

TODOS LOS CAMINOS LLEVAN A LA EDUCACIÓN: UN PRIMER ANÁLISIS DE LA ROBOTIZACIÓN, LA EDUCACIÓN Y EL EMPLEO

María Teresa BALLESTAR

ESIC Business and Marketing School

Aida GARCÍA-LÁZARO

University of Bath, Reino Unido

Jorge SAINZ

Universidad Rey Juan Carlos

Resumen

El crecimiento económico moderno ya no se encuentra en la productividad total de los factores (PTF), ya que hay ganancias por el cambio tecnológico que nunca se registran en los rendimientos de la innovación ni en las cuentas nacionales. La existencia de complementariedades entre las tecnologías derivadas del uso de la robótica, el comercio electrónico o la innovación son difíciles de evaluar a través de los registros de los países. Dado que las investigaciones anteriores se han centrado principalmente en la robotización a nivel agregado o industrial, las investigaciones centradas en el análisis de empresas y complementariedades son todavía bastante escasas. Para llenar este vacío, en este trabajo pretendemos aportar nuevas evidencias sobre los efectos de la robotización, la digitalización y la innovación en la productividad y el empleo en las empresas, utilizando una amplia muestra de 4.354 empresas manufactureras españolas y para el período 1990-2016. Estos datos captan el resultado de las altas tasas de inversión necesarias para mejorar la tecnología de producción de las empresas en un nuevo marco de competencia mundial.

Palabras clave: robotización, digitalización, productividad, educación.

Abstract

Modern economic growth is no longer found in total factor productivity (TFP), as there are gains from the technological change that have never been recorded in innovation yields or national accounts. The existence of complementarities between technologies derived from the use of robotics, e-commerce or innovation is difficult to assess through the countries' records. Previous research has focused mainly on robotization at the aggregate or industrial level however, the research on the analysis of firms and its complementarities is still relatively scarce. In this work, we attempt to fill that gap by providing new evidence using firm-level data on the effects of robotization, digitalization and innovation on productivity and employment. The study comprehends a large sample of 4,354 Spanish manufacturing companies and for the period 1990-2016. These data capture the result of the high investment rates required to improve the production technology of companies in a new framework of global competition.

Keywords: robotization, digitalization, productivity, education.

JEL classification: I26, O33.

I. INTRODUCCIÓN

EN 1875, el ingeniero español José Alcover, al describir la Exposición Universal de Viena del año anterior reconocía la importancia de los «...conocimientos humanos...» que permiten a los países que cuentan con ellos desarrollar «...progresos y adelantos de todo género...» de «... inmenso porvenir...» para la industria. Tras ciento cincuenta años de innovación tecnológica y tres revoluciones industriales después, nadie pone en duda que la participación del conocimiento sea todavía más importante en el desarrollo de la industria. Casi un siglo y medio después ya no es la máquina de vapor o la siderurgia, sino la automatización y robotización, la tecnología que está modificando la industria y cómo esa modificación afecta a las condiciones de empleo de los trabajadores (OCDE, 2019;

Pérez, 2003). Nuestro objetivo en esta contribución es analizar cómo se está produciendo la adopción de los robots en la industria y el papel que juega el nivel educativo de los trabajadores en la inclusión de robots en el proceso productivo.

La creciente literatura sobre los efectos de tecnologías como la robotización, la inteligencia artificial (IA) o *big data* intenta explicar la transformación experimentada por las empresas en las últimas tres décadas. Sin embargo, no existe una única explicación que indique cómo se producen los cambios en la productividad y las complementariedades entre algunas de esas tecnologías. Las explicaciones que se han dado van desde una medición errónea de la destrucción creativa, que sugiere que los nuevos productos ofrecen una mayor calidad y utilidad a los consumidores (Aghion *et al.*, 2019), a los errores

de cálculo en los índices de precios de los bienes de comercio electrónico, especialmente si se tienen en cuenta los productos gratuitos en los mercados *online* (Brynjolfsson y Oh, 2012; Goolsbee y Klenow, 2018) o que el desempleo aparece en las economías desarrolladas como consecuencia de la reversión en el crecimiento de la productividad de las empresas tras las ganancias iniciales de la revolución de las tecnologías de la información y las comunicaciones (TIC) (Syverson, 2011).

No obstante, existe un acuerdo entre la literatura sobre el hecho de que este crecimiento de la productividad en la economía de los países occidentales fue debido a las empresas de las TIC que comenzó a finales del siglo pasado y, a pesar de todo, se mantuvo durante la Gran Recesión y, aunque no aparece claramente en los datos, si permitió la mejora de los niveles de vida (Jorgenson, Ho y Samuels, 2011). El impacto fue tanto directo, a través de las empresas innovadoras, como indirecto, a través de los efectos que la innovación tecnológica tuvo en el sector tradicional y fue mayor cuando se produjo una transformación digital de la empresa (Kraus *et al.*, 2019). Dicha transformación requiere una buena cantidad de complementariedades relacionadas con el conocimiento y las habilidades de los trabajadores, tanto en empresas nuevas como en las ya establecidas, para crear lo que Soriano, Martínez-Climent y Tur-Porcar (2018) denominan un «círculo virtuoso» que mejora el bienestar y el crecimiento mediante la mejora de la organización dentro de la empresa.

La evaluación de este «círculo virtuoso», incluido el impacto de la adopción de una de estas formas de transformación digital, el uso de la robótica industrial, es cada vez más importante en el contexto actual de la nueva revolución industrial. A pesar de que muchas empresas están deseosas de adoptar estas tecnologías como una forma de aumentar la productividad, se han planteado algunas preocupaciones sobre el impacto en los costes de la transformación y su efecto en la fuerza de trabajo debido a la capacitación y a las nuevas formas de organización del trabajo (Acemoglu *et al.*, 2019; Díaz-Chao, Sainz-González, J., & Torrent-Sellens, 2015).

Una de las piezas que faltan en la evaluación del crecimiento es el papel de la educación en un momento como el actual, cuando la pandemia de la COVID-19 ha mostrado las debilidades de las economías industriales. Tal como establecen Carlaw y Lipsey (2002), las nuevas tecnologías dinamizaron

el crecimiento mediante la creación de conocimiento, que no se mide adecuadamente por la productividad total de los factores (PTF). Los beneficios del cambio tecnológico se producirán por medio de efectos indirectos y externalidades que no se muestran fácilmente en las estadísticas nacionales. Para salvar esa brecha utilizaremos los datos de la *Encuesta sobre estrategias empresariales (ESEE)*. La *ESEE* proporciona un panel de datos representativo de empresas industriales españolas, recolectados a través de una encuesta anual realizada por el Ministerio de Hacienda y Administración Pública de España. Abarca un amplio período desde 1990 hasta 2016, incluyendo la recesión económica de 2008 y el posterior período de recuperación (Torrent-Sellens, 2018).

La base de datos de nuestra investigación contiene información de 4.354 empresas españolas, incluyendo pequeñas y medianas empresas (74,7 por 100 con 3.251 empresas) y grandes empresas (25,3 por 100 con 1.103 empresas), y abarca tres áreas principales. En primer lugar, la toma de decisiones estratégicas sobre precios, costes, mercados e inversiones; en segundo lugar, el proceso de valor que implica el capital humano (nivel educativo, formación dentro de la compañía, etc.), la organización, la innovación empresarial, la investigación y desarrollo (I+D) y las TIC y, por último, los indicadores y ratios más importantes de los balances y las pérdidas y ganancias.

Una de las diferencias de este trabajo se halla en que, a diferencia de una gran parte de las investigaciones anteriores que se han centrado principalmente en el efecto de la robotización a nivel agregado o industrial, la nuestra, se centra en el análisis individual de cada empresa, permitiendo responder muchas preguntas sobre el efecto en el empleo según la cualificación (Seamans y Raj, 2018b).

Una segunda novedad de este artículo es la estrategia empírica empleada. Para aportar nuevas pruebas sobre los efectos de la robotización, la digitalización y la innovación en la productividad y el empleo en las empresas hemos decidido optar por una estrategia de triangulación mediante la utilización de métodos paramétricos y no paramétricos (Kankaraš Feron y Renbarger, 2019). Nuestro objetivo será analizar empíricamente el efecto que la robotización tiene en el empleo de las empresas según la formación de los trabajadores a través de un modelo econométrico tradicional y un modelo supervisado de *machine learning (ML)* y analizar las

similitudes en los resultados. Nuestro objetivo es avanzar en el uso de las estrategias empíricas con modelos supervisados y no supervisados de *ML* y su intersección con la econometría tradicional en aspectos como la estimación óptima de políticas o los modelos de elección del consumidor (Athey e Imbens, 2019, 2017).

El resto del documento se estructura de la siguiente manera: en la sección segunda se presenta la revisión de la literatura y marco teórico; en la sección tercera se describe la estrategia empírica combinando modelos paramétricos y no paramétricos de *machine learning* para el análisis de los datos de las empresas de la base de datos y, finalmente, en la sección cuarta se presentan las conclusiones de esta investigación y propuestas de investigación futura.

II. REVISIÓN DE LA LITERATURA

La formación de los trabajadores y sus habilidades recaen en lo que Carlaw y Lipsey (2002) definen como una complementariedad, ya que a través de ellas «...los agentes iniciadores ... afectan al valor de las tecnologías existentes de los agentes receptores y/o sus empresas que tienden a seguir estrategias de innovación proactivas, especialmente en lo que respecta a la adopción de procesos y tecnologías, como la comercialización de nuevos productos, la mejora de su calidad, el aumento de la capacidad o la inversión en procesos innovadores producen esas complementariedades».

Pero para aprovechar dichas complementariedades, los agentes necesitan una mayor variedad de recursos tecnológicos que mejoren su rendimiento. Actualmente, la robótica, la inteligencia artificial, el aprendizaje automático y los grandes volúmenes de datos están inseparablemente vinculados dentro de esa economía de innovación en general, pero especialmente en el sector industrial. A medida que han ido evolucionando, las empresas se han centrado en la necesidad de ofrecer mejores productos de mayor calidad, hacer que la tecnología mejore su rendimiento y abrir espacio para nuevos desarrollos (Dickson y Hadjimanolis, 1998).

Los efectos de esta complementariedad son diversos y aún no se comprenden plenamente. Algunas de las cuestiones, como el efecto sobre el empleo, han suscitado gran interés porque el problema no es simplemente una cuestión de los trabajadores que van a ser sustituidos por la auto-

matización, definida como la acción que sustituye la actividad laboral humana por el trabajo realizado por las máquinas, con dos objetivos: aumentar la calidad y reducir el coste unitario (Muro *et al.*, 2019). El impacto va a ser espantoso: La OCDE (2019) estima que el 14 por 100 de los puestos de trabajo actuales pueden desaparecer debido a la automatización y que el 35 por 100 puede verse seriamente afectado por el mismo problema.

El efecto que tendrá en el mercado de trabajo es absoluto, ya sea directa o indirectamente debido al desplazamiento sucesivo del capital humano expulsado de sus puestos anteriores a otros puestos con necesidad de otras cualificaciones. Chui, Manyika y Miremadi (2016) estiman que, en las actividades físicas previsibles, como las líneas de producción, el envasado, etc., y en función de las complementariedades, la sustitución podría llegar al 78 por 100, mientras que, en las no previsibles, las labores forestales o los agricultores, por ejemplo, el nivel de sustitución se sitúa en el 25 por 100, mientras que la sustitución en los servicios dependerá de sus características.

Graetz y Michaels (2018) se centran en la productividad analizando, a través de datos de la Federación Internacional de Robótica (IFR), el papel de los robots. Sus estimaciones no presentan ningún cambio en el número de horas trabajadas tras el aumento de la densidad de los robots, pero sin cambio en su composición, presenta un sesgo tecnológico a favor de los empleados con cualificaciones altas y medias. Faber (2018), obtiene un resultado similar, pero en el impacto en el *outsourcing*: el aumento de la densidad de automatización en los Estados Unidos tiene un efecto negativo en la deslocalización de la producción a México planteando el efecto de la asimetría en las habilidades y el conocimiento.

Acemoglu y Restrepo, (2017, 2019), utilizando también datos de la IFR, estiman el efecto de la competencia entre los robots y la mano de obra humana en el empleo y los salarios en los Estados Unidos. Ellos encuentran una reducción neta del empleo por sustitución entre ambos, lo que puede estar relacionado con las diferencias de uso entre los sectores. Con datos similares, Frey y Osborne (2017) señalan una reducción del empleo en sectores que tradicionalmente ya son usuarios de robots, como el de la automoción, pero también en otros hasta ahora protegidos, como el de los servicios.

Utilizando datos de empresas, una muestra más corta de la misma base de datos que nosotros, Doraszelski y Jaumandreu (2018) encuentran un sesgo en el cambio tecnológico y descubren que efectivamente hay uno tecnológico, pero que, al contrario de lo que hemos visto anteriormente, tiene un signo positivo hacia el empleo. Con los mismos datos, y a través de metodologías *machine learning*, Ballestar *et al.*, (2020) confirman que existe un factor significativo (5 por 100) y creciente en las ganancias de productividad ligadas a una mayor presencia de capital humano de calidad en la empresa, pero sin dar detalles sobre cómo se logran esas ganancias.

Confirmando también la intuición de Seamans y Raj (2018) de que los datos de las empresas a nivel individual recogerán más información sobre la automatización, Blanas, Gancia y Lee (2019) avanzan esta idea utilizando el panel de datos KLEMS de la Unión Europea. Así, mediante un modelo relativamente sencillo de demanda de empleo, descubren que los trabajadores que realizan tareas rutinarias y de baja cualificación, que normalmente corresponden a mujeres y jóvenes, serán los más afectados por la introducción de robots debido a las ganancias de productividad, mientras que determinan que es la entrada de los robots la que determina el tamaño de la fuerza laboral de las empresas.

Estos análisis suponen un avance, pero no detallan las características de este cambio que corresponde a la formación de los empleados, la forma en que las empresas han cambiado la gestión de los conocimientos de su capital humano y los efectos de estos cambios vinculados a la formación. Para ello, es importante determinar las complementariedades del cambio en el empleo y la productividad para poder determinar las políticas de formación necesarias para evitar el desempleo y centrar el apoyo en los grupos más desfavorecidos.

A diferencia de Graetz y Michaels (2018) o Acemoglu y Restrepo (2020), Autor y Salomons (2017), Doraszelski y Jaumandreu (2018), Ballestar *et al.* (2021) muestran que el aumento de la productividad es uno de los principales beneficios en la expansión de la concurrencia con los resultados de que las industrias pueden beneficiarse del crecimiento de la productividad que aumenta la mano de obra mediante las complementariedades, mientras que el empleo parece caer dentro de una industria a medida que aumenta la productividad específica de la misma. Los efectos indirectos posi-

tivos para algunos sectores compensan con creces los efectos negativos en otros.

Como sugieren Seamans y Raj (2018), el trabajo empírico que utiliza principalmente estadísticas agregadas por industria o país no logra encontrar cómo hay complementariedades que se esconden bajo las estructuras del mercado. Nuestra investigación se centrará, por primera vez, en estimar algunos de los efectos de las complementariedades educativas con datos a nivel de empresa en una serie temporal de 26 años.

III. ESTRATEGIA EMPÍRICA

Los avances de la estadística desde mediados del siglo pasado han estado ligados al incremento en la capacidad de procesamiento de datos y, hasta cierto punto, a la separación de los métodos tradicionales de inferencia estadística y el uso de algoritmos computacionales cuya base estadística tiene, o por lo menos debería tener, un sustrato en esa teoría estadística. La capacidad de predicción de los modelos de *big data* y su aplicación comercial ha creado una barrera invisible entre la ortodoxia académica y los algoritmos de uso común en la industria (Ballestar, Ribeiro y Sainz, 2018; Efron y Hastie, 2016). Hal Varian (2014) señala que la existencia de bases de datos de mayor tamaño permite el empleo de métodos distintos a los tradicionales por el volumen de los datos. Así se puede realizar una selección, supervisada o no, de los datos o la utilización de las ventajas que aporta la no linealidad. Es en ese momento donde entran en juego modelos de *machine learning* que aportan flexibilidad a través del pensamiento algorítmico más que en su justificación inferencial.

Ambas estrategias empíricas tienen ventajas e inconvenientes. En el caso del *ML*, el efecto «caja negra» de algunos modelos ya representa por sí mismo una difícil elección entre ellos para los investigadores, lo que puede llevar a veces a la necesidad de verificar los resultados con análisis fuera de muestra o entre distintos modelos para asegurar la robustez de, por ejemplo, el análisis de políticas públicas (Athey e Imbens, 2019; Ballestar *et al.*, 2019; Kankaraš, Feron y Renbarger, 2019). Nuestra propuesta para esta investigación va en la misma línea, utilizar un método econométrico tradicional y un modelo de *ML* supervisado sobre la misma base de datos para hacer una evaluación del efecto de la robotización en el empleo según los distintos niveles educativos.

1. Datos

En esta investigación se ha utilizado la *Encuesta sobre estrategias empresariales (ESEE)* con datos a nivel de empresa para el período 1990 a 2016. La muestra del estudio contiene 4.354 empresas en diez industrias, de las cuales 3.251 (74,7 por 100) son pequeñas y medianas empresas (pyme) mientras que 1.103 (25,3 por 100) son empresas grandes. Pymes son aquellas con menos de 200 empleados, empresas grandes son aquellas con más de 200 empleados. Debido al extenso período de la muestra que compone el panel –26 años–, existen algunas empresas que salen de la encuesta y son remplazadas por otra para mantener el tamaño de la muestra. La gestión de la muestra de empresas que compone el panel es realizada directamente por el Ministerio de Hacienda y Administración Pública de España.

Uno de los beneficios de la *ESEE* es que contiene información sobre el uso o adopción de robots cada cuatro años. De esta manera contamos con datos sobre la robotización de las empresas y su grado de adopción para los años 1990, 1994, 1998, 2010 y 2014. El porcentaje promedio de empresas grandes usando robots entre empresas grandes y PME, es de 13,1 por 100 y 12,9 por 100 respectivamente. En contraste, las empresas no robóticas se encuentran primordialmente entre las empresas pequeñas y medianas (véase cuadro n.º 1).

2. Estrategia econométrica

Para la estimación del modelo se utilizan los indicadores de empleo, salarios, el valor agregado y se construye un indicador del coste marginal. La variable dependiente prueba los efectos de la robotización por nivel educativo. En este caso se utiliza como variables dependientes la participación de trabajadores titulados superiores sobre el total de empleados; la participación de titulados medios

sobre el total de empleados y la participación de no titulados sobre el empleo total.

Como variables explicativas se ha introducido el uso de robots. Para ello se construyen tres indicadores con la información disponible en la *ESEE*, que presenta la variable *dummy* para el uso de robots donde 1 representa que la empresa *i* usa robots y 0 en otro caso. Los tres indicadores construidos denotan la frecuencia de uso de robots en cada empresa.

El primer indicador es una variable *dummy* con valor de 1 para las empresas que en todos los períodos han sido robóticas y 0 si es de otra manera. La segunda variable captura a las empresas no robóticas a partir de una variable *dummy* que indica con el valor de 1 a aquellas empresas que no han sido robóticas en ninguno de los períodos y con cero a aquellas que al menos en algún período reportaron el uso de robots. El tercer indicador es una variable categórica que revela el quintil en el que se encuentra una empresa de acuerdo con el número de períodos que ha usado robots en los años para los cuales está disponible la información. Se espera que el uso de robots tenga alguna incidencia sobre el empleo y la participación de trabajadores en un determinado nivel educativo. Estudios previos sobre robots sugieren el efecto negativo de los robots sobre los trabajadores con calificación media (Graetz y Michaels, 2018). Mientras que (Arntz, Gregory y Zierahn, 2020; Gregory, Salomons y Zierahn, 2016) encuentran que existen efectos compensatorios que conducen al incremento de la demanda de trabajo en los sectores más expuestos a la rutina del trabajo.

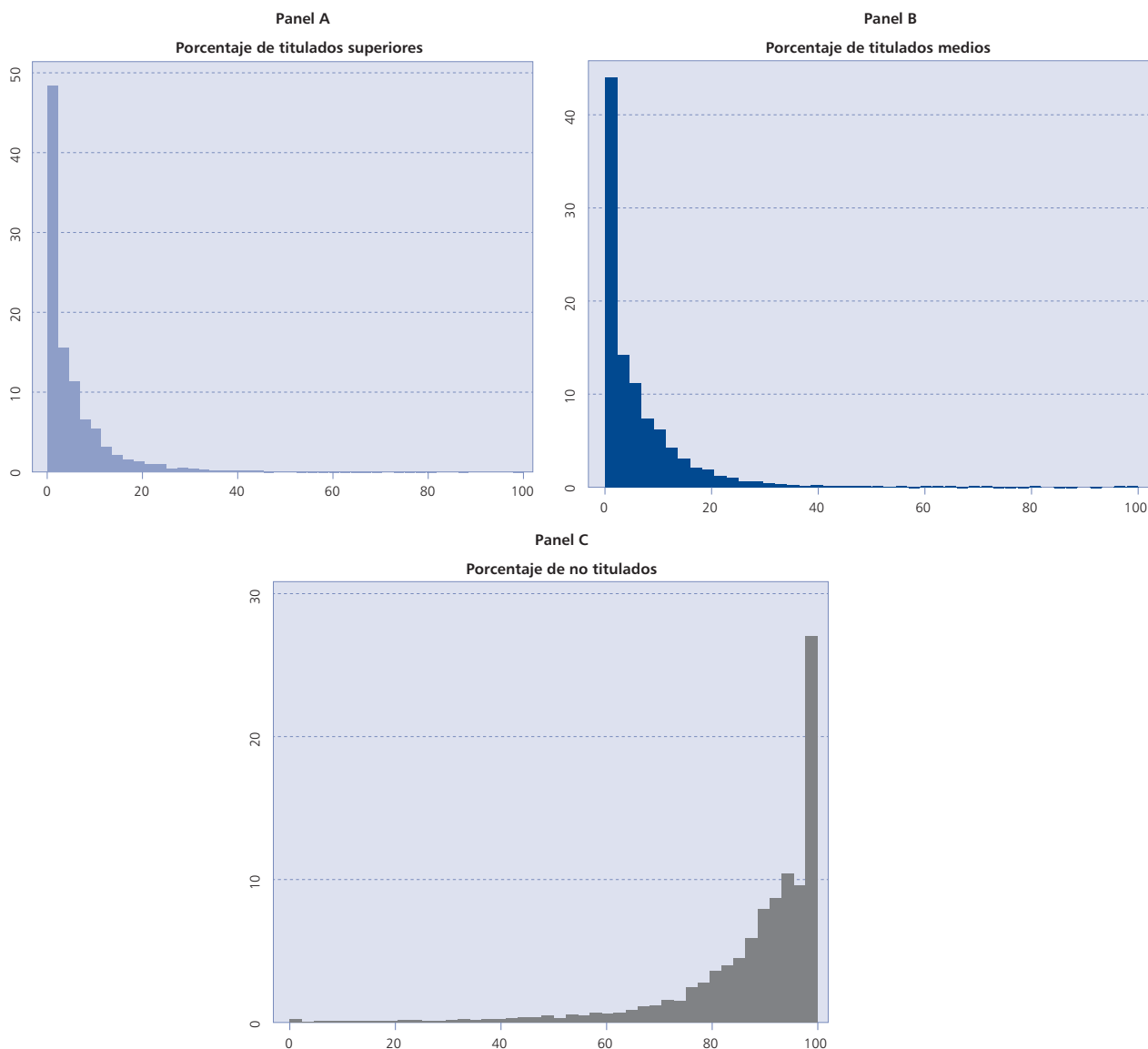
En la estimación se incluyeron variables explicativas de control como salarios reales, el valor agregado de cada empresa, los costes marginales reales aproximado por la razón de costes del personal más costes de capital sobre el valor agregado. Se ha tomado como referencia la aportación de Gregory, Salomons y Zierahn (2019) en la formulación de la estimación econométrica en esta investigación. Sin embargo, no se ha examinado el papel de la rutina, sino el impacto de la robotización en la dinámica de la demanda de trabajo. Asimismo, este estudio presenta datos a nivel de empresa siguiendo la propuesta de (Seamans y Raj, 2018a) con datos de la *EEE*, mientras que Gregory, Salomons y Zierahn (2019) hacen un estudio a nivel sectorial sin desagregar para el mercado de trabajo europeo.

CUADRO N.º 1
ADOPCIÓN DE ROBOTS POR TAMAÑO DE LA EMPRESA
(En porcentaje)

ESTADO DE ROBOTIZACIÓN	SMEs	GRANDES EMPRESAS	TOTAL
Robóticas	12,9	13,1	26,0
No robóticas	59,4	14,6	74,0
Total	72,3	27,7	100,0

Fuente: Elaboración propia con datos de la *ESEE*.

GRÁFICO 1
HISTOGRAMA DE LAS VARIABLES DEPENDIENTES



Fuente: Elaboración propia con datos de la ESEE.

En el gráfico 1 se observa cómo la variable dependiente por nivel educativo está censurada por un lado de la distribución y por tanto sesgada. Esto implica que el uso de los estimadores de esta investigación con el método de mínimos cuadrados ordinarios no sería la mejor aproximación. Por tanto, se ha decidido implementar un modelo Tobit, diseñado para este tipo de variables censuradas. Las estimaciones se realizaron con errores robustos.

Siguiendo (Gregory, Salomons y Zierahn, 2019), se propone la siguiente formulación:

$$l_{it}^k = \beta_0^k + \beta_r^k r_{it}^k * t + \beta_t^k t + \beta_{mc}^k mc_{it}^k + \beta_{wv}^k w_{it}^k + \beta_y^k y_{it}^k + v_{it}^k + u_t^k + \varepsilon_{it}^k \quad [1]$$

donde l_{it}^k es la participación de trabajadores de la empresa i en el período t por nivel educativo $k = \{hs, ms, ls\}$; hs denota titulados superiores; ms

titulados medios y *ls* no titulados. La robotización se modela usando uno de los indicadores construidos sobre robots, los cuales son interactuados con una tendencia lineal $r_{jt}^k * t$ donde $j = \{siempre_r, nunca_r, quintil_r\}$; *siempre_r* indica la variable *dummy* sobre las firmas que siempre han sido robóticas. La variable *dummy nunca_r* captura a las empresas que nunca han sido robóticas y *quintil_r* es la variable categórica que indica el quintil en el que se encuentra una empresa de acuerdo con el número de años que reportó usar robots.

Adicionalmente, incluimos una tendencia lineal representada por t ; mc_{it}^k son los costes marginales; w_{it}^k denota los salarios reales y y_{it}^k es el valor agregado de la empresa i . Se han introducido los efectos fijos de empresa y tiempo respectivamente, v_i^k y u_t^k . Estos permiten controlar por las heterogeneidades entre firmas y en el tiempo, los cuales puede introducir ruido en las estimaciones de otra manera. Finalmente, ε_{it}^k representa el término de los residuos.

En el cuadro n.º 2 condensamos los resultados base. Las primeras tres columnas muestran las estimaciones por nivel educativo al introducir la variable *dummy* sobre las variables que siempre han sido robóticas. Columnas (4)-(6) presentan los resultados de las estimaciones cuando se introduce la

variable quintílica. Finalmente, las columnas (7)-(9) reportan los resultados usando la variable *dummy* para empresas que nunca han sido robóticas.

Los coeficientes en las columnas (1) a (3) muestran que la variable *dummy* para empresas que siempre han sido robóticas es significativa y afecta positivamente la participación de los empleados con educación intermedia. Para el caso de la participación de trabajadores titulados superiores y no titulados el efecto es negativo, pero no significativo. Estos resultados sugieren que las empresas que siempre han sido robóticas demandan 6,8 por 100 más trabajadores con educación intermedia que las que no lo son. Tomando las columnas (4)-(6) encontramos que las empresas con más años usando robots también demandan una proporción mayor de trabajadores titulados medios mayor. El coeficiente es menor que en la columna (2) pero igualmente positivo y significativo. Es decir que por cada quintil adicional en el que se encuentra una empresa con uso de robots, la demanda de titulados medios se incrementa en 1,4 por 100. Los coeficientes para titulados superiores y no titulados no son significativos.

En las columnas (7)-(9) se muestran los resultados de las estimaciones cuando la variable *dummy* de

CUADRO N.º 2

REGRESIONES TOBIT POR NIVEL EDUCATIVO

VARIABLES	(1) HS	(2) MS	(3) LS	(4) HS	(5) MS	(6) LS	(7) HS	(8) MS	(9) LS
y_{it}^k	0,335*** (0,064)	0,282*** (0,081)	-1.787*** (0,144)	0,313*** (0,066)	0,255*** (0,083)	-1.793*** (0,147)	0,300*** (0,066)	0,294*** (0,083)	-1.816*** (0,147)
Tendencia lineal	0,228*** (0,009)	0,148*** (0,012)	-0,478*** (0,020)	0,222*** (0,010)	0,136*** (0,013)	-0,478*** (0,023)	0,233*** (0,010)	0,167*** (0,013)	-0,482*** (0,022)
w_{it}^k	3.504*** (0,242)	2.748*** (0,310)	-8.064*** (0,550)	3.527*** (0,242)	2.760*** (0,310)	-8.052*** (0,551)	3.536*** (0,242)	2.721*** (0,310)	-8.035*** (0,550)
mc_{it}^k	0,132 (0,143)	0,084 (0,186)	-1.354*** (0,313)	0,124 (0,143)	0,064 (0,187)	-1.352*** (0,314)	0,117 (0,143)	0,081 (0,187)	-1.364*** (0,314)
Siempre_r	-0,018 (0,016)	0,068*** (0,020)	-0,036 (0,033)						
Quintil_r				0,002 (0,004)	0,014*** (0,005)	-0,004 (0,008)			
Nunca_r							-0,016 (0,011)	-0,021 (0,014)	-0,003 (0,024)
Constante	-39.862*** (2.106)	-31.210*** (2.681)	211.189*** (4.858)	-39.741*** (2.110)	-30.889*** (2.686)	211.150*** (4.867)	-39.633*** (2.110)	-31.125*** (2.687)	211.344*** (4.870)
Observaciones	8.690	8.686	8.687	8.690	8.686	8.687	8.690	8.686	8.687
Número de empresas	4.086	4.085	4.085	4.086	4.085	4.085	4.086	4.085	4.085

Nota: Errores estándar entre paréntesis, *** $p < 0,01$. ** $p < 0,05$. * $p < 0,1$.

robots indica con el valor igual a 1 para aquellas empresas que nunca han sido robóticas, en este caso. Se encuentra, sin embargo, que los coeficientes para la variable de interés no son significativos en ninguna de las participaciones de empleados por nivel educativo. Por tanto, esto podría indicar que mientras existe incremento del empleo cuando se introducen robots en el proceso productivo, no se observa un efecto penalizador en términos de la demanda de trabajo para aquellas empresas que no los utilizan.

Resta decir que en todas las especificaciones de la participación de empleo de titulados superiores y medios del cuadro n.º 1, los coeficientes de valor agregado y la tendencia lineal son en todos los casos positivos y significativos tal como se esperaba, mientras que resultan negativos para el caso de no titulados. Adicionalmente, los salarios reales son significativos y negativos solo en el caso de los trabajadores no titulados y positivos para el resto. Por su parte, el coste marginal es significativo y negativo solo para la participación de trabajadores no titulados.

Tamaño de la empresa. Desde que el tamaño de la empresa es una variable clave para explicar la dinámica y la capacidad de una empresa para tomar decisiones en términos de cambio tecnológico. Esto por el acceso diferenciado al financiamiento, el contacto y relación con proveedores y mercados en el exterior, lo cual es más recurrente en el caso de las empresas grandes. Por tanto, se ha decidido controlar las estimaciones por el tamaño de la empresa. Debido a que la variable que recoge el tamaño está definida con base en el número de empleados, decidimos dividir la muestra entre grandes y pequeñas empresas en lugar de introducir esta variable como explicativa, para evitar potenciales problemas de endogeneidad.

En el cuadro n.º 3 introducimos los resultados con la misma estructura que en el cuadro n.º 2, pero considerando como muestra solo aquellas empresas grandes (+200 empleados). Las columnas (1)-(3) presentan los coeficientes de las empresas grandes y que siempre han sido robóticas. En la columna (1) observamos que cuando las empresas siempre han sido robóticas, la participación de trabajadores titulados superiores es menor en 5,2 por 100. En contraste, la columna (2) muestra que el efecto es positivo y significativo, en 7,7 por 100 para la participación de trabajadores con títulos medios en el caso de las empresas que siempre han sido robóticas. Esto confirma los resultados obtenidos en el cuadro n.º 2. Por otro lado, la par-

ticipación de trabajadores no titulados cuando una empresa es grande no resulta afectada si la empresa siempre ha sido robótica, de acuerdo con nuestros resultados en la columna (3).

Las columnas (4)-(6) muestran los coeficientes para el caso de empresas grandes que han adoptado el uso de robots en algunos períodos de la muestra, pero no en todos. Los estimados sugieren que un mayor número de años usando robots contrae la participación de trabajadores titulados superiores en 1,6 por 100 e incrementa la participación de titulados medios en 2,8 por 100. Mientras que los coeficientes son no significativos para los trabajadores no titulados. Estos resultados coinciden los encontrados en las tres primeras columnas. Finalmente, no existe una penalización significativa para aquellas empresas grandes que nunca han usado robots en términos de la participación de trabajadores, tal como se aprecia en las columnas (7)-(9), dado que los coeficientes no son diferentes de cero.

En el cuadro n.º 4 presentamos el mismo ejercicio, pero únicamente para las empresas pequeñas y medianas, es decir aquellas con menos de 200 empleados. Como se puede apreciar, ninguno de los indicadores sobre el uso de robots resulta significativo. Estos resultados son consistentes con (Acemoglu y Restrepo, 2019; Graetz y Michaels, 2018), mostrando un efecto de desplazamiento de trabajo en algunos niveles de cualificación, pero no en todos.

Los resultados anteriores nos permiten argumentar que el impacto de la robotización en el empleo está localizado en las empresas grandes primordialmente. Encontramos un incremento en el empleo entre los trabajadores con educación intermedia simultánea a una caída para los trabajadores con títulos superiores. Esto puede indicar que el uso de robots en la economía española esté concentrado en las tareas de alta cualificación, lo cual hace que exista un efecto de sustitución entre los trabajadores más cualificados y los robots. La relación con los trabajadores con educación media resulta complementaria, lo que es coherente si consideramos que la adopción de robots, aunque desplaza a trabajadores altamente cualificados, demanda técnicos capaces de operar con los robots a un nivel más operativo. Esto reduce los costes de personal y aumenta la productividad laboral, dado por la introducción de nuevo capital. En balance, la introducción de robots en las grandes empresas incrementa el empleo de forma diferenciada según el nivel educativo.

CUADRO N.º 3

REGRESIONES TOBIT POR NIVEL EDUCATIVO – GRANDES EMPRESAS (+200 TRABAJADORES)

VARIABLES	(1) HS	(3) MS	(5) LS	(7) HS	(9) MS	(11) LS	(13) HS	(15) MS	(17) LS
lnVAreal	0,768*** (0,208)	0,301 (0,240)	-1.166*** (0,369)	0,789*** (0,209)	0,242 (0,241)	-1.147*** (0,371)	0,758*** (0,208)	0,325 (0,240)	-1.176*** (0,370)
Timetrend	0,297*** (0,018)	0,239*** (0,025)	-0,533*** (0,035)	0,324*** (0,025)	0,190*** (0,034)	-0,515*** (0,048)	0,282*** (0,017)	0,264*** (0,025)	-0,541*** (0,034)
lncrealper	5.444*** (0,568)	4.707*** (0,699)	-10.498*** (1,046)	5.449*** (0,568)	4.713*** (0,698)	-10.500*** (1,046)	5.475*** (0,568)	4.637*** (0,699)	-10.478*** (1,045)
lcmgreales	-0,063 (0,307)	-0,283 (0,407)	0,264 (0,586)	-0,044 (0,308)	-0,327 (0,407)	0,280 (0,586)	-0,057 (0,308)	-0,287 (0,407)	0,267 (0,586)
Always_robotictrend	-0,052** (0,026)	0,077** (0,032)	-0,031 (0,047)						
q_periods_is_robottrend				-0,016** (0,008)	0,028*** (0,010)	-0,011 (0,014)			
Never_robotictrend							0,022 (0,031)	-0,030 (0,037)	0,006 (0,056)
Constante	-68.777*** (5.335)	-52.533*** (6.260)	226.471*** (9.522)	-69.204*** (5.343)	-51.589*** (6.259)	226.152*** (9.534)	-68.856*** (5.347)	-52.364*** (6.276)	226.490*** (9.536)
Observaciones	2.386	2.386	2.387	2.386	2.386	2.387	2.386	2.386	2.387
Número de empresas	1.103	1.103	1.103	1.103	1.103	1.103	1.103	1.103	1.103

Nota: Errores estándar entre paréntesis, *** $p < 0,01$. ** $p < 0,05$. * $p < 0,1$.

CUADRO N.º 4

REGRESIÓN TOBIT POR NIVEL DE EDUCACIÓN – SMES (<200 TRABAJADORES)

VARIABLES	(1) HS	(3) MS	(5) LS	(7) HS	(9) MS	(11) LS	(13) HS	(15) MS	(17) LS
lnVAreal	0,287*** (0,097)	0,425*** (0,127)	-2.576*** (0,240)	0,267*** (0,099)	0,439*** (0,130)	-2.622*** (0,245)	0,262*** (0,098)	0,466*** (0,129)	-2.634*** (0,244)
Timetrend	0,198*** (0,011)	0,114*** (0,014)	-0,453*** (0,026)	0,195*** (0,012)	0,115*** (0,015)	-0,461*** (0,028)	0,200*** (0,013)	0,115*** (0,017)	-0,437*** (0,030)
lncrealper	3.217*** (0,272)	2.345*** (0,352)	-7.436*** (0,678)	3.235*** (0,272)	2.328*** (0,352)	-7.404*** (0,679)	3.238*** (0,272)	2.308*** (0,352)	-7.398*** (0,678)
lcmgreales	0,304* (0,178)	0,375 (0,229)	-2.454*** (0,423)	0,293 (0,178)	0,382* (0,230)	-2.483*** (0,424)	0,289 (0,178)	0,398* (0,230)	-2.491*** (0,424)
Always_robotictrend	-0,015 (0,019)	0,036 (0,025)	-0,002 (0,044)						
q_periods_is_robottrend				0,001 (0,004)	0,003 (0,005)	0,007 (0,010)			
Never_robotictrend							-0,006 (0,012)	0,005 (0,016)	-0,028 (0,029)
Constante	-35.495*** (2.351)	-28.672*** (3.059)	215.680*** (6.002)	-35.394*** (2.357)	-28.693*** (3.066)	216.020*** (6.018)	-35.355*** (2.357)	-28.884*** (3.066)	216.147*** (6.018)
Observaciones	6.304	6.300	6.300	6.304	6.300	6.300	6.304	6.300	6.300
Número de empresas	3.251	3.250	3.250	3.251	3.250	3.250	3.251	3.250	3.250

Nota: Errores estándar entre paréntesis, *** $p < 0,01$. ** $p < 0,05$. * $p < 0,1$.

3. Estrategia machine learning

Metodología. En esta investigación se utiliza una red neuronal artificial perceptrón multicapa (ANN-MLP, por sus siglas en inglés) con el doble propósito de clasificación y predicción. Las redes neuronales son utilizadas para analizar relaciones complejas existentes entre las variables de entrada (variables independientes) y la variable de salida (variable dependiente). Estas relaciones complejas pueden ser de carácter no lineal. Además, las redes neuronales artificiales son capaces de manejar una gran cantidad de datos, lo que las convierte en una poderosa alternativa frente a métodos más tradicionales como la regresión logística, que es una de las metodologías más habituales cuando se desarrollan modelos predictivos en los que la variable de salida es dicotómica (Paliwal y Kumar, 2009; Tu, 1996). Sin embargo, las redes neuronales artificiales no se recomiendan cuando el objetivo es analizar la relación existente entre las diferentes variables o explicar los motivos subyacentes existentes en el resultado de la predicción realizada (Chung et al., 2015).

La ANN-MLP es un método supervisado que utiliza una técnica de aprendizaje de propagación hacia atrás (*backpropagation*) que tiene como finalidad minimizar el error entre la predicción realizada por la red y el valor real observado. Si bien es cierto que existen diversos tipos de redes neuronales, este es uno de los tipos más ampliamente adoptados por su flexibilidad respecto al tipo de variables que permite utilizar, así como su robustez y facilidad de aplicación. Las variables soportadas pueden ser cualitativas o cuantitativas, discretas o continuas (Aad et al., 2012). Destacar como principal inconveniente, que, al tratarse de un método supervisado, la precisión de las predicciones de la red estará sujeta a la calidad de la muestra con la que se la entrena y valida (Aad et al., 2012; Ballestar, Grau-Carles y Sainz, 2018; Li y Eastman, 2006).

En este caso, la red neuronal es entrenada para buscar unos patrones y relaciones entre variables que permita la clasificación de las empresas evaluadas como robotizadas o no robotizadas. Las variables de entrada utilizadas para el entrenamiento de la red son las relativas a la composición y características de la plantilla de trabajadores de las compañías. Estas variables son: número de empleados, distribución de dichos empleados en función de tu relación profesional con la compañía y nivel educativo y, finalmente, costes de personal asociados. La identificación de estas relaciones permitirá también a la red neuronal

artificial predecir, tan solo utilizando las estas variables de entrada para una determinada empresa, si ha realizado su transición a la robotización o no.

Esta metodología no solo nos permite clasificar y predecir si una empresa debiera ser robótica o no en función de la composición y características de su plantilla de trabajadores, sino que también permite identificar desviaciones respecto al grado de robotización idóneo de las compañías. Por un lado, empresas que la red neuronal identifica que deberían ser robóticas, pero aún no han realizado el tránsito, y podrían estar perdiendo oportunidades y ventajas competitivas. Mientras que otras han realizado la transición demasiado pronto y podrían estar poniendo en riesgo el proceso de transformación por una situación de inmadurez y precipitación. Corroborando que este tipo de metodologías *machine learning* ampliamente utilizadas en comercio electrónico (Ballestar, Grau-Carles y Sainz, 2019), también son aplicables en otros sectores como el industrial.

Como anteriormente se ha indicado, se utiliza la *Encuesta sobre estrategias empresariales*. En el cuadro n.º 5 se describen las variables de entrada y la variable de salida que forman parte de la red neuronal artificial perceptrón multicapa (ANN-MLP).

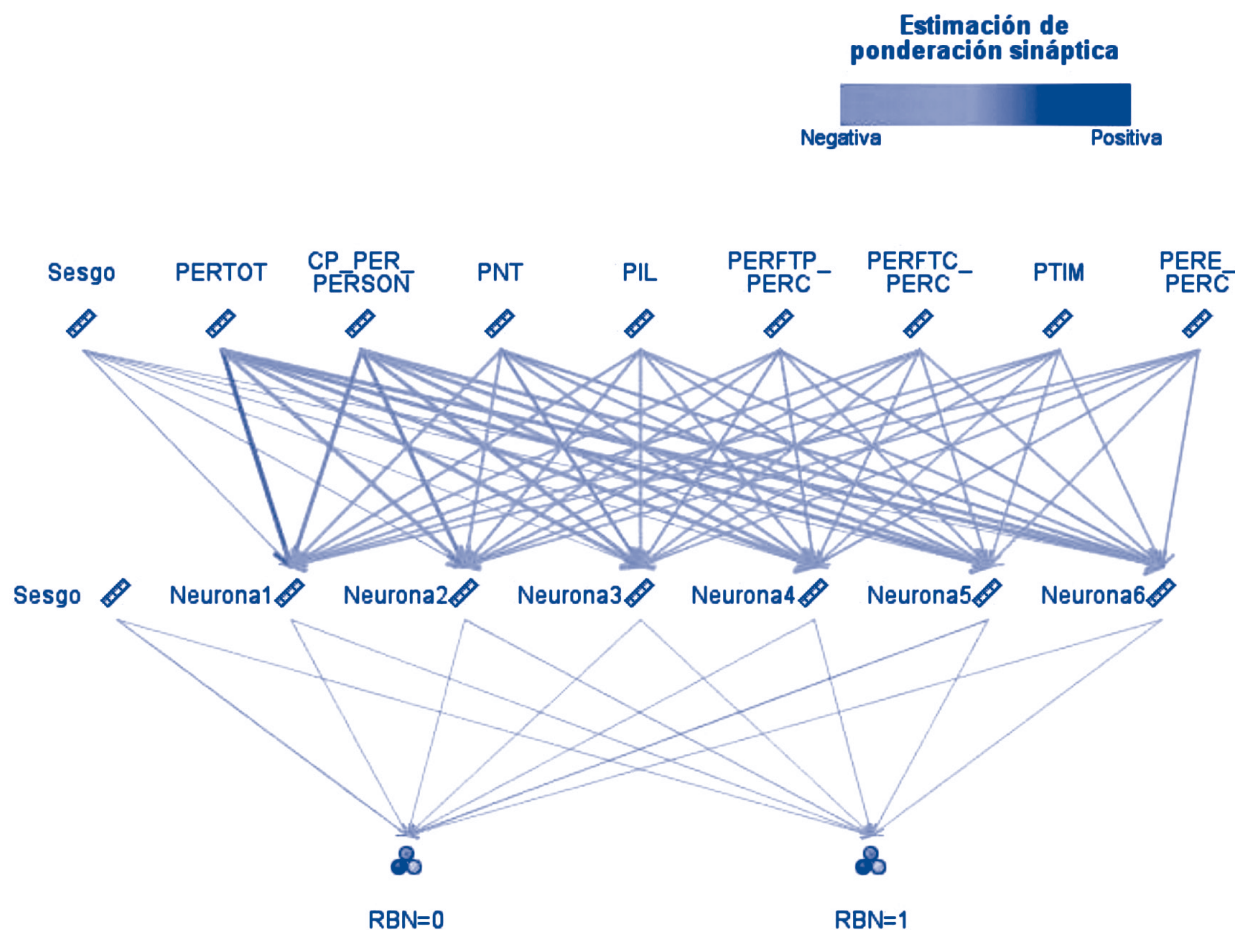
CUADRO N.º 5

DESCRIPCIÓN DE LAS VARIABLES DEL MODELO ANN-MLP

VARIABLES DE ENTRADA	DESCRIPCIÓN
Relación profesional de los empleados con la compañía	
PERFTC_PERC	Variable numérica continua. Porcentaje de asalariados fijos a tiempo completo.
PERFTP_PERC	Variable numérica continua. Porcentaje de asalariados fijos a tiempo parcial.
PERE_PERC	Variable numérica continua. Porcentaje de asalariados eventuales.
Nivel de estudios	
PNT	Variable numérica continua. Porcentaje de empleados no graduados.
PTIM	Variable numérica continua. Porcentaje de titulados medios.
PIL	Variable numérica continua. Porcentaje de ingenieros y titulados superiores.
Tamaño de la empresa y costes de personal	
PERTOT	Variable numérica continua. Número de empleados.
CP_PER_PERSON	Variable numérica continua. Costes de personal por empleado
VARIABLES DE SALIDA	DESCRIPCIÓN
RBN	Variable dicotómica. Valor 1 si la empresa es robótica y 0 si no lo es.

Fuente: Elaboración propia.

GRÁFICO 2
ARQUITECTURA DE LA ANN-MLP



Fuente: Elaboración propia.

Arquitectura de la ANN-MLP. En esta investigación la estructura de nuestra red neuronal consta de tres capas. Una primera capa de entrada, una intermedia denominada capa oculta, y, finalmente, una capa de salida con la clasificación o predicción de la red según sea el caso. Estas capas se encuentran interconectadas en una sola dirección entre sí por los pesos sinápticos. La conexión entre capas es secuencial, comienza en la capa de entrada, transcurre a través de la capa oculta y finaliza en la capa de salida. Por tanto, en este modelo se efectúa un mapeo de las ocho unidades de entrada o nodos pasivos, correspondiente a las ocho variables numéricas continuas independientes de la capa de entrada, con las dos unidades correspondientes a la variable de salida o dependiente, que determina si una empresa es robótica o no. El número óptimo de

seis unidades en la capa oculta es automáticamente calculado durante el proceso de entrenamiento utilizando el algoritmo de aprendizaje de propagación hacia atrás (*backpropagation*).

En el gráfico 2 se muestra la arquitectura de esta red neuronal artificial perceptrón multicapa, en la que el tipo de función de activación de la capa oculta y la capa de salida utilizados corresponde con una tangente hiperbólica y softmax respectivamente.

Artificial neural network (ANN); entrenamiento supervisado y procesos de validación. En esta investigación se ha utilizado la metodología de entrenamiento-test-validación, que consiste en partir la muestra de aleatoriamente en tres grupos

(entrenamiento, test y validación) para entrenar el modelo de clasificación y predictivo y ser capaces de evaluar su precisión y estabilidad (Tu, 1996). La red neuronal artificial perceptrón multicapa se crea y entrena con el 74,15 por 100 (10,959) de los registros de la base de datos, correspondientes a datos históricos de 3.228 empresas. Mientras que un 17,33 por 100 (2.562 registros históricos de 755 empresas) adicional de la muestra se utiliza en la fase de testeo para testar los posibles errores ocurridos durante la etapa de entrenamiento y evitar, especialmente, fenómenos de sobreentrenamiento del modelo. Finalmente, la etapa de validación se realiza con el restante 8,52 por 100 (1.259 registros históricos de 371 empresas) de la muestra, siendo el propósito validar la capacidad de clasificación y predictiva del modelo, así como su estabilidad.

Debido a que las empresas que han implementado la robotización al menos un año durante el período en observación representan tan solo el 26 por 100 del total de la muestra, podemos concluir que se trata de una muestra desbalanceada. Esto significa que una de las categorías, en este caso las empresas robóticas, se encuentran menos representada en la muestra. Este sesgo natural del fenómeno en estudio podría tener un impacto negativo en el proceso de construcción y entrenamiento del modelo si no se realiza un tratamiento previo para balancear la muestra. Por este motivo, se aplicó un método de *oversampling* sobre el grupo minoritario (empresas que han sido robóticas al menos un año del período en observación) para entrenar a la red neuronal con una muestra balanceada entre empresas robóticas y no robóticas (Ganganwar, 2012; Guo y Viktor, 2004).

Al mismo tiempo, las muestras de test y validación se mantienen desbalanceadas con el propósito de verificar la capacidad predictiva y estabilidad de la ANN en un escenario real, no balanceado, con posterioridad a la etapa de entrenamiento de la red.

Análisis empírico y resultados. La precisión en la clasificación, sensibilidad, especificidad, el área debajo de la curva ROC y el coeficiente de Gini son los indicadores más relevantes a la hora de evaluar la precisión en la clasificación y capacidad predictiva de la red neuronal.

En nuestra investigación, la red neuronal artificial tiene una precisión del 69,66 por 100 (una tasa de error del 30,34 por 100). En el cuadro n.º 6 se muestra tanto la precisión del modelo como la matriz de confusión que contiene el porcentaje de casos clasificados tanto correcta como incorrectamente para las dos categorías de la variable dependiente. Las empresas robóticas en el año de la observación están representadas por el valor 1, mientras que las no robóticas lo están con el valor 0. Estos porcentajes son muy similares entre las tres muestras (entrenamiento, test y validación), lo que confirma que el modelo no está sobreentrenado.

Se observa en el cuadro n.º 6 que los porcentajes para la muestra de entrenamiento, test y validación son muy similares, por lo que se puede descartar un sobreentrenamiento de la red neuronal artificial.

El porcentaje de positivos verdaderos también denominado *sensibilidad*, es del 67,69 por 100. Este

CUADRO N.º 6

PRECISIÓN DEL MODELO Y MATRIZ DE CONFUSIÓN

PRECISIÓN DEL MODELO		MATRIZ DE CONFUSIÓN					
MUESTRA	PORCENTAJE CORRECTO	MUESTRA	OBSERVADO	PREDICCIÓN			
				TAMAÑO MUESTRA		PORCENTAJE	
				0	1	0	1
Entrenamiento	69,66	Entrenamiento*	0	4.568	1.852	71,15	28,85
			1	2.022	4.236	32,31	67,69
Test	69,43	Test	0	1.331	535	71,33	28,67
			1	210	486	30,17	69,83
Validación	68,55	Validación	0	640	284	69,26	30,74
			1	101	234	30,15	69,85

* Oversampled para balancear la muestra.

Fuente: Elaboración propia.

valor explica el porcentaje de empresas robóticas que han sido correctamente clasificadas en base a las variables de relación profesional de los empleados con la compañía, nivel de estudios de los empleados, tamaño de la empresa y costes de personal. Por otro lado, el porcentaje de negativos verdaderos, también denominado *especificidad*, es del 71,15 por 100. Este valor se corresponde con el porcentaje de empresas no robóticas que han sido correctamente clasificadas en base a estas mismas variables de entrada. Los valores complementarios son el porcentaje de falsos positivos, con un valor de 28,85 por 100 que son las empresas no robóticas que han sido clasificadas incorrectamente como robóticas. Finalmente, el porcentaje de falsos negativos es 32,31 por 100. Este valor responde a las empresas robóticas que han sido incorrectamente clasificadas como no robóticas.

En esta investigación se utiliza el área debajo de la curva (AUC) ROC como principal medida de precisión de la red neuronal artificial, ya que se considera un indicador de precisión más robusto que el indicador de precisión de la clasificación cuando se trabaja con muestras no balanceadas (María Teresa Ballestar,

Grau-Carles y Sainz, 2019; Cheng, Hua y Tan, 2019; Dželićhodžić y Djonko, 2016; Yin et al., 2013).

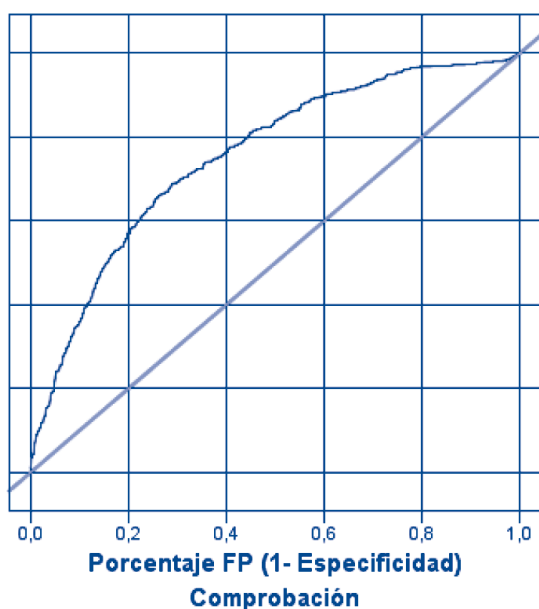
Esta evaluación de precisión de la red neuronal artificial utilizando el AUC ROC se efectúa sobre las muestras de test y validación. Los posibles valores de AUC se encuentran entre 0,50 y 1, siendo 0,5 indicativo de que el modelo está realizando clasificaciones al azar y 1 indicativo de que el modelo es capaz de realizar clasificaciones perfectas.

Las curvas ROC para las muestras de test y validación en esta investigación son muy similares tal como se muestra en el gráfico 3. Las correspondientes AUC tienen valores superiores a 0,7. (0,759 para la muestra de test y 0,752 para la muestra de validación); por tanto, la calidad de la red neuronal artificial perceptrón multicapa es buena (Hosmer Jr, Lemeshow y Sturdivant, 2013).

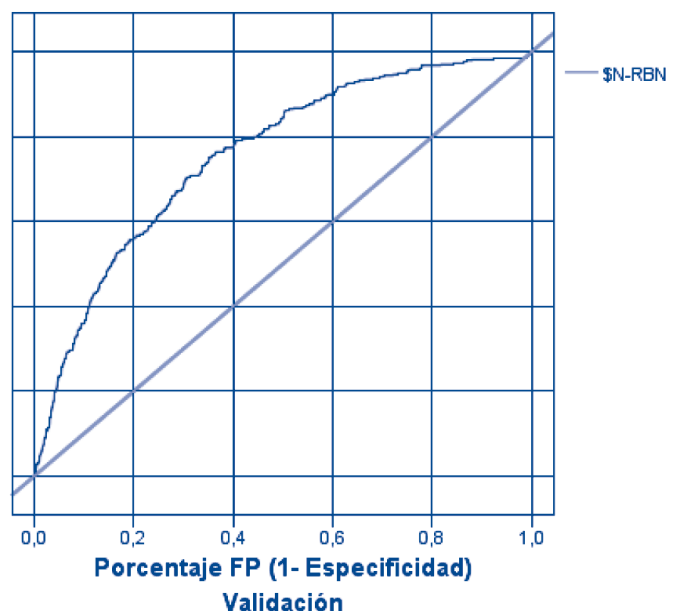
El coeficiente de Gini está relacionado con el AUC, ya que representa dos veces el área existente entre la curva ROC y la diagonal. Los coeficientes de Gini, por tanto, son 0,518 para la muestra de test y 0,504 para la de validación.

GRÁFICO 3
CURVAS ROC PARA LAS MUESTRAS DE TEST Y VALIDACIÓN

Muestra de Comprobación/Test



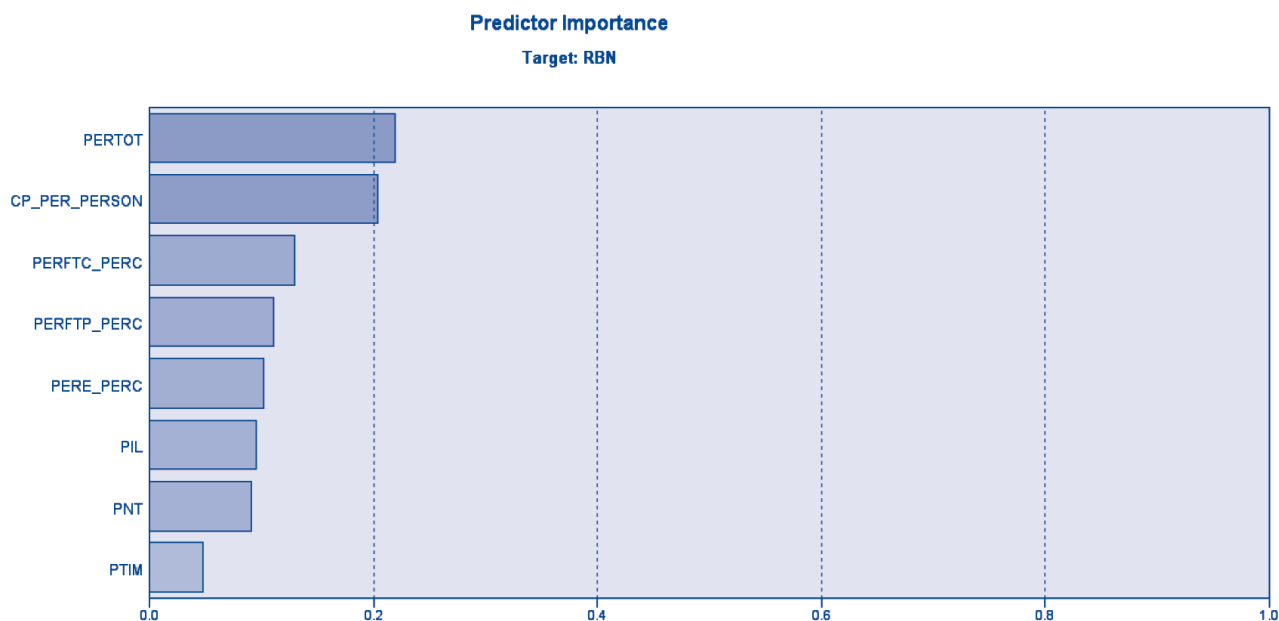
Muestra de Validación



Fuente: Elaboración propia.

GRÁFICO 4

IMPORTANCIA NORMALIZADA DE LAS VARIABLES DE ENTRADA DE LA RED NEURONAL ARTIFICIAL



Fuente: Elaboración propia.

Los resultados obtenidos corroboran que esta *ANN-MLP* es un modelo robusto para clasificar empresas en función de si han realizado su transición a la robotización o no, tan solo analizando variables que caracterizan a su plantilla laboral como relación profesional de los empleados con la compañía, nivel de estudios de los empleados, número total de empleados de la compañía y costes de personal. Adicionalmente, permite realizar predicciones acerca de si una compañía en particular se encuentra en el punto de madurez idóneo, en cuanto a transformación y configuración de su plantilla de empleados, para abordar el proceso de robotización o, incluso, realizar evaluaciones de si está demorando en exceso esta transformación o realizándolo de forma prematura. De esta forma, este modelo *ANN-MLP* constituye una herramienta de recomendación para minimizar la incertidumbre en la toma de decisiones de los directivos en los procesos de transformación a la robotización y digitalización de las compañías, tanto pymes como grandes.

En el gráfico 4 se muestra la importancia normalizada de estas variables de entrada en la red neuronal artificial. La suma de todos estos valores es igual a 1, ya que se trata de valores relativos. Cabe destacar que la importancia de las variables

no está relacionada con la calidad o precisión del modelo, sino que proporciona información sobre la importancia de cada variable de entrada de la red neuronal artificial a la hora de realizar la clasificación o predicción, pero sin entrar a valorar la precisión de dicha predicción.

En este caso, la importancia normalizada de las variables de entrada del modelo *ANN-MLP* es crucial a la hora de validar y corroborar los resultados también obtenidos usando la estrategia econométrica y es, por tanto, cuando se lleva a cabo el ejercicio de triangulación metodológica. Una misma base de datos analizada utilizando dos metodologías diferentes da lugar a conclusiones y resultados que se soportan y a la vez se complementan.

La *ANN-MLP* concluye que las dos variables con la importancia relativa normalizada más elevada son el número de empleados de la compañía y el coste de personal por cada empleado, corroborando con la estrategia econométrica la importancia de estas variables a la hora de estimar diferencias entre la constitución de la estructura de la fuerza de trabajo de las compañías robóticas comparadas con las no robóticas. Por tanto, se justifica la necesidad de dividir las muestras para su análisis

individualizado entre empresas grandes y pequeñas para evitar problemas de endogeneidad en modelos econométricos tradicionales.

A continuación, las siguientes tres variables con mayor importancia normalizada hacen referencia a la relación profesional de los empleados con la compañía y, finalmente, las tres variables que explican el nivel de estudios de los empleados. Es aquí donde se observa por primera vez en esta investigación la relación que existe entre el nivel de estudios de los empleados con el tipo de contrato, es decir, relación profesional de los empleados con la compañía. Dando más importancia a la hora de discriminar si una empresa es robótica o no al porcentaje de empleados fijos a tiempo completo que tiene la compañía, sobre los fijos a tiempo parcial o eventuales.

Por último, nuevamente se corroboran los resultados obtenidos en el modelo de estrategia econométrica, ya que la importancia normalizada del porcentaje de empleados con estudios de ingeniería o titulaciones superiores, seguidos de los de titulaciones medias es mayor que los de empleados no titulados a la hora de discriminar si una empresa es robótica o no.

Es decir, la presencia de titulados superiores y medios es la que diferenciará una fuerza de trabajo de una empresa que ha realizado su transición a la robotización frente a una que aún no ha afrontado dicho proceso, independientemente de si es una empresa grande o pequeña.

En consecuencia, la *ANN-MLP*, como sucede aplicando modelos Tobit, también es capaz de capturar las diferencias que la robotización ha generado en la configuración de la fuerza de trabajo en las empresas tanto pequeñas y medianas como grandes. Siendo estas diferencias tan significativas que la red puede agrupar e incluso predecir si las empresas han realizado su transformación a la robotización solamente con información relativa a la composición y característica de su fuerza de trabajo.

IV. CONCLUSIONES

Uno de los muchos efectos indirectos de la crisis de la COVID-19 va ser el crecimiento de la automatización en los países occidentales. La tendencia a la digitalización está afectando a las condiciones de los trabajadores y cuestiones como si existe o no la dicotomía trabajador máquina, y si ya se

está produciendo ese cambio qué habilidades son las necesarias para preservar el nivel de empleo en las economías occidentales, son cuestiones que han cobrado una relevancia crítica. A pesar del interés que existe por este tema, sobre el que se ha pronunciado, por ejemplo, el Parlamento Europeo (2017), no existen evidencias concluyentes sobre cuál es el efecto y mucho menos sobre qué competencias deben tener los trabajadores para seguir manteniendo el empleo en la industria. Además, dichos estudios se han realizado mayoritariamente sobre agregados sectoriales o de país y no han tenido en cuenta la evolución y la adaptación de la industria al proceso de automatización. Para completar en ese ámbito la literatura en este trabajo se han aportado nuevas evidencias sobre los efectos de la robotización, la digitalización y la cualificación del empleo en las empresas, utilizando una amplia muestra de empresas manufactureras españolas para el período 1990-2016.

Los resultados de esta investigación señalan que, tal como se esperaba, las empresas que más utilizan robots emplean trabajadores con mayor nivel de educación, pero también muestra que las compañías que más tiempo llevan participando en el proceso de robotización alcanzan un nivel de saturación, ya que contratan más trabajadores con una formación intermedia. Es decir, de estos resultados se colige que durante la implantación de robots industriales las empresas incrementan el número de empleados cualificados y que una vez que la empresa se ha transformado en robótica vuelve a realizar contrataciones de personal de cualificación intermedia, reduciendo las titulaciones superiores. Esto hace pensar que, en la gran industria española, mucha de ella participada por capital internacional, existe un efecto de sustitución entre los trabajadores más cualificados y la automatización, mientras los trabajadores con cualificación media son complementarios. Esto indica que la adopción de robots desplaza trabajadores altamente cualificados, pero demanda técnicos que sean capaces de operar los robots.

Los resultados de esta investigación son consistentes tanto en la especificación del modelo econométrico que se ha planteado como en el análisis a través de un modelo de *machine learning* supervisado; en este caso empleamos una red neuronal artificial perceptrón multicapa (*ANN-MLP*) con el objetivo de capturar los mismos efectos a través de dos metodologías distintas. Los resultados son consistentes entre los modelos, lo que demuestra que ambas estrategias son complementarias y pueden

aportar riqueza al análisis empírico (Athey y Imbens, 2019; Efron y Hastie, 2016; Varian, 2014).

El empleo de datos desagregados por empresa permite avanzar en el conocimiento de cómo se produce la automatización de un país completando estudios anteriores. Pero nuestro análisis muestra solo una imagen preliminar. Creemos que hay análisis causales que todavía están por explorar y que aspectos como la definición de nivel educativo, cualificación y formación o la definición del sector deberían añadir riqueza al análisis. Además, como plantean (Gregory, Salomons y Zierahn, 2019) solo capturamos uno de los muchos aspectos de la digitalización de las economías.

BIBLIOGRAFÍA

- AAD, G., ABBOTT, B., ABDALLAH, J., ABDELALIM, A. A., ABDESSELAM, A., ABDINOV, O., ABI, B., ABOLINS, M., ABRAMOWICZ, H., ABREU, H., ACERBI, E., ACHARYA, B. S., ADAMS, D. L., ADDY, T. N., ADELMAN, J., ADERHOLZ, M., ADOEIT, S., ADRAGNA, P., ADYE, T., ... ZWALINSKI, L. (The ATLAS Collaboration Group). (2012). Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 7$ TeV with ATLAS. *European Physical Journal C*, 72(1), pp. 1-35. <https://doi.org/10.1140/epjc/s10052-011-1844-6>
- ACEMOGLU, D., MAKHDOUNI, A., MALEKIAN, A. y OZDAGLAR, A. (2019). *Learning From Reviews: The Selection Effect and the Speed of Learning* *.
- ACEMOGLU, D. y RESTREPO, P. (2017). *Automation and local labor markets*.
- (2019). Robots and Jobs: Evidence from US Labor Markets. *Journal of Political Economy*. <https://doi.org/10.1086/705716>
- AGHION, P., BERGEAUD, A., BOPPART, T., KLENOW, P. J. y LI, H. (2019). Missing Growth from Creative Destruction. *American Economic Review*, 109(8), pp. 2795-2822. <https://doi.org/10.1257/aer.20171745>
- ALCOVER, J. (1875). *La industria en 1874*. Manuel Tello. https://books.google.es/books/about/La_industria_en_1874.html?id=Z0WkQwAACAAJ&redir_esc=y
- ARNTZ, M., GREGORY, T. y ZIERAHN, U. T. (2020). Digitization and the Future of Work: Macroeconomic Consequences. *Handbook of Labor, Human Resources and Population Economics*.
- ATHEY, S. e IMBENS, G. W. (2019). Machine Learning Methods That Economists Should Know About. *Annual Review of Economics*, 11(1), pp. 685-725. <https://doi.org/10.1146/annurev-economics-080217-053433>
- (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*, 31(2), pp. 3-32. <https://doi.org/10.1257/jep.31.2.3>
- AUTOR, D. y SALOMONS, A. (2018). Is automation labor-displacing? Productivity growth, employment, and the labor share. En: *Brookings Papers on Economic Activity*. <https://www.brookings.edu/bpea-articles/is-automation-labor-displacing-productivity-growth-employment-and-the-labor-share/>
- BALLESTAR, M. T., DÍAZ-CHAO, Á., SAINZ, J. y TORRENT-SELLENS, J. (2020). Knowledge, robots and productivity in SMEs: Explaining the second digital wave. *Journal of Business Research*, 108, pp. 119-131. <https://doi.org/10.1016/j.jbusres.2019.11.017>
- (2021). Impact of robotics on manufacturing: A longitudinal machine learning perspective. *Technological Forecasting and Social Change*, 162, 120348. <https://doi.org/10.1016/j.techfore.2020.120348>
- BALLESTAR, M. T., DONCEL, L. M., SAINZ, J. y ORTIGOSA-BLANCH, A. (2019). A novel machine learning approach for evaluation of public policies: An application in relation to the performance of university researchers. *Technological Forecasting and Social Change*, 149. <https://doi.org/10.1016/j.techfore.2019.119756>
- BALLESTAR, M. T., GRAU-CARLES, P. y SAINZ, J. (2018). Customer segmentation in e-commerce: Applications to the cashback business model. *Journal of Business Research*, 88, pp. 407-414. <https://doi.org/10.1016/j.jbusres.2017.11.047>
- (2019). Predicting customer quality in e-commerce social networks: a machine learning approach. *Review of Managerial Science*. <https://doi.org/10.1007/s11846-018-0316-x>
- BALLESTAR, M. T., RIBEIRO, D. y SAINZ, J. (2018). ¿Es el big data el siguiente paso en la digitalización de la empresa? *Economía Industrial*, 409, pp. 47-56.
- BRYNJOLFSSON, E. y OH, J. (2012). The Attention Economy: Measuring the Value of Free Digital Services on the Internet. *ICIS 2012 Proceedings*. <https://aisel.aisnet.org/icis2012/proceedings/EconomicsValue/9>
- CARLAW, K. I. y LIPSEY, R. G. (2002). Externalities, technological complementarities and sustained economic growth. *Research Policy*, 31(8-9), pp. 1305-1315. [https://doi.org/10.1016/S0048-7333\(02\)00065-3](https://doi.org/10.1016/S0048-7333(02)00065-3)
- CHENG, C., HUA, Y. y TAN, D. (2019). Spatial dynamics and determinants of sustainable finance: Evidence from venture capital investment in China. *Journal of Cleaner Production*, 232, pp. 1148-1157. <https://doi.org/10.1016/j.jclepro.2019.05.360>
- CHUI, M., MANYIKA, J. y MIREMADI, M. (2016). Where machines could replace humans – and where they can't (yet). *McKinsey Quarterly*, 30(2), pp. 1-9. <http://www.oregon4biz.com/assets/e-lib/Workforce/MachReplaceHumans.pdf>
- CHUNG, P., SOLTOGGIO, A., DAWSON, C. W., MENG, Q. y PAIN, M. (2015). *Proceedings of the 10th International Symposium on Computer Science in Sports (ISCSS)* (Vol. 392). Springer.

- DÍAZ-CHAO, Á., SAINZ-GONZÁLEZ, J. y TORRENT-SELLENS, J. (2015). ICT, innovation, and firm productivity: New evidence from small local firms. *Journal of Business Research*, 68(7), pp. 1439-1444. <https://doi.org/10.1016/j.jbusres.2015.01.030>
- DICKSON, K. E. y HADJIMANOLIS, A. (1998). Innovation and networking amongst small manufacturing firms in Cyprus. *International Journal of Entrepreneurial Behaviour & Research*, 4(1), pp. 5-17. <https://doi.org/10.1108/13552559810203939>
- DORASZELSKI, U. y JAUMANDREU, J. (2018). Measuring the Bias of Technological Change. *Journal of Political Economy*. <https://doi.org/10.1086/697204>
- DŽELIHODŽIĆ, A. y DJONKO, D. (2016). Comparison of ensemble classification techniques and single classifiers performance for customer credit assessment. *Modeling of Artificial Intelligence*, 3, pp. 140-150.
- EFFRON, B. y HASTIE, T. (2016). Computer age statistical inference: Algorithms, evidence, and data science. En *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science*. <https://doi.org/10.1017/CBO9781316576533>
- FABER, M. (2018). *Robots and reshoring: Evidence from Mexican local labor markets*. WWZ.
- FREY, C. B. y OSBORNE, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 114, pp. 254-280. <https://doi.org/10.1016/j.techfore.2016.08.019>
- GANGANWAR, V. (2012). An overview of classification algorithms for imbalanced datasets. *International Journal of Emerging Technology and Advanced Engineering*, 2(4), pp. 42-47.
- GOOLSBEE, A. D. y KLENOW, P. J. (2018). Internet Rising, Prices Falling: Measuring Inflation in a World of E-Commerce. *AEA Papers and Proceedings*, 108, pp. 488-492. <https://doi.org/10.1257/pandp.20181038>
- GRAETZ, G. y MICHAELS, G. (2018). Robots at work. En *Review of Economics and Statistics*, 100(5), pp. 753-768. MIT Press Journals. https://doi.org/10.1162/rest_a_00754
- GREGORY, T., SALOMONS, A. y ZIERAHN, U. (2016). Racing with or Against the Machine? Evidence from Europe. *SSRN Electronic Journal*, 16. <https://doi.org/10.2139/ssrn.2815469>
- (2019). Racing With or Against the Machine ? Evidence from Europe. In *IZA Institute of Labor Economics (IZA DP n.º 12063; Discussion Papers*, vol. 2019, issue 12063).
- GUO, H. y VIKTOR, H. L. (2004). Learning from imbalanced data sets with boosting and data generation: the databoost-im approach. *ACM Sigkdd Explorations Newsletter*, 6(1), pp. 30-39.
- HOSMER JR, D. W., LEMESHOW, S. y STURDIVANT, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.
- JORGENSEN, D. W., HO, M. S. y SAMUELS, J. D. (2011). Information technology and U.S. productivity growth: Evidence from a prototype industry production account. *Journal of Productivity Analysis*, 36(2), pp. 159-175. <https://doi.org/10.1007/s11123-011-0229-z>
- KANKARAŠ, M., FERON, E. y RENBARGER, R. (2019). *Assessing students' social and emotional skills through triangulation of assessment methods* (Issue 208).
- KRAUS, S., PALMER, C., KAILER, N., KALLINGER, F. L. y SPITZER, J. (2019). Digital entrepreneurship: A research agenda on new business models for the twenty-first century. In *International Journal of Entrepreneurial Behaviour and Research*, 25(2). Emerald Group Publishing Ltd. <https://doi.org/10.1108/IJEBR-06-2018-0425>
- LI, Z. y EASTMAN, J. R. (2006). Commitment and typicality measurements for fuzzy ARTMAP neural network. *Geoinformatics 2006: Geospatial Information Science*, 6420, 642011.
- MURO, M., MAXIM, R., WHITON, J. y HATHAWAY, I. (2019). *Automation and Artificial Intelligence: How Machines are Affecting People and Places*.
- OCDE. (2019). *Determinants and impact of automation: an analysis of robots' adoption in OECD countries*. <https://doi.org/https://doi.org/10.1787/ef425cb0-en>
- PALIWAL, M. y KUMAR, U. A. (2009). Neural networks and statistical techniques: A review of applications. *Expert Systems with Applications*, 36(1), pp. 2-17.
- EUROPEAN PARLIAMENT RESOLUTION ON DIGITALISING EUROPEAN INDUSTRY (2017) (testimony of European Parliament).
- PÉREZ, C. (2003). *Technological revolutions and financial capital*. Edward Elgar Publishing.
- SEAMANS, R. y RAJ, M. (2018a). AI, Labor, Productivity, and the Need for Firm-Level Data. *National Bureau of Economic Research*. <https://doi.org/10.3386/w24239>
- (2018b). *NBER Working Paper Series AI, Labor, Productivity and the Need for Firm-Level Data*. <http://www.nber.org/papers/w24239>
- SYVERSON, C. (2011). What determines productivity. En *Journal of Economic Literature*, 49(2), pp. 326-365. <https://doi.org/10.1257/jel.49.2.326>
- TU, J. V. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of Clinical Epidemiology*, 49(11), pp. 1225-1231.
- VARIAN, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2), pp. 3-28. <https://doi.org/10.1257/jep.28.2.3>
- YIN, L., GE, Y., XIAO, K., WANG, X. y QUAN, X. (2013). Feature selection for high-dimensional imbalanced data. *Neurocomputing*, 105, pp. 3-11.