



ASOCIACION ARGENTINA
DE ECONOMIA POLITICA

ANALES | ASOCIACION ARGENTINA DE ECONOMIA POLITICA

LIII Reunión Anual

Noviembre de 2018

ISSN 1852-0022

ISBN 978-987-28590-6-0

Las capacidades de las firmas argentinas: una
aproximación con análisis de cluster

Martínez Correa Julián
Pereira Mariano

Las capacidades de las firmas argentinas: una aproximación con análisis de cluster

Julián Martínez Correa^{*}

Mariano Pereira^{**}

31 de agosto de 2018

Resumen

Las capacidades de las firmas son un input importante a la hora de pensar el desempeño innovativo, productivo y exportador de las mismas. Utilizamos la ENDEI que provee información acerca de distintas dimensiones de las capacidades de las empresas industriales en Argentina para el período 2010-2012. Mediante la aplicación del análisis de cluster para agrupar las firmas argentinas en función de sus capacidades, buscamos relacionar las capacidades con otros resultados del desempeño empresarial. Identificamos la existencia de dos cluster: uno ampliamente mayoritario de bajas capacidades relativas (64 % de la muestra) y otro minoritario de altas capacidades relativas. Encontramos que las empresas de mayores capacidades son de mayor tamaño y en este grupo tienen un mayor peso relativo las empresas farmacéuticas y de otros productos químicos y productos de caucho y plástico. Asimismo, gastan una mayor proporción de las ventas en actividades de innovación y en I+D, tienen una mayor probabilidad de innovar, de exportar, exportan a una mayor cantidad de destinos y una mayor productividad del trabajo. Estos hallazgos alertan sobre las potencialidades de crecimiento de gran parte de las firmas industriales argentinas.

Códigos JEL: C49, O32, L2

Palabras clave: análisis de cluster, capacidades de las firmas, gestión, organización

1. Introducción

La innovación y la mejora tecnológica de las firmas son de los principales motores del crecimiento de la productividad de las firmas (impulsando el proceso de destrucción creativa Schumpeteriana) que contribuye a la mejora de la productividad agregada (Aghion y Howitt, 1992; Aghion et al., 2014;

^{*}Universidad Nacional de La Plata (UNLP) y Centro Interdisciplinario de Estudios en Ciencia, Tecnología e Innovación (CIECTI). E-mail: julian.martinezcorrea@gmail.com, tel: (+54) 11 50230006

^{**}Centro Interdisciplinario de Estudios en Ciencia, Tecnología e Innovación (CIECTI) y Universidad Nacional General Sarmiento (UNGS). E-mail: lic.mpereira@gmail.com

Grossman y Helpman, 1991; Romer, 1990)¹. Tradicionalmente, se ha puesto el foco en la investigación y desarrollo (I+D) formal como propulsora de esta innovación y los bajos niveles gastos en I+D en los países de bajos y medianos ingresos han justificado la intervención pública mediante una amplia gama de instrumentos. Sin embargo, como sugiere hace tiempo la literatura de sistemas nacionales de innovación (SNI), otros factores son relevantes y complementarios para entender la dinámica innovativa de las empresas².

Uno de los más relevantes *inputs* complementarios de la I+D son las capacidades multidimensionales de las firmas asociados a cuestiones productivas, tecnológicas y organizacionales³. Entre ellas, se destacan las capacidades de gestión y organizativas⁴, las cuales van de capacidades organizacionales básicas hasta plan de rutinas y sistemas de gestión de los recursos humanos; algunos ejemplos son *lean manufacturing*, monitoreo del desempeño, supervisión de la calidad, fijación de objetivos claros, incentivos, etc. Por otro lado, las capacidades de absorción (Cohen y Levinthal, 1990) son necesarias para el proceso de transferencia tecnológica y a la vez pueden determinar las capacidades previamente mencionadas. En conjunto, las altas capacidades permiten a las firmas identificar nuevas oportunidades tecnológicas, desarrollar un plan para explotarlas y reunir los recursos humanos necesarios para llevarlo a cabo. De esta forma, son claves para desarrollar nuevos productos, procesos y tecnologías (Cirera y Maloney, 2017; Barletta et al., 2016).

La evidencia empírica reciente, limitada debido a la escasez de datos, muestra que las capacidades de las firmas están fuertemente asociadas con la intensidad de I+D y otros *inputs* de la innovación (Barletta et al., 2016; Bloom et al., 2017; Cirera y Maloney, 2017) así también como con *outputs* de la innovación (controlando por I+D) (Bartz et al., 2016; Bloom et al., 2017; Cirera y Maloney, 2017; Rammer et al., 2009). Adicionalmente, la I+D y las capacidades interactúan positivamente en el proceso innovativo: la inversión en gastos de I+D tiene un mayor impacto en la innovación en las firmas mejor gestionadas que el mismo esfuerzo en empresas pobremente gestionadas (Cirera y Maloney, 2017).

Además del efecto indirecto de las capacidades sobre la productividad al afectar la innovación de las firmas, las prácticas de gestión y organizativas afectan directamente la productividad posibilitando un uso más eficiente de los recursos⁵. Bloom et al. (2013) encuentran un efecto causal importante de estas

¹ Geroski et al. (2009) genera evidencia de que la innovación es el motor más importante de los cambios en la productividad en firmas de Reino Unido. Hall (2011) y Mohnen y Hall (2013) repasan la literatura que encuentra una asociación positiva y significativa entre innovación y productividad.

² Cirera y Maloney (2017) argumentan que la falta de estos otros *inputs* explica lo que ellos denominan la “paradoja de la innovación”: aparentemente altos retornos a los gastos en innovación en los países en desarrollo que no se efectivizan debido a la falta de estos factores claves.

³ La literatura de capacidades se remonta al trabajo seminal de Penrose (1959). El concepto de capacidades de las firmas ha sido estudiado detalladamente por la teoría de la organización, *management* y la literatura de la innovación aunque naturalmente con diferencias (Freeman, 1974; Cohen y Levinthal, 1990 (capacidades de absorción); Rosenberg, 1982; Jacovides y Winter, 2005; (rutinas) Nelson y Winter, 1982; (capacidades dinámicas) Teece y Pisano, 1994; Nelson, 1991). Recientemente, ha cobrado más importancia en la literatura más *mainstream* (por ejemplo, Sutton, 2005, 2012; Maloney, 2017; Bloom et al., 2016, 2017). En este trabajo tomamos un enfoque integrador contemplando las capacidades en un sentido amplio. Entendemos las distintas dimensiones de las capacidades en un modo sistémico, donde todas se retroalimentan y determinan entre sí.

⁴ Ver en Cirera y Maloney (2017) el esquema conceptual de las capacidades productivas y tecnológicas y el subconjunto de capacidades de gestión y organizativas.

⁵ Bartz et al. (2016) encuentran que el efecto directo de las prácticas de gestión y de la innovación sobre la productividad

prácticas sobre la productividad total de los factores (PTF) mediante un experimento de campo en empresas textiles en India⁶. En México, [Bruhn et al. \(2018\)](#) mediante un RCT en pequeñas y medianas empresas (pymes) muestran que el acceso a servicios asesoramiento en gestión durante un año tiene un impacto positivo y grande sobre la PTF, el empleo y la masa salarial. Asimismo, [McKenzie y Woodruff \(2017\)](#) encuentran que las prácticas empresariales predicen mayores tasas de supervivencia y rápido crecimiento de las ventas⁷. Por último, también existen trabajos que las relacionan directamente con el desempeño exportador ([Hallak, 2006](#); [Artopoulos et al., 2013](#)).

Desgraciadamente, la *World Management Survey* (WMS) recientemente introducida por [Bloom y Van Reenen \(2007\)](#)⁸ muestra que las prácticas de gestión están muy ausentes en los países de bajos y medianos ingresos: sus empresas están muy atrasadas en la fijación de objetivos de largo plazo, organización de los recursos humanos y monitoreo del desempeño. Sin embargo, la evidencia sobre el diagnóstico de las capacidades de las firmas al interior de un país es todavía muy escasa⁹.

En este trabajo proponemos aprovechar el análisis de cluster para ahondar en la caracterización las capacidades empresariales de las firmas argentinas. Específicamente, utilizando la Encuesta Nacional del Empleo y la Innovación (ENDEI) queremos explorar la existencia de grupos de empresas (y su magnitud) que cuentan con capacidades similares dentro del grupo pero significativamente distintas a las de otros grupos. La ENDEI recaba información necesaria para este análisis para el período 2010-2012: no solo indaga sobre aspectos tecnológicos de las empresas (I+D formal, recursos humanos calificados, etc.) sino sobre otro conjunto de capacidades de gestión y organizativas como los procesos de mejora continua, trazabilidad, calidad, participación del personal en la toma de decisiones, incentivos, entre otros. Estas múltiples dimensiones de las capacidades son utilizadas por el análisis de cluster para identificar potencialmente grupos de empresas que son esencialmente diferentes de los otros grupos. La caracterización de los grupos deviene de analizar ex-post las variables relevantes de cada uno de ellos. Una vez encontrados los clusters, analizamos si estos grupos definidos en función de *inputs* están asociados a variables estructurales de las empresas (tamaño, capital internacional, etc.) así también como ciertos resultados del desempeño innovativo y de mercado.

Los resultados sugieren la existencia de dos clusters: uno ampliamente mayoritario de bajas capacidades (64 % de la muestra) y otro minoritario de altas capacidades. La diferencias entre los grupos se intensifican en ciertas dimensiones: si bien una porción de las firmas de bajas parecen realizar algunas

difiere según el grado de desarrollo del país: en los países de bajos ingresos el efecto de las prácticas es más fuerte que el de introducir una innovación de producto o proceso.

⁶ [Bloom et al. \(2018\)](#) muestran que las prácticas adoptadas por estas mismas empresas persisten en el tiempo.

⁷ También existen meta-análisis sobre este tipo de impacto. En uno, [McKenzie y Woodruff \(2014\)](#) hallan que en la mayoría de los casos las firmas beneficiarias incorporaron nuevas prácticas y lograron incursionar en nuevos negocios luego de las intervenciones. [Cravo y Piza \(2016\)](#) revisan cuarenta evaluaciones rigurosas de apoyo empresarial a pymes en países de bajos ingresos y encontraron que estas ayudan a mejorar el desempeño de las empresas y crear empleos.

⁸ Los autores diseñan una encuesta con prácticas identificadas por expertos como claves para explicar el desempeño a nivel de la firma. Estas están relacionadas con tres dimensiones: (1) monitoreo del desempeño (*lean manufacturing*, mejora continua, indicadores para medir desempeño, etc.), (2) fijación de objetivos adecuados y actuar cuando surgen problemas, y (3) uso de incentivos para atraer y retener personal talentos.

⁹ La referencia a los países suele ser con la distribución de un promedio estandarizado de las dimensiones de las prácticas de gestión

rutinas de monitoreo del proceso productivo, tanto la calificación de los empleados como la capacitación del personal es muy pobre. Adicionalmente, encontramos que las empresas de mayores capacidades son de mayor tamaño y en este grupo tienen un mayor peso relativo las empresas farmacéuticas y de otros productos químicos, productos de caucho y plástico y un menor peso las de confecciones, madera, edición y muebles. Asimismo, gastan una mayor proporción de las ventas en actividades de innovación y en I+D, tienen una mayor probabilidad de innovar, de exportar, exportan a una mayor cantidad de destinos y una mayor productividad del trabajo.

Este trabajo contribuye a la literatura existente de diversas formas. En primer lugar, proponemos una metodología novedosa (análisis de cluster y sus últimos desarrollos) para analizar las múltiples dimensiones de las capacidades de las firmas; creemos que puede ser una herramienta útil para estudiar más en profundidad la distribución de capacidades de un país o región. En este sentido, también es una aplicación de las novedosas técnicas de *machine learning*, este caso una técnica no supervisada, en el campo de la economía¹⁰. En segundo lugar, Argentina es un país de ingresos medios en los que, como en muchos países de la región, la política pública de innovación se focaliza hace tiempo en fomentar la I+D a través instrumentos convencionales (Crespi et al., 2014). Este trabajo contribuye a una caracterización lo más exhaustiva posible factores complementarios a la I+D¹¹ lo que permite comenzar a pensar la adecuación de aquella política. En tercer lugar, contribuye a la reciente y creciente literatura de las prácticas empresariales y su relación con la dinámica innovativa de la empresa y el desempeño en el mercado brindando evidencia para Argentina.

Los resultados tienen algunas implicancias de política. Dada la mencionada relevancia de los factores complementarios a la I+D, el solo hecho de hacer un diagnóstico de este en las empresas argentinas tiene valor por si mismo. Además, el hecho de que el grueso de las empresas argentinas tengan capacidades (relativas) bajas pone de manifiesto la necesidad de actuar sobre este grupo de empresas con la política pública mejorando las capacidades de gestión y organización básicas de las empresas. Por último, sucesivos análisis de cluster como este permitirían monitorear la evolución en el tiempo de este grupo de bajas capacidades.

El resto del trabajo se organiza de la siguiente manera. En la segunda sección se describen los datos utilizados en el trabajo. En la tercera sección, se presenta la estrategia empírica del análisis de cluster. En la sección cuatro se muestran los grupos resultantes y la asociación entre estos y variables estructurales y de desempeño de la firma. Finalmente, en la última sección concluimos.

2. Datos y estadísticas descriptivas

Los datos utilizados para el análisis de cluster provienen de la Encuesta Nacional de Innovación (ENDEI), realizada por el Ministerio de Ciencia y Tecnología (MINCyT) y el Ministerio de Trabajo y

¹⁰ Una interesante aplicación relacionada es la de Caruso et al. (2015) donde encaran el problema de la medición de la pobreza multidimensional con el análisis de cluster.

¹¹ Barletta et al. (2016) miran correlaciones entre capacidades y perfiles de I+D.

Seguridad Social (MTSS). La encuesta abarca el período 2010-2012 y es representativa de las firmas de la industria manufacturera argentina con al menos diez empleados. Contiene características relevantes básicas tales como localidad, rama de actividad, tamaño, inserción externa, etc. Asimismo, brinda información detallada de las actividades y resultados de innovación y variables asociadas al desempeño económico de las empresas (empleo, valor agregado, ventas, etc.).

Además, de suma importancia a los fines de este trabajo, la encuesta contiene variables relacionadas con distintas dimensiones las capacidades productivas y tecnológicas de las firmas. La elección de las variables se desprende tanto de cuestiones prácticas como conceptuales. Como mencionamos en la introducción, y específicamente en la nota al pie 3, la literatura de capacidades es amplia y con distintas vertientes, aunque naturalmente con un hilo común. En este trabajo tomamos un enfoque integrador para la elección de las variables. En primer lugar, aprovechamos el trabajo de [Barletta et al. \(2016\)](#), quienes con la misma base de datos utilizada en este trabajo (la ENDEI) construyen variables de capacidades de la firma justificando cada una de ellas desde la literatura evolucionista. La gran mayoría (todas las relacionadas a las capacidades de gestión y organizativas) son comunes a las de los trabajos de [Bloom y Van Reenen \(2007, 2010\)](#), cuyo cuestionario sobre prácticas de gestión y organizativas está asentado en la literatura (ver nota 7). Adicionalmente, en base a esta última literatura de *management* incorporamos otras dimensiones como la de incentivos a los recursos humanos. Específicamente, consideramos cuatro dimensiones: (i) monitoreo del desempeño productivo; (ii) organización del trabajo; (iii) absorción; y (iv) incentivos¹².

Para la primera dimensión asociada a las prácticas de gestión, tenemos un conjunto de indicadores que dan cuenta de actividades y herramientas relacionadas al monitoreo del desempeño en la producción (especificación de materias primas e insumos críticos, especificación de características críticas del proceso productivo, trazabilidad, equipos para mejora de procesos, herramientas de mejora continua, implementación de normas de calidad, etc.).

La dimensión de la organización del trabajo se relaciona con las capacidades asociadas a mecanismos sistemáticos de generación y circulación de conocimiento en la organización. Así, contemplamos la existencia de esquemas rotación planificada entre puestos de trabajo y el grado de participación del personal en los procesos de evaluación y planificación de mejoras.

Incluimos además las capacidades de absorción del conocimiento externo para incorporarlo en la generación de nuevo conocimiento (transferencia tecnológica) ([Cohen y Levinthal, 1990](#)). Las capacidades de absorción acumuladas las medimos con el stock de recursos humanos calificados (porcentaje de profesionales y con calificación técnica en el total de empleados) y la presencia de un departamento formal de I+D. Por otra parte, las capacidades de absorción potenciales las aproximamos con la capacitación de los trabajadores (existencia de área responsable de organizar actividades de capacitación y el porcentaje de personas que recibieron cursos a nivel jerárquico y no jerárquico) y la existencia de vinculaciones con otras empresas y con el sector público ([Lall, 1992](#)).

¹² Cabe aclarar que la clasificación en dimensiones es puramente descriptiva ya que la técnica de clustering no utiliza ninguna clasificación inicial de las variables.

Por último, en la dimensión de incentivos, analizamos si la firma aplica algún sistema de evaluación de desempeño para el personal. Una descripción más detallada de las variables utilizadas puede encontrarse en el Apéndice A.

La muestra final contiene 3.240 empresas¹³. La Tabla 1 presenta estadísticas descriptivas básicas de las variables utilizadas en el análisis de cluster para la muestra considerada. Por ejemplo, muestra que un 68 % de las empresas realiza especificaciones de los insumos críticos y/o características críticas del proceso/producto. El 52 % de las empresas implementa normas de calidad, sectoriales y/o de productos. En promedio, del total de empleados de la firma más de un 7 % son profesionales y un 8 % son técnicos. Las firmas en las que hay una rotación planificada del personal son el 32 %. Aproximadamente un 30 % tiene vinculaciones con empresas y con el sector público. Por último, un 38 % tiene incentivos para el personal.

Tabla 1: Estadísticas descriptivas

	Media	Mediana	SD	N
Especificaciones	0.683	1.000	0.465	3,240
Trazabilidad	0.646	1.000	0.478	3,240
Equipos para solucionar problemas	0.523	1.000	0.500	3,240
Mejora continua	0.378	0.000	0.485	3,240
Gestión de proyecto y diseño	0.326	0.000	0.469	3,240
Normas	0.521	1.000	0.500	3,240
Rotación del personal	0.327	0.000	0.469	3,240
Grado participación del personal	0.633	0.000	0.901	3,240
Departamento formal I+D	0.294	0.000	0.456	3,240
Profesionales sobre total empleo (%)	7.170	3.966	10.646	3,240
Técnicos sobre total empleo (%)	8.375	4.169	12.586	3,240
Área para actividades capacitación	0.452	0.000	0.498	3,240
Personal capacitado a nivel jerárquico (%)	20.202	0.000	33.472	3,240
Personal capacitado a nivel no jerárquico (%)	24.622	0.000	38.723	3,240
Vinculaciones con firmas	0.351	0.000	0.477	3,240
Vinculaciones con sector público	0.287	0.000	0.452	3,240
Incentivos al personal	0.377	0.000	0.485	3,240

Nota: La tabla presenta estadísticas de la muestra utilizada en base a la ENDEI(2010-2012). Una descripción detallada de las variables se encuentra en el Apéndice A.

¹³ Se eliminan *outliers* según las recomendaciones de uso de la encuesta y las observaciones sin información en alguna de las variables utilizadas en el presente trabajo.

3. Análisis de cluster

La estrategia empírica consiste en caracterizar el entramado industrial argentino identificando grupos de empresas que son similares entre sí y distintas a las de los otros grupos en base a sus capacidades. Un método apropiado es el análisis de cluster no jerárquico, como algoritmo de K-medias, ya que particiona la base de datos en K clusters mutuamente excluyentes¹⁴. El método resuelve el trade-off de la similitud dentro/entre: asigna las observaciones a grupos de forma tal que estas son cercanas (definiendo una medida de distancia) a las del mismo grupo y lejanas a aquellas de los otros grupos. En otras palabras, divide a las observaciones entre grupos homogéneos (baja varianza dentro del cluster) mientras que los grupos son muy distintos entre sí (alta varianza entre clusters).

Sea una matriz X de $N \times P$, donde las filas se corresponden con N observaciones y las columnas a P variables, las cuales representan conjuntamente las capacidades multidimensionales de las firmas. Cada fila es un punto de p dimensiones y un cluster es una colección/grupo de estos puntos. El mecanismo de clusters está caracterizado por una función o “encoder” $C(i) : (1, \dots, N) \rightarrow (1, \dots, K)$ que, para K clusters, asigna cada punto a un solo cluster. El objetivo es dividir los puntos en clusters de modo que los puntos dentro de un cluster sean similares y a su vez distintos a los de cualquier otro cluster (Caruso et al., 2015).

Así, la noción de similitud es de suma relevancia. Sea $d(x_i, x_j)$ una función de distancia entre dos puntos x_i y x_j de la matriz X . Consideremos el siguiente problema:

$$\min_C W(C) = \frac{1}{2} \sum_{k=1}^K \left[\sum_{i,j/C(i)=C(j)=k} d(x_i, x_j) \right] \quad (1)$$

Los aglortimos de cluster resuelven este problema, es decir, minimizan la función de pérdida $W(C)$ (la cual agrega las similitudes *dentro* de cada cluster) definida sobre el conjunto de todos los *encoders* posibles. Es interesante notar que minimizar $W(C)$ es equivalente a maximizar la agregación de las distancias *entre* clusters. Una posibilidad es minimizar $W(C)$ chequeando todas las posibles clusterizaciones pero es computacionalmente impracticable. El algoritmo de K-medias resuelve este problema de una forma simple encontrando un mínimo local. La función de distancia más comunmente utilizada es el cuadrado de la distancia euclidiana¹⁵ $d(x_i, x_j) = \|x_i - x_j\|^2$. El algoritmo de K-medias comienza asignando aleatoriamente las observaciones a clusters iniciales, computa el vector de medias de las p variables para las observaciones del cluster k , reasigna las observaciones al cluster más cercano de acuerdo a las medias computadas e itera este proceso hasta que no haya más reasignaciones. Definiciones alternativas de la función de distancia como de los *centroides* (las medias en el caso anterior)

¹⁴ Los algoritmos de cluster pueden categorizarse principalmente en dos grupos: (i) clustering jerárquico: arma un árbol con los datos cada partición son grupos cada vez más similares; no requieren que se especifique de antemano el número de clusters y (ii) clustering de particiones. Métodos de evaluación de los clusters sugieren que K-medias es un método adecuado para nuestros datos. Ver discusión más abajo y Apéndice C.

¹⁵ En este caso: $W(C) = \sum_{k=1}^K N_k \left[\sum_{i/C(i)=k} \|x_i - \bar{x}_k\|^2 \right]$, donde N_k es el número de observaciones del cluster k y $\bar{x}_k$ es el vector de medias del cluster k .

determinan distintos resultados.

De esta forma, los métodos de cluster son sensibles a la función de distancia, la estandarización y la clusterización inicial, lo que requiere tomar decisiones previo al análisis de cluster¹⁶. En primer lugar, como nuestra base de datos cuenta con variables categóricas/cuantitativas discretas y continuas (proporciones) con distintos rangos es necesario estandarizar las variables para evitar efectos de escala. A cada observación x_{il} de la variable l se le resta el mínimo de la variable y es dividida por su rango, obteniendo su versión estandarizada z_{il} :

$$z_{il} = \frac{x_{il} - \min_l(x_{il})}{\max_l(x_{il}) - \min_l(x_{il})} \quad (2)$$

Mientras que las variables binarias y proporciones toman los mismos valores, las categóricas si se modifican. Todas las variables quedan con un rango $[0, 1]$.

La elección de la función de distancia también es un aspecto relevante del análisis de cluster. Para variables continuas una decisión usual es trabajar con el cuadrado de la euclidiana (como se comentó anteriormente en el algoritmo tradicional de K-medias). Sin embargo, dada la combinación de variables categóricas (binarias y cuantitativas discretas) y continuas, una mejor opción es trabajar con una medida aditiva como la distancia de Manhattan¹⁷. Para dos observaciones z_i y z_j , esta se define como:

$$d_{ij} = \sum_{l=1}^P |z_{il} - z_{jl}| \quad (3)$$

Cabe notar que para el caso de las variables binarias la distancia mide el número de variables en las que las observaciones toman el mismo valor (al igual que el índice de Jaccard).

Como mencionamos, las condiciones iniciales del algoritmo pueden afectar los resultados (el algoritmo encuentra un óptimo local en vez de uno global). Por ello, consideramos 50 inicializaciones y tomamos la de menor suma de cuadrados intra-cluster.

Una de las decisiones más importantes a tomar ex-ante es la cantidad de clusters. Es importante remarcar que nada garantiza la existencia efectiva de clusters (menos un número determinado de ellos). Por ello, por un lado, utilizamos el índice de Calinski-Harabasz ([Caliński y Harabasz, 1974](#)), ampliamente utilizado en la literatura, para encontrar el número óptimo de clusters. Por otro lado, ante la presencia de una multiplicidad de métodos para tomar esta decisión, calculamos 30 índices que sostienen la elección de un dado K y elegimos el número de clusters óptimo en base a la regla de la mayoría¹⁸.

Por último, los clusters encontrados pueden ser simplemente el resultado de agrupar “ruido” y no representar verdaderos subgrupos. Por ello, analizamos de diversas maneras (índice Hopkins, test

¹⁶ Dada la naturaleza similar de los datos, en lo que sigue tomamos como referencia a [Caruso et al. \(2015\)](#).

¹⁷ Además, esta medida se ve menos afectada por la presencia de *outliers*.

¹⁸ Utilizamos la función *NbClust* implementada en R.

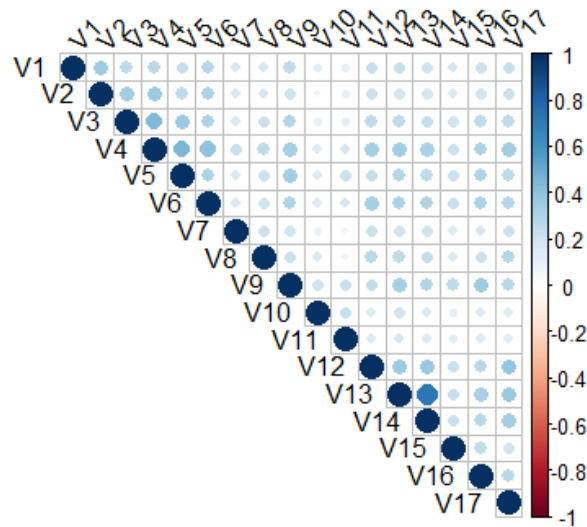
Hotelling, evaluación visual de la tendencia de agrupamiento, etc.) si efectivamente puede decirse que en nuestros datos existe una tendencia al agrupamiento que justifique un análisis de cluster.

4. Resultados

4.1. Identificando los clusters

Las variables incluidas en el análisis de clusters esperamos que capten distintas dimensiones de las capacidades de las empresas. La Figura 1 ilustra las correlaciones entre las variables utilizadas. Encontramos que no existe una correlación excesivamente alta entre ellas lo que aporta evidencia sobre la característica multidimensional de las capacidades¹⁹. En el Apéndice B presentamos la matriz de correlaciones.

Figura 1: Correlaciones



Nota: Los nombres se corresponden con el orden de las variables del Apéndice A. Ver en Apéndice B la matriz de correlaciones.

La elección del número óptimo de grupos es un paso crucial en el algoritmo de K-medias. La Tabla 2 muestra que el índice de Calinski-Harabasz decrece monotonímicamente con un máximo en $K = 2$. Por otro lado, la Figura 2 presenta la frecuencia de los resultados de 30 índices para determinar el número de clusters: por regla de la mayoría (11 métodos), el número óptimo es dos²⁰. En suma, los datos sugieren que existen dos grupos de empresas cuando las intentamos caracterizar en función de sus capacidades.

Es importante remarcar que el número de clusters no se define independientemente de los datos,

¹⁹ La mayor correlación se da naturalmente entre el porcentaje del personal capacitado a nivel jerárquico y no jerárquico.

²⁰ El método de silhouette, otro de los más conocidos, también recomienda dos.

Grupos (K)	Índice CH
1	
2	896.06
3	625.77
4	525.21
5	459.90
6	380.91
7	372.23
8	349.63

Tabla 2: Índice Calinski-Harabasz.

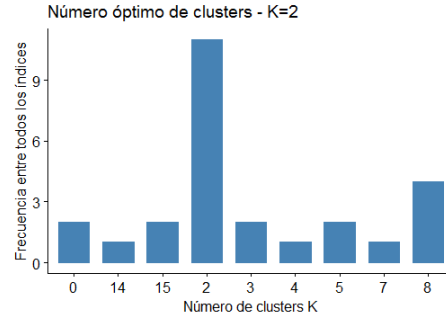


Figura 2: Frecuencia K óptimo según los resultados de 30 índices.

de otro modo, cualquier número es necesariamente arbitrario. De hecho, podría ser el caso en que efectivamente no existan grupos y que las capacidades se distribuyan uniformemente en un solo conjunto de empresas. Cobra relevancia en este marco intentar definir la cantidad de grupos con algún criterio de objetividad. Haber encontrado dos grupos de firmas implica que hacer más particiones no agrega suficiente información.

Complementariamente, analizamos si efectivamente existe una tendencia al agrupamiento (es decir, si los datos son efectivamente agrupables). En primer lugar, calculamos el estadístico de Hopkins el cual evalúa la tendencia de *clustering* mediante el cálculo de la probabilidad de que los datos procedan de una distribución uniforme (es decir, testea la aleatoriedad espacial de los datos). El valor del indicador se encuentra cerca de cero (0,25) lo que sugiere que la estructura de los datos contiene algún tipo de agrupación²¹. Asimismo, evaluamos la tendencia visualmente con el método VAT²² (*Visual Assessment of cluster Tendency*). La Figura C1 en el Apéndice C confirma, con sus formas rectangulares, que en nuestro set de datos sí hay una estructura de grupos. Por otro lado, los resultados para el test de Hotelling para las diferencias de medias muestran claramente diferencias estadísticamente significativas entre los dos grupos ($p = 0,000$). Esto es esperable: el algoritmo de k-medias maximiza la diferencia entre los grupos. Por último, analizamos si la elección del algoritmo de K-medias (en contraposición con el *clustering* jerárquico) y el número de clusters es pertinente dados nuestros datos. En concreto, calculamos para el algoritmo de K-medias y para *clustering* jerárquico y para un rango de número de grupos los valores que toman determinadas medidas de validación interna (conectividad, silhouette y Dunn) y de estabilidad (APN, AD, ADM, FOM)²³. En la Tabla C1 del Apéndice C mostramos los

²¹ Lo calculamos con la función *get_clust_tendency()* del paquete *factoextra* de R. El estadístico es computado como $1 - H$, donde $H = \frac{\sum_{i=1}^N y_i}{\sum_{i=1}^N x_i + \sum_{i=1}^N y_i}$ con x_i la distancia del punto i a su vecino más cercano y y_i la distancia de del punto i de una distribución uniforme simulada a su vecino más cercano de los datos originales. Valores cercanos a cero sugieren datos clustereables

²² Representa gráficamente la matriz de distancias ordenada (observaciones similares situadas cerca unas de otras) empleando un gradiente de color para el valor de las distancias.

²³ Las medidas de validación interna miden la compactibilidad y separación de los clusters. En particular, la conectividad indica el grado de conexión de los clusters determinado según si los vecinos más cercanos se encuentran o no en el mismo cluster. Toma valores de 0 a infinito y debe ser minimizado. El índice de silhouette define la compactibilidad basado en las distancias de todos los pares de observaciones en un cluster y la separación basado en todas las distancias entre todos los puntos de un cluster y los de su cluster más cercano. Son deseables valores cercanos a 1. El índice de

resultados. Todas las medidas coinciden en que el mejor método es K-medias (*clustering* jerárquico no resulta óptimo en ninguna de las medidas) y, además, en casi todas el número óptimo de grupos es dos. Estos hallazgos parecen corroborar la elección de dos grupos.

La Tabla 3 presenta las medias de los dos grupos que surgen del análisis de cluster. Encontramos, por un lado, un grupo de altas capacidades (relativas) compuesto por menos del 36 % de la muestra y, por el otro, un grupo de bajas capacidades (relativas) con el 64 % restante²⁴. En concreto, el primer grupo exhibe medias en todas dimensiones de las capacidades significativamente mayores a las del segundo. Por ejemplo, en la dimensión productiva, más de un 90 % del grupo de altas capacidades realiza las especificaciones y trazabilidad contra aproximadamente un 50 % en el grupo de bajas. Además, un 80 % de las empresas del primer grupo utiliza herramientas de mejora continua mientras que menos de un 15 % lo hace en el segundo grupo. La implementación de normas de calidad, sectoriales o de producto también es una característica común a la mayoría de las empresas de altas capacidades: un 90 % de las firmas este grupo tiene estas normas contra un 30 % del grupo de bajas.

En lo que refiere a la dimensión de organización del trabajo, más de un 50 % de las empresas identificadas como de altas capacidades implementan rotaciones planificadas del personal y un 20 % del otro grupo lo hace. Adicionalmente, el grado de participación del personal en el desarrollo de actividades también es sustancialmente mayor en el primer grupo.

Por otro lado, el grupo de altas capacidades se destaca en la dimensión de capacidades de absorción. Un 60 % tiene un departamento formal de I+D (contra un 12 % del otro grupo) y el porcentaje de de profesionales y técnicos sobre el total de empleo es 10 % en ambos casos mientras que en el grupo de bajas capacidades este porcentaje es despreciable. Por otro lado, el personal capacitado a nivel jerárquico y no jerárquico es una porción relevante del personal en las empresas de altas capacidades mientras que es despreciable en el grupo de bajas capacidades. Asimismo, las redes de vinculaciones también son más fuerte en el grupo de mayores capacidades: mientras que cerca de un 60 % tiene vinculaciones con el sector público y con otras empresas en el de altas, solo un 11 % y un 23 % lo hace en el de bajas, respectivamente.

Por último, los incentivos al personal se dan en un 71 % de las empresas del primer grupo y solo en un 19 % de las del segundo.

En resumen, estos resultados muestran que los grupos resultantes del análisis de cluster son económica y significativamente diferentes. La porción mayoritaria (64 %) de las empresas industriales argen-

Dunn representa el ratio entre la menor distancia entre observaciones que no pertenecen a un mismo cluser y la mayor distancia intra-cluster y tiene que ser maximizado. Por su parte, las medidas de estabilidad testean qué tan sensible es la cluserización a perturbaciones en la base de datos. *Average proportion of non-overlap* (APN) mide la proporción promedio de observaciones que no se ubican en el mismo cluster basado en clustering con todos los datos y clustering habiendo removido una columna. El *average distance* (AD) mide la distancia promedio entre las observaciones ubicadas en el mismo cluster también bajo los dos casos anteriores. El *average distance between means* (ADM) mide la distancia promedio entre los centros de los clusters para observaciones situadas en el mismo cluster bajo los dos casos. Por último, el *figure of merit* (FOM) mide la varianza promedio intra-cluster de la columna removida, donde el clustering se hace con las columnas restantes. Todas estas deben ser minimizadas.

²⁴ El hecho de encontrar dos grupos estadísticamente diferentes no necesariamente implica que uno tiene que tener capacidades relativas más altas que el otro.

tinas está compuesta por firmas con escasas capacidades y tienen prácticas de gestión rudimentarias en relación al grupo de mejores capacidades relativas. En este sentido, los resultados alertan sobre las potencialidades de crecimiento del grueso de las empresas industriales argentinas. Sin embargo, las diferencias entre los dos grupos no son homogéneas en todas las dimensiones. Específicamente, las firmas de bajas parecen realizar ciertas rutinas de monitoreo del proceso productivo (concretamente, especificaciones de las materias primas, insumos críticos y/o características del proceso y trazabilidad) pero tanto la calificación de los empleados como la capacitación del personal jerárquico y no jerárquico es muy magra.

Tabla 3: Composición de los clusters

	Altas	Bajas
Especificaciones	0.925	0.547
Trazabilidad	0.911	0.497
Equipos para solucionar problemas	0.857	0.336
Mejora continua	0.792	0.146
Gestión de proyecto y diseño	0.651	0.143
Normas	0.889	0.314
Rotación del personal	0.523	0.216
Grado participación del personal	0.350	0.133
Departamento formal I+D	0.603	0.120
Profesionales sobre total empleo (%)	0.103	0.054
Técnicos sobre total empleo (%)	0.125	0.061
Área para actividades capacitación	0.799	0.257
Personal capacitado a nivel jerárquico (%)	0.445	0.066
Personal capacitado a nivel no jerárquico (%)	0.514	0.096
Vinculaciones con firmas	0.563	0.232
Vinculaciones con sector público	0.593	0.115
Incentivos al personal	0.714	0.188
N	1,166	2,074

Nota: Medias de las variables resultantes del análisis de cluster.

4.2. Capacidades, características y desempeño de las empresas

Las capacidades pueden estar asociadas a diversas características de las empresas, tanto estructurales como indicadores del desempeño de las firmas. En esta sección analizamos si aquellos grupos clasificados únicamente en función de las capacidades de gestión, organizativas y de absorción tienen

un correlato en otras variables relevantes.

La Tabla 4 muestra la distribución de distintas características de las firmas (tamaño²⁵, antigüedad, origen del capital y rama de actividad) para cada uno de los grupos resultantes (frecuencias relativas condicionadas por columna). Encontramos que las empresas con mejores capacidades en promedio, de mayor tamaño, de mayor edad y tienen una mayor proporción de capitales extranjeros. Esto último va en línea con lo hallado por Bloom y Van Reenen (2010) de que las multinacionales tienen buenas prácticas de gerenciamiento en todos los países. En cuanto a la rama de actividad²⁶, se destacan diferencias en el peso relativo de determinados sectores. En particular, las empresas de altas capacidades relativas tienen mayor peso en las ramas de farmacéuticas y otros productos químicos y productos de caucho y plástico y menor peso relativo en las de confecciones, madera, edición y muebles.

²⁵ La ENDEI utiliza el número de empleados para clasificar a las firmas por tamaño: las pequeñas son las que tienen entre 10 y 25 empleados, las medianas entre 26 y 99 empleados y las grandes con 100 o más.

²⁶ La desagregación en la ENDEI por rama de actividad es a dos dígitos del CIIU Rev.3, y una apertura a cuatro dígitos en algunos sectores de especial interés (Alimentos y bebidas, Químicos, Maquinaria y equipo y Automotores).

Tabla 4: Características estructurales de los clusters

	Altas	Bajas
Tamaño		
Pequeña	22.5 %	53.3 %
Mediana	36.7 %	37.8 %
Grande	40.8 %	8.9 %
Edad		
Menos de 10 años	18.5 %	30.0 %
Más de 10 años	81.5 %	70.0 %
Origen capital		
Capital nacional	82.4 %	96.8 %
Capital internacional	17.6 %	3.2 %
Rama de actividad		
Confecciones	1.5 %	5.4 %
Madera	1.5 %	4.3 %
Edición	1.8 %	4.6 %
Productos químicos	8.0 %	3.4 %
Productos de caucho y plástico	7.1 %	4.2 %
Muebles	2.1 %	4.6 %
Farmacéuticas	7.5 %	1.6 %

Nota: Composición de variables estructurales para los clusters de la Tabla 3. En la rama de actividad, por simplicidad solo se presentan aquellas en donde existen diferencias relevantes en el peso relativo entre clusters. En el Apéndice B se presentan la desagregación completa.

La Tabla 5 muestra cual es el la actividad innovadora de las empresas en cada uno de los grupos identificados. Podemos ver que prácticamente la totalidad de las empresas del grupo de altas capacidades realizan gastos en actividades de innovación (AI) mientras que aproximadamente la mitad del otro grupo lo hace. Específicamente en I+D, aproximadamente un 50 % del primer grupo realiza este tipo de gastos contra menos de un 17 % en el segundo. Esto sugiere que, a diferencia del grupo de altas capacidades, en el segundo grupo los gastos en AI se concentran en gastos que no son de I+D²⁷. En lo que refiere a la intensidad de los gastos en AI y en I+D los resultados van en la misma dirección: las empresas con mejores capacidades gastan una proporción de las ventas en AI y en I+D que es más del

²⁷ Estos pueden ser: adquisición de bienes de capital; hardware; software; contratación de tecnología; capacitación; diseño Industrial y actividades de ingeniería; diseño organizacional y gestión; y consultorías.

doble y triple de la que gastan las de bajas, respectivamente. Por último, estos mayores gastos en AI y I+D parecen traducirse en resultados de innovación: un 73 % de las empresas de mayores capacidades lograron innovaciones en productos o procesos en el período contra un 34 % del otro grupo. Estas correlaciones son consistentes con la evidencia presentada por [Cirera y Maloney \(2017\)](#) y [Bloom et al. \(2017\)](#)²⁸.

Tabla 5: Indicadores de innovación y clusters

	Altas	Bajas
Gasta en AI (=1 si gasta)	0.90	0.51
Gasta en I+D (=1 si gasta)	0.49	0.17
AI /ventas (%)	3.13	1.47
I+D/ventas (%)	0.63	0.21
Innovó (=1 si innovó en producto o proceso)	0.73	0.34

Nota 1: Las variables binarias se construyeron en función de si en el período 2010-2012 se dio el evento mientras que los niveles son promedios del mismo período.

Nota 2: Medias de las variables de insumos y resultados de innovación para cada cluster obtenido (Tabla 3). Para todas las variables, un test de diferencias de medias entre los grupos arroja diferencias significativas al nivel del 1 %.

La Tabla 6 se concentra en algunos indicadores del desempeño de las empresas (*outputs*). Encontramos que empresas de altas capacidades relativas son a su vez las que no solo tienen mayor empleo sino también una mayor productividad laboral (valor agregado por trabajador). Asimismo, este grupo más reducido de empresas presenta un mejor desempeño exportador: más de un 60 % exporta contra un 26 % del grupo de bajas capacidades. Al considerar solo las empresas exportadoras, la cantidad de destinos²⁹ y la proporción de empresas que exportan a destinos de altos ingresos es en ambos casos mayor en las empresas mejor gestionadas y con mayores capacidades³⁰.

²⁸ [Bloom et al. \(2017\)](#) dividen a las empresas en quintiles según el score en las prácticas y muestran que aquellas con mejores prácticas de gestión son las que tienen mayor intensidad de I+D y registran más patentes por trabajador.

²⁹ la ENDEI considera seis destinos: MERCOSUR, otros países de América Latina, Estados Unidos y Canadá, Europa, Asia y África y Oceanía.

³⁰ Con una reciente encuesta y análisis de regresión, [Bloom et al. \(2017\)](#) encuentran resultados similares en la relación más específica de desempeño y prácticas de gestión: los establecimientos utilizando mejores prácticas de gestión tienen mayor productividad, rentabilidad, innovación (medidas como gastos de I+D y la intensidad de patentamiento) y crecimiento.

Tabla 6: Performance productiva y clusters

	Altas	Bajas
Empleo	246.09	48.30
Valor agregado por trabajador (USD)	63,923.73	47,829.84
Exporta	0.61	0.26
Cantidad de destinos (si exporta=1)	2.60	1.92
Exporta destinos de altos ingresos (si exporta=1)	0.38	0.25

Nota 1: Consideramos destinos de altos de altos ingresos son Europa, Estados Unidos y Canadá.

Nota 2: Medias de las variables para cluster obtenido (Tabla 3). Para todas las variables, un test de diferencias de medias entre los grupos arroja diferencias significativas al nivel del 1 %.

En síntesis, encontramos que los clusters encontrados en función de las capacidades están asociados a diversas características tanto estructurales de las firmas como a su desempeño innovativo y productivo. El grupo minoritario de mejores prácticas de gestión, organizativas y de gestión también presenta mejores resultados en la dinámica de innovación, productiva y exportadora mientras que el grupo mayoritario presenta resultado más pobres. Si bien no puede atribuirse causalidad en las correlaciones encontradas, al menos sugiere que la información de de las capacidades puede ser muy relevantes para analizar las potencialidades de desarrollo de las firmas³¹.

5. Conclusiones

Este trabajo presentó evidencia acerca del estado de las múltiples dimensiones de las capacidades de las firmas industriales en Argentina. Para ello, hicimos uso del análisis de cluster a los fines de caracterizar estas facultades en función de grupos de empresas con capacidades de gestión, organizativas y de absorción similares dentro de un grupo pero significativamente distintas a las de otro grupo.

Los resultados utilizando la ENDEI sugieren la existencia de dos grupos de empresas esencialmente diferentes. Por un lado, un grupo mayoritario (64 % de la muestra) con bajas capacidades en todas las dimensiones consideradas. Por el otro, un grupo menor compuesto por las firmas restantes que presenta altas capacidades (relativas). Las diferencias entre los grupos se exacerban en la calificación y la capacitación de personal. Asimismo, las capacidades parecen estar asociadas a características estructurales y el desempeño innovativo, productivo y exportador de las firmas. Específicamente, encontramos que las empresas de mayores capacidades son de mayor tamaño, gastan una mayor proporción de las ventas en actividades de innovación y en I+D, tienen una mayor probabilidad de innovar, de exportar, exportan a una mayor cantidad de destinos y destinos de altos ingresos y una mayor productividad laboral.

³¹ Otros trabajos ya mencionados como Bloom et al. (2013) aportan evidencia causal que apoya los resultados de este trabajo.

Dada la asociación entre las capacidades y el desempeño, los resultados de este trabajo alertan sobre las potencialidades de desarrollo de gran parte de las firmas del país: la mayoría de las empresas industriales argentinas presenta capacidades muy rudimentarias. Por otro lado, el hecho de que solo hayan resultado dos grupos del análisis de cluster puede estar sugiriendo la complementariedad entre las múltiples dimensiones de las capacidades donde todas las variables juegan un rol importante para caracterizar a las firmas. Como la acumulación de capacidades es un proceso acumulativo y local, la información diferencial puede ser clave para explicar por qué las las empresas quedan rezagadas en sus capacidades³². Creemos que desarrollar las capacidades de las firmas tiene que ser un elemento central de la política de innovación, y de desarrollo en general, sobre todo para países como Argentina que están lejos de la frontera tecnológica.

Teniendo presente que los resultados muestran que una gran proporción de empresas argentinas parece tener capacidades bajas, incluso muy pobres (en términos absolutos) en lo que hace a la calificación de los recursos humanos, llama la atención que [Goñi y Maloney \(2017\)](#) encuentren que el retorno a la inversión en I+D se maximizaría en un país a una distancia de la frontera tecnológica como la de la Argentina actual (2006-2010). Esto es, según las estimaciones de los autores el país estaría a una distancia óptima de la frontera tecnológica contando con los factores complementarios para maximizar los retornos de la inversión en I+D. Sin embargo, si bien sería deseable una comparación internacional, nuestros resultados sugerirían que aún hay un largo camino por recorrer en las capacidades de las firmas.

El trabajo no está exento de limitaciones. En primer lugar, como mencionábamos es necesario tener una visión comparativa de las capacidades entre países, por ejemplo, incluyendo todas las empresas de los países de América Latina. Adicionalmente, una de las desventajas del algoritmo de K-medias es que las observaciones son necesariamente clasificadas en un grupo. Sin embargo, las firmas pueden pertenecer a los grupos con determinada probabilidad. Otros tipos de algoritmos como *fuzzy clustering* o *model-based clustering* contemplan esto. A modo de ejemplo, en nuestro caso aquellas empresas que parecen estar en la frontera entre los clusters de bajas y altas capacidades podrían ser un grupo interesante para la intervención pública. Por último, no necesariamente todas las variables son relevantes para la discriminación entre clusters. Existen desarrollos muy recientes para encontrar las variables realmente informativas en la clusterización aunque existen complicaciones con datos con variables categóricas y continuas como en nuestro caso. Estas son áreas promisorias para la investigación futura.

Referencias

Aghion, P., Akcigit, U., y Howitt, P. (2014). What do we learn from schumpeterian growth theory? In *Handbook of economic growth*, volume 2, pages 515–563. Elsevier.

³² En cuanto a las prácticas de gestión y organizativas, [Bloom et al. \(2013\)](#) y [Bloom et al. \(2017\)](#) encuentran que muchas veces las empresas no están al tanto de cuales son las mejores prácticas (no saben cómo se comparan las suyas con las mejores) o piensan que no son relevantes

- Aghion, P. y Howitt, P. (1992). A model of growth through creative destruction. *Econometrica*, 60(2):323–51.
- Artopoulos, A., Friel, D., y Hallak, J. C. (2013). Export emergence of differentiated goods from developing countries: Export pioneers and business practices in argentina. *Journal of Development Economics*, 105:19–35.
- Barletta, M. F., Pereira, M., Suárez, D., y Yoguel, G. (2016). Construcción de capacidades en las firmas argentinas. más allá de los laboratorios de i+ d. *Pymes, Innovación y Desarrollo*, 4(3):39–56.
- Bartz, W., Mohnen, P., y Schweiger, H. (2016). The role of innovation and management practices in determining firm productivity in developing economies. MERIT Working Papers 034, United Nations University - Maastricht Economic and Social Research Institute on Innovation and Technology (MERIT).
- Bloom, N., Brynjolfsson, E., Foster, L., Jarmin, R. S., Patnaik, M., Saporta-Eksten, I., y Van Reenen, J. (2017). What drives differences in management? Technical report, National Bureau of Economic Research.
- Bloom, N., Eifert, B., Mahajan, A., McKenzie, D., y Roberts, J. (2013). Does management matter? evidence from india. *The Quarterly Journal of Economics*, 128(1):1–51.
- Bloom, N., Mahajan, A., McKenzie, D., y Roberts, J. (2018). *Do management interventions last? evidence from India*. The World Bank.
- Bloom, N., Sadun, R., y Van Reenen, J. (2016). Management as a technology? Technical report, National Bureau of Economic Research.
- Bloom, N. y Van Reenen, J. (2007). Measuring and explaining management practices across firms and countries. *The Quarterly Journal of Economics*, 122(4):1351–1408.
- Bloom, N. y Van Reenen, J. (2010). Why do management practices differ across firms and countries? *Journal of economic perspectives*, 24(1):203–24.
- Bruhn, M., Karlan, D., y Schoar, A. (2018). The impact of consulting services on small and medium enterprises: Evidence from a randomized trial in mexico. *Journal of Political Economy*, 126(2):635–687.
- Caliński, T. y Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3(1):1–27.
- Caruso, G., Sosa-Escudero, W., y Svarc, M. (2015). Deprivation and the dimensionality of welfare: a variable-selection cluster-analysis approach. *Review of Income and Wealth*, 61(4):702–722.

- Cirera, X. y Maloney, W. F. (2017). *The innovation paradox: Developing-country capabilities and the unrealized promise of technological catch-up*. The World Bank.
- Cohen, W. M. y Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, 35:128–152.
- Cravo, T. A. y Piza, C. (2016). *The impact of business support services for small and medium enterprises on firm performance in low-and middle-income countries: a meta-analysis*. The World Bank.
- Crespi, G., Fernández-Arias, E., y Stein, E. (2014). Investing in ideas: policies to foster innovation. In *Rethinking Productive Development*, pages 61–106. Springer.
- Geroski, P., Kretschmer, T., y Walters, C. (2009). Corporate productivity growth: champions, leaders, and laggards. *Economic Inquiry*, 47(1):1–17.
- Goñi, E. y Maloney, W. F. (2017). Why don't poor countries do r&d? varying rates of factor returns across the development process. *European Economic Review*, 94:126 – 147.
- Grossman, G. M. y Helpman, E. (1991). Quality ladders in the theory of growth. *The Review of Economic Studies*, 58(1):43–61.
- Hall, B. H. (2011). Innovation and productivity. Technical report, National bureau of economic research.
- Hallak, J. C. (2006). Product quality and the direction of trade. *Journal of international Economics*, 68(1):238–265.
- Lall, S. (1992). Technological capabilities and industrialization. *World development*, 20(2):165–186.
- Maloney, W. F. (2017). *Revisiting the national innovation system in developing countries*. The World Bank.
- McKenzie, D. y Woodruff, C. (2014). What are we learning from business training and entrepreneurship evaluations around the developing world? *The World Bank Research Observer*, 29(1):48–82.
- McKenzie, D. y Woodruff, C. (2017). Business practices in small firms in developing countries. *Management Science*, 63(9):2967–2981.
- Mohnen, P. y Hall, B. H. (2013). Innovation and productivity: An update. *Eurasian Business Review*, 3(1):47–65.
- Penrose, E. T. (1959). The theory of the growth of the firm (1995).
- Rammer, C., Czarnitzki, D., y Spielkamp, A. (2009). Innovation success of non-r&d-performers: substituting technology by management in smes. *Small Business Economics*, 33(1):35–58.

- Romer, P. M. (1990). Endogenous technological change. *Journal of political Economy*, 98(5, Part 2):S71–S102.
- Sutton, J. (2005). Competing in capabilities: an informal overview. *London: London School of Economics*.
- Sutton, J. (2012). *Competing in capabilities: the globalization process*. Oxford University Press.

Apéndice A

Dimensión productiva/monitoreo del desempeño productivo

1. Realiza especificaciones de las materias primas e insumos críticos y/o de las características críticas del proceso y/o producto. Variable binaria
2. Realiza trazabilidad (seguimiento del producto durante el proceso productivo). Variable binaria.
3. Existen equipos para solucionar problemas o lograr mejoras en el proceso. Variable binaria.
4. Utiliza herramientas internas de mejora continua. Variable binaria.
5. Utiliza herramientas específicas de gestión de proyecto y diseño. Variable binaria.
6. Implementa normas ISO de calidad, sectoriales o de producto. Variable binaria.

Dimensión de organización del trabajo

7. Implementa una rotación planificada del personal. Variable binaria.
8. Grado de participación del personal para el desarrollo de actividades (participación nula; evaluación de eficiencia; evaluación y plan de mejoras; evaluación, planificación e implementación del plan de mejoras). Variable de números enteros de 0 a 3.

Dimensión de absorción

9. Tiene departamento formal de I+D. Variable binaria.
10. Porcentaje de profesionales en el personal total. Variable continua de 0 a 100.
11. Porcentaje de personal con calificación técnica en el personal total. Variable continua de 0 a 100.
12. Tiene un área responsable de organizar las actividades de capacitación. Variable binaria.
13. Porcentaje de personas de la empresa que recibieron cursos de formación a nivel jerárquico durante el año 2012. Variable continua de 0 a 100.
14. Porcentaje de personas de la empresa que recibieron cursos de formación a nivel no jerárquico durante el año 2012. Variable continua de 0 a 100.
15. Tiene vinculaciones con otras firmas (estos pueden ser para: capacitación de RRHH; I+D; intercambio tecnológico; pruebas y ensayos; desarrollo o mejoras de productos/ procesos; gestión y/o certificación de calidad; cambios organizacionales; y actividades de diseño industrial y/o ingeniería). Variable binaria.

16. Tiene vinculaciones con el sector público (estos pueden ser para: capacitación de RRHH; I+D; intercambio tecnológico; pruebas y ensayos; desarrollo o mejoras de productos/ procesos; gestión y/o certificación de calidad; cambios organizacionales; y actividades de diseño industrial y/o ingeniería). Variable binaria.

Incentivos

17. Aplica algún sistema de evaluación de desempeño para el personal (directivo, nivel medio o no jerárquico). Variable binaria.

Apéndice B

Tabla B1: Matriz de correlaciones

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15
V1	1.00	0.35	0.26	0.28	0.24	0.28	0.17	0.18	0.26	0.13	0.11	0.23	0.22	0.21	0.18
V2	0.35	1.00	0.34	0.36	0.25	0.31	0.16	0.18	0.21	0.09	0.10	0.21	0.19	0.19	0.16
V3	0.26	0.34	1.00	0.44	0.36	0.32	0.17	0.22	0.30	0.11	0.13	0.26	0.26	0.25	0.20
V4	0.28	0.36	0.44	1.00	0.45	0.41	0.23	0.26	0.34	0.15	0.18	0.34	0.34	0.33	0.21
V5	0.24	0.25	0.36	0.45	1.00	0.32	0.17	0.21	0.35	0.15	0.22	0.24	0.30	0.25	0.21
V6	0.28	0.31	0.32	0.41	0.32	1.00	0.18	0.20	0.31	0.15	0.18	0.33	0.32	0.30	0.22
V7	0.17	0.16	0.17	0.23	0.17	0.18	1.00	0.22	0.19	0.10	0.07	0.21	0.20	0.20	0.15
V8	0.18	0.18	0.22	0.26	0.21	0.20	0.22	1.00	0.22	0.16	0.08	0.27	0.26	0.24	0.16
V9	0.26	0.21	0.30	0.34	0.35	0.31	0.19	0.22	1.00	0.21	0.22	0.25	0.34	0.30	0.25
V10	0.13	0.09	0.11	0.15	0.15	0.15	0.10	0.16	0.21	1.00	0.23	0.15	0.19	0.15	0.11
V11	0.11	0.10	0.13	0.18	0.22	0.18	0.07	0.08	0.22	0.23	1.00	0.15	0.19	0.15	0.13
V12	0.23	0.21	0.26	0.34	0.24	0.33	0.21	0.27	0.25	0.15	0.15	1.00	0.37	0.38	0.22
V13	0.22	0.19	0.26	0.34	0.30	0.32	0.20	0.26	0.34	0.19	0.19	0.37	1.00	0.73	0.23
V14	0.21	0.19	0.25	0.33	0.25	0.30	0.20	0.24	0.30	0.15	0.15	0.38	0.73	1.00	0.21
V15	0.18	0.16	0.20	0.21	0.21	0.22	0.15	0.16	0.25	0.11	0.13	0.22	0.23	0.21	1.00
V16	0.22	0.20	0.25	0.30	0.30	0.31	0.19	0.21	0.35	0.15	0.14	0.29	0.32	0.28	0.26
V17	0.24	0.21	0.25	0.35	0.27	0.30	0.22	0.29	0.28	0.14	0.14	0.40	0.38	0.35	0.21

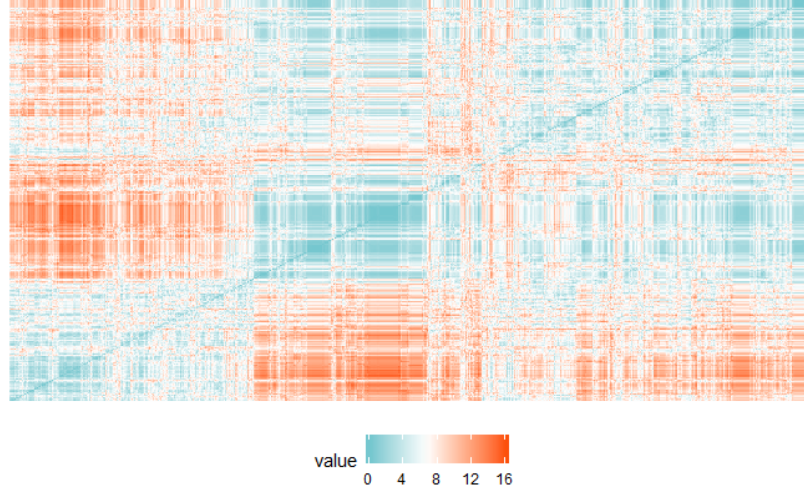
Nota: Coeficientes de correlación de Pearson. Las variables siguen el orden del Apéndice A.

Tabla B2: Distribución clusters por rama de actividad

Rama de actividad	Altas	Bajas
Alimentos	7.89 %	9.55 %
Productos textiles	4.29 %	5.83 %
Confecciones	1.54 %	5.35 %
Cuero	2.23 %	4.48 %
Madera	1.54 %	4.34 %
Papel	2.83 %	4.24 %
Edición	1.80 %	4.58 %
Productos químicos	7.98 %	3.38 %
Productos de caucho y plástico	7.12 %	4.19 %
Otros minerales no metálicos	2.74 %	4.29 %
Metales comunes	3.26 %	3.28 %
Otros productos de metal	6.52 %	6.12 %
Maquinaria y equipo	3.86 %	3.28 %
Instrumentos médicos	3.00 %	1.59 %
Otros equipo de transporte	1.54 %	1.83 %
Muebles	2.06 %	4.58 %
Maquina herramienta en general	3.60 %	3.47 %
Frigoríficos	3.43 %	5.54 %
Productos lácteos	2.74 %	3.57 %
Vinos y otras bebidas fermentadas	2.92 %	2.46 %
Farmacéuticas	7.46 %	1.59 %
Maquinaria Agropecuaria y Forestal	1.29 %	2.41 %
Aparatos de uso doméstico	2.14 %	2.17 %
Material eléctrico, radio, televisión	4.80 %	3.23 %
Carrocerías, remolques y semirremolques	1.97 %	0.58 %
Autopartes	5.32 %	2.94 %
Otras	4.12 %	1.11 %
Total	100 %	100 %

Apéndice C

Figura C1: Evaluación visual de la tendencia de *clustering*



Nota: Representa gráficamente la matriz de distancias ordenada empleando un gradiente de color para el valor de las distancias. VAT realizado con función *fviz_dist* con matriz de distancias calculada con la distancia de *manhattan*.

Tabla C1: Evaluación métodos de clustering: *Scores* óptimos

	Score	Método	Clusters
APN	0.05	K-medias	2
AD	4.60	K-medias	10
ADM	0.08	K-medias	2
FOM	0.36	K-medias	10
Connectivity	554.79	K-medias	2
Dunn	0.04	K-medias	2
Silhouette	0.37	K-medias	2

Nota: Evaluación de los métodos de clustering comparando, de forma simultánea, múltiples algoritmos (K-means y jerárquico), empleando medidas de validación internas (connectividad, silhouette y Dunn) y de estabilidad (APN, AD, ADM y FOM). Ver nota al pie 23. La tabla muestra los *scores* del método y número de clusters óptimo. La mayoría coincide en con el algoritmo de K-medias y con $K = 2$. Realizado con el paquete *clValid* en R.