

**PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ**

**FACULTAD DE CIENCIAS SOCIALES**



**PUCP**

**Comparación entre métodos de emparejamiento de evaluación de impacto**

**TRABAJO DE SUFICIENCIA PROFESIONAL PARA OPTAR EL TÍTULO DE  
LICENCIADA EN ECONOMÍA**

**AUTORA**

Carmen Raquel Rosario Armas Montalvo

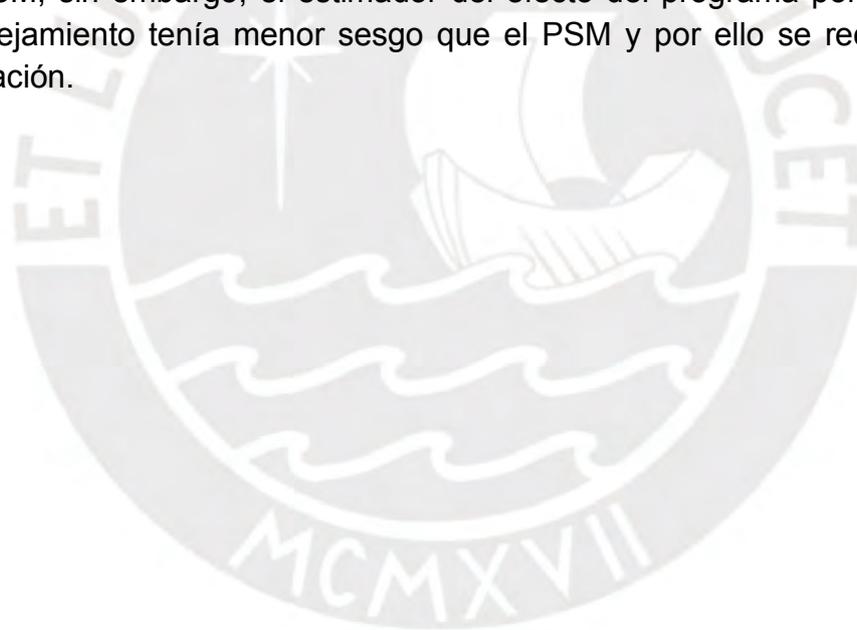
**ASESOR**

José Carlos Orihuela Paredes

**Setiembre, 2021**

## Resumen

La experiencia profesional que abarca este informe es sobre la que obtuve al desempeñar el cargo de especialista cuantitativo de evaluación de impacto en la Oficina de evaluación y seguimiento del Minedu. Este tipo de experiencia está comprendido en la temática de políticas efectivas o con resultados que se comprende en el curso de economía pública. Las herramientas metodológicas son las relacionadas a evaluación de impacto. Debido a que es un análisis sobre la efectividad de métodos de evaluación de impacto, se utilizan varios métodos de emparejamiento tales como Covariate Balancing propensity score (CBPS), Coarsened exact matching (CEM), Coarsened exact matching (CEM) y Emparejamiento genético. Durante la evaluación, se evidencio que CBPS, CEM y Frontera de Emparejamiento tenían resultados similares en balanceo y muestra que PSM; sin embargo, el estimador del efecto del programa por Frontera de Emparejamiento tenía menor sesgo que el PSM y por ello se recomienda su exploración.



## Contenido

INTRODUCCIÓN .....	I
EL MÉTODO DE EMPAREJAMIENTO EN LA EVALUACIÓN DE IMPACTO - 1 -	
LITERATURA DE REFERENCIA.....	- 4 -
BASE DE DATOS .....	- 8 -
METODOLOGÍA .....	- 10 -
RESULTADOS.....	- 15 -
CONCLUSIONES .....	- 17 -
BIBLIOGRAFÍA.....	- 19 -
ANEXO 1: GRÁFICOS Y TABLAS.....	- 24 -
ANEXO 2: BIOGRAFÍA DE LA AUTORA.....	- 33 -
ANEXO 3: CV RESUMIDO .....	- 34 -



## **INTRODUCCIÓN**

En los últimos años, el Ministerio de Educación (Minedu) se ha posicionado en ser uno de los ministerios con mayor evidencia de efectividad de sus programas a través de varios tipos de evaluaciones que incluyen a las evaluaciones de impacto. Desde el año 2015, la Oficina de Seguimiento y Evaluación Estratégica (OSEE) y la Unidad de Medición y Calidad (UMC) han venido realizando evaluaciones e investigaciones sobre los programas educativos de las diferentes oficinas del Minedu para fortalecer y mantener el hábito de formular políticas en base a evidencia.

Dentro de la OSEE, la Unidad de Seguimiento y Evaluación (USE) tiene la coordinación de evaluación de diseño, implementación e impacto de los programas educativos. La primera y segunda se encargan de hacer la evaluación del diseño de los programas en su fase piloto y, también, del funcionamiento del mismo al corto plazo de haber sido implementado, respectivamente.

En cambio, la evaluación de impacto es un análisis del efecto del programa sobre los resultados de interés, generalmente los logros académicos de los estudiantes, dependiendo de la disponibilidad de data secundaria. Es decir, la evaluación de impacto es una herramienta que permite estimar la efectividad de los programas por los ministerios y de esta forma tomar decisiones de política sobre la continuidad o modificación de estos para la mejora de resultados. Sin embargo, la amplia gama de métodos de evaluación de impacto causa dificultad al momento de utilizar alguno de sus métodos

Debido a que la evaluación de impacto básicamente es identificar un grupo de individuos similar al de las personas que recibió el tratamiento, existen una amplia variedad de métodos de evaluación de impacto. Además, para algunos de ellos, no existen diferencias teóricas, pero sí empíricas. De esta forma, es necesario comparar estos métodos similares con bases de datos reales o simuladas para comprobar su rendimiento ante diferentes características de las bases de datos.

Una categoría de métodos muy conocida y usada en evaluación de impacto es la de emparejamiento, ya que consiste en encontrar un grupo de observaciones con características observables y similares al grupo que recibió el tratamiento. En esta categoría, el método más conocido es el que se basa en una probabilidad de estar en el grupo de tratamiento condicionado a un número de características observables que es el Propensity Score Matching (PSM). Éste era el método más utilizado en el equipo de evaluación de impacto de la OSEE en Minedu y por ello, era necesario validar si tenía ventajas frente a los otros métodos de emparejamiento.

Si bien El PSM es un método muy utilizado en evaluación de impacto, también tiene muchas críticas en cuanto a sus supuestos y sensibilidad a la especificación o a la muestra e incluso algunos señalan que tiene más desventajas que ventajas. Por ello, se decidió hacer un análisis comparativo del “Propensity Score Matching” (PSM) y otros métodos de emparejamiento que será el centro de este informe profesional. Para ello, se utilizó la base del piloto del programa Expande tu Mente (EtM) que ya contaba con una evaluación de impacto (Outes, Sánchez y Vakis, 2017). Con estos resultados que han sido validados como los más confiables debido a su diseño experimental, se hace una comparación de los resultados con métodos de emparejamiento, incluyendo el PSM. De esta forma, se validará si el PSM es mejor o no frente a otros métodos de emparejamiento y si es necesario hacer cambios en las metodologías utilizadas por el equipo de evaluación de impacto.

Con la finalidad de organizar la información, el presente informe se divide en las siguientes secciones (1) El método de emparejamiento en la evaluación de impacto; (2) Literatura de referencia; (3) Base de datos; (4) Metodología; (5) Resultados; (6) Conclusión; y (7) Bibliografía.

## **EL MÉTODO DE EMPAREJAMIENTO EN LA EVALUACIÓN DE IMPACTO**

El método de emparejamiento consiste en comparar los resultados promedio de los participantes en el tratamiento con el resultado promedio de los no participantes emparejados que fueron escogidos en base a la similitud en características observables. En términos prácticos, el método consiste en seleccionar, duplicar o selectivamente borrar observaciones de nuestros datos para que las distribuciones de los grupos de tratamiento y control sean lo más similares posibles (Rubin, 1976a y Todd, 2010). El objetivo es eliminar o reducir la correlación entre las variables exógenas y la decisión de participación sin introducir sesgo o sin incrementar ineficiencia para que de esta forma se aproxime lo más posible al resultado de un experimento aleatorio y controlado (Ho et al 2011).

Existen dos supuestos que deben cumplirse: el supuesto de independencia condicional (CIA) y el de superposición. El primer supuesto señala que las diferencias entre los grupos de tratamiento y control se deben a características observables, únicamente; mientras que el segundo restringe que la probabilidad de participación o no participación sea positiva. Sin embargo, debido a que hay varias razones por las que estos supuestos no se pueden cumplir en economía (Abadie & Imbens, 2002; Heckman, Ichimura & Todd, 1998; Imbens, 2004 y Todd, 2010), se propone una versión más débil de estos y, por tanto, más plausible que funciona igual de bien cuando el parámetro de interés es el efecto promedio del tratamiento para los tratados (ATT) ( Heckman, Ichimura & Todd, 1998).

En la tabla 1, se muestra cómo se debilita los supuestos básicos de los estimadores de emparejamiento. El primer supuesto (CIA) implica la existencia de “*selection into treatment*” solo a través de que los componentes no observables que afectan la decisión de participación y la variable resultado sean del grupo de tratamiento, pero no del control. Es decir, no existen componentes no observables del grupo control que causan un problema de endogeneidad o “*selection into treatment*”. En cuanto al segundo supuesto, los parámetros SATT

y PATT solo necesitan garantizar que exista una persona no participante similar para cada persona participante.

Cada método de emparejamiento contiene cuatro partes: selección de variables, definición de la métrica, definición de la estrategia de emparejamiento y el algoritmo de emparejamiento (Steiner & Cook, 2013; Stuart & Rubin 2007; y Stuart, 2010). Por el lado de las variables, un pobre desempeño de los métodos de emparejamiento generalmente se debe al número de variables que se utilizan (Stuart, 2010). La métrica cuantifica la disimilitud entre un par de observaciones y se basan en el valor observado de las covariables en la línea de base. La definición de la estrategia de emparejamiento consiste en tres preguntas: ¿Cuántas unidades se deberían emparejar a cada unidad tratada?, ¿Cómo se define el valor umbral de la métrica (caliper) en que una observación deja de ser similar a la unidad tratada? Y ¿Se debería hacer el emparejamiento con reemplazo o sin reemplazo? Por último, el algoritmo es una decisión importante cuando se decide hacer un emparejamiento sin reemplazo. La diferencia entre los métodos de emparejamiento de la siguiente sección será principalmente por su forma de cuantificar la disimilitud entre 2 observaciones o la métrica.

Una vez que se termina de emparejar, se debe revisar la calidad del emparejamiento (Stuart, 2010). Un buen emparejamiento causará que las distribuciones de las variables que se utilizaron en los grupos de tratamiento y control sean iguales o, en otras palabras, tengan un buen *balancing score* (Rosenbaum & Rubin, 1983). En base a la teoría de Rubin y Thomas (1996), Rubin (2001) presenta tres medidas de balanceo: (1) la diferencia de medias estandarizada; (2) La ratio de varianzas del PSM entre los grupos de tratamiento y control; y (3) La ratio de varianza de los residuos de la ecuación del PSM entre los grupos tratado y control para cada variable. Sin embargo, la mayoría de las veces, la métrica de balanceo dependerá de la métrica de emparejamiento (Stuart, 2010).

Por último, como todo método, tiene sus ventajas y desventajas. Entre sus principales ventajas, puede trabajar con una fuente de datos de un solo tiempo, no requiere supuestos de la especificación de los resultados, aunque la mayoría de los estimadores de emparejamiento imponen supuestos de la forma funcional

cuando modelan las decisiones de participación (Glewwe, 2017); mejora el desempeño cuando se utiliza con otro estimador (Cochran y Rubin, 1973); y se puede aplicar a diversos contextos independientemente de las reglas de asignación del programa (Gertler et al, 2017). Por el lado de las desventajas, se basa en características o variables observables únicamente (Glewwe, 2017 y Gertler et al, 2017); en algunos contextos el DID tiene menor sesgo (Chabé-Ferret, 2015); y requiere un conjunto de datos amplios sobre grandes muestras de unidades (Gertler et al 2017, Smith & Todd 2005 y Heckman Ichimura & Todd 1998).



## LITERATURA DE REFERENCIA

Los métodos de emparejamiento tradicionales provienen de una sola clase: *Equal Percent Bias Reducing* (EPBR). Se conocen con este nombre porque tienen como objetivo mantener una proporcionalidad entre la reducción porcentual del sesgo entre los tratados y controles emparejados y la reducción porcentual del sesgo del total de tratados y controles. Si se define  $u_{i*}$  como la media esperada de las covariables en la muestra aleatoria del grupo  $G_{i*}$ , se tiene que un método de emparejamiento será EPBR si:

$$u_{1*} - u_{2*} = \gamma(u_1 - u_2) \quad \text{donde } \gamma \text{ es un escalar constante}$$

Los supuestos de los métodos EPBR son: (1) las covariables son de una distribución aleatoria; (2) Las covariables tienen una distribución elipsoidal simétrica; y (3) Los algoritmos de emparejamiento son invariantes a transformaciones afines de las covariables. De esta forma, la mayoría de métodos clásicos son del tipo EPBR con el PSM como el más conocido. Sin embargo, Rubin (1976b) señala que, si bien el EPBR es el punto inicial para estudiar los métodos de emparejamiento multivariados por la relación lineal entre la variable dependiente y las covariables, las relaciones no lineales también son importantes.

Según la definición de métrica, los métodos de emparejamiento tradicionales (mayormente EPBR) se pueden dividir en dos ramas: valores originales de las covariables y *propensity score* (Zhao, 2004). Los métodos de emparejamiento por exactitud, distancia euclídeana (EuM) y distancia mahalanobis (MDM) son del primer tipo; mientras que Propensity Score Matching (PSM), General Propensity Score (GPS) y Covariate Balancing Propensity Score (CBPS) son los basados en el propensity score. El primer tipo de métodos tiene problemas tales como dimensionalidad, sensibilidad a la estructura de correlaciones de las variables y a la presencia de outliers (Ho et al 2007 & 2011; Steiner & Cook, 2013; y Stuart, 2010 ) y, por ello, los basados en *propensity score* son más útiles pues al haber un número grande de covariables el PSM funciona mejor (Gu & Rosenbaum, 1993; y Rubin & Thomas, 2000).

El PSM es el método de emparejamiento más popular pues es simple de entender, elimina el problema de dimensionalidad y tiene pocos supuestos sobre la forma funcional así como la distribución de los errores (Bernal & Peña, 2011; Gertler et al, 2017; Khandker, Koolwal & Samad, 2009; y Zhao, 2008). Básicamente consiste en calcular la probabilidad de participar en el programa de cada individuo para ambos grupos dada las características observables (Bernal & Peña, 2011). Sus propiedades son principalmente dos: (1) La distribución de las covariables es la misma entre el grupo tratado y control en cada valor del PSM<sup>1</sup>(Rosembaum y Rubin, 1993); y (2) El estimador PSM cumple el supuesto de condicionalidad independiente (Stuart, 2010).

Debido a su amplio uso, el PSM ha sido criticado en varias ocasiones: supuestos poco plausibles (Steiner & Cook 2013); sensibilidad a la especificación (Zhao, 2008; y Crump et al, 2009), sensibilidad al supuesto de independencia condicional (Ichino et al 2008 y Nannicini 2007), sensibilidad al conjunto de variables y a la muestra (Smith & Todd, 2005<sup>2</sup>). Además, King & Nielsen (2016) señalan que el PSM incrementa el imbalance, ineficiencia, dependencia al modelo y sesgo. Incluso King y Nielsen (2016) argumentan que el *propensity score* (PS) no debería ser usado para emparejamiento, pero hay autores que señalan su utilidad para muchos otros métodos pues ayuda a reducir el sesgo (Cochran & Rubin, 1973; y Ho et al, 2011) pero se debería probar antes de combinarlo (King & Nielsen, 2016).

Sin embargo, hay trabajos que han mejorado el desempeño del estimador en: optimización del balance de covariables (Sant'Anna, Song& Xu, 2018; y Sekon, 2011); reemplazo del modelo logístico por uno no paramétrico (boosted cart o generalized boosted model) que han mostrado un buen desempeño (McCaffrey et al., 2004; Setoguchi et al., 2008; Lee et al., 2009); o semiparamétrico (Kordas & Lehrer, 2004) que tiene menor sesgo; o por métodos

---

<sup>1</sup> Esta propiedad es llamada balancing score en la literatura y según Rosenbaum and Rubin (1983), un balancing score ( $b(x)$ ) es una función de las covariables observadas ( $x$ ) tal que la distribución condicional de  $x$  dado  $b(x)$  es la misma para tratados ( $z=1$ ) y controles ( $z=0$ ). El PSM es el menos refinado balancing score que existe; mientras que el  $x$  es el balancing score más refinado.

<sup>2</sup> Dehejia (2005) respondió a esta crítica señalando que los resultados de los autores eran inválidos por tres temas: la elección de la muestra, la especificación del PSM y las pruebas de balanceo.

de aprendizaje estadístico que no asumen formas funcionales como árboles de clasificación (Steiner & Cook 2013).

En la literatura actual de métodos de emparejamiento, se está proponiendo nuevos métodos que solucionen los principales problemas de estos métodos: número de variables, balanceo y especificación del modelo. De los métodos revisados, CBPS es uno de los más estudiados, ya que logra el balance en el mismo momento que se crea el grupo control; sin embargo, están apareciendo otros que acaparan el interés por sus propiedades de balance tales como la frontera de emparejamiento (King & Nielsen, 2017) emparejamiento genético (Diamond & Sekhon, 2005; y Sekhon & Grieve, 2011); el estimador basado en el problema de optimización de 2 niveles o BLOP (Diaz, Rau & Rivera 2015); etc. En cuanto al número de variables, ahora hay métodos de emparejamiento con muchas covariables como Topical Inverse Regression Matching o TIRM (Robert, Stewart & Nielsen, 2015), Randomized Nearest Neighborhood Matching (Li et al, 2016)<sup>3</sup>.

Por último, hay un grupo de métodos distintos al EPBR. Los métodos anteriores se basaban en que la variable resultado es una combinación lineal de las covariables; sin embargo, Iacus et al (2008) flexibiliza la relación lineal entre las covariables y proponen una nueva clase de métodos conocidos como monotonic imbalance bounding (MIB).

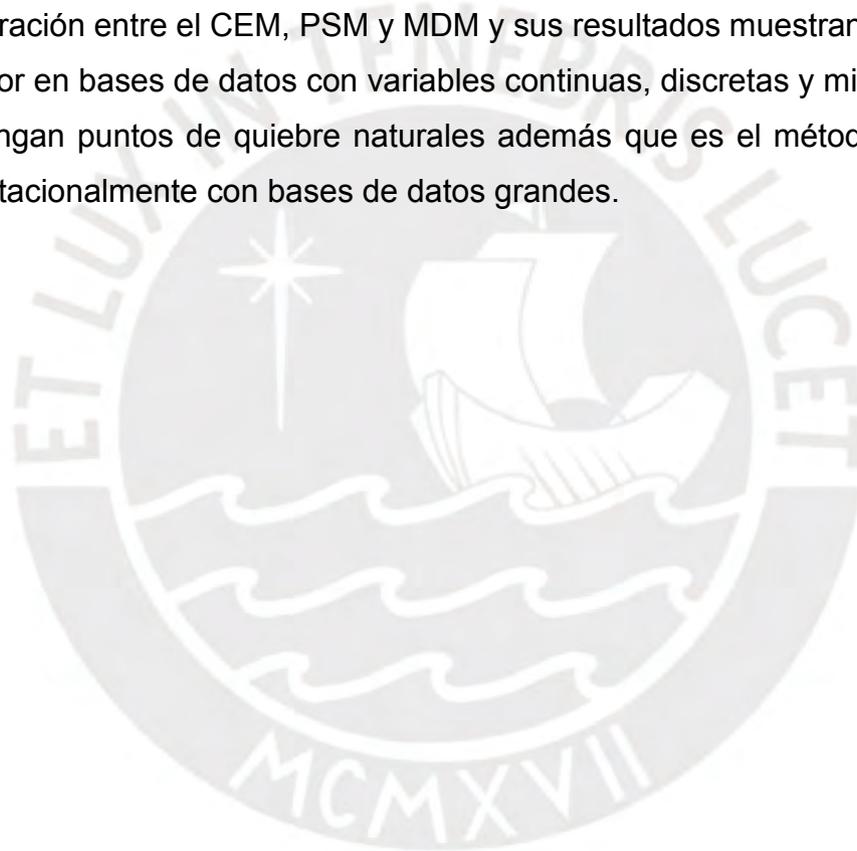
Esta clase de métodos tienen las siguientes diferencias con los EPBR (Iacus et al, 2011): (1) Solo necesita el supuesto de la propiedad invariante de los algoritmos de emparejamiento ante transformaciones afines de las variables; (2) El objetivo de los métodos MIB es balancear las covariables dentro de las muestras de tratamiento y control más que incrementar el balanceo como es en el caso de los EPBR; (3) La reducción del sesgo es menor a una proporción fija; (4) Se utiliza una función genérica como medida de desbalance en lugar de solo usar el promedio. Los autores y Rubin (1976b) señalan que los EPBR son el caso

---

<sup>3</sup> Esta literatura proviene de los emparejamientos que se quieren hacer con datos de texto. Para una revisión general de los métodos de emparejamiento para este tipo especial de datos, revisar Mozer et al (2018).

más simple de los MIB. El método MIB más conocido es el Coarsened exact matching (CEM).

El “coarsening” depende de los datos y del problema a analizar, por lo que es criterio de cada investigador. Además, tiene muchas propiedades: sigue el principio de la congruencia, automatiza el proceso del soporte empírico común, minimiza la dependencia a la especificación del modelo, minimiza el error de estimación produce un buen balanceo y considerable número de observaciones emparejadas, y eficiencia computacional. King y Nielsen (2016) hace una comparación entre el CEM, PSM y MDM y sus resultados muestran que el CEM es mejor en bases de datos con variables continuas, discretas y mixtas siempre que tengan puntos de quiebre naturales además que es el método más veloz computacionalmente con bases de datos grandes.



## **BASE DE DATOS**

La base de datos que se consideró para las pruebas de emparejamiento es la que se utilizó para la evaluación de impacto del piloto Expande tu Mente (EtM). Esta base se generó a partir de la información del Censo Escolar 2014. El universo de esta evaluación son las Instituciones Educativas (IIEE) de nivel Secundaria, de gestión pública, ubicadas en los sectores urbanos de las regiones de Áncash, Junín y Lima dentro del Censo Escolar 2014.

En la evaluación de impacto de EtM realizada por Outes, Sánchez y Vakis (2016), los autores aplican tres filtros. Primero, se excluyó a la IIEE que se ubicaban en los distritos del quintil inferior de pobreza (el más rico) a nivel nacional. Segundo, optaron por considerar solo IIEE que tengan turno mañana y tarde, pero no horario nocturno. Tercero, solo se consideran IIEE con más 15 alumnos para los grados de primero y segundo de secundaria. Dado estos filtros, Outes, Sánchez y Vakis (2016) tienen un marco muestral de 946 IIEE, las cuales se distribuyen de la siguiente manera: 226 en Áncash, 211 en Junín y 509 en Lima. De estas, seleccionaron una muestra aleatoria de 800 IIEE para que participen de la evaluación. Asimismo, se asignó aleatoriamente a 400 IIEE como grupo de tratamiento y otras 400 como grupo de control.

Debido a que se aplicará técnicas de emparejamiento, se necesita el universo de las IIEE del piloto de EtM que serían las IIEE de primero y segundo de secundaria, de gestión pública, en área urbana y de los departamentos de Áncash, Junín y Lima del censo escolar 2014. Por ello, se replicó el proceso explicado anteriormente. De este modo, se obtuvo 1351 IIEE como universo de IIEE. Luego, se utilizó la base de 800 IIEE para validar las variables filtro de pobreza, horario nocturno y alumnos por grado pues la base piloto debería ser contenida en el universo de IIEE luego de los tres filtros (946 IIEE). Sin embargo, el quintil de pobreza no se pudo asignar a 33 IIEE de las 1351 disponibles y se utilizó el mapa de pobreza 2013 para asignarles un quintil de pobreza. De este modo, el marco muestral es de 1,351 IIEE, pero su tamaño se reduce a 1,316 IIEE cuando se elimina las observaciones sin información de las variables adicionales de control.

Por lo tanto, el marco muestral contiene al grupo de tratamiento (400) y del resto de IIEE (916) se obtendrá el grupo control definido por los métodos de emparejamiento. El objetivo de las simulaciones es encontrar las parejas más cercanas de estas 400 IIEE tratadas y que sean similares a las 400 IIEE que se escogieron como grupo de control de manera aleatoria en el diseño del piloto de EtM.



## **METODOLOGÍA**

Debido a que el objetivo es comparar otros métodos de emparejamiento con el PSM, se usará este último como el estándar para el resto de metodologías. Asimismo, se utiliza la especificación del equipo de evaluación de impacto para hacer estas evaluaciones que es el PSM con vecino más cercano 1 a 1, sin reemplazo y un límite (caliper) igual a 0.05 d.e.. Las metodologías a comparar son las siguientes: Covariate Balancing propensity score (CBPS), Coarsened exact matching (CEM), Frontera de emparejamiento y Emparejamiento genético.

Asimismo, se tiene dos grupos de variables: Un grupo reducido de 4 variables y uno extenso de 19 variables que incluye las 4 variables del primer grupo. La diferencia es que el primer grupo solo tiene las variables que se utilizaron para el diseño aleatorio del piloto de EtM<sup>4</sup>; mientras que el segundo grupo contiene, además del grupo de variables anteriores, las comúnmente utilizadas por la USE en su emparejamiento<sup>5</sup>. Se hará comparaciones con respecto al experimento EtM<sup>6</sup> en tres ámbitos: (1) emparejamiento; (2) balance de covariables; y (3) resultados.

### i. Covariate Balancing propensity score (CBPS)

Imai & Ratkovic (2014) propone este método y a diferencia del PSM, el CBPS modela la asignación al tratamiento mientras optimiza el balance de covariables. Otra diferencia es que explota al máximo la propiedad de balanceo a través del marco del método general de momentos (GMM). Básicamente, el método consiste en maximizar la siguiente función:

---

<sup>4</sup> Las variables son (1) El número de alumnos en primero de secundaria; (2) El número de alumnos en segundo de secundaria; (3) El turno de horario del colegio; y (4) El quintil de pobreza del distrito del colegio.

<sup>5</sup> Son las siguientes: (1) Porcentaje de aprobados; (2) Porcentaje de retirados; (3) Ratio de estudiantes por sección; (4) Total de docentes; (5) Ratio de estudiantes por docente; (6) Porcentaje de estudiantes con lengua nativa distinta al castellano; (7) Ratio de computadoras por estudiante; (8) Tiempo de la IE a la UGEL, en horas; (9) Porcentaje de hogares con material noble en los techos en el distrito; (10) Porcentaje de hogares con material noble en las paredes en el distrito; (11) Porcentaje de hogares con alumbrado en el distrito; (12) Porcentaje de hogares con agua potable en el distrito; (13) Porcentaje de hogares en hacinamiento en el distrito; (14) Número de hogares rurales en el distrito; y (15) Porcentaje de mujeres en el distrito.

<sup>6</sup> Los resultados presentados en Outes, Sánchez y Vakis (2017).

$$\hat{\beta}_{MLE} = \arg \max_{\beta \in \Theta} \sum_{i=1}^N T_i \log\{\pi_{\beta}(X_i)\} + (1 - T_i) \log\{1 - \pi_{\beta}(X_i)\}$$

Siendo la condición de primer orden es:

$$\frac{1}{N} \sum_{i=1}^N s_{\beta}(T_i, X_i) = 0$$

Donde

$$s_{\beta}(T_i, X_i) = \frac{T_i \pi_{\beta}'(X_i)}{\pi_{\beta}(X_i)} - \frac{(1 - T_i) \pi_{\beta}'(X_i)}{1 - \pi_{\beta}(X_i)}$$

Si esta condición se satisface, también lo hace la condición de balanceo. El método generaliza esta condición reemplazando la  $\pi_{\beta}'(X_i)$  por una función medible de X ( $\tilde{X}_i$ ) y utiliza los momentos de esta función en el marco del GMM:

$$E \left\{ \frac{T_i \tilde{X}_i}{\pi_{\beta}(X_i)} - \frac{(1 - T_i) \tilde{X}_i}{1 - \pi_{\beta}(X_i)} \right\} = 0$$

Sus principales características son: (1) Mitiga los efectos de una potencial mala especificación de un modelo paramétrico para la probabilidad; (2) Es posible extenderlo para incorporar el GPS y la estimación generalizada de un tratamiento para una población objetivo; (3) Hereda todas las propiedades y metodologías que existen en la literatura del GMM y función de verosimilitud. Debido a que solo mejora la estimación del PS, varios métodos de emparejamiento y ponderación basados en PS pueden ser implementados sin modificación. Encuentran que este método tiene mejor performance que el PSM o los métodos de ponderación (weighting). El CBPS construye pesos directamente del PS pues determinar el mecanismo de la asignación del tratamiento mientras balancea las covariables. Existen extensiones de este modelo: estimar modelos estructurales marginales (Imai & Ratkovic, 2015); tratamiento no binario (Fong, Hazlett & Imai, 2014 y 2018).

ii. Coarsened exact matching (CEM)

El método de esta clase más conocido es el Coarsened Exact Matching (CEM) que consiste básicamente en tres pasos (Iacu et al., 2009 y 2012): (1)

Temporalmente, convierte los valores continuos de las covariables en categorías y forma las versiones “coarsen” de las covariables; (2) A todas las unidades asigna el valor de las covariables en su versión coarse; (3) Elimina los estratos que no incluyan al menos una unidad de control y tratado. Luego de ello, se utilizan los estratos creados para asignar pesos:

Tenemos la siguiente notación:

$T^s$ : Unidades tratadas en el estrato  $s$

$m_t^s$ : # Unidades tratadas en el estrato  $s$

$C^s$ : Unidades controles en el estrato  $s$

$m_c^s$ : # Unidades controles en el estrato  $s$

$m_t = \sum_{s \in S} m_t^s$ : # Unidades tratadas en total

$m_c = \sum_{s \in S} m_c^s$ : # Unidades controles en total

Dada la notación, se tiene que los pesos por cada unidad:

$$w_i = \begin{cases} 1, & i \in T^s \\ \frac{m_c}{m_t} \frac{m_t^s}{m_c^s}, & i \in C^s \end{cases}$$

A diferencia del PSM, no asume una relación lineal entre la probabilidad de ser tratado y las covariables y permite mayor número de emparejamientos que el PSM.

### iii. Frontera de emparejamiento

Similar al CBPS, la frontera de emparejamiento también tiene como objetivo el balanceo de covariables, pero además identifica el tamaño muestral con un procedimiento automático (King et al, 2017). Primero, se plantea una ratio de tratados y controles (R), número de la muestra de tratamiento y control (U), definición del resultado de interés (Q)<sup>7</sup>. Dado estos 3 parámetros, se define  $XX_n$  como todas las combinaciones posibles de las bases de datos (una matriz de

<sup>7</sup> En el trabajo de King et al (2017), se mencionan dos: SATT y FSATT. El SATT es el efecto promedio del tratamiento basado en la muestra de tratados y el FSATT y el FSATT es el SATT pero solo para los que tienen un buen emparejamiento.

datos de número de filas igual a  $N$  y columnas igual al número de parámetros  $K$ ) formadas por la combinatoria  $\binom{N}{n}$ . Luego:

$$XX = \{XX_n | n \in \{N, N - 1, \dots, 1\}\}$$

Donde  $n(XX) = 2^n$

Asimismo, se define un elemento (una base de datos) dentro de cada conjunto  $XX_n$  de la siguiente forma:

$$x_n = \arg_{x_0 \in XX_n} \min d(x_0, x_1)$$

Se define como frontera de emparejamiento:

$$F = \{x_n | (n \in \{N, N - 1, \dots, 1\}) / (d_{n-1} \leq d_n)\}$$

El problema de los algoritmos actuales es trabajar con un número grande de bases de datos (en total  $2^n$ ), y por ello una contribución adicional del trabajo es desarrollar algoritmos para las dos cantidades definidas anteriormente que sean rápidos, pero no necesariamente óptimos.

#### iv. Emparejamiento genético

Fue propuesto por Diamond & Sekhon (2005) con el fin de automatizar el proceso de maximizar el balance entre covariables observables de la muestra emparejada. Utiliza una forma generalizada la métrica utilizada por MDM:

$$d(X_i, X_j) = \left\{ (X_i - X_j)' (S_x^{-1/2})' W S_x^{-1/2} (X_i - X_j) \right\}^{\frac{1}{2}}$$

Según Sekhon & Grieve (2012),  $W$  es una matriz diagonal de una dimensión igual al número de covariables utilizadas. Es decir, las covariables son ajustadas por esta matriz que maximiza el balance de covariables de acuerdo a la función de pérdida definida por el usuario. La función de pérdida es la métrica de no balance que el usuario define. Los autores señalan que el PSM y el MDM son casos límites del emparejamiento genético.

Los autores sugieren utilizar una métrica que sea sensible a otras formas de desbalance<sup>8</sup> y por ello no recomiendan utilizar la diferencia de media como única métrica. Sugieren utilizar la métrica Kolmogorov-Smirnov o el estadístico D. Este último provee información valiosa en el valor de la variable donde la discrepancia máxima ocurre.

Formalmente, una distribución empírica se define como:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{X_i \leq x}$$

Donde  $I_{X_i \leq x}$  es una función que indica 1 si  $X_i \leq x$  y cero en otro caso. El estadístico D para dos muestras es definido como:

$$D_{n,n'} = \sup_x |F_n(x) - F_{n'}(x)|$$

El emparejamiento genético escoge los pesos para W cuando minimiza una función de pérdida definida por el usuario. Dentro de la función de pérdida, diferentes métricas de balance pueden ser definidas. Con la función de métrica definida, se tiene que el problema de optimización es difícil de conducir; por ello, la otra novedad es el algoritmo evolutivo que Sekhon y Mebane (1998) plantearon. Sin embargo, este método no automatiza dos procesos: (1) escoger las covariables y (2) la métrica del desbalance de covariables. Estos dos temas son temas abiertos a futuras mejoras del método.

---

<sup>8</sup> Rosembaun y Rubin (1993) sugieren que uno debería revisar el balance no solo en las medias sino también en productos cruzados, términos cuadráticos y varianzas de las covariables.

## **RESULTADOS**

El primer grupo de resultados es sobre el emparejamiento de cada método. La primera columna de las tablas 2 y 3 muestran los resultados del emparejamiento en el experimento EtM; la segunda, resultados del PSM, y las restantes son los resultados de los métodos en el encabezado de las columnas. Existen tres indicadores: número de tratados, número de controles y número de controles del RCT (que están en el grupo control del experimento EtM).

Se evidencia que los que tienen mayor número de coincidencias son: Frontera de Emparejamiento, PSM, CBPS y CEM. Sin embargo, como se muestra en la tabla N°4, existe una reducción significativa de número de controles coincidentes con el experimento EtM en los métodos de CEM y Frontera de emparejamiento cuando se aumenta el número de variables; sin embargo, la reducción no es tanta en CBPS y PSM e incluso el número de controles coincidentes en fronteras es similar al de PSM. Por último, se observa un deterioro mayor en los casos de Mahalanobis y CEM cuando aumenta el número de variables, siendo el emparejamiento por Mahalanobis imposible de implementar.

El segundo tipo de resultado es sobre el balanceo de las variables propuesto por cada técnica. Para ello, se utilizan varios tipos de pruebas que muestren el balance de las variables en los grupos de tratamiento y control: pruebas t-test para medias pruebas de ratio de varianza, pruebas de diferencias de colas de la distribución en el percentil 95 y que no deberían ser mayores a 0.25 d.e.<sup>9</sup>

Los resultados fueron que PSM, CBPS, CEM y frontera de emparejamiento fueron muy superiores en las pruebas de diferencias de medias cuando el grupo de variables de emparejamiento son las variables filtro del diseño del piloto (tabla 6); sin embargo, los resultados de balance de frontera de emparejamiento empeoran cuando aumentan el número de variables de emparejamiento a 19 variables (tabla 7). Mahalanobis y genético también tienen

---

<sup>9</sup> Los resultados se encuentran en el anexo 1

un pobre desempeño en esta prueba de balance. Similares resultados ocurren con la prueba de balance de ratio de varianza (tablas 8 y 9).

Por otro lado, la prueba de balance de las diferencias estandarizadas del percentil 95 presenta algunas diferencias a las pruebas anteriores (tablas 10 y 11). En primer lugar, ningún método de emparejamiento presenta un balance adecuado, incluyendo al diseño aleatorio del experimento, según esta prueba. En segundo lugar, Frontera de Emparejamiento y el algoritmo genético tienen más variables balanceadas que el resto de los métodos (tabla 10) y este número aumenta cuando se aumenta el número de variables por emparejamiento para el caso del método genético (tabla 11).

Se comparó todas las técnicas con los resultados del diseño aleatorio para saber si hubo una ganancia con estos métodos y se encontró que los métodos de PSM, Genético y Frontera de Emparejamiento son mejores al diseño aleatorio cuando las variables de emparejamiento son las variables filtro; mientras que los métodos Genético, CEM y Frontera de Emparejamiento son los mejores cuando se añaden las 15 variables típicas en las evaluaciones realizadas por USE.

Por último, los resultados de las estimaciones del efecto del tratamiento sobre la variable de aprendizaje en matemáticas muestran que CEM y Frontera de emparejamiento tienen resultados más similares al experimento que PSM y CBPS (tablas 4 y 5). A pesar de que el PSM y el CBPS tienen resultados positivos en la mayoría de las pruebas y superiores en las pruebas de diferencias de medias, no resulta ser suficiente en cuanto a la eficiencia de la estimación. En cambio, Frontera de emparejamiento y CEM tienen resultados superiores en las pruebas que están relacionadas a los otros momentos de las variables (media y kurtosis) más que las pruebas de balance de medias y obtienen un resultado más similar al experimento.

## **CONCLUSIONES**

Como se señaló al inicio, el gran número de métodos de emparejamiento no permite identificar el mejor entre todos. Lo más común es utilizar el PSM pues es el método estándar en la literatura internacional y nacional; sin embargo, tiene muchas desventajas y posibles sesgos que sus mejoras no han logrado solucionar. Por ello, en este trabajo se ha buscado compararlo con otros métodos de emparejamiento que son relativamente nuevos, pero que teóricamente tienen un mejor balanceo de covariables y tamaño de muestra: CBPS, CEM, Frontera de Emparejamiento y Emparejamiento genético.

Para realizar esta comparación se utilizó la base de datos del programa Expande tu mente (EtM) pues existe la evaluación de Otis, Sanchez y Vakis (2017) que estiman el efecto de este programa con una regresión sencilla ya que el diseño del programa fue experimental. Debido a ello, los efectos encontrados son los más cercanos a los verdaderos y son el punto de comparación con los métodos cuasiexperimentales del CBPS, CEM, Frontera de Emparejamiento, PSM y Emparejamiento genético. A partir de ello, se obtiene dos tipos de resultados: el emparejamiento y el efecto estimado del programa.

En cuanto al emparejamiento, el número de coincidencias es alto en los métodos de emparejamiento de CBPS, PSM, CEM y Frontera que incluso al aumentar el número de variables, el número de coincidencias sigue siendo alto. En cuanto al balanceo, se muestra que todos los métodos tienen buen desempeño en las pruebas de balanceo a excepción del emparejamiento genético. Asimismo, las diversas pruebas de balanceo, que corresponden a los diversos momentos<sup>10</sup> de una variable, no tienen resultados similares entre todos los métodos. Por último, las estimaciones calculadas con las especificaciones del PSM utilizada en la oficina muestran resultados muy inferiores a otros métodos a pesar de que este método obtuvo un buen resultado en las pruebas de balanceo y emparejamiento.

---

<sup>10</sup> Los principales momentos de una variable aleatoria son: media o promedio (primer momento), varianza (segundo momento), skewness (tercer momento) y curtosis .

Es necesario señalar que no hay un consenso de la literatura, tanto teórica y empírica, sobre el mejor método de emparejamiento. Sin embargo, los trabajos empíricos sobre sus propiedades reducen la falta de información. Por ello, en base al análisis exploratorio anterior, la recomendación final es explorar el método de Frontera de Emparejamiento pues este tiene el mejor desempeño cuando el grupo de variables de emparejamiento es las que se utilizaron como filtro en el diseño del experimento EtM. Por otro lado, sería útil explorar el número y tipo de variables que se usan en el emparejamiento. Sería útil hacer una evaluación del tipo de variable que se utiliza en el emparejamiento ya que existe una diferencia entre los dos grupos de variables que se utilizaron

En esta experiencia, utilicé muchas capacidades técnicas que aprendí de la carrera de economía en la PUCP. Debido a que era un trabajo prácticamente de investigación, utilicé mucha de mi experiencia como asistente de investigación, así como los conocimientos adquiridos de los cursos de estadística y econometría a lo largo de la carrera. Por ello, creo que en esta experiencia pude utilizar los conocimientos de la carrera y aplicarlos para mejorar la metodología utilizada en la oficina del equipo de evaluación de impacto.

## **BIBLIOGRAFÍA**

Abadie, A., & Imbens, G. W. (2002). Simple and bias-corrected matching estimators [Technical Report]. Department of Economics. *University of California, Berkeley*. Retrieved March, 6, 2016.

Abadie, A., & Imbens, G. W. (2009). Matching on the estimated propensity score (No. w15301). National Bureau of Economic Research.

Abadie, A., & Imbens, G. W. (2011). Bias-corrected matching estimators for average treatment effects. *Journal of Business & Economic Statistics*, 29(1), 1-11.

Abadie, A., & Imbens, G. W. (2016). Matching on the estimated propensity score. *Econometrica*, 84(2), 781-807.

Austin, P. C., & Small, D. S. (2014). The use of bootstrapping when using propensity-score matching without replacement: a simulation study. *Statistics in medicine*, 33(24), 4306-4319.

Bernal, R., & Peña, X. (2011). Guía práctica para la evaluación de impacto. Ediciones Uniandes-Universidad de los Andes.

Bodory, H., Camponovo, L., Huber, M., & Lechner, M. (2018). The finite sample performance of inference methods for propensity score matching and weighting estimators. *Journal of Business & Economic Statistics*, (just-accepted), 1-43.

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys*, 22(1), 31-72.

Chabé-Ferret, S. (2015). Analysis of the bias of matching and difference-in-difference under alternative earnings and selection processes. *Journal of Econometrics*, 185(1), 110-123.

Cochran, W. G., & Rubin, D. B. (1973). Controlling bias in observational studies: A review. *Sankhyā: The Indian Journal of Statistics, Series A*, 417-446.

Crump, R. K., Hotz, V. J., Imbens, G. W., & Mitnik, O. A. (2009). Dealing with limited overlap in estimation of average treatment effects. *Biometrika*, 96(1), 187-199.

Dehejia, R. H., & Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and statistics*, 84(1), 151-161.

Díaz, J., Rau, T., & Rivera, J. (2015). A Matching Estimator Based on a Bilevel Optimization Problem. *Review of Economics and Statistics*, 97(4), 803-812.

Diamond, A., & Sekhon, J. S. (2013). Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies. *Review of Economics and Statistics*, 95(3), 932-945.

Fong, C., & Imai, K. (2014). Covariate balancing propensity score for general treatment regimes. *Princeton Manuscript*, 1-31.

Fong, C., Hazlett, C., & Imai, K. (2018). Covariate balancing propensity score for a continuous treatment: application to the efficacy of political advertisements. *The Annals of Applied Statistics*, 12(1), 156-177.

Frölich, M. (2004). Finite-sample properties of propensity-score matching and weighting estimators. *Review of Economics and Statistics*, 86(1), 77-90.

Gertler, P. J., Martínez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. (2017). *La evaluación de impacto en la práctica*, Segunda edición. The World Bank.

Glewwe, P. (2017) *Impact Evaluation Technical Course 2017: How to Design, Manage, and Conduct Impact Evaluations* [Material de clase]. Sesiones del 4 al 13 de enero, 2017.

Gu, X. S., & Rosenbaum, P. R. (1993). Comparison of multivariate matching methods: Structures, distances, and algorithms. *Journal of Computational and Graphical Statistics*, 2(4), 405-420.

Heckman, J. J., Ichimura, H., & Todd, P. (1998). Matching as an econometric evaluation estimator. *The review of economic studies*, 65(2), 261-294.

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). MatchIt: nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, 42(8), 1-28.

Iacus, S. M., King, G., & Porro, G. (2009). Matching for causal inference without balance checking.

Iacus, S. M., King, G., & Porro, G. (2011). Multivariate matching methods that are monotonic imbalance bounding. *Journal of the American Statistical Association*, 106(493), 345-361.

Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political analysis*, 20(1), 1-24.

Ichino, A., Mealli, F., & Nannicini, T. (2008). From temporary help jobs to permanent employment: What can we learn from matching estimators and their sensitivity?. *Journal of applied econometrics*, 23(3), 305-327.

- Imai, K., & Ratkovic, M. (2014). Covariate balancing propensity score. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1), 243-263.
- Imbens, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and statistics*, 86(1), 4-29.
- Khandker, S., B. Koolwal, G., & Samad, H. (2009). *Handbook on impact evaluation: quantitative methods and practices*. The World Bank.
- King, G., Blackwell, M., Iacus, S., & Porro, G. (2010). cem: Coarsened exact matching in Stata.
- King, G., & Nielsen, R. (2016). Why propensity scores should not be used for matching. Copy at <http://j.mp/1sexgVw> Download Citation BibTex Tagged XML Download Paper, 378.
- King, G., Lucas, C., & Nielsen, R. A. (2017). The balance-sample size frontier in matching methods for causal inference. *American Journal of Political Science*, 61(2), 473-489.
- Kordas, G., & Lehrer, S. F. (2004). Matching using semiparametric propensity scores. Unpublished manuscript, Queen's University.
- Lee, B. K., Lessler, J., & Stuart, E. A. (2009). Improving propensity score weighting using machine learning. *Statistics in medicine*, 29(3), 337-346.
- Li, S., Vlassis, N., Kawale, J., & Fu, Y. (2016, July). Matching via Dimensionality Reduction for Estimation of Treatment Effects in Digital Marketing Campaigns. In *IJCAI* (pp. 3768-3774).
- McCaffrey, D. F., Ridgeway, G., & Morral, A. R. (2004). Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological methods*, 9(4), 403.
- Mozer, R., Miratrix, L., Kaufman, A. R., & Anastasopoulos, L. J. (2018). Matching with Text Data: An Experimental Evaluation of Methods for Matching Documents and of Measuring Match Quality. arXiv preprint arXiv:1801.00644.
- Nannicini, T. (2007). Simulation-based sensitivity analysis for matching estimators. *Stata Journal*, 7(3), 334.
- Outes, I., Sánchez, A., & Vakis, R. (2017). Cambiando la mentalidad de los estudiantes: evaluación de impacto de ¡Expande tu Mente! sobre el rendimiento académico en tres regiones del Perú. MISC.
- Roberts, M. E., Stewart, B. M., & Nielsen, R. (2015). Matching methods for high-dimensional data with applications to text. Working paper.

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.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39(1), 33-38.

Rubin, D. B. (1976a). Multivariate matching methods that are equal percent bias reducing, I: Some examples. *Biometrics*, 109-120.

Rubin, D. B. (1976b). Multivariate matching methods that are equal percent bias reducing, II: Maximums on bias reduction for fixed sample sizes. *Biometrics*, 121-132.

Rubin, D. B. (2001) Using propensity scores to help design observational studies: application to the tobacco litigation. *Health Services & Outcomes Research Methodology*, 2, 169–188.

Rubin, D. B., & Thomas, N. (1992). Affinely invariant matching methods with ellipsoidal distributions. *The Annals of Statistics*, 1079-1093.

Rubin, D. B., & Thomas, N. (1996). Matching using estimated propensity scores: relating theory to practice. *Biometrics*, 249-264.

Rubin, D. B., & Thomas, N. (2000). Combining propensity score matching with additional adjustments for prognostic covariates. *Journal of the American Statistical Association*, 95(450), 573-585.

Sant'Anna, P. H., Song, X., & Xu, Q. (2018). Covariate Distribution Balance via Propensity Scores. *Available at SSRN 3258551*.

Sekhon, J. S. (2011). Multivariate and propensity score matching software with automated balance optimization: the matching package for R.

Sekhon, J. S., & Grieve, R. D. (2011). A matching method for improving covariate balance in cost-effectiveness analyses. *Health economics*, 21(6), 695-714.

Sekhon, J. S., & Mebane, W. R. (1998). Genetic optimization using derivatives. *Political Analysis*, 7, 187-210.

Setoguchi, S., Schneeweiss, S., Brookhart, M. A., Glynn, R. J., & Cook, E. F. (2008). Evaluating uses of data mining techniques in propensity score estimation: a simulation study. *Pharmacoepidemiology and drug safety*, 17(6), 546-555.

Smith, J. A., & Todd, P. E. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators?. *Journal of econometrics*, 125(1-2), 305-353.

Steiner, P. M., & Cook, D. (2013). Matching and propensity scores. *The Oxford handbook of quantitative methods in psychology*, 1, 237.

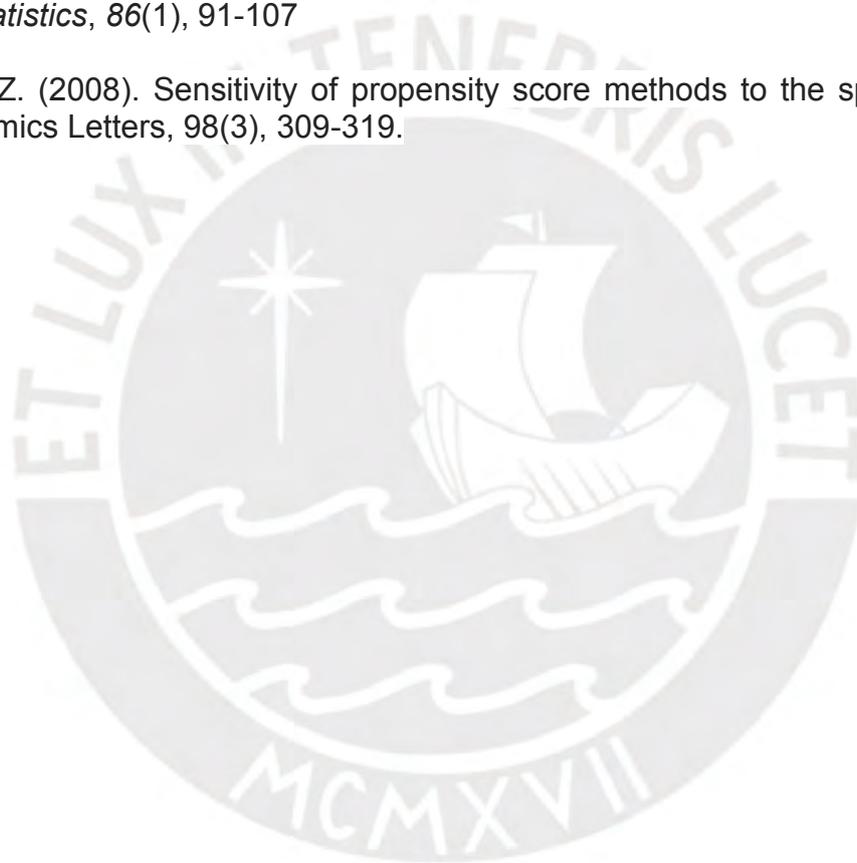
Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 25(1), 1.

Stuart, E. A., & Rubin, D. B. (2007). Best practices in quasi-experimental designs. *Best practices in quantitative methods*, 155-176.

Todd, P. E. (2010). Matching estimators. In *Microeconometrics*(pp. 108-121). Palgrave Macmillan, London.

Zhao, Z. (2004). Using matching to estimate treatment effects: Data requirements, matching metrics, and Monte Carlo evidence. *review of economics and statistics*, 86(1), 91-107

Zhao, Z. (2008). Sensitivity of propensity score methods to the specifications. *Economics Letters*, 98(3), 309-319.



### ANEXO 1: GRÁFICOS Y TABLAS

**Tabla 1: Supuestos básicos de los estimadores de emparejamiento**

Supuestos	Versión original	Versión flexibilizada
CIA	$(Y_0, Y_1)    D/Z$	$E(Y_0/Z, D = 1) = E(Y_0/Z, D = 0)$
Superposición	$0 < \Pr(D = 1/Z) < 1$	$\Pr(D = 1/Z) < 1$

Fuente: Todd 2010 y Heckman, Ichimura & Todd 1998

**Tabla 2: Tamaños de los grupos de control y tratamiento según técnica de emparejamiento (4 variables)**

Indicadores	Experimento EtM	PSM	Mahalanobis	CBPS	Genético	CEM	Frontera
Tratados	400	365	296	368	400	394	327
Controles	399	365	131	368	400	394	473
# Controles RCT	399	249 (62%)	92 (23%)	242 (61%)	177 (44%)	239 (60%)	297 (74%)

**Tabla 3: Tamaños de los grupos de control y tratamiento según técnica de emparejamiento (19 variables)**

Indicadores	Experimento EtM	PSM	Mahalanobis	CBPS	Genético	CEM	Frontera
Tratados	400	364	-	357	400	23	260
Controles	399	364	-	357	400	23	540
# Controles RCT	399	242 (61%)	-	231 (58%)	166 (42%)	16 (4%)	244 (61%)

**Tabla 4: Estimación por región y método de emparejamiento (4 variables)**

Regiones	Universo	Experimento	PSM	Mahalanobis	CBPS	Genético	CEM	Frontera
Total	3.01*	2.87	1.05	0.81	2.33	1.10	3.21*	2.82
Lima	0.14	0.26	1.81	1.61	-1.82	-1.21	1.34	1.05
No Lima	6.30	5.37*	4.04	-0.14	6.74**	4.11	4.84	4.59

**Tabla 5: Estimación por región y método de emparejamiento (19 variables)**

Región	Universo	Experimento	PSM	CBPS	Genético	CEM*	Frontera
<b>Total</b>	3.01*	2.87	0.67	1.81	1.90	5.98	3.38*
<b>Lima</b>	0.14	0.26	-0.24	0.58	-0.06	5.98	0.61
<b>No Lima</b>	6.30	5.37*	1.63	3.07	4.07	-	9.11**

\*46 observaciones: T: 23 y C: 23



**Tabla 6: Prueba de balance (diferencia de medias) entre grupos tratado y control según técnica de emparejamiento (4 variables)**

Indicadores	Experi- mento	PSM	Mahala nobis	CBPS	Genético	CEM	Fronter a
Quintil de pobreza	-0.007	-0.027	-0.246**	-0.024	-0.905***	0.292***	-0.062
Código del turno en que el colegio funciona	-	0.112**	-	-	-0.305***	-0.124**	0.008
Número de alumnos en primero de secundaria	-	-	-	-	-	-	9.653**
	-2.265	15.882*	14.673*	14.853*	-15.163***	-8.617	*
Número de alumnos en segundo de secundaria	-	-	-	-	-	-	9.766**
	-1.535	14.890*	13.108*	14.622*	-14.523***	-7.365	*
Porcentaje de aprobados	0.004	0.002	0.010	0.005	0.007	0.008	0.004
Porcentaje de retirados	0.003	0.003	0.005	0.002	0.007*	0.003	0.004
Ratio de estudiantes por sección	-0.155	-0.468	-1.239*	-0.341	0.046	-0.073	2.382**
			-				*
Total de docentes	-0.358	-3.077*	3.648***	-3.736**	-5.860***	-2.774*	2.298**
Ratio de estudiantes por docente	0.219	0.030	-0.159	0.170	0.531	0.480	1.620**
Porcentaje de estudiantes con lengua nativa distinta al castellano	-2.188	-1.544	1.479	-2.797	4.384***	-1.019	0.024
							-
Ratio de computadoras por estudiante	-0.006	-0.013	0.028	0.004	-0.053	0.002	0.126**
							*
Tiempo de la IE a la UGEL, en horas	3.626	6.655	13.588*	1.899	12.350***	6.319	-2.517
Porcentaje de hogares con material noble en los techos en el distrito	-0.014	-0.007	-0.020	-0.011	-0.095***	-0.031	0.017
Porcentaje de hogares con material noble en las paredes en el distrito	0.011	0.009	-0.046	0.013	-0.161***	-0.032	0.033
Porcentaje de hogares con alumbrado en el distrito	-0.001	0.005	-0.016	0.009	-0.054***	-0.012	0.005
Porcentaje de hogares con agua potable en el distrito	-0.010	-0.001	-0.032	-0.002	-0.082***	-0.021	0.002
Porcentaje de hogares en hacinamiento en el distrito	0.010*	0.011**	0.012	0.014**	0.020***	0.010*	0.016**
							*
Número de hogares rurales en el distrito	435.85	403.211	121.678	445.736	1,982.125**	610.002	965.340
	0				*		*
Porcentaje de mujeres en el distrito	-0.002	-0.001	0.000	-0.001	-0.001	-0.002	0.001
<b>Resumen de diferencias</b>							
<b>Para las 4 variables filtro</b>							

# Variables con diferencias al 1%	0	0	3	0	4	1	2
# Variables con diferencias al 5%	1	2	1	2	0	1	0
# Variables con diferencias al 10%	0	0	0	0	0	0	0
<b>Para las 15 variables adicionales</b>							
# Variables con diferencias al 1%	0	0	1	0	9	0	4
# Variables con diferencias al 5%	0	1	0	2	0	0	1
# Variables con diferencias al 10%	1	1	2	0	1	2	1

**Tabla 7: Prueba de balance (diferencia de medias) entre grupos tratado y control según técnica de emparejamiento (19 variables)**

Indicadores	Experimento	PSM	CBPS	Genético	CE M*	Frontera
Quintil de pobreza	-0.007	0.000	-0.017	-0.860***	7	-0.768***
Código del turno en que el colegio funciona	-0.112**	-0.019	0.003	-0.293***	0	-0.262***
Número de alumnos en primero de secundaria	-2.265	-2.126	-3.339	-23.448***	3	-10.078*
Número de alumnos en segundo de secundaria	-1.535	-2.198	-3.042	-21.853***	5	-8.201*
Porcentaje de aprobados	0.004	0.000	0.001	0.004	7	0.006
Porcentaje de retirados	0.003	0.000	-0.002	0.005	2	-0.002
Ratio de estudiantes por sección	-0.155	-0.184	-0.165	0.217	2	-0.516
Total de docentes	-0.358	0.060	-1.014	-7.235***	8	-3.410***
Ratio de estudiantes por docente	0.219	-0.151	-0.164	0.363	4	0.044
Porcentaje de estudiantes con lengua nativa distinta al castellano	-2.188	0.120	-0.862	3.565**	4	0.304
Ratio de computadoras por estudiante	-0.006	-0.002	-0.005	-0.062	2	0.023*
Tiempo de la IE a la UGEL, en horas	3.626	4.175	3.152	12.573***	6	6.339***
Porcentaje de hogares con material noble en los techos en el distrito	-0.014	-0.006	0.003	-0.097***	1	-0.112***
Porcentaje de hogares con material noble en las paredes en el distrito	0.011	-0.005	-0.003	-0.151***	2	-0.127***
Porcentaje de hogares con alumbrado en el distrito	-0.001	-0.006	-0.003	-0.048***	0	-0.033***
Porcentaje de hogares con agua potable en el distrito	-0.010	-0.014	0.005	-0.077***	0	-0.074***

Porcentaje de hogares en hacinamiento en el distrito	0.010*	0.004	0.005	0.017***	0.00	0.013***
Número de hogares rurales en el distrito	435.850	12.044	105.933	1,961.243***	0.00	1,551.06
Porcentaje de mujeres en el distrito	-0.002	-0.001	0.001	-0.003	0.00	-0.003**
<b>Resumen de diferencias</b>						
<b>Para las 4 variables filtro</b>						
# Variables con diferencias al 1%	0	0	0	4	0	2
# Variables con diferencias al 5%	1	0	0	0	0	0
# Variables con diferencias al 10%	0	0	0	0	0	2
<b>Para las 15 variables adicionales</b>						
# Variables con diferencias al 1%	0	0	0	8	0	8
# Variables con diferencias al 5%	0	0	0	1	0	1
# Variables con diferencias al 10%	1	0	0	0	0	1

\*46 observaciones: T: 23 y C: 23

**Tabla 8: Prueba de balance (ratio varianza) entre grupos tratado y control según técnica de emparejamiento (4 variables)**

Indicadores	Experi- mento	PSM	Mahala nobis	CBPS	Genét ico	CEM	Front era
Quintil de pobreza	0.995	0.953	1.023	0.917	0.735	0.724	1.077
Código del turno en que el colegio funciona	0.829	0.823	0.724	0.739	0.733	0.830	0.976
Número de alumnos en primero de secundaria	0.953	0.505	0.673	0.567	0.732	0.627	1.164
Número de alumnos en segundo de secundaria	0.914	0.525	0.688	0.565	0.665	0.670	1.154
Porcentaje de aprobados	0.774	0.701	0.872	0.681	1.009	0.837	0.660
Porcentaje de retirados	1.295	1.130	1.393	0.908	1.557	1.160	0.828
Ratio de estudiantes por sección	1.006	0.808	0.765	0.824	0.804	0.971	0.690
Total de docentes	0.943	0.611	0.692	0.599	0.582	0.705	1.129
Ratio de estudiantes por docente	1.203	1.034	1.041	1.118	1.097	1.240	0.971
Porcentaje de estudiantes con lengua nativa distinta al castellano	0.763	0.818	1.124	0.700	2.038	0.865	0.968
Ratio de computadoras por estudiante	0.899	0.849	1.211	0.934	0.090	1.309	0.103
Tiempo de la IE a la UGEL, en horas	1.087	1.552	2.016	1.280	2.179	1.126	1.015
Porcentaje de hogares con material noble en los techos en el distrito	1.049	0.988	1.260	0.962	0.973	0.972	0.979

Porcentaje de hogares con material noble en las paredes en el distrito	0.967	0.977	1.062	0.959	0.978	0.954	1.000
Porcentaje de hogares con alumbrado en el distrito	1.157	1.160	1.512	1.022	1.919	1.198	1.146
Porcentaje de hogares con agua potable en el distrito	1.064	1.060	1.531	1.001	1.336	1.074	0.961
Porcentaje de hogares en hacinamiento en el distrito	1.007	1.035	1.144	1.037	1.363	0.986	1.109
Número de hogares rurales en el distrito	1.065	1.139	0.843	1.186	1.684	1.090	1.420
Porcentaje de mujeres en el distrito	1.010	1.009	1.053	0.903	1.095	1.020	0.818
<b>Resumen de diferencias</b>							
<b>Para las 4 variables filtro</b>							
Entre 0.95 y 1.05	2	1	1	0	0	0	1
Entre 0.90 y 1.10	3	1	1	1	0	0	2
Entre 0.80 y 1.20	4	2	1	1	0	1	4
<b>Para las 15 variables adicionales</b>							
Entre 0.95 y 1.05	5	5	1	5	3	5	6
Entre 0.90 y 1.10	9	6	3	8	5	7	6
Entre 0.80 y 1.20	11	12	7	11	6	12	11

**Tabla 9: Prueba de balance (ratio varianza) entre grupos tratado y control según técnica de emparejamiento (19 variables)**

Indicadores	Experim ento	PSM	CBPS	Genético	CEM*	Frontera
Quintil de pobreza	0.995	0.906	0.848	0.685	0.850	0.804
Código del turno en que el colegio funciona	0.829	0.912	0.922	0.720	1.039	0.902
Número de alumnos en primero de secundaria	0.953	1.015	1.097	0.524	1.516	1.130
Número de alumnos en segundo de secundaria	0.914	0.987	1.056	0.506	1.294	1.100
Porcentaje de aprobados	0.774	0.840	0.639	0.915	0.927	1.332
Porcentaje de retirados	1.295	0.828	0.989	1.057	0.836	1.134
Ratio de estudiantes por sección	1.006	0.713	0.769	0.761	0.993	1.126
Total de docentes	0.943	1.016	0.876	0.502	0.976	1.046
Ratio de estudiantes por docente	1.203	0.926	0.890	0.970	0.984	1.212
Porcentaje de estudiantes con lengua nativa distinta al castellano	0.763	0.968	0.845	1.727	-	1.251
Ratio de computadoras por estudiante	0.899	0.852	0.589	0.083	1.206	1.346
Tiempo de la IE a la UGEL, en horas	1.087	1.122	1.075	2.105	1.087	1.760

Porcentaje de hogares con material noble en los techos en el distrito	1.049	0.983	0.927	0.942	1.008	1.151
Porcentaje de hogares con material noble en las paredes en el distrito	0.967	0.981	0.997	0.954	0.977	1.112
Porcentaje de hogares con alumbrado en el distrito	1.157	1.339	1.292	1.580	0.994	1.757
Porcentaje de hogares con agua potable en el distrito	1.064	1.152	1.025	1.278	1.005	1.501
Porcentaje de hogares en hacinamiento en el distrito	1.007	1.004	0.948	1.261	1.000	1.581
Número de hogares rurales en el distrito	1.065	0.819	0.860	1.858	1.000	2.106
Porcentaje de mujeres en el distrito	1.010	0.866	0.812	1.018	1.046	1.884
<b>Resumen de diferencias</b>						
<b>Para las 4 variables filtro</b>						
Entre 0.95 y 1.05	2	2	0	0	1	0
Entre 0.90 y 1.10	3	4	3	0	1	1
Entre 0.80 y 1.20	4	4	4	0	2	4
<b>Para las 15 variables adicionales</b>						
Entre 0.95 y 1.05	5	5	3	3	10	1
Entre 0.90 y 1.10	9	6	6	6	12	1
Entre 0.80 y 1.20	11	13	11	6	13	5

\*46 observaciones: T: 23 y C: 23

**Tabla 10: Prueba de balance (percentil 95) entre grupos tratado y control según técnica de emparejamiento (4 variables)**

Indicadores	Experi- mento	PSM	Mahalan obis	CBPS	Genét ico	CEM	Front era
Quintil de pobreza	0.252	0.153	0.259	0.148	0.510	0.835	0.265
Código del turno en que el colegio funciona	0.750	0.712	1.071	0.652	0.551	0.725	1.172
Número de alumnos en primero de secundaria	0.984	1.128	1.561	1.177	1.031	1.252	0.849
Número de alumnos en segundo de secundaria	0.964	1.101	1.696	1.204	1.143	0.965	0.941
Porcentaje de aprobados	0.239	0.212	0.368	0.208	0.344	0.280	0.123
Porcentaje de retirados	0.730	0.745	0.372	0.698	0.737	0.611	0.689
Ratio de estudiantes por sección	0.618	0.776	0.760	0.678	0.625	0.738	0.473
Total de docentes	0.940	1.194	1.557	1.259	1.136	1.335	1.095
Ratio de estudiantes por docente	0.620	0.570	0.840	0.538	0.571	0.512	0.458
Porcentaje de estudiantes con lengua nativa distinta al castellano	2.345	2.541	2.444	2.537	1.146	2.458	2.443
Ratio de computadoras por estudiante	0.604	0.760	0.911	0.888	0.189	0.651	0.091

Tiempo de la IE a la UGEL, en horas	0.434	0.289	0.594	1.157	0.213	0.433	1.198
Porcentaje de hogares con material noble en los techos en el distrito	0.593	0.576	0.611	0.549	0.472	0.599	0.566
Porcentaje de hogares con material noble en las paredes en el distrito	0.345	0.291	0.245	0.294	0.296	0.481	0.407
Porcentaje de hogares con alumbrado en el distrito	0.214	0.231	0.213	0.222	0.271	0.181	0.187
Porcentaje de hogares con agua potable en el distrito	0.078	0.072	0.064	0.062	0.023	0.018	0.072
Porcentaje de hogares en hacinamiento en el distrito	0.837	0.917	0.773	0.943	0.859	0.850	0.878
Número de hogares rurales en el distrito	1.206	0.178	1.586	0.214	0.083	1.191	0.119
Porcentaje de mujeres en el distrito	0.528	0.512	0.297	0.528	0.382	0.487	0.451
<b>Resumen de diferencias</b>							
<b>Para las 4 variables filtro</b>							
Entre de 0.10 d.e y 0.20 d.e	0	1	0	1	0	0	0
Más de 0.20 d.e y 0.25 d.e.	0	0	0	0	0	0	0
Más de 0.25 d.e.	4	3	4	3	4	4	4
<b>Para las 15 variables adicionales</b>							
Menos de 0.10 d.e	1	1	1	1	1	1	2
Entre de 0.10 d.e y 0.20 d.e	0	1	0	0	1	1	3
Más de 0.20 d.e y 0.25 d.e.	2	2	2	3	1	0	0
Más de 0.25 d.e.	12	11	12	11	11	13	10

**Tabla 11: Prueba de balance (percentil 95) entre grupos tratado y control según técnica de emparejamiento (19 variables)**

Indicadores	Experi- mento	PSM	CBPS	Genétic o	CEM*	Front era
Quintil de pobreza	0.252	0.148	0.996	0.509	0.569	0.292
Código del turno en que el colegio funciona	0.750	0.801	0.785	0.553	0.454	0.383
Número de alumnos en primero de secundaria	0.984	0.761	0.970	1.621	0.584	0.802
Número de alumnos en segundo de secundaria	0.964	0.898	0.852	1.256	0.475	0.807
Porcentaje de aprobados	0.239	0.239	0.164	0.293	0.119	0.311
Porcentaje de retirados	0.730	0.620	0.897	0.669	0.522	0.729
Ratio de estudiantes por sección	0.618	0.729	0.706	0.496	0.926	0.610
Total de docentes	0.940	0.802	1.093	1.363	0.817	0.803
Ratio de estudiantes por docente	0.620	0.676	0.693	0.532	0.776	0.647
Porcentaje de estudiantes con lengua nativa distinta al castellano	2.345	2.574	2.625	0.552	-	1.034
Ratio de computadoras por estudiante	0.604	0.638	0.671	0.036	0.971	0.767

Tiempo de la IE a la UGEL, en horas	0.434	0.814	0.939	0.250	0.610	0.008
Porcentaje de hogares con material noble en los techos en el distrito	0.593	0.593	0.631	0.463	0.029	0.317
Porcentaje de hogares con material noble en las paredes en el distrito	0.345	0.288	0.348	0.308	0.014	0.085
Porcentaje de hogares con alumbrado en el distrito	0.214	0.232	0.239	0.272	0.226	0.357
Porcentaje de hogares con agua potable en el distrito	0.078	0.056	0.030	0.008	0.033	0.033
Porcentaje de hogares en hacinamiento en el distrito	0.837	0.929	0.912	0.902	0.312	0.380
Número de hogares rurales en el distrito	1.206	0.439	1.314	0.068	1.674	0.186
Porcentaje de mujeres en el distrito	0.528	0.559	0.424	0.611	1.079	0.719
<b>Resumen de diferencias</b>						
<b>Para las 4 variables filtro</b>						
Entre de 0.10 d.e y 0.20 d.e	0	1	0	0	0	0
Más de 0.20 d.e y 0.25 d.e.	0	0	0	0	0	0
Más de 0.25 d.e.	4	3	4	4	4	4
<b>Para las 15 variables adicionales</b>						
Menos de 0.10 d.e	1	1	1	3	3	2
Entre de 0.10 d.e y 0.20 d.e	0	0	1	0	1	1
Más de 0.20 d.e y 0.25 d.e.	2	2	1	0	1	0
Más de 0.25 d.e.	12	12	12	12	10	11

\*46 observaciones: T: 23 y C: 23

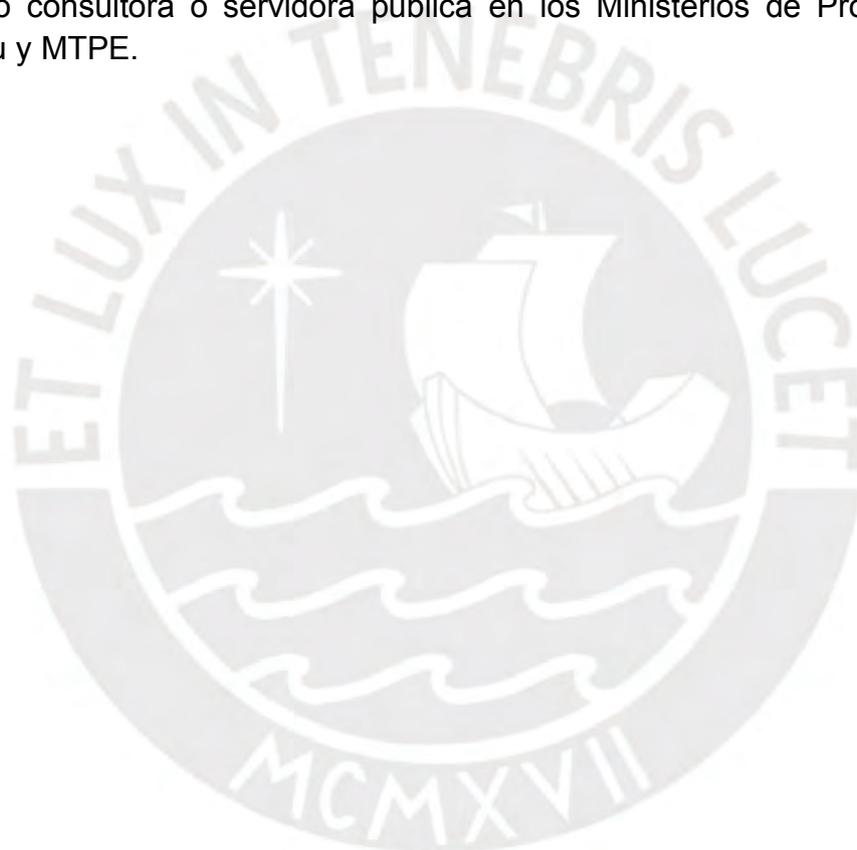
**Tabla 12: Número de variables con mejor balance en el percentil 95 que con el diseño aleatorio de EtM según método de emparejamiento (19 variables)**

Grupo de variables de emparejamiento	PSM	Mahalanobis	CBPS	Genético	CEM*	Frontera
Grupo de 4 variables	10	6	9	10	8	12
Grupo de 19 variables	8	-	5	11	10	13

\*En el caso del CEM, con emparejamiento con 19 variables, son solo 46 observaciones: T: 23 y C: 23

## **ANEXO 2: BIOGRAFÍA DE LA AUTORA**

Nací, crecí y estudié mi primaria y secundaria en el distrito pujante de Los Olivos del Departamento de Lima, Perú. Curse la carrera de Economía en la facultad de Ciencias Sociales de la Pontificia Universidad Católica del Perú. Fui décimo superior de mi promoción. Durante mi pregrado, trabajé como profesora de matemática de la CEPREPUC y asistente de investigación de profesores de econometría y del Cisepa de la PUCP. Al egresar del pregrado, llevé el curso del BCRP y también la maestría en Economía con especialización en Desarrollo Económico de la PUCP. Durante ese mismo tiempo, trabajé en la ONG GRADE y como consultora o servidora pública en los Ministerios de Produce, PCM, Minedu y MTPE.



### **ANEXO 3: CV RESUMIDO**

Magistra en Economía de la Pontificia Universidad Católica del Perú (PUCP). Décimo superior de la clase 2016, con habilidades en programación y conocimiento en econometría aplicada. Además, me considero de rápido aprendizaje, mente abierta, responsable, curiosidad por el conocimiento y una trabajadora empeñosa.

#### **ESTUDIOS**

2020-2021  
2018  
Económico. Especialización en Estadística Aplicada- INEI  
Magistra en Economía- Mención en Desarrollo  
y construcción de encuestas.  
2008-2013 Bachiller en Economía (PUCP)

#### **EXPERIENCIA LABORAL**

06/06/2020- 30/04/2021 MTPE- Coordinadora de la Política Nacional de Empleo Decente. *Coordinación con las direcciones del MTPE, Sectores y CEPLAN para la formulación de Objetivos prioritarios, lineamientos y servicios para la Política Nacional de Empleo Decente.*

30/10/2019- 05/06/2020 MTPE- Especialista en la formulación de la Política Nacional de Empleo Decente. *Revisión de literatura, estadísticas descriptivas sobre empleo y análisis adicionales en el diseño de la Política Nacional de Empleo Decente.*

18/09/2018-30/10/2019 MINEDU- Especialista Cuantitativo en Técnicas de Evaluación de Políticas Educativas. *Coordinación con direcciones del MINEDU para proponer evaluaciones de impacto de los programas de interés de dichas direcciones.*

27/12/2017-27/08/2018 PCM-Analista de estudios y Especialista de la Oficina de Cumplimiento del Gobierno e Innovación Sectorial (OCGIS). *Coordinación con MININTER para el seguimiento de las políticas priorizadas para el tema de seguridad ciudadana. Coordinación con INEI para la elaboración de los indicadores mensuales de seguimiento a los temas prioritarios del gobierno (Empleo, Seguridad ciudadana, Salud, Simplificación de trámites, Educación, Agua y Saneamiento, y Reconstrucción con Cambios)*

02/06/2014- 23/09/2017 GRADE- Asistente de Investigación con Javier Escobal. *Revisión y análisis de temas relacionados a pobreza, pobreza rural, agricultura, etc.,*

## EXPERIENCIA EN DOCENCIA

- 17/03/2014-12/07/2014 PUCP-Jefe de prácticas de Introducción a la Macroeconomía.
- Agosto 2012- junio 2014 Jefe de prácticas en USMP: Econometría Series de Tiempo; Econometría-corte transversal; Macroeconomía, Matemática para economistas y Economía Internacional.
- 12/03/2012-07/07/2012 Jefe de prácticas en PUCP: Introducción a la Macroeconomía.

## PROGRAMAS

- Stata
- R
- Eviews
- Povmap
- Python
- Microsoft Office
- Arcgis/Qgis
- SPSS
- Scientific Workplace
- Geoda o Geodaspace
- Matlab
- Gauss

## INVESTIGACIONES Y PUBLICACIONES

- 2018 **Prevalencia de delitos en la población penal peruana: desorganización social versus aprendizaje social.** Proyecto breve ganador del CIES 2017.  
URL: <http://www.cies.org.pe/es/investigaciones/prevalencia-de-delitos-en-la-poblacion-penal-peruana-desorganizacion-social>
- 2018 **De la Parada al mercado mayorista de Santa Anita: efectos sobre los márgenes, competitividad y eficiencia de la cadena de producción y comercialización de la papa.** Capítulo en el libro *Perú: el problema agrario en debate SEPIA XVII*.  
URL: <http://www.grade.org.pe/publicaciones/de-la-parada-al-mercado-mayorista-de-santa-anita-efectos-sobre-los-margenes-competitividad-y-eficiencia-de-la-cadena-de-produccion-y-comercializacion-de-la-papa/>
- 2018 **Inclusión económica y tributación territorial: el caso de las exoneraciones altoandinas.**  
URL: <http://www.grade.org.pe/publicaciones/inclusion-economica-y-tributacion-territorial-el-caso-de-las-exoneraciones-altoandinas/>
- 2017 **¿El lugar importa en la victimización?: el caso de Lima Metropolitana.** Tesis de Maestría de Economía de Pontificia Universidad Católica del Perú.  
URL: <http://files.pucp.edu.pe/departamento/economia/DDD451.pdf>

2017 **Interacción social y crimen: análisis provincial en Perú.**  
Trabajo ganador del concurso de Think Thank Initiative  
de GRADE para jóvenes investigadores 2016.

URL: <http://repositorio.grade.org.pe/handle/GRADE/361>

2015 **El uso de encuestas y censos agropecuarios para  
desarrollar una tipología de la pequeña y  
mediana agricultura familiar en el Perú.**

Capítulo en “Agricultura Peruana: Nuevas miradas desde el  
Censo Agropecuario”.

URL: <http://www.grade.org.pe/en/publicaciones/agricultura-peruana-nuevas-miradas-desde-el-censo-agropecuario/>

2015 **Estructura agraria y dinámica de pobreza rural en el  
Perú.** Parte de la serie de “Documento de investigación.  
Desarrollo Rural y Agricultura”

URL: <http://www.grade.org.pe/publicaciones/estructura-agraria-y-dinamica-de-pobreza-rural-en-el-peru/>

