k in a scatter plot. The causal effect of treatment (subjects in

blue) on Y will only be estimated with subjects with X_i scores in the range $(-0.85, 0.45)$ as the cut-point is -0.2 and the interval width ± 0.65 .

Figure 5. Control of endogeneity using instrumental variables



Establishing the values of h and k is therefore essential in RDD. Statistical procedures are available to estimate optimum h width; the most popular is the procedure proposed by Imbens & Kalyanaraman (2012). In relation to k , this is normally established by the researcher under the fundamental assumption that the density of X around k is continuous. Setting the k value is simple when, as in figure 4, all subjects not assigned to treatment are to the left and all members assigned to treatment are to the right; this is known as sharp RDD. However, sometimes despite observing a clear trend, there are subjects assigned and not assigned to treatment both to the right and left of k . In other words, X does not always determine assignment to treatment for all subjects, rather there is an error in T_i assignment for some subjects; this is known as fuzzy RDD.

The general linear model of sharp RDD is as follows¹:

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 (X_i - k) + \beta_3 [T_i * (X_i - k)] + \varepsilon_i$$

where Y_i is the dependent variable; T_i a dummy variable with values 1 and 0 for subjects receiving treatment or not; $(X_i - h)$ scores of subjects in the continuous variable that determines T ; and ε_i the error term. Scores X_i are seen to centre around k so that subjects to the left of k will have negative scores in the model $(X_i - k)$ and subjects to the right will have positive scores $(X_i + k)$. Thus, the model parameter of interest for estimating the causal effect of treatment is β_1 , considering the average treatment effect (ATE).

In the case of fuzzy RDD, a two-stage procedure similar to IV must be applied as T_i has no exact dependence on X_i . In the first stage, dependent variable assignment or non-assignment to treatment is estimated based on $(X_i - k)$:

$$\hat{T}_i = \beta_0 + \beta_1 D_i + \beta_2 (X_i - k) + \beta_3 [D_i * (X_i - k)]$$

¹ It is important to bear in mind that the regression model is only calculated in the subset of the sample within the score

range $(h - k, h + k)$, using the estimation method commonly known as *Local Ordinary Least Squares*.

where \hat{T}_i is the estimate indicating whether the subject is treated or not and D_i is a dummy variable with values 0 and 1 for subjects below and above k respectively.

The final model for obtaining ATE will be applied in the second stage:

$$Y_i = \gamma_0 + \gamma_1 \hat{T}_i + \gamma_2 (X_i - k) + \gamma_3 [D_i * (X_i - k)] + \varepsilon_i$$

In this case, ATE is parameter γ_1 .

Method

Research objective and hypotheses

The **objective** of this study is to provide a methodological proposal for applying causal inference techniques to panel data analysis as part of *ex post facto* research.

Results present a brief example of each of the three techniques proposed based on the analysis of secondary data from the PISA 2018 assessment (OECD, 2019). The following research hypotheses are proposed for each model:

H1. Early childhood education exerts a positive causal effect on a student's academic performance during secondary education, regardless of key socio-demographic and economic factors (PSM).

H2. Future academic expectations of secondary students have direct and significant effects on their academic performance, even controlling endogeneity with family socio-economic status (IV).

H3. The size of the school's municipality (rural or urban) has no causal effect on secondary student performance (RDD); correlational effects found are due to family socio-economic status.

Participants

The reference population was Spanish secondary education students aged 15 at the time of the PISA 2018 assessment. The main sample comprised $n=35,943$ students and $m=1,089$ schools assessed in Spain.

Variables

Table 1 shows the variables used in each model presented. To simplify the interpretation of results, skills or *performance in mathematics* was included as a dependent variable in all cases. *Family socio-economic status* (SES) was included as a covariate in the three models. Independent variables and covariates were selected according to compliance with the prior assumptions of each model.

Table 1. Variables included in models

	PSM	IV	RDD
Dependent variables	Performance in mathematics (<i>PVIMATH</i>)		
Independent variable	Early Childhood Education (<i>DURECEC</i>)	Academic expectations (<i>ST225Q</i>)	Municipality size (<i>SC001Q01TA</i>)
	Socio-economic status (<i>ESCS</i>)		
Covariates	Month of birth (<i>ST003D02T</i>) Immigrant status (<i>IMMIG</i>) Grade repetition (<i>REPEAT</i>) Gender (<i>ST004D01T</i>)	-	-

Covariate *academic expectations* was recoded from the six variables provided in PISA: each student was given a score ranging from 1 to 6² based on completion of highest studies expected by the student (1=Compulsory Secondary Education; 2=Intermediate VT; 3=Baccalaureate; 5=Advanced VT; 6=University). The *Immigrant status* covariate was recoded into a dichotomous variable: all students born in Spain were considered native (native and second-generation immigrants in PISA) and the rest immigrants. In PISA, the *gender* covariate only includes the categories of “male” and “female”.

Procedure and data analysis

With a level $\alpha=5\%$, the following analytical procedures were applied:

- **Sample weights:** as the PSM and RDD models select a subsample based on the initial sample, no sample weights were included as a weighting variable.
- **Lost values:** no lost values were attributed and listwise deletion was applied in the models. The proportion of lost values in these variables justifies this decision:
 - No lost value in performance, month of birth and gender
 - Grade repetition (1.4%), ESCS (1.8%), municipality size (3.2%), immigrant status (3.1%), expectations (4.3%), early childhood education (14.7%)
- **Plausible values:** the OECD (2009) points out that, although the use of plausible value sets and replicate sample weights is the most efficient alternative for estimating parameters and standard errors, the direct use of one of the plausible values can provide unbiased parameter estimates. A plausible value was selected in line with the decision regarding the use of sample weights. Thus, the three models used the first

plausible value of performance in mathematics.

The following procedure was used when applying each of the techniques:

- **PSM:** SMD is used as an adjustment statistic of the balance or distance between groups after matching. According to Zhang et al. (2019), $SMD < .1$ is an indicator of good balance. Distance of squares and interactions between covariates is also verified (Belitser et al., 2011). After checking balance, the performance of each group is compared using a t-test for related groups, including effect size (Cohen's d).
- **IV:** the assumption of endogeneity is verified with the Wu-Hausman test (Hill et al., 2021); the alternative hypothesis (H_0) is that there is no endogeneity. The weak instrument test is applied for the assumption of relevance (Maydeu-Olivares et al., 2020); in this case, the alternative hypothesis is that the instrument is weak. Relevance analysis is complemented with the F statistic (Stock & Yogo, 2005), where instruments with $F > 10$ are strong. Exogeneity is not verified as the model includes a single IV. After validating these assumptions, the parameter associated with academic expectations is compared in the OLS and IV models.
- **RDD:** the h value is established using the Imbens-Kalyanaraman (2012) procedure and the assumption of continuity of the X variable around k is validated with McCrary's test (2008); the alternative hypothesis is that the variable is continuous. Finally, fuzzy RDD is applied to obtain the average effect of treatment T (school in a rural or urban environment).

² Value 4 was not included in this variable as there are no ISCED 4 studies (post-secondary non-tertiary, standardised

classification CINE97) in Spain, a level that is included in PISA.

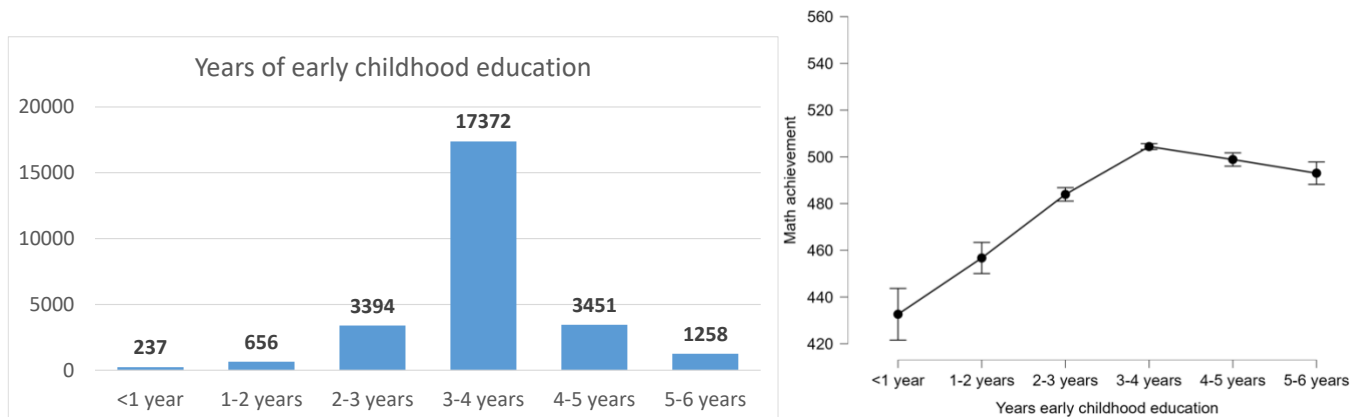
Results

PSM - Causal effect of early childhood education on performance

As shown in figure 6, PISA includes early childhood education in number of years. Based

on the descriptive and inferential analysis, early childhood education is seen to have significant effects on performance, with a direct relationship of small effects ($\eta^2=0.018$).

Figure 6. Distribution of years of early childhood education and relationship with performance



The benefits of early childhood education stabilises from 2-3 years, so the PSM will divide the sample into two experimental conditions:

- Control group (CG): students with at least 2 years of education.
- Treatment or experimental group (EG): students with less than 2 years of education.

Table 2 indicates the differences between the CG and EG before and after applying PSM. While the initial sample shows significant differences with moderate effect sizes, after PSM both groups are even in terms of the covariates. After applying PSM, matched samples are distributed randomly in the control and treatment groups. In fact, while the SMD value of the distance in the initial sample was .563, in the matched sample it is .001.

In the initial database, EG students have clearly lower socio-economic status, double

the grade repetition rate of the other group, a higher proportion of boys than girls, and have higher immigrant student rates (observed in both immigrant status and language at home). Thus, PSM has selected CG subjects with similar characteristics to each member of the EG, rejecting the remaining subjects with no suitable matching.

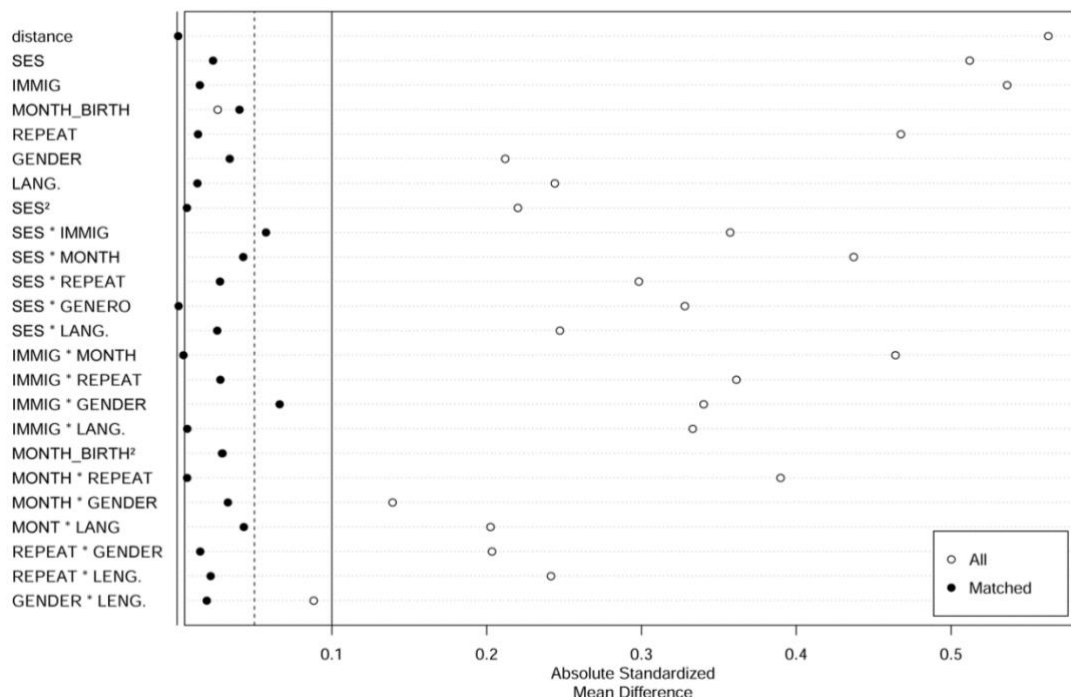
Before continuing, note that as the EG had $n_{EG} = 893$ subjects, when matching the selected sample of CG students is also 893 subjects, discarding the remaining sample subjects in this group ($n_{CG} = 25,475$). This procedure is key to PSM since pairs have been selected in the CG that are most similar or *twins* with their peers in the EG based on the covariates incorporated. Figure 7 confirms that distances between EG and CG in the PSM sample are adequate in terms of variables and their interactions and squares. Undesirable spurious effects have been controlled and causal effects can be estimated by considering both groups to be homogeneous.

Table 2. Balancing groups using PSM

		Ave. CG	Ave. EG	SMD	p.	r
SES ^a	Initial	-0.511	0.033	-0.512	<.001	.295
	PSM	-0.511	-0.536	0.023	.794	.010
Month of birth ^a	Initial	6.642	6.551	0.026	.429	-.015
	PSM	6.642	6.780	-0.040	.505	-.027
Immigrant status ^b	Initial	29.68%	5.18%	0.536	<.001	.186
	PSM	29.68%	29.00%	0.015	.755	.008
Grade repetition ^b	Initial	44.01%	20.80%	0.468	<.001	.102
	PSM	44.01%	43.34%	0.014	.775	.007
Gender ^b	Initial	42.22%	52.69%	-0.212	<.001	-.038
	PSM	42.22%	40.54%	0.034	.471	.017
Language at home ^b	Initial	23.85%	13.45%	0.244	<.001	.055
	PSM	23.85%	24.41%	-0.013	.782	-.007

^a Scale variable: p-value and effect size based on Mann-Whitney U test in initial sample and Wilcoxon in PSM sample; ^b Dichotomous variable: p-value and effect size based on Chi-square

Figure 7. Distance between CG and EG in covariates and their interactions



Analysing the causal effects of the difference in performance between both groups is pertinent as the EG and CG distributions are equal in terms of covariates (table 3). Without applying PSM, significant differences favourable to students spending more time in early childhood education with

average effect sizes can be observed; significant differences but of lesser magnitude are maintained in the sample matched by PSM. In fact, the difference in averages between both groups drops from just over 50 points in the initial sample to slightly under 20 in the matched sample.

Table 3. Differences in performance by years of education. Initial sample and PSM

	CG mean	EG mean	t	p.	d
Initial sample^a	450.32	500.35	17.26	<.001	0.587
PSM matched sample^b	450.32	470.11	5.25	<.001	0.176

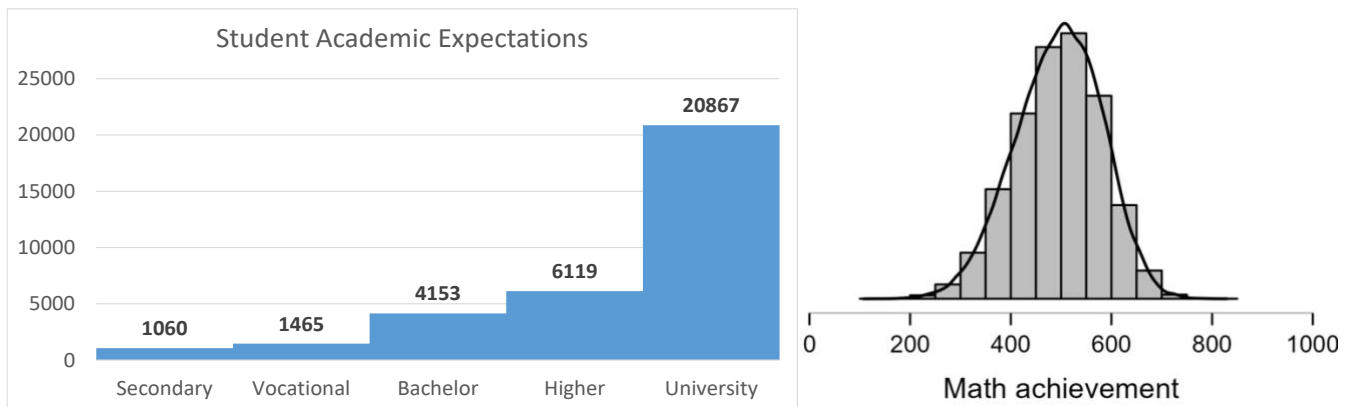
^a t-test for independent groups; ^b t-test for related groups

Therefore, after selecting CG subjects truly comparable to their EG peers using the PSM method, assuming that all key covariates have been included, we can state that there is a direct cause-effect relationship with small effects sizes between early childhood education and student academic performance at 15 years old.

Figure 8 shows the distribution of performance in mathematics (Y) and student academic expectations (X) variables. Most students (62%) say they want to go on to university studies, while a minimal proportion (3.1%) only want to complete compulsory secondary education level.

IV - Causal effect of academic expectations on performance

Figure 8. Distribution of performance in mathematics (Y) and expectations (X)



Compliance with the prior assumptions of the IV model is shown in table 4. Firstly, there is endogeneity as H_0 is rejected in the Wu-Hausman test. Evidence also shows that the

instrument is strong (rejection of H_0 on weak instruments and $F > 10$) and there is a high-intensity significant correlation with variable X, so it can be considered relevant.

Table 4. Verification of prior assumptions of the IV model

	Statistic	p.	F
Wu-Hausman (Endogeneity)	2093	<.001	-
Relevance	Weak instruments	<.001	4084.97
	Correlation (r_{xz})	.414	<.001

Table 5 presents the OLS and IV (2SLS estimate) models obtained. After controlling student academic expectations with their family SES, the direct effects of X over Y are seen to clearly increase. A one-unit increase in

X increases performance by over 86 points in the IV model, while it only rises 34 points in the OLS model.

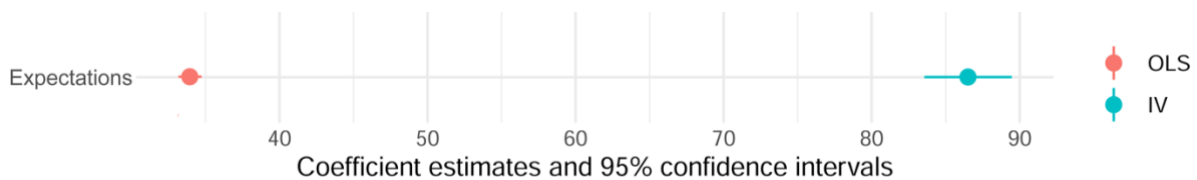
Table 5. OLS and IV regression model comparison

	OLS				IV (2SLS)			
	β	ET	t	p.	β	ET	t	p.
Intercept	347.97	1.81	192.51	<.001	121.18	7.00	17.31	<.001
Academic expectations	33.95	0.41	83.40	<.001	86.51	1.60	54.24	<.001

This change in academic expectation weight from one model to the other is more clearly illustrated in figure 9. Despite the confidence interval around the parameter being wider in IV due to an increase in standard error, the direct influence of student expectations on

performance is clearly superior. A one-point increase in student academic expectations in the IV model accounts for a rise of almost 87 points in performance, while it only rose 34 points in the OLS model.

Figure 9. Parameter associated with student expectations in the IV and OLS models

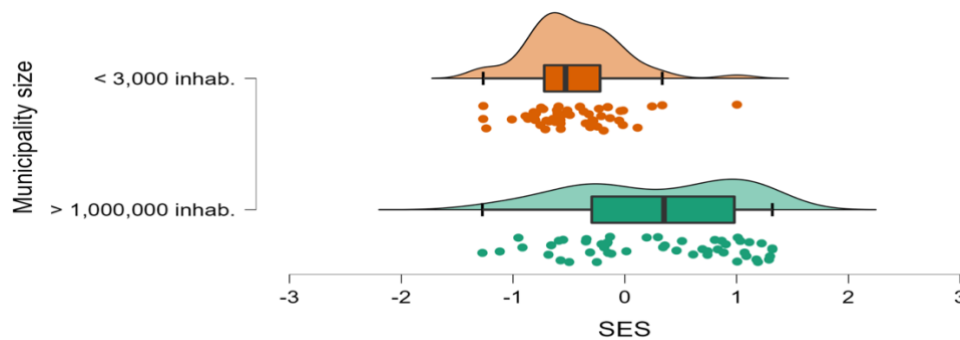


In conclusion, considering that endogeneity was controlled adequately (family SES is a cause variable of both student academic expectations and performance), the IV models shows that the causal effect of academic expectations on performance is direct and significant, even more intense than indicated in the OLS correlational model.

RDD - Causal effect of education in urban VS rural schools on performance

Figure 10 shows the distribution of schools by SES based on municipality size. An intense relationship is observed between both variables: while SES distribution in rural schools is extremely homogeneous and generally with scores under 0, large cities are more scattered and have higher SES values. Applying the *t*-test for independent variables shows significant differences with a very high effect size ($t = 6.19, p < .001, d = 1.23$).

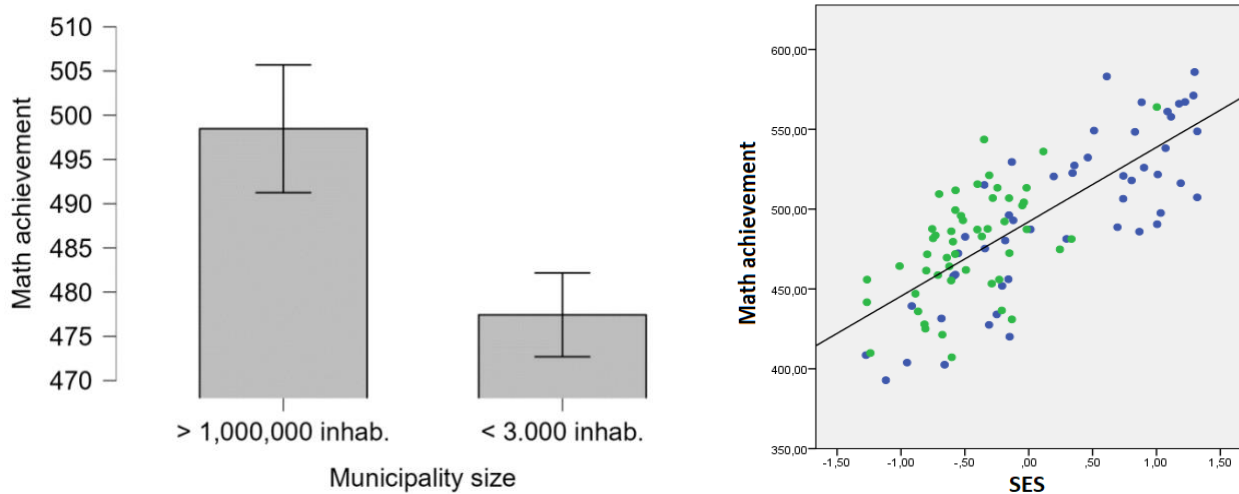
Figure 10. Rural and large city school distribution by SES



A significant relationship with moderate effect size can also be observed between performance and municipality size ($t = 2.43$, $p = .017$, $d = .49$), although these effects may be due to the spurious relationship of SES (figure 11). Applying the regression model with

municipality size and SES as predictor variables, municipality size β parameter is still significant ($\beta = 18.96$, $t = 2.82$, $p = .006$), as is SES ($\beta = 53.72$, $t = 11.25$, $p < .001$).

Figure 11. Performance in mathematics by school size and SES



We can thus affirm that there is a clear linear relationship between municipality size and performance, even controlling for SES. Now the issue is: does municipality size (large city VS village) have a causal effect on performance in mathematics?

Given the intense relationship between SES and municipality size, the fuzzy RDD technique seems to be a good choice. The model will include SES as variable X, municipality size as variable T, and performance as variable Y.

Firstly, k and h values are established. Given the SES distribution observed in both groups in figures 8 and 9, cut-point $k = 0$ appears to be adequate. Table 6 shows compliance with the continuity assumption of the SES variable and the estimated h value for $k = 0$. The null hypothesis of continuity around 0 is accepted and the SES score range is established (-.741, .714), so RDD can be applied. The model will be applied to the 66 schools within this score range.

Table 6. Compliance with continuity assumption and h value of the RDD model

McCrary test ($k=0$)		Cut-point h (Imbens-Kalyanaraman)
Z	p	
-1.917	.055	$h = .741$

Table 7 shows results obtained in both stages of the RDD regression model. The average effect of treatment is seen to be insignificant ($\gamma_1 = -32.10$, $p = .801$).

Table 7. Fuzzy RDD regression model

	Stage one					Stage two		
	β	ET	t	p	γ	ET	t	p
Intercept	0.631	0.12	5.24	<.001	508.06	76.57	6.64	<.001
D (dummy treat.)	-0.170	0.28	-0.60	.550	-	-	-	-
\hat{T} (instrument)	-	-	-	-	-32.10	126.86	-0.25	.801
SES	-0.047	0.391	-0.12	.905	25.61	32.44	0.79	.433
D*SES	-0.534	0.84	-0.63	.529	47.30	124.13	0.38	.704

Assuming endogeneity is adequately controlled in RDD, we can conclude that municipality size has no causal effects on student performance. These results contrast with the results of the correlational-inferential approach, which find significant effects in the relationship between both variables even controlling for SES.

Discussion and conclusions

Leading researchers in multivariate statistics stress the potential of causal inference techniques in testing causal effects in *ex post facto* samples (e.g., Imbens & Rubin, 2015; Rosenbaum & Rubin, 2022). Authors in this line of research even received the Nobel Prize for Economics³ in 2021. For years, international authors in the field of education have also been highlighting the use of these techniques in large-scale educational

assessments (e.g., Kaplan, 2016; Rutkowski & Delandshere, 2016). Given the current limited impact these techniques have in educational research, our objective was to show the possibilities of causal inference to analyse panel data, specifically with the Spanish PISA 2018 sample.

PSM results show that key covariates were selected to control endogeneity (Rosenbaum & Rubin, 2022). The use of PISA data (OECD, 2019) allowed this thorough selection of covariates and obtaining truly comparable treatment and control groups. Results confirm the first hypothesis (H1): early childhood education has a causal effect on Spanish student performance during secondary education. While systematic reviews and international meta-analyses point to this result (Barnett, 1998; McCoy et al., 2017), some previous studies with smaller, regional

³<https://www.nobelprize.org/uploads/2021/10/advanced-economicsprize2021.pdf>

samples reached similar conclusions in other countries also using PSM (Amadon et al., 2022; Barnett & Jung, 2021; Courtney et al., 2023).

The use of IV included the SES variable as an instrument to control endogeneity, a fundamental covariate in studies on factors associated with performance (Gamazo & Martínez-Abad, 2020; Martínez-Abad et al., 2020). Even though previous correlational and multivariate studies already observed direct and significant relationships between expectations and performance in secondary education (Levi et al., 2014; Sanders et al., 2001; Suárez-Álvarez et al., 2014), our results confirm that this relationship is causal (H2), with more significant causal effects than expected: the effects of academic expectations obtained in the IV are higher than with OLS, pointing to the bias associated with attributing causality based on one-stage regression models.

In response to H3, RDD was applied to estimate the effects of rural-urban school type (treatment) on the performance (variable Y) of Spanish students, considering how closely SES (covariate X) is associated with school type. Results once again point to the bias associated with using OLS regression; even with SES as a covariate in the OLS model, the effects of school type obtained in RDD are significantly different. Previous studies in other countries already pointed in this direction (Amini & Nivorozhkin, 2015; Song & Tan, 2022).

Given its educational nature, the main limitation of this paper is that it presents examples that are simple to apply without adding other covariates to control covariance. Education is complex and multivariate, which sometimes requires the use of broader, more comprehensive models than those used. Future research specifically focusing on each of the techniques studied would therefore be of interest so as to address the specific possibilities of each technique in depth and to explore the construction of more complex and, therefore, more comprehensive statistical models.

This paper shows that causal inference techniques are not excessively complex on a conceptual or technical level and proves that their use in applied research is simple with freely available statistical packages. Therefore, we understand that the marginal use of causal inference techniques by researchers applied to education science is due to a profound lack of knowledge. Specific, easily applicable examples of the three main causal inference techniques (PSM, IV and RDD) are presented so they can be easily replicated and transferred by other researchers in non-experimental studies.

Acknowledgements

Grant PID2021-125775NB-I00 funded by MCIN/AEI/10.13039/501100011033/ and by ERDF A Way of Making Europe

References

- Ali, M. S., Groenwold, R. H. H., Pestman, W. R., Belitser, S. V., Roes, K. C. B., Hoes, A. W., de Boer, A., & Klungel, O. H. (2014). Propensity score balance measures in pharmacoepidemiology: A simulation study. *Pharmacoepidemiology and Drug Safety*, 23(8), 802-811. <https://doi.org/10.1002/pds.3574>
- Altman, M. (2020). A more scientific approach to applied economics: Reconstructing statistical, analytical significance, and correlation analysis. *Economic Analysis and Policy*, 66, 315-324. <https://doi.org/10.1016/j.eap.2020.05.006>
- Amadon, S., Gormley, W. T., Claessens, A., Magnuson, K., Hummel-Price, D., & Romm, K. (2022). Does early childhood education help to improve high school outcomes? Results from Tulsa. *Child Development*, 93(4), e379-e395. <https://doi.org/10.1111/cdev.13752>
- Amini, C., & Nivorozhkin, E. (2015). The urban-rural divide in educational outcomes: Evidence from Russia. *International Journal of Educational Development*, 44, 118-133.

- <https://doi.org/10.1016/j.ijedudev.2015.07.006>
- Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434), 444-455. <https://doi.org/10.1080/01621459.1996.10476902>
- Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly*, 21(6), 1086-1120. <https://doi.org/10.1016/j.leaqua.2010.10.010>
- Austin, P. C. (2011). Comparing paired vs non-paired statistical methods of analyses when making inferences about absolute risk reductions in propensity-score matched samples. *Statistics in Medicine*, 30(11), 1292-1301. <https://doi.org/10.1002/sim.4200>
- Barnett, W. S. (1998). Long-Term Cognitive and Academic Effects of Early Childhood Education on Children in Poverty. *Preventive Medicine*, 27(2), 204-207. <https://doi.org/10.1006/pmed.1998.0275>
- Barnett, W. S., & Jung, K. (2021). Effects of New Jersey's Abbott preschool program on children's achievement, grade retention, and special education through tenth grade. *Early Childhood Research Quarterly*, 56, 248-259. <https://doi.org/10.1016/j.ecresq.2021.04.001>
- Belitser, S. V., Martens, E. P., Pestman, W. R., Groenwold, R. H. H., de Boer, A., & Klungel, O. H. (2011). Measuring balance and model selection in propensity score methods. *Pharmacoepidemiology and Drug Safety*, 20(11), 1115-1129. <https://doi.org/10.1002/pds.2188>
- Campbell, D. T., & Stanley, J. (1963). *Experimental and quasi-experimental designs for research*. Wadsworth Publishing.
- Castro Aristizabal, G., Giménez, G., & Pérez Ximénez-De-Embún, D. (2017). Educational inequalities in latin america, PISA 2012: Causes of differences in school performance between public and private schools. *Revista de Educación*, 2017(376), 33-59. Scopus. <https://doi.org/10.4438/1988-592X-RE-2017-376-343>
- Choi, A., Calero, J., & Escardíbul, J.-O. (2012). Private tutoring and academic achievement in Korea: An approach through PISA-2006. *KEDI Journal of Educational Policy*, 9(2), 299-322. Scopus.
- Cinelli, C., Forney, A., & Pearl, J. (2022). A crash course in good and bad controls. *Sociological Methods & Research*, 00491241221099552. <https://doi.org/10.1177/00491241221099552>
- Cordero, J. M., & Gil-Izquierdo, M. (2018). The effect of teaching strategies on student achievement: An analysis using TALIS-PISA-link. *Journal of Policy Modeling*, 40(6), 1313-1331. Scopus. <https://doi.org/10.1016/j.jpolmod.2018.04.003>
- Courtney, J. R., Garcia, J. T., Rowberry, J., Eckberg, N., Dinces, S. M., Lobaugh, C. S., & Tolman, R. T. (2023). Measuring impact of New Mexico prekindergarten on standardized test scores and high school graduation using propensity score matching. *International Journal of Child Care and Education Policy*, 17(1), 9. <https://doi.org/10.1186/s40723-023-00112-9>
- Crespo-Cebada, E., Pedraja-Chaparro, F., & Santín, D. (2014). Does school ownership matter? An unbiased efficiency comparison for regions of Spain. *Journal of Productivity Analysis*, 41(1), 153-172. Scopus. <https://doi.org/10.1007/s1123-013-0338-y>
- Gamazo, A., & Martínez-Abad, F. (2020). An Exploration of Factors Linked to Academic Performance in PISA 2018 Through Data

- Mining Techniques. *Frontiers in Psychology*, 11. <https://doi.org/10.3389/fpsyg.2020.575167>
- García-Pérez, J.I., & Hidalgo-Hidalgo, M. (2017). No student left behind? Evidence from the Programme for School Guidance in Spain. *Economics of Education Review*, 60, 97-111. Scopus. <https://doi.org/10.1016/j.econedurev.2017.08.006>
- Hill, A. D., Johnson, S. G., Greco, L. M., O'Boyle, E. H., & Walter, S. L. (2021). Endogeneity: A Review and agenda for the methodology-practice divide affecting micro and macro research. *Journal of Management*, 47(1), 105-143. <https://doi.org/10.1177/0149206320960533>
- Huenermund, P., Louw, B., & Rönkkö, M. (2022). The choice of control variables: How causal graphs can inform the decision. *Academy of Management Proceedings*, 2022(1), 15534. <https://doi.org/10.5465/AMBPP.2022.294>
- Imai, K., Keele, L., Tingley, D., & Yamamoto, T. (2011). Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies. *American Political Science Review*, 105(4), 765-789. <https://doi.org/10.1017/S0003055411000414>
- Imbens, G. W., & Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *Review of Economic Studies*, 79(3), 933-959. Scopus. <https://doi.org/10.1093/restud/rdr043>
- Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2), 615-635. Scopus. <https://doi.org/10.1016/j.jeconom.2007.05.001>
- Imbens, G. W., & Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*.
- Jin, S. (2022). On inconsistency of the overidentification test for the model-implied instrumental variable approach. *Structural Equation Modeling*. Scopus. <https://doi.org/10.1080/10705511.2022.2122978>
- Kaplan, D. (2016). Causal inference with large-scale assessments in education from a Bayesian perspective: A review and synthesis. *Large-Scale Assessments in Education*, 4(1). Scopus. <https://doi.org/10.1186/s40536-016-0022-6>
- Kerlinger, F. N., & Lee, H. (1999). *Foundations of behavioral research* (004 ed.). Wadsworth Publishing.
- Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, 48(2), 281-355. Scopus. <https://doi.org/10.1257/jel.48.2.281>
- Levi, U., Einav, M., Ziv, O., Raskind, I., & Margalit, M. (2014). Academic expectations and actual achievements: The roles of hope and effort. *European Journal of Psychology of Education*, 29(3), 367-386. <https://doi.org/10.1007/s10212-013-0203-4>
- Lopez-Agudo, L. A., González-Betancor, S. M., & Marcenaro-Gutierrez, O. D. (2021). Language at home and academic performance: The case of Spain. *Economic Analysis and Policy*, 69, 16-33. Scopus. <https://doi.org/10.1016/j.eap.2020.11.003>
- Maydeu-Olivares, A., Shi, D., & Fairchild, A. J. (2020). Estimating causal effects in linear regression models with observational data: The instrumental variables regression model. *Psychological Methods*, 25(2), 243-258. <https://doi.org/10.1037/met0000226>
- Martínez-Abad, F., Gamazo, A., & Rodríguez-Conde, M.-J. (2020). Educational Data Mining: Identification of factors associated with school effectiveness in PISA assessment. *Studies in Educational Evaluation*, 66, 100875. <https://doi.org/10.1016/j.stueduc.2020.100875>

- McCoy, D. C., Yoshikawa, H., Ziol-Guest, K. M., Duncan, G. J., Schindler, H. S., Magnuson, K., Yang, R., Koepf, A., & Shonkoff, J. P. (2017). Impacts of Early Childhood Education on Medium- and Long-Term Educational Outcomes. *Educational Researcher*, 46(8), 474-487. <https://doi.org/10.3102/0013189X17737739>
- McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142(2), 698-714. <https://doi.org/10.1016/j.jeconom.2007.05.005>
- OECD. (2009). *PISA Data Analysis Manual: SPSS, Second Edition*. Organisation for Economic Co-operation and Development. <https://doi.org/10.1787/19963777>
- OECD. (2019). *PISA 2018 Assessment and Analytical Framework*. OECD Publishing. <https://doi.org/10.1787/b25efab8-en>
- Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. Falta editorial
- Pokropek, A. (2016). Introduction to instrumental variables and their application to large-scale assessment data. *Large-Scale Assessments in Education*, 4(1). <https://doi.org/10.1186/s40536-016-0018-2>
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41-55. <https://doi.org/10.2307/2335942>
- Rosenbaum, P. R., & Rubin, D. B. (2022). Propensity scores in the design of observational studies for causal effects. *Biometrika*, asac054. <https://doi.org/10.1093/biomet/asac054>
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66, 688-701. <https://doi.org/10.1037/h0037350>
- Rutkowski, D., & Delandshere, G. (2016). Causal inferences with large scale assessment data: Using a validity framework. *Large-scale Assessments in Education*, 4(1), 6. <https://doi.org/10.1186/s40536-016-0019-1>
- Sanders, C. E., Field, T. M., & Diego, M. A. (2001). Adolescents' academic expectations and achievement. *Adolescence*, 36(144), 795-802.
- Song, Q., & Tan, C. Y. (2022). The association between family socioeconomic status and urban-rural and high-school attainment gaps: A logistic regression analysis of the China Family Panel Studies data. *British Educational Research Journal*, 48(6), 1102-1124. <https://doi.org/10.1002/berj.3817>
- Stock, J., & Yogo, M. (2005). Testing for weak instruments in linear iv regression. En D. W. K. Andrews, *Identification and Inference for Econometric Models* (pp. 80-108). Cambridge University Press.
- Suárez-Álvarez, J., Fernández-Alonso, R., & Muñoz, J. (2014). Self-concept, motivation, expectations, and socioeconomic level as predictors of academic performance in mathematics. *Learning and Individual Differences*, 30, 118-123. <https://doi.org/10.1016/j.lindif.2013.10.019>
- Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data, second edition*. Revisar
- Zhang, Z., Kim, H. J., Lonjon, G., & Zhu, Y. (2019). Balance diagnostics after propensity score matching. *Annals of Translational Medicine*, 7(1), 16. <https://doi.org/10.21037/atm.2018.12.10>

Apéndice. Código R modelos.

Propensity Score Matching

```
# Instalar y cargar paquete MatchIt
install.packages("MatchIt")
library(MatchIt)

# Obtener el modelo PSM e imprimir resultados en pantalla
m.out=matchit(ED_INF ~ ESCS + IMMIG_REC + ST003D02T + REPEAT + ST004D01T + LANG,
data=DDBB, mehtod = 'nearest')
summary(m.out, interactions=TRUE)

# Obtener gráfico de distancias en datos iniciales y PSM
graf<-summary(m.out, interactions=TRUE)
plot(graf)

# Obtener gráficos comparativos de densidades entre datos iniciales y PSM
plot(m.out, type="density")

# Guardar los resultados del PSM
psm <- match.data(m.out)
write.csv(psm, "C:\\Users\\Plasti-Atia\\Desktop\\psm.csv", row.names = TRUE, dec=".", sep="\t", na="")
```

Variables Instrumentales

```
# Instalar y cargar paquetes ivreg y AER
install.packages("ivreg")
install.packages("AER")
library(ivreg)
library(AER)

# Obtener el modelo IV e imprimir resultados en pantalla
IV<-ivreg(PV1MATH~ST_EXP|ESCS, data=DDBB)
summary(IV, df=Inf, vcov=sandwich, diagnostics=TRUE, test="Chisq")

# Obtener el modelo MCO e imprimir resultados comparativos MCO e IV
OLS<- OLS<-lm(DDBB~PV1MATH ~ ST_EXP, data=DDBB)
m_list <- list(OLS = OLS, IV = IV)
msummary(m_list)

# Obtener gráfico comparativo de los parámetros MCO e IV con intervalo de confianza
modelplot(m_list, coef_omit = "Intercept")
```

Diseño de Regresión Discontinua

```
# Instalar y cargar paquetes rdd y rdrobust
install.packages("rdd")
install.packages("rdrobust")
library(rdd)
library(rdrobust)


# Obtener el valor de la amplitud del intervalo h
IKbandwidth(DDBB$ESCS, DDBB$PV1MATH, cutpoint = 0, verbose = FALSE, kernel = "triangular")

# Comprobar supuesto de continuidad (test McCrary)
DCdensity(DDBB$NSE, 0, htest = TRUE)

# Obtener el modelo RDD e imprimir resultados completos
MRDD <- RDestimate(REND_MAT ~NSE+TAM_MUNICIPIO, data=DDBB, cutpoint =0,
verbose=TRUE, model=TRUE)

# Imprimir resultados resumidos del modelo RDD
summary(MRDD)
```

Authors / Autores

Martínez-Abad, Fernando (fma@usal.es)  0000-0002-1783-8198

Associate Professor in the Area of Research Methods and Diagnosis in Education at the University of Salamanca and Associate Director of the Master's Degree in Evaluation and Research in Learning Organizations and Contexts (MEVINAP). His teaching profile focuses on the field of research methodology in education and quantitative data analysis. His main lines of research deal with factors associated with academic achievement and the analysis of educative large-scale international assessments: school effectiveness and educational equity.

León, Jaime (Jaime.leon@ulpgc.es)  0000-0001-8400-2801

Professor of Research Methods in Education at the University of Las Palmas de Gran Canaria. His concern as a professor is that primary and secondary teachers get to optimize their students learning, for this he focuses on evidence-based education. As a researcher his concern is the same, to optimize the learning and performance of students, especially in secondary. To do so, he focuses on identifying factors that can be modified: teaching quality, classroom language, passion for knowledge, etc. In order to get the teacher to change in the classroom, he is focusing on designing a method that allows the teacher to obtain feedback on his or her teaching practice. Some of his publications and projects can be consulted in jaimeleon.es/ULPGC



Revista ELectrónica de Investigación y EValuación Educativa
E-Journal of Educational Research, Assessment and Evaluation

[ISSN: 1134-4032]



Esta obra tiene [licencia de Creative Commons Reconocimiento-NoComercial 4.0 Internacional](https://creativecommons.org/licenses/by-nc/4.0/).

This work is under a [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by-nc/4.0/).