



XXXIV SEMINARIO ANUAL
DE INVESTIGACIÓN 2023

Coorganiza:



FACULTAD DE
**CIENCIAS
ECONÓMICAS**

Cambios en la globalización y los retos
del desarrollo sostenible en el Perú



¿LAS PRÁCTICAS DE GESTIÓN DE CALIDAD INFLUYEN EN LA PRODUCTIVIDAD LABORAL EN LAS EMPRESAS FORMALES EN EL PERÚ?: Un Análisis Basado en un Diseño No Experimental y Técnicas de Machine Learning Causal

PhD. Mario D. Tello

Departamento de Economía
Pontificia Universidad Católica del Perú y Facultad
de Ciencias Económicas de la Universidad
Nacional Mayor de San Marcos

PhD. Daniel S. Tello-Trillo

Frank Batten School of Leadership and Public
Policy
University of Virginia e Investigator del
National Bureau of Economic Research, NBER-
USA

Instituciones comprometidas con el conocimiento

Platinum

FUNDACIÓN
M. J. Bustamante De La Fuente

Oro



BANCO CENTRAL DE RESERVA DEL PERÚ



GRUPO BANCO MUNDIAL

Plata



Naciones
Unidas

Con el apoyo de



PERÚ Ministerio
de Economía y Finanzas

Aliados mediáticos

El Comercio



GESTIÓN



Índice

1. [Objetivo](#)

2. [Base de Datos](#)

3. [El problema de las regresiones](#)

4. [Pruebas de exogenidad](#)

5. [Resultados](#)

6. [Conclusiones](#)



Índice

1. [Objetivo](#)

2. Base de Datos

3. El problema de las regresiones

4. Pruebas de exogenidad

5. Resultados

6. Conclusiones



1. Objetivo

El estudio estima los impactos de las herramientas de gestión de calidad sobre el desempeño de las empresas del Perú, específicamente la productividad laboral, para el periodo 2014-2019 basados en técnicas de Aprendizaje Automático (o Machine Learning, ML, en inglés) causal (MLC).



Índice

1. Objetivo

2. [Base de Datos](#)

3. El problema de las regresiones

4. Pruebas de exogenidad

5. Resultados

6. Conclusiones



2. Base de datos

Las variables del estudio se obtienen de la Encuesta Nacional de Empresas del INEI-ENE (2022) para el período 2014-2019. Se usa dos períodos de las encuestas, uno del período de publicación 2015-2018 que corresponde al periodo real del 2014-2017; el otro con periodo de publicación 2014-2018, y 2020 que corresponde al período real 2014-2017, y 2019.

Luego de una limpieza de la información se construye dos tipos de bases de datos:

- **Una base conjunta o 'pool'** de empresas para todos los años sin distinción de ellas a lo largo del periodo. En esta base se registra 1224 observaciones y 245 empresas, en promedio, por año.
- **La segunda es una base de datos de panel** (aunque incompleto) con 1044 observaciones y un promedio de 209 empresas de por lo menos dos años de información del periodo. El panel fue formado por el código de la empresa a través de los años en la encuesta y ciertas características de éstas (tales como año de inicio de la empresa, productividad laboral y nivel empleo).

También se ha seleccionado varias variables binarias y continuas como **variables instrumentales de la variable de interés *DCAL***.



Índice

1. Objetivo

2. Base de Datos

3. [El problema de las regresiones](#)

4. Pruebas de exogenidad

5. Resultados

6. Conclusiones



3. El problema de las regresiones (Confounding variables – Sobreajustes 'overfitting')

Usualmente en métodos cuantitativos existen diversas formas de 'medir-estimar-evaluar' los impactos de 'políticas-tratamientos-intervenciones' sobre variables 'objetivos-resultados' que se desea afectar, supuestamente a través de políticas. Estos métodos se basan en diseños (usualmente muestrales) experimentales (RCT, randomized controlled trials) y no experimentales (con datos secundarios, o encuestas predeterminadas). En el trabajo se aplica diseños no experimentales dada la base de datos descrita.



3. El problema de las regresiones (Confounding variables – Sobreajustes 'overfitting')

Aun con diseños no experimentales los métodos econométricos son muchos (ejemplos, regresiones simples, y regresiones 'complejas' como difference & differences. Propensity matching scores, regresión discontinuity, synthetic control, etc.) y la aplicación dependen, entre otros factores, de la información que se dispone, y los problemas econométricos que emergen de las especificaciones a estimar.



3. El problema de las regresiones (Confounding variables – Sobreajustes 'overfitting')

En el estudio se aplica una técnica nueva con herramientas de 'machine learning casual, MLC'. La técnica se denomina Double Debiased ML desarrollados por desarrollado por Belloni, Chernozhukov, y Hansen (2014a, b), Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey, y Robins (2018 y 2017), y Baiardi & Naghi (2020).



3. El problema de las regresiones (Confounding variables – Sobreajustes 'overfitting')

Por límite de tiempo, solo presentaré 'la filosofía' de MLC. Una excelente discusión de los métodos tradicionales y los de ML lo desarrolla Imbens & Athey (2019, Machine Learning Methods Economists Should Know About) y Felix Chan & László Mátyás (2022, Econometrics with Machine Learning, capítulo 3). Aquí presentaré de manera breve y simple el método DML.



3. El problema de las regresiones (*Confounding variables – Sobreajustes 'overfitting'*)

Estimadores de Mínimos Cuadrado Ordinarios

$$[3.1] \quad PL_{it} = \beta \cdot DCAL_{it} + \vec{X}_{it}' \cdot \vec{\delta} + \varepsilon_{it}; i = 1, \dots, N; t = 2014 - 2019$$

$$\hat{\varepsilon}_1 = PL_{it} - \vec{X}_{it}' \hat{\delta}_1, \dots, N;$$

$$\hat{\varepsilon}_2 = DCAL_{it} - \vec{X}_{it}' \hat{\delta}_2, \dots, N;$$

Se demuestra que el estimador MCO de la variable de interés $DCAL_{it}$, $\hat{\beta}$, es la regression $\hat{\varepsilon}_1 = F(\hat{\varepsilon}_2)$.

El problema de MCO es que no se sabe el número óptimo de variables de control, \vec{X}_{it} , (problema de 'overfitting= sobreajuste') y tampoco se conoce las variables confusas (confounding) que afectan a PL_{it} y $DCAL_{it}$ cuando es a través de $DCAL_{it}$ que se afecta a PL_{it} . Lo cierto es que la variable confusa (desconocida) es la causa de PL_{it} y no $DCAL_{it}$. Ambos problemas originan 'sesgos' en el estimador MCO de β . EL método DML resuelve ambos problemas con herramientas de MLC.



3. El problema de las regresiones (Confounding variables – Sobreajustes 'overfitting')

- [3.2] $PL_{it} = \beta \cdot DCAL_{it} + go(\vec{X}_{it}) + U_{it}; \quad i = 1, \dots, N; t = 2014 - 2019$
- [3.3] $DCAL_{it} = mo(\vec{X}_{it}) + V_{it};$

Primero se estima \hat{U}_{it} de [3.2] con técnicas ML (por ejemplo, Lasso). Luego se estima \hat{V}_{it} de [3.3] también con técnicas ML y finalmente el coeficiente de la regresión $\hat{U}_{it}(\hat{V}_{it})$ es el estimador $\hat{\beta}$ del parámetro de interés β . En términos de estimación, las técnicas ML se usan para predicción y las usuales técnicas econométricas para relaciones de causalidad con ciertos supuestos del modelo de los datos (data modelling).



Cuadro 1:

Regresiones MCO de variables de interés y las instrumentales: Pool 2014-2017

No	Variables	MCO (DCAL) _{it}	MCO (DCALDIF) _{it}	MCO (DCAL&DIF CIU) _{it}	MCO (DMIL) _{it}	MCO (DMIL CIU) _{it}	MCO (DIMB) _{it}	MCO (DIMB&DIF CIU) _{it}	MCO (DOCD) _{it}	MCO (DOCD&DIF CIU) _{it}
1.	DCAL	0.263*** (0.0893)	-	-	-	-	-	-	-	-
2.	DCALDIF	-	-0.00189 (0.0738)	-	-	-	-	-	-	-
3.	DCALDIF& CIU	-	-	0.00545** (0.00216)	-	-	-	-	-	-
4.	DMIL	-	-	-	0.108* (0.0652)	-	-	-	-	-
5.	DMIL&CIU	-	-	-	-	0.00857*** (0.00283)	-	-	-	-
6.	DIMB	-	-	-	-	-	0.148 (0.122)	-	-	-
7.	DIMB&CIU	-	-	-	-	-	-	0.00618*** (0.00154)	-	-
8.	DOCDg	-	-	-	-	-	-	-	0.338*** (0.118)	-
9.	DOCDg&CIU	-	-	-	-	-	-	-	-	0.0115*** (0.00209)



Cuadro 2:
Regresiones TSLS de variables instrumentales: Pool: 2014 -2017

No	Variables	TSLS (DCALDIF) _{it}	TSLS (DCAL&DIF CIU) _{it}	TSLS (DMIL) _{it}	TSLS (DMIL CIU) _{it}	TSLS (DIMB) _{it}	TSLS (DIMB&DIF CIU) _{it}	TSLS (DOCDg) _{it}	TSLS (DOCDg&DIF CIU) _{it}
1	\bar{X}	-0.0523 (2.036)	-3.440* (1.778)	-2.629 (2.088)	-13.31 (11.91)	-4.719 (5.761)	-6.710** (3.040)	1.961** (0.764)	6.627*** (1.954)
2	DL_{MED}	0.981 (0.917)	2.501*** (0.812)	2.137** (0.950)	6.928 (5.365)	3.192 (2.691)	4.119*** (1.447)	0.0779 (0.386)	-2.015** (0.934)
3	DL_{Gr}	0.421 (1.031)	2.180** (0.928)	1.759 (1.086)	7.306 (6.212)	2.846 (3.043)	3.891** (1.619)	-0.625 (0.436)	-3.048*** (1.070)
4	lnk	0.0378 (0.0342)	0.0947*** (0.0347)	0.0811** (0.0398)	0.261 (0.208)	0.124 (0.117)	0.163** (0.0668)	0.00396 (0.0205)	-0.0744 (0.0471)
5	DC	0.667*** (0.177)	0.857*** (0.239)	0.811*** (0.230)	1.410 (0.900)	0.854** (0.430)	0.965** (0.382)	0.554*** (0.157)	0.292 (0.316)
6	$DTEC_1$	0.199* (0.102)	0.305*** (0.117)	0.280** (0.110)	0.615 (0.454)	0.341* (0.200)	0.395** (0.178)	0.136* (0.0819)	-0.0106 (0.148)
7	$DNET$	0.241 (0.372)	0.628* (0.370)	0.535 (0.372)	1.754 (1.748)	0.735 (0.768)	0.947 (0.728)	0.0116 (0.430)	-0.521 (0.834)
8	$DFIN$	0.175 (0.186)	0.260 (0.261)	0.240 (0.233)	0.510 (0.852)	0.332 (0.356)	0.381 (0.436)	0.124 (0.220)	0.00568 (0.439)
9	$DTurn_2$	0.255*** (0.0878)	0.148 (0.115)	0.174 (0.115)	-0.163 (0.477)	0.0322 (0.313)	-0.0630 (0.213)	0.319*** (0.0905)	0.466*** (0.173)
10	$CINST$	0.00291 (0.00203)	0.00438 (0.00277)	0.00403 (0.00250)	0.00866 (0.00855)	0.00541 (0.00434)	0.00643 (0.00424)	0.00204 (0.00217)	2.02e-05 (0.00369)
11	$DCAP$	0.0766 (0.317)	0.569** (0.282)	0.451 (0.328)	2.005 (1.755)	0.801 (0.855)	1.092** (0.482)	-0.216 (0.147)	-0.895*** (0.345)
12	DX	0.579** (0.288)	1.019*** (0.264)	0.913*** (0.291)	2.301 (1.586)	1.190 (0.784)	1.456*** (0.449)	0.317** (0.132)	-0.289 (0.283)
13	$DCIF(\%)$	-0.120 (0.0761)	-0.153 (0.104)	-0.145 (0.0943)	-0.249 (0.288)	-0.157 (0.129)	-0.168 (0.156)	-0.101 (0.0819)	-0.0554 (0.144)
	$F(13, 1841)$	41.64***	20.21***	25.57***	2.72***	12.70***	7.90***	37.09***	10.55***
	\bar{R}^2	0.180	-0.594	-0.290	-10.293	-1.201	-2.525	0.022	-2.117
	NT	1,855	1,855	1,855	1,855	1,756	1,756	1,855	1,855



Índice

1. Objetivo

2. Base de Datos

3. El problema de las regresiones

4. Pruebas de exogenidad

5. Resultados

6. Conclusiones



4. Pruebas de exogenidad

Cuadro 3:

Pruebas estadísticas de variables instrumentales, Pool: 2014 - 2019

Variables	2014-2017			2014-2019		
	F-Test de Stock & Yogo (2005) ¹	t-F Test de Lee et al (2022) ²	Test de Exogeneidad ³	F-Test de Stock & Yogo (2005) ¹	t-F Test de Lee et al (2022) ²	Test de Exogeneidad ³
<i>DCALDIF</i>	3.471	-0.013	0.024327	0.570	0.018	0.022272
			0.024125			0.022115
<i>DCALDIF & CIU</i>	11.640	-1.096	7.50093***	0.442	-0.066	4.87598*
			7.52676***			4.85411**
<i>DMIL</i>	4.940	-0.280	3.2951	0.378	-0.061	4.79925*
			3.27715*			4.78122**
<i>DMIL & CIU</i>	1.301	-0.117	9.43276***	5.503	0.181	6.95757**
			9.57196***			6.98927***
<i>DIMB</i>	1.170	-0.086	1.66686	0.208	-0.046	1.54076
			1.66099			1.53492
<i>DIMB & CIU</i>	7.957	-1.038	17.3006***	1.469	0.125	16.7946***
			17.612***			16.985***
<i>DOCDg</i>	32.662*	2.026**	17.3006***	21.801*	1.882*	11.8115***
			17.612***			11.855***
<i>DOCDg & CIU</i>	16.876*	2.339**	26.7364***	23.625*	2.762***	28.7314***
			27.8163***			29.6865***



Índice

1. Objetivo

2. Base de Datos

3. El problema de las regresiones

4. Pruebas de exogenidad

5. Resultados

6. Conclusiones



5. Resultados

Cuadro 4:
Pruebas estadísticas de variables instrumentales, Pool: 2014 - 2019

Variables	2014-2017		2014-2019	
	K=2	K=5	K=2	K=5
DML-IV				
<i>DCAL</i>	0.266*** (0.086)	0.272*** (0.087)	0.228*** (0.064)	0.240*** (0.064)
<i>IV – DCALDIF</i>	0.0273 (2.219)	-0.0309 (1.997)	-0.209 (4.881)	0.915 (3.927)
<i>IV – DCALDIF&CIU</i>	-3.802* (1.970)	-3.698* (1.948)	-10.79 (16.01)	-8.877 (10.86)
<i>IV – DMIL</i>	-2.965 (2.272)	-2.800 (2.227)	-14.98 (27.63)	-13.31 (25.67)
<i>IV – DMIL&CIU</i>	-12.51 (10.91)	-15.53 (15.64)	3.916 (2.479)	5.096 (3.193)
<i>IV – DIMB</i>	-3.488 (3.812)	-3.952 (4.411)	-7.621 (13.49)	-9.671 (21.17)
<i>IV – DIMB&CIU</i>	-6.219** (2.896)	-6.443** (2.884)	11.59 (8.937)	13.34 (11.70)
<i>IV – DOCDg</i>	1.631** (0.729)	1.981** (0.884)	2.771*** (0.914)	2.688*** (0.936)
<i>IV – DOCDg&CIU</i>	5.716*** (1.585)	6.315*** (1.806)	4.700*** (1.272)	4.928*** (1.403)



Índice

1. Objetivo

2. Base de Datos

3. El problema de las regresiones

4. Pruebas de exogenidad

5. Resultados

6. Conclusiones



6. Conclusiones

Este trabajo utiliza técnicas modernas para determinar la inferencia causal del uso de las herramientas de gestión de calidad sobre las empresas peruanas en el período 2014-2019. La base de datos proviene de la Encuesta Nacional de Empresas del INEI del periodo 2015-2020 (INEI-ENE 2023). Los resultados que a continuación se detallan están sujetos a las limitaciones de estas encuestas. Por un lado, del total de empresas encuestadas para todas regiones y sectores luego del proceso de limpieza, **la muestra se redujo en por lo menos 75% del tamaño original de cada encuesta. Más aún, cerca del 75% de las empresas de la muestra reducida son medianas o grandes, dominando las empresas grandes en cerca el 50% del total de la muestra.** Esto sugiere que **los resultados tienen una mayor aplicabilidad a las empresas grandes (y en menor medida a empresas medianas)** que a las empresas pequeñas. Una segunda limitación de la base de datos es que en muchas encuestas no existe información sobre los aspectos productivos como el ratio capital-trabajo lo cual limita los resultados obtenidos.

Se requiere una mejora drástica de las encuestas conteniendo aspectos productivos y que se recoja más información sobre las empresas pequeñas.

Luego de la batería de métodos y pruebas estadísticas implementadas, la conclusión más robusta para los dos períodos considerados, particularmente para la definición del tamaño de las empresas por el número de trabajadores, **es que las herramientas de gestión de calidad sirven para incrementar la productividad laboral de las empresas, particularmente para las empresas medianas y grandes y del sector manufacturero las cuales dominan en la muestra utilizada.**

Teóricamente existen una serie de mecanismos mediante los cuales los instrumentos de gestión de calidad pueden mejorar la productividad laboral de las empresas. Entre los más relevantes figuran, i) incremento o establecimiento de reputación del producto por los consumidores lo cual incrementa o sostiene las ventas de la empresa; ii) reducción de los costos de producción debido a que las herramientas de gestión de calidad pueden producir productos menos defectuosos y con menos demoras en el proceso de producción; y iii) el hecho que la calidad reduce la reelaboración de los productos incrementando así la productividad.



GRACIAS

