



Universidad de Oviedo



Universidad de Oviedo

Grado en Economía

4º Curso

Trabajo de fin de Grado

**ANÁLISIS DEL ENTORNO EMPREDEDOR:
VALIDACIÓN DEL MODELO GEM**

ÁNGEL HERRERO GARCÍA

OVIEDO, ENERO 2024



Universidad de Oviedo

DECLARACIÓN RELATIVA AL ARTÍCULO 8.3 DEL REGLAMENTO SOBRE LA ASIGNATURA TRABAJO FIN DE GRADO

(Acuerdo de 5 de marzo de 2020, del Consejo de Gobierno de la Universidad de Oviedo).

Yo ÁNGEL HERRERO GARCÍA, con DNI

DECLARO

Que el TFG titulado ANALISIS DEL ENTORNO EMPRENDEDOR: VALIDACIÓN DEL MODELO GEM es una obra original y que he citado debidamente todas las fuentes utilizadas.

Enero del 2024



Universidad de Oviedo

ANÁLISIS DEL ENTORNO EMPRENDEDOR: VALIDACIÓN DEL MODELO GEM

RESUMEN:

El proyecto GEM surge en 1997 como una respuesta a la necesidad de medir y comparar el fenómeno emprendedor entre países, actividad que representa uno de los principales motores de cambio en la economía. Frente a este objetivo, el proyecto propone una metodología sólida y homogénea. Nosotros, a través de esta investigación, tratamos de verificar si es capaz de medir correctamente el emprendimiento.

Mediante un panel compuesto por observaciones para las 17 CCAA, desde 2003 hasta 2021, utilizamos una metodología de panel con estimaciones por efectos fijos. Tratamos de buscar correlaciones significativas y positivas entre los índices de emprendimiento elaborados por el GEM y el emprendimiento directamente observado, que se traduce en nuevas sociedades mercantiles constituidas y nuevos afiliados al Régimen Especial de Trabajadores Autónomos. Finalmente, se verifica que, aunque pequeña, existe una relación positiva entre los emprendedores estimados y la creación de sociedades.

ANALYSIS OF THE ENTREPRENEURIAL ENVIRONMENT: VALIDATION OF THE GEM MODEL

ABSTRACT:

The GEM project was created in 1997 in response to the need to measure and compare the entrepreneurial phenomenon between countries, an activity that represents one of the main driving force of change in the economy. With this objective in mind, the project proposes a solid and homogeneous methodology. We, through this research, try to verify if it is able to measure entrepreneurship correctly.

Using a panel composed of observations for the 17 autonomous communities, from 2003 to 2021, we use a panel methodology with fixed effects estimations. We try to find significant and positive correlations between the entrepreneurship index elaborated by the GEM and the directly observed entrepreneurship, which is shown into new mercantile companies constituted and new affiliates to the Special Regime for Self-Employed Workers. Finally, it is verified that, although small, there is a positive relationship between the estimated number of entrepreneurs and the creation of companies.



ÍNDICE

1. INTRODUCCIÓN	5
2. CONCEPTO DE EMPRENDIMIENTO	6
2.1 CÓMO MEDIR EL FENÓMENO EMPRENDEDOR: EL GEM	7
3. PANEL Y DATOS	8
4. METODOLOGIA DE PANEL	10
4.1 ESTIMACION DE MCO COMBINADOS	11
4.2 ESTIMACION POR EFECTOS FIJOS	12
4.3 ESTIMACION POR EFECTOS ALEATORIOS	13
4.4 SELECCIÓN DE ESTIMADORES.....	13
4.5 CONTRASTES DE DIAGNOSTICO	14
5. ESTIMACION DE MODELOS.....	15
5.1 MODELOS INICIALES.....	16
5.2 MODELOS EN DIFERENCIAS	19
5.3 MODELOS CON RETARDOS.....	20
5.4 MODELOS EN LOGARITMOS	22
6. ANÁLISIS DE RESULTADOS	24
6.1 EFECTOS INDIVIDUALES	24
6.2 EFECTOS TEMPORALES.....	29
7. CONCLUSIONES	30
8. BIBLIOGRAFÍA	32



1. INTRODUCCIÓN

Es ampliamente reconocido que el emprendimiento es uno de los factores determinantes en los cambios del panorama económico, (Baumol, 1968; Birch, 1979; Acs y otros 1999). Se trata de una actividad generadora de riqueza, empleo y progreso técnico, factor fundamental en el crecimiento económico y la innovación de países y regiones. Sin embargo, su medición no es una tarea sencilla, pero sí necesaria. El fenómeno emprendedor abarca una gran variedad de actividades, desde la creación de pequeñas empresas hasta la innovación de las grandes, por lo que cuantificarlo es una tarea compleja. Además, es un concepto con cierta subjetividad y que sufre una gran variación entre países y regiones, lo que también dificulta la comparación.

Por estas razones, surge la necesidad de una metodología homogénea y confiable para medir el fenómeno emprendedor, que pueda ayudar a los distintos agentes sociales y económicos a disponer de una imagen fiel del emprendimiento que guíe sus actuaciones. En este punto, entra el proyecto Global Entrepreneurship Monitor (GEM), que propone una metodología sólida, no sólo cuantifica el fenómeno, sino, que permite la comparación entre países o regiones.

La pregunta que cabe plantearse es si la metodología GEM sirve para estimar correctamente el fenómeno emprendedor.

En respuesta a esta cuestión, se propone como finalidad de esta investigación verificar la siguiente hipótesis: “los índices de emprendimiento TEA y TEP elaborados por el GEM capturan de manera efectiva y significativa la dinámica del emprendimiento en la población analizada”. Estos dos índices, la tasa de actividad emprendedora total (TEA) y la tasa de actividad emprendedora potencial (TEP), se han elegido por ser la medida de emprendimiento más representativa de las que elabora el proyecto. Para capturar el fenómeno emprendedor, se han elegido variables como la creación de nuevas sociedades mercantiles y las nuevas afiliaciones al Régimen Especial de Trabajadores Autónomos (RETA), que son directamente observadas y disponibles en estadísticas oficiales.

Nuestra hipótesis de trabajo se va a contrastar con un panel de datos compuesto por observaciones para las 17 CCAA desde 2003 hasta 2021. Mediante técnicas econométricas, que se describirán en el apartado de metodología, se van a estimar modelos que expliquen, a través del emprendimiento estimado, las variaciones en la constitución de nuevas sociedades mercantiles, por un lado, y las variaciones en las nuevas afiliaciones al RETA por otro. La gran ventaja de utilizar metodología de panel es que comprende una doble dimensión, temporal e individual, lo que proporciona una mayor información en los resultados, midiendo efectos que modelos más simples son incapaces de detectar. También permite aislar efectos individuales y temporales, lo que nos posibilita medir qué comunidades tienen un efecto mayor y cuáles un efecto menor en el nivel de emprendimiento, o qué años presentan tendencias más marcadas.

El trabajo partirá en el siguiente capítulo con algunas aproximaciones de autores destacados a la definición de emprendimiento, y su importancia en la sociedad y la economía, no sólo como actividad creadora de riqueza, sino como motor del cambio e innovación. El papel del proyecto GEM como respuesta a la necesidad de medirlo y compararlo. En el tercer capítulo se presentará el panel de datos y las variables que lo componen, y en el cuarto, los distintos métodos de estimación que utilizaremos, así como los contrastes para seleccionar los más adecuados y diagnosticar problemas que



puedan presentar los modelos. El quinto capítulo está dedicado a plantear y estimar modelos que buscan verificar la hipótesis de trabajo, y en el sexto se medirán e interpretarán los efectos individuales y temporales. Por último, se presentarán las conclusiones.

2. CONCEPTO DE EMPRENDIMIENTO

El emprendimiento es una actividad que implica la creación de valor a partir de la identificación y explotación de oportunidades de negocio, la innovación, la asunción de riesgos y el uso de recursos. Esta actividad tiene una gran importancia para el desarrollo económico y social de los países y regiones, ya que contribuye a la generación de empleo, ingresos, competitividad y crecimiento. Su estudio se remonta a los orígenes de la economía, autores como Richard Cantillon, pionero en la disciplina, introduce por primera vez el término emprendedor en su obra *“Ensayo sobre la naturaleza del comercio en general”*, publicada póstumamente en 1755. En ella, define al emprendedor como el agente que compra los medios de producción a ciertos precios y los combina de forma ordenada para obtener de allí un nuevo producto, (Cantillon y Jevons, 1950). Posteriormente, Adam Smith también reflexiona sobre la importancia del fenómeno emprendedor. En su libro *“Una investigación sobre la naturaleza y las causas de la riqueza de las naciones”*, publicado por primera vez en 1776. Smith entiende el emprendimiento como el resultado de la división del trabajo, la especialización, la acumulación de capital y la libre competencia. Actividad que repercute en beneficio de la sociedad con el aumento de la productividad, la innovación, el bienestar y el progreso, (Smith y otros, 1996).

Ya en el siglo XX, el emprendimiento adquirió una mayor relevancia académica y práctica, debido a los cambios tecnológicos, sociales y políticos que favorecieron el surgimiento de nuevas formas de emprender. Uno de los autores más destacados de esta época fue Joseph Schumpeter, cuya obra *“Teoría del desenvolvimiento económico”*, publicada por primera vez en 1911 y considerada una de las más importantes de la economía. El emprendedor es capaz de crear nuevas combinaciones productivas que generan un proceso de destrucción creativa, es decir, la sustitución de las viejas estructuras por otras más eficientes y rentables (Schumpeter, 1976). A lo largo de este siglo, se han dado numerosos ejemplos de emprendedores fundadores de empresas que han revolucionado la forma de vivir de las personas, tanto en el ámbito industrial como en el tecnológico. Algunos ejemplos destacables son Henry Ford, quien introdujo la producción en serie y el automóvil de bajo coste; George Eastman, quien popularizó la fotografía con su cámara Kodak; Charles Francis Brush, quien desarrolló la primera central eléctrica comercial; Bill Gates, quien fundó Microsoft y lideró la revolución del software; y Steve Jobs, quien creó Apple y transformó la industria de la informática, la música y la telefonía. Estos emprendedores no solo crearon productos y servicios innovadores, sino que también contribuyeron al desarrollo económico y social de sus países y del mundo. El emprendimiento en el siglo XX estuvo influenciado por diversos factores, como la segunda revolución industrial, la globalización, la digitalización, la emergencia de nuevos mercados, etc. Estos factores crearon oportunidades y desafíos para los emprendedores, que tuvieron que adaptarse a las nuevas demandas y competencias.



2.1 CÓMO MEDIR EL FENÓMENO EMPRENDEDOR: EL GEM

A finales de siglo, en 1997, nace el proyecto GEM, una iniciativa internacional creada por el Babson College y la London Business School, cuyos objetivos principales son medir y comparar el nivel y las características del emprendimiento en diferentes países y regiones del mundo. Utiliza una metodología común y una amplia base de datos, que permite la comparación entre países. El proyecto GEM es una fuente de información valiosa para entender el fenómeno del emprendimiento y sus implicaciones para el desarrollo de los países.

Esta metodología se vertebra en tres ejes: la encuesta de población adulta (APS), las entrevistas a expertos y un conjunto de fuentes secundarias, como bases de datos internacionales o nacionales (INE, Naciones Unidas, el Banco Mundial, etc) con el fin de establecer hipótesis de partida en el ámbito internacional, como se señala en el informe ejecutivo GEM 1999, (Reynolds y otros, 1999) y en el informe ejecutivo GEM España 2005, (Navarro y otros, 2007). La APS es una encuesta estandarizada que se realiza a una muestra de, al menos, 2000 personas que se encuentren entre 18 y 64 años. Mediante esta encuesta se obtienen los datos para confeccionar los índices, como la tasa de actividad emprendedora (TEA), la tasa de emprendimiento establecido (EE) o la tasa de abandono de negocios (BA), entre otros. Las entrevistas a expertos son conversaciones estructuradas en las que participan personas con conocimiento y experiencia en el entorno emprendedor, se seleccionan 36 por país. Estas entrevistas pretenden evaluar las condiciones del entorno emprendedor en cada país, como el apoyo financiero, las políticas de fomento del emprendimiento, la educación, la transferencia de I+D, etc. Con la información de estas entrevistas se obtiene el índice de condiciones del entorno para el emprendimiento (EFC).

Estos dos pilares, la APS y las encuestas, son las herramientas que dotan a la metodología GEM de esa homogeneidad que permite comparaciones entre países u otras unidades geográficas, gracias a que todos los equipos pertenecientes al consorcio recogen los datos aplicando el mismo procedimiento, mediante las mismas preguntas y cuestionarios. Luego esto se complementa con las fuentes secundarias mencionadas.

Una vez tratados los datos, las medias e índices se publican en informes anuales, realizados por los equipos nacionales. En la actualidad participan 115 países, que representan el 69% de la población mundial y el 83% del PIB mundial, datos del informe global GEM 2021-2022 (Hill y otros, 2022). Toda esta información es muy valiosa para los distintos agentes sociales y económicos. A los académicos les proporciona datos rigurosos y comparables sobre las características, las motivaciones, las actividades y el impacto de los emprendedores, así como sobre las condiciones del entorno que favorecen o dificultan el emprendimiento. Para los empresarios ofrece información relevante sobre las oportunidades y los desafíos que existen para emprender en diferentes mercados y sectores. Y para los políticos aporta evidencia sobre el estado y la evolución del fenómeno emprendedor, también recomendaciones y orientaciones para diseñar políticas públicas que fomenten el emprendimiento.

De los índices elaborados por el GEM, el más relevante y referenciado es la TEA, el cual se calcula como la población adulta entre 18 y 64 en cada país o región, propietarios o copropietarios fundadores de empresas de nueva creación que hayan persistido en el mercado por un periodo de 0 a 42 meses.



Figura 2.1.1: Índices de emprendimiento elaborados por el GEM. Fuente: (Reynolds y otros, 2005).

En la figura 2.1.1 encontramos las fases en las que la metodología del proyecto GEM descompone el proceso emprendedor descrito en el artículo “GEM: Data Collection Design Implementation 1998-2003”, (Reynolds y otros, 2005). La primera fase del proceso es la potencial, en ella se encuadran adultos de 18 a 64 años que han expresado la intención de emprender en los próximos 3 años. La segunda fase, emprendimiento naciente, se da cuando se emprende un nuevo negocio y se hayan pagado salarios por menos de 3 meses, estas son las iniciativas en fase de despegue. De la segunda fase se pasa a la tercera, emprendimiento nuevo, con el pago de salarios por más de tres meses y hasta 42, son negocios en consolidación. Y, por último, la cuarta fase, empresas consolidadas. Son empresas que han superado las fases anteriores y llevan más de 42 meses operando en el mercado.

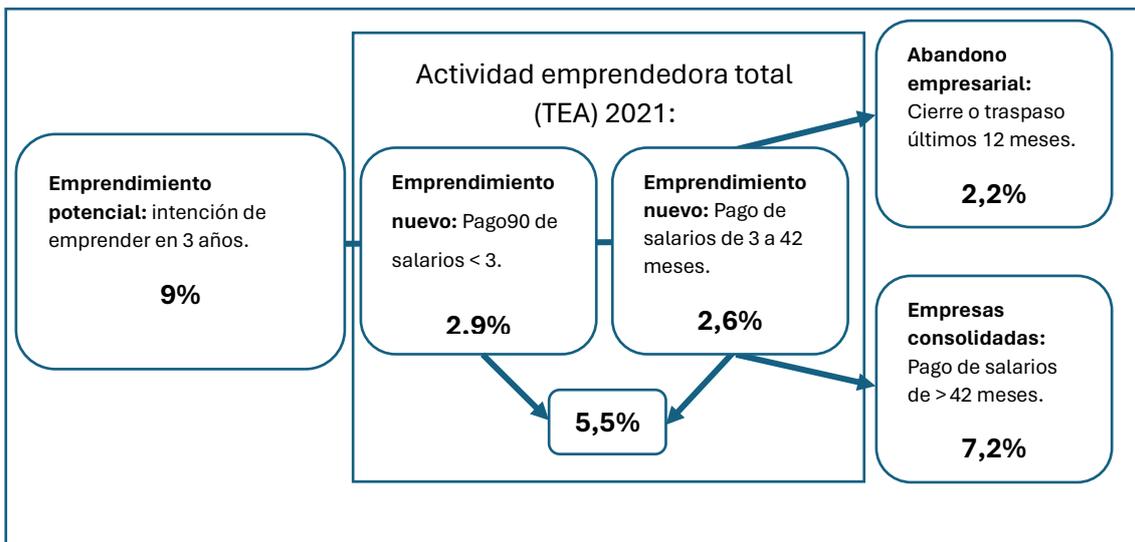


Figura 2.1.2: Las fases del proceso emprendedor en el modelo GEM. Fuente: (Laviada y otros 2022).

La metodología GEM considera a los emprendedores que se encuadran en la fase 1 como emprendedores potenciales, de donde se calcula la TEP. Los encuadrados en las fases 2 y 3 son los llamados emprendedores nacientes y nuevos, respectivamente. La



Universidad de Oviedo

suma de ambos sirve para calcular la TEA. De la fase 4 se calcula la tasa de empresas consolidadas. También se calcula el índice de abandono empresarial a partir de las empresas que se han cerrado o traspasado en los últimos 12 meses. En la figura 2.1.2 vemos el gráfico de índices sobre el proceso emprendedor que se publica en los informes anuales.

3. PANEL Y DATOS

En este apartado se presentan los datos recopilados para construir el panel. El objetivo central del trabajo es verificar si existe correlación entre los índices TEA y TEP elaborados por el observatorio GEM en España y otras variables en las que se traduce el fenómeno emprendedor. En otras palabras, se van a estimar las correlaciones que existen entre los datos reales que pueden aproximar el fenómeno emprendedor (número de autónomos o empresas) y los datos estimados por el GEM a través de su estudio de campo anual. Se han recopilado datos de 17 individuos (los datos proporcionados por los 17 equipos regionales GEM en sus informes anuales, correspondientes a las 17 comunidades autónomas), y los referentes a sociedades mercantiles extraídos de las bases de datos del Instituto Nacional de Estadística (INE) y los de afiliación al Régimen Especial de Trabajadores Autónomos provenientes de la (RETA) Seguridad Social. Estos datos se han recopilado para 19 periodos, desde 2003 hasta 2021.

Aunque la metodología GEM es común para todos los países, el trabajo de campo se efectúa de forma descentralizada por los equipos locales. Esto provocó que, desafortunadamente, no se cuente con las mismas observaciones para todas las comunidades, ya que algunos de los equipos se constituyeron antes que otros. En el caso del índice TEA, se cuenta con observaciones desde 2003 para Andalucía, Cataluña, Extremadura y Galicia; desde 2004 para Castilla y León, Comunidad de Madrid, Comunidad Valenciana, Islas Canarias y País Vasco; desde 2005 para Navarra; y desde 2006 para las restantes. Para el índice TEP se cuentan con datos desde 2003 para Andalucía, Cataluña y Extremadura; desde 2004 para Castilla y León, Comunidad de Madrid, Comunidad Valenciana, Islas Canarias y País Vasco; desde 2005 para Galicia y Navarra; y desde 2006 para los restantes.

Adicionalmente, para disponer de variables expresadas en las mismas unidades, se han convertido las tasas en número de emprendedores. Para ello se multiplicaron dichas tasas por el número de habitantes (datos de población extraídos de las bases de datos del INE), obteniendo así las variables *Emprendedores*, que es la TEA en número de habitantes, y *EmprendedoresPot*, que es la TEP también en número de habitantes. Estas dos últimas variables serán las explicativas de los modelos.

Respecto a las variables explicadas, se han elegido como representativas del fenómeno emprendedor: las nuevas sociedades mercantiles constituidas, *NuevasSoc*, con datos extraídos del INE, y las nuevas afiliaciones al RETA, *NuevosSSAut*, datos provenientes de la Seguridad Social. Para ambas variables se cuentan con observaciones para las 17 comunidades desde 2003 hasta 2021.



Una vez hemos recopilado los datos, estas son las 6 variables resultantes:

1. TEA: Tasa de actividad emprendedora total.
2. TEP: Tasa de actividad emprendedora potencial.
3. Emprendedores: personas en edad adulta que manifiestan desarrollar una actividad emprendedora en un periodo de 0 a 42 meses.
4. EmprendedoresPot: personas en edad adulta que manifiestan la intención de desarrollar una actividad emprendedora en un plazo de 3 años.
5. NuevasSoc: nuevas sociedades mercantiles constituidas.
6. NuevosSSAut: nuevos afiliados al RETA.

4. METODOLOGIA DE PANEL

En el apartado anterior se han presentado las variables y características del panel con el que se va a trabajar para estimar los modelos regresión. En esta sección se expondrán la metodología de panel que se va a aplicar y los supuestos teóricos, así como los contrastes que nos ayudarán a elegir qué tipo de estimación es la más adecuada: MCO combinados, efectos fijos o efectos aleatorios. También interpretar y analizar los resultados y detectar problemas estructurales que puedan surgir, como son la autocorrelación y la heterocedasticidad.

Al contrario que los conjuntos de datos de series temporales o de corte transversal, los modelos de panel son una herramienta que permite capturar la variabilidad a lo largo del tiempo y entre individuos. Estos modelos nos aportan una serie de ventajas. Entre las expuestas por (Hsiao, 2003) y (Klevmarken, 1989) destacan:

1. Controlan la heterogeneidad individual no observada que podría influir en las variables de interés. Los modelos de panel permiten aislar y analizar tanto las variaciones temporales como transversales.
2. Los paneles, al combinar información temporal y transversal, proporcionan una mayor información, más variabilidad, menos colinealidad entre las variables, más grados de libertad y mayor eficiencia.
3. Permite medir e identificar efectos que en los datos de series temporales o corte transversal no son detectables. Esto se debe a que los modelos de panel aprovechan la variabilidad temporal y transversal de las variables y controlan la heterogeneidad inobservable de las unidades.

Teóricamente, un conjunto de datos de panel se compone de una dimensión $N \times T$, donde N son el número de individuos y se representan con el subíndice i ($i=1, \dots, N$), y T el número de periodos, subíndice t ($t=1, \dots, T$), las variables se denotan como k ($k=1, \dots, K$). Si la dimensión temporal es mayor que la individual ($T > N$) se trata de un panel largo. En cambio, si es al revés ($T < N$), se denomina panel corto. Además, si existen observaciones ausentes, se tratará de un panel no balanceado, mientras que, si tenemos observaciones completas para las dimensiones transversal y temporal en todas las variables, se tratará de un panel balanceado. En este caso, nuestro panel es de 17 individuos ($N=17$) y 19 periodos ($T=19$). Al haber algunas observaciones ausentes para las variables *TEA*, *TEP*, *Emprendedores* y *EmprendedoresPot* hay que señalar que este es un panel incompleto o no equilibrado y al tener una dimensión temporal mayor que individual ($T > N$) se trata de un panel largo.



La formulación para el modelo de panel general se puede expresar de la siguiente manera:

$$y_{it} = \beta_1 + X_{kit} * \beta_2 + u_{it}$$

Para que las estimaciones a través de mínimos cuadrados ordinarios (MCO) sean eficientes se deben cumplir los siguientes supuestos para las perturbaciones:

1. Esperanza nula: $E(u_{it}) = 0$
2. No autocorrelación: $Cov(u_{it}, u_{is}) = 0$
3. Homocedasticidad: $Var(u_{it}) = \sigma^2$

Al encontrarnos en un modelo de panel, en principio, las perturbaciones del error se descompondrán en:

$$u_{it} = \alpha_i + \phi_t + \varepsilon_{it}$$

Donde α_i y ϕ_t son los componentes individuales y temporales, componen la heterogeneidad no observada. Esto son las características no observadas ni medibles que son comunes entre los individuos o a lo largo del tiempo.

En principio, u_{it} no verifica los supuestos mencionados al no tener en cuenta los efectos temporales e individuales, por lo que la estimación por MCO no sería eficiente. En cambio, ε_{it} es una perturbación de ruido blanco, lo que verifica las 3 hipótesis.

Según las hipótesis que se establezcan sobre α_i y ϕ_t , las alternativas para estimar los modelos de panel son las tres siguientes:

1. Modelo de coeficientes constantes.
2. Modelo de efectos fijos.
3. Modelo de efectos aleatorios.

4.1 ESTIMACION DE MCO COMBINADOS

Partimos de la estimación más simple para los modelos de panel, donde se ignora la heterogeneidad no observable, es decir, los efectos individuales y temporales. No es la más adecuada para estimar conjuntos de panel, pero sí servirá en el análisis para comparar los resultados con las estimaciones por efectos fijos o aleatorios. Solamente se realiza una estimación MCO normal, donde se asume que los coeficientes son idénticos para cada unidad transversal. Algebraicamente se expresa como:

$$y_{it} = \beta_1 + X_{kit} * \beta_2 + u_{it}$$

De esta manera tenemos k parámetros a estimar, los cuales se consideran constantes dentro de todas las unidades espaciales y temporales de la muestra.

Como se comenta, no es el método de estimación ideal para paneles, ya que viola alguno de los supuestos mencionados, como homocedasticidad o no autocorrelación. Es fácil que entre las secciones cruzadas aparezca heterocedasticidad, ya que la varianza del error es diferente para cada una de ellas. También la autocorrelación, de manera que para un mismo periodo t aparezca correlación en los términos del error entre diferentes secciones cruzadas, o dentro de una misma aparezca esta correlación en diferentes periodos t.



4.2 ESTIMACION POR EFECTOS FIJOS

La estimación mediante efectos fijos parte del supuesto de que existen variaciones entre las diferentes secciones transversales, lo que implica la presencia de un coeficiente único para cada unidad individual que abarca dichos efectos individuales. Este método es especialmente ventajoso cuando se busca capturar y controlar las particularidades específicas de cada unidad en el análisis, crucial en contextos donde la heterogeneidad no observada puede tener un impacto significativo en las variables de interés. Además, al incluir efectos fijos, se puede mitigar el problema de la endogeneidad y sesgo potencial resultante de la omisión de variables no observadas.

En las estimaciones de efectos fijos tenemos la misma expresión general de los modelos de panel, como en el apartado anterior:

$$y_{it} = \beta_1 + X_{kit} * \beta_2 + u_{it}$$

En este caso, al contemplar los efectos individuales, el término del error se descompone:

$$u_{it} = \alpha_i + \varepsilon_{it}$$

Siendo:

$$y_{it} = \beta_1 + X_{kit} * \beta_2 + \alpha_i + \varepsilon_{it}$$

donde α_i es el componente específico de cada unidad transversal y constante en el tiempo y ε_{it} el error específico de cada observación. Por tanto, cada α_i se trata como un parámetro fijo que debe ser estimado. Para este fin existen tres principales procedimientos:

1. Método de los mínimos cuadrados de variables ficticias (MCFV) o “*least square dummy variables*” (LSDV). En este caso el procedimiento es incluir una variable ficticia por cada sección transversal y se elimina la constante global.
2. Método entre (o “*between*”). En este caso se resta la media de cada grupo para cada variable y se estima el modelo sin constante.

Estos dos métodos son matemáticamente equivalentes. Gretl, el software utilizado para este trabajo utiliza el segundo método.

3. Estimador de efectos fijos (o “*within*”): este enfoque se centra en eliminar las variaciones individuales constantes a lo largo del tiempo para cada unidad, enfocándose en las diferencias dentro de cada unidad a través del tiempo.

Un problema destacable de utilizar estimadores de efectos fijos es que las variables constantes en el tiempo se eliminan, mientras que en las estimaciones por efectos aleatorios que se explican a continuación si se capturan esas variables.



4.3 ESTIMACION POR EFECTOS ALEATORIOS

Para este tipo de estimadores volvemos otra vez a la expresión inicial:

$$y_{it} = \beta_1 + X_{kit} * \beta_2 + u_{it}$$

En este caso, el término del error se descompone como:

$$u_{it} = v_i + \varepsilon_{it}$$

A diferencia de la estimación por efectos fijos, el parámetro v_i no es fijo, si no que varía aleatoriamente entre las secciones siguiendo un cierto modelo de probabilidad.

Conceptualmente hablando, se considera que esta clase de modelos son adecuados cuando las N unidades transversales son una muestra aleatoria de una población mayor.

Como se menciona en el apartado 4.1, el método de MCO no es el estimador más eficiente para los modelos de panel, en cambio, para las estimaciones de efectos aleatorios se utiliza el método de los mínimos cuadrados generalizados (MCG), que son una extensión de los MCO relajando alguno de los supuestos, homocedasticidad y correlación no serial de los errores.

Para estimar por efectos fijos el software Gretl nos ofrece tres alternativas:

1. Método de Swamy y Arora (1972), que se utiliza en el contexto de un panel balanceado.
2. Variante de Baltagi y Chang (1994) del método de Swamy y Arora, en cambio, si tenemos un panel no balanceado, se recomienda esta vía, que modifica algunos supuestos para ajustarlo.
3. Método de Nerlove (1971): del que la literatura econométrica no ofrece orientación. Gretl utiliza el promedio ponderado de los efectos fijos como extensión natural del método original.

4.4 SELECCIÓN DE ESTIMADORES

Conceptualmente, la elección del método de estimación depende de la naturaleza del conjunto de datos: si el panel se compone de observaciones sobre un conjunto de individuos relativamente pequeño, en principio, lo más adecuado es la estimación por efectos fijos. En cambio, si se trata de observaciones de un conjunto mayor de individuos seleccionadas aleatoriamente, entonces optaríamos por una estimación de efectos aleatorios.

Sin embargo, para este fin, existen una serie de pruebas para los modelos de panel que nos permiten identificar la naturaleza de nuestros datos para poder elegir el método más adecuado. En primer lugar, partiendo de la estimación por efectos aleatorios, se realiza el contraste propuesto por Breusch y Pagan (1979), también llamada de Multiplicadores de Lagrange (LM). Analiza la existencia de un término de error específico para cada unidad transversal. Las hipótesis nula y alternativa son:

- $H_0: Var(\alpha_i) = 0$ Supone la nulidad del error específico
- $H_1: Var(\alpha_i) \neq 0$ Supone la existencia de un componente inobservable de la varianza que este asociado a cada individuo.



El contraste se realiza mediante una distribución chi cuadrado. Si no se rechaza la hipótesis nula, se concluye que no existen efectos específicos para cada individuo y, por tanto, la estimación “pool” MCO es la más adecuada. En cambio, si se rechaza la hipótesis nula, se acepta la existencia de efectos específicos para cada unidad transversal y la estimación “pool” MCO es sesgada. En este caso pasamos a la siguiente fase, el *test* de Hausman.

Éste supone decidir qué estimación, por efectos fijos o aleatorios, es la mejor. Sus hipótesis nula y alternativa se definen:

- $H_0: E(\alpha_i/x_{it}) = 0$ Supone que no existe una correlación sistemática entre los efectos individuales no observados y las variables explicativas.
- $H_1: E(\alpha_i/x_{it}) \neq 0$ Supone la existencia de una correlación sistemática entre los efectos individuales no observados y las variables explicativas.

Este contraste se realiza también mediante una distribución chi cuadrado. El no rechazo de la hipótesis nula equivale a admitir que los efectos aleatorios no están correlacionados con las variables explicativas y por tanto la estimación por efectos aleatorios es consistente y eficiente. Del mismo modo, rechazar la hipótesis nula implica admitir que existe correlación entre los efectos aleatorios y las variables explicativas, y se concluye que los efectos aleatorios pueden ser inconsistentes y es preferible la estimación por efectos fijos.

4.5 CONTRASTES DE DIAGNOSTICO

A parte de los contrastes de Breusch y Pagan, y Hausman para elegir entre los diferentes estimadores que ofrece la metodología de panel, Gretl ofrece otra serie de contrastes que también nos indican la necesidad o no de estimar a través de efectos fijos. También de validación de los modelos, para detectar la presencia de heterocedasticidad y de autocorrelación. Se presentan a continuación:

1. Contrastes para la validez del modelo:
Son el contraste de significación individual, en cuya hipótesis nula el coeficiente de la variable seleccionada es igual a 0, y en la alternativa distinto de 0, es decir, la variable tiene algún efecto sobre la dependiente. El estadístico de contraste es una t de Student con n-k grados de libertad.
Y el contraste conjunto de los regresores. Su hipótesis nula establece que los regresores conjuntos no son significativos. El rechazo, en favor la hipótesis alternativa, equivale a afirmar que existe algún regresor en el modelo que si tiene un efecto significativo para explicar la variable dependiente. El estadístico de contraste es una F de Snedecor.
2. Contrastes de efectos fijos:
Gretl, además de los nombrados en el apartado anterior, en las estimaciones por efectos fijos, nos proporciona las siguientes pruebas en relación con la necesidad de estimar por efectos fijos:
 - a. Contraste de diferentes interceptos por grupos:
Establece como hipótesis nula la existencia de un intercepto común para todos los individuos, su rechazo equivale a asumir que existen diferentes



interceptos por grupos y el modelo debe ser estimado a través de efectos fijos. Su estadístico de contraste es una F de Snedecor.

b. *Test F MCVF*:

Este es muy similar al anterior, contrasta también en su hipótesis nula la existencia de un intercepto común entre las secciones transversales. El estadístico de contraste es una F de Snedecor.

3. Contraste de Wald de heterocedasticidad:

Este es el contraste disponible en Gretl para contrastar la presencia de heterocedasticidad en las estimaciones de panel. Su hipótesis nula establece que las unidades tienen la misma varianza de la perturbación, es decir, la presencia de homocedasticidad. El rechazo en favor de la hipótesis alternativa concluye la existencia de heterocedasticidad. El estadístico de contraste es una chi-cuadrado.

4. Contraste de Woolridge de autocorrelación:

Este último contraste sirve para contrastar la existencia de autocorrelación de primer orden. La hipótesis nula es la no autocorrelación de primer orden y rechazarla establece la existencia de autocorrelación. El estadístico de contraste es una F de Snedecor.

5. ESTIMACION DE MODELOS

En esta sección se van a proponer y estimar, mediante las técnicas expuestas anteriormente, una serie de modelos que buscan capturar correlaciones significativas entre las variables del GEM: *Emprendedores* y *EmprendedoresPot* y las que reflejan el emprendimiento: *NuevasSoc* y *NuevosSSAut*. Para evitar posibles sesgos por el mayor peso poblacional de las comunidades con más habitantes, se incluye como variable de control *Pob18_64*, que se define como la población de 18 a 64 años, el rango de edad que utiliza el GEM para elaborar los índices, de cada comunidad y en cada periodo. El software econométrico para estimar los modelos es Gretl, de libre distribución, que es el que se ha utilizado en las asignaturas de estadística y econometría a lo largo del Grado en Economía.

Para elegir el tipo de estimación: MCO combinados, efectos fijos o aleatorios, nos guiaremos por los resultados de los contrastes de Breusch-Pagan y Hausman siguiendo el esquema de la figura 5.1, extraída del libro “Econometría aplicada con Gretl”, (Pérez y López, 2019).

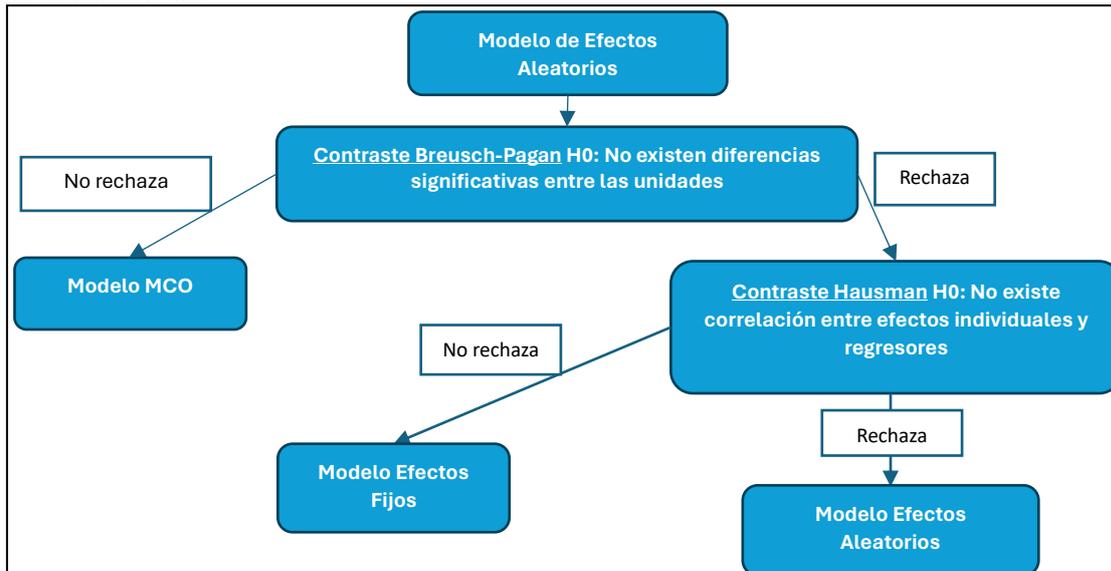


Figura 5.1: Metodología de panel

Fuente: (Pérez y López, 2019).

En los sucesivos apartados, se les aplicarán transformaciones a los modelos con la intención de mejorar la capacidad explicativa y validez de los modelos, así como abordar los problemas de heterocedasticidad y autocorrelación que puedan surgir. En el apartado 5.2 se realizarán las estimaciones con las variables explicativas en primeras diferencias, en el 5.3 se les aplicará un retardo a estas primeras diferencias y, finalmente, en el 5.4 se aplicarán logaritmos a las variables explicadas y explicativas de los modelos inicialmente propuestos.

5.1 MODELOS INICIALES

Se proponen los siguientes modelos, cuyos resultados e indicadores se presentan en la tabla 5.1.1:

- **Modelo 1.1:** $NuevasSoc_{it} = \beta_1 + Emprendedores_{it} * \beta_2 + Pob18_{64} * \beta_3 + u_{it}$
- **Modelo 2.1:** $NuevosSSAut_{it} = \beta_1 + Emprendedores_{it} * \beta_2 + Pob18_{64} * \beta_3 + u_{it}$
- **Modelo 3.1:** $NuevasSoc_{it} = \beta_1 + EmprendedoresPot_{it} * \beta_2 + Pob18_{64} * \beta_3 + u_{it}$
- **Modelo 4.1:** $NuevosSSAut_{it} = \beta_1 + EmprendedoresPot_{it} * \beta_2 + Pob18_{64} * \beta_3 + u_{it}$

Los modelos, tanto estos, como todos los sucesivos, se han estimado teniendo en cuenta desviaciones típicas robustas mediante el método de Arellano que es el más adecuado de los que nos ofrece Gretl para los modelos planteados (Arellano, 2003). Esto se ha hecho por la presencia de heterocedasticidad y autocorrelación en los datos, lo que puede afectar a la validez de los estimadores, con este método de estimación robusta se pretende tratar este problema.

En primer lugar, siguiendo el procedimiento descrito en la figura 5.1, debemos fijarnos en los resultados de los tests de Breusch-Pagan y Hausman. En ambos, los valores críticos son prácticamente 0, para los modelos 1.1, 3.1 y 4.1, por lo que podemos



rechazar las hipótesis nulas, concluyendo que existen diferencias significativas entre las unidades y no hay correlación entre los efectos individuales y los regresores, con lo cual todos los modelos se han estimado por efectos fijos. Para el modelo 2.1, sin embargo, el *test* de Hausman tiene un valor crítico lo suficientemente alto como para rechazar la hipótesis nula y concluir que es conveniente su estimación por efectos aleatorios.

Los modelos 3.1 y 4.1 no son coherentes con nuestra hipótesis de trabajo, la variable *EmprendedoresPot* tiene asociado en ambos un coeficiente negativo para explicar *NuevasSoc* y *NuevosSSAut*, cuando cabría esperar que a mayor número de emprendedores potenciales hubiera un mayor número de sociedades y/o autónomos. Además, en el primero, la variable no es estadísticamente significativa para explicar *NuevasSoc*.

Variable	Modelo 1.1	Modelo 2.1	Modelo 3.1	Modelo 4.1
const	33748,90*** (4652,07)	21933,30** (9399,03)	34543,60*** (5289,44)	26753,10 (70935,80)
Emprendedores	0,02*** (0,005)	0,09* (0,05)	-	-
EmprendedoresPot	-	-	-0,002 (0,003)	-0,06*** (0,01)
Poblacion18_64	-0,02*** (0,002)	0,09*** (0,09)	-0,02*** (0,003)	0,09** (0,04)
R2 MCVF	0,97	-	0,96	0,99
R2 “intra”	0,50	-	0,38	0,18
Breusch-Pagan (Chi-cuadrado)	363,08***	1731,90***	450,16***	1716,46***
Conclusión	Rechaza H0	Rechaza H0	Rechaza H0	Rechaza H0
Hausman (Chi-cuadrado)	64,03***	0,78	43,63***	52,43***
Conclusión	Rechaza H0	No rechaza H0	Rechaza H0	Rechaza H0
Conjunto de regresores (F Snedecor)	34,91***	484,05***	26,18***	10,32***
Conclusión	Rechaza H0	Rechaza H0	Rechaza H0	Rechaza H0
Diferentes interceptos por grupos (F Snedecor)	24,81***	-	35,25***	139,32***
Conclusión	Rechaza H0	-	Rechaza H0	Rechaza H0
Heterocedasticidad de Wald (Chi-cuadrado)	553293***	-	78839,60***	585514***
Conclusión	Rechaza H0	-	Rechaza H0	Rechaza H0
Autocorrelación de Wooldridge (F Snedecor)	133,80***	91,05***	177,53***	77,09***
Conclusión	Rechaza H0	Rechaza H0	Rechaza H0	Rechaza H0
Individuos	17	17	17	17
Periodos	16-19	16-19	16-19	16-19
Observaciones	295	295	293	293
Notas: i) Los * se refieren al nivel de significación de los contrastes individuales, *** 1%, ** 5% y *1%, ii) Debajo de cada contraste figura la conclusión, rechazo o no rechazo de la hipótesis nula, iii) El estadístico en el c. conjunto de regresores en efectos aleatorios es una Chi-cuadrado.				

Tabla 5.1.1: Modelos iniciales.

Fuente: elaboración propia.



Los dos modelos iniciales si presentan una relación positiva entre *Emprendedores* y la variable dependiente. En el 2.1, sin embargo, la variable no es significativa para explicar *NuevosSSAut*, cuyo nivel crítico asociado está por encima de 0,05, que se establece como el nivel de significación límite para rechazar la hipótesis nula, por tanto, tampoco ofrece garantías de cara a validar nuestra hipótesis. El 1.1 sí es estadísticamente relevante para explicar *NuevasSoc*. El coeficiente asociado a *Emprendedores* interpreta como que el aumento en una unidad de los emprendedores estimaos supone un crecimiento de 0,02 en *NuevasSoc*, *ceteris paribus* el resto de variables. En cambio, la variable de control *Pob18_64*, tiene un coeficiente negativo, -0,02, lo que se traduce como la disminución en 0,02 nuevas sociedades por cada unidad que aumenta la población de 18 a 64 años, *ceteris paribus* el resto de variables. Esto no tiene mucho sentido en nuestro planteamiento, pero hay que coger este resultado con precaución por la simplicidad de los modelos utilizados.

En cuanto al coeficiente de determinación lineal, en el caso de los modelos de efectos fijos, los resultados que nos arroja Gretl son diferentes a los resultantes de las estimaciones por MCO combinados. En la salida, el programa da dos medidas: R^2 MCVF (mínimos cuadrados de variables ficticias) y R^2 "intra". El primero, indica qué parte de la variación total en la variable dependiente se explica por todas las variables en el modelo, incluyendo las variables ficticias por individuo. El segundo, representa la proporción de la variación en la variable dependiente que es explicada por las variables dentro de cada individuo a lo largo del tiempo.

En este caso, el modelo 1.1 arroja un 0,50 aproximadamente en el R^2 "intra". Esto se puede traducir en que el 50% de las variaciones de *NuevasSoc* son explicadas por la variable *Emprendedores* y *Pob18_64* dentro de cada individuo a lo largo del tiempo. Muy distinto es el valor del R^2 MCVF, que es de 0,97 aproximadamente. Por tanto, se puede afirmar que las variables *Emprendedores* y *Pob18_64*, junto con las variables *dummy* de las comunidades, explican el 97% de las variaciones de *NuevasSoc* entre todas las comunidades y a lo largo del tiempo.

Esta disparidad entre los coeficientes se debe a la manera en la que se plantean los modelos. La gran diferencia se debe, probablemente a las características propias (efectos fijos) de cada comunidad, que afectan significativamente a la variable explicada. Al incluir las estimaciones de 17 variables ficticias, se captura esa variación, que parece elevar el resultado del coeficiente. Sin embargo, a nivel temporal dentro de cada sección transversal, las variables independientes explican una menor variación. Esta situación se va a dar casi en la totalidad de los modelos planteados, con R^2 MCVF por encima de 0,9 y R^2 "intra" bastante más bajos, en torno al 0,1.

En el resultado del contraste de diferentes interceptos por grupos, se rechaza la hipótesis nula al ser el p-valor lo suficientemente bajo. Esto se interpreta como la no existencia de un intercepto común para todos los individuos, lo que refuerza la decisión de estimar el modelo por efectos fijos. Se detecta también, mediante las pruebas de heterocedasticidad de Wald y de autocorrelación de Woolridge, la presencia de heterocedasticidad y autocorrelación de primer orden, cuyos efectos sobre los estimadores se abordan mediante la estimación robusta.



5.2 MODELOS EN DIFERENCIAS

En este apartado se modifican los 4 modelos anteriores y se plantean con la introducción de primeras diferencias en las variables *Emprendedores* y *EmprendedoresPot*. Es decir, se considera el incremento anual de emprendedores y emprendedores potenciales, con lo que las nuevas variables del modelo serán: *NuevosEmp* y *NuevosEmpPot*. La intención de esta transformación es poder estimar estas variables dándoles la misma unidad de medida que las variables explicadas, lo que nos permitiría interpretar los coeficientes de estas variables como el aumento en nuevas sociedades mercantiles constituidas o nuevos afiliados al epígrafe RETA cuando se incrementan en una unidad los emprendedores o emprendedores potenciales que estima el GEM. Así, los modelos a estimar en este apartado son:

- Modelo 1.2: $NuevasSoc_{it} = \beta_1 + NuevosEmp_{it} * \beta_2 + Pob18_{64} * \beta_3 + u_{it}$
- Modelo 2.2: $NuevosSSAut_{it} = \beta_1 + NuevosEmp_{it} * \beta_2 + Pob18_{64} * \beta_3 + u_{it}$
- Modelo 3.2: $NuevasSoc_{it} = \beta_1 + NuevosEmpPot_{it} * \beta_2 + Pob18_{64} * \beta_3 + u_{it}$
- Modelo 4.2: $NuevosSSAut_{it} = \beta_1 + NuevosEmpPot_{it} * \beta_2 + Pob18_{64} * \beta_3 + u_{it}$

En la tabla 5.2.1 vemos los resultados para los 4 modelos planteados. De ellos, los 1.2 y 3.2, en base a los resultados del *test* de Hausman, se han estimado por efectos fijos, mientras que los 2.2 y 4.2 por efectos aleatorios. Destaca, que, de nuevo, sólo es estadísticamente significativo el modelo que relaciona a los emprendedores estimados con las nuevas sociedades, el 1.2.

Este modelo 1.2 nos dice que, por cada nuevo emprendedor, *NuevosEmp*, se constituyen 0,01 nuevas sociedades, *NuevasSoc*, *ceteris paribus* el resto de variables. Tanto individualmente, como en su conjunto, todas las variables del modelo son significativas, añadir que de nuevo la variable de control de la población nos arroja un coeficiente negativo, poco coherente con el planteamiento como se ha señalado anteriormente. Los coeficientes de significación arrojan resultados muy cercanos a los del modelo 1.1, el R^2 "intra" es un 20% menor aproximadamente, de 0,40. Se detecta presencia de heterocedasticidad y autocorrelación de primer orden.



Variable	Modelo 1.2	Modelo 2.2	Modelo 3.2	Modelo 4.2
const	36679,40*** (8051,12)	25291,20*** (8356,82)	39430,90*** (7596,51)	23983,80*** (8039,96)
Emprendedores	0,01*** (0,002)	0,005 (0,03)	-	-
EmprendedoresPot	-	-	-0,002 (0,002)	0,005 (0,007)
Poblacion18_64	-0,02*** (0,004)	0,09*** (0,005)	-0,02*** (0,004)	0,09*** (0,005)
R2 MCVF	0,96	-	0,96	-
R2 "intra"	0,40	-	0,36	-
Breusch-Pagan (Chi-cuadrado)	390,19***	1611,98***	364,99***	1624,94***
Conclusión	Rechaza H0	Rechaza H0	Rechaza H0	Rechaza H0
Hausman (Chi-cuadrado)	54,69***	1,78	28,87***	5,07*
Conclusión	Rechaza H0	No rechaza H0	Rechaza H0	No rechaza H0
Conjunto de regresores (F Snedecor)	52,54***	376,85	12,07***	547,45***
Conclusión	Rechaza H0	No rechaza H0	Rechaza H0	Rechaza H0
Diferentes interceptos por grupos (F Snedecor)	30,97***	-	36,61***	-
Conclusión	Rechaza H0	-	Rechaza H0	-
Heterocedasticidad de Wald (Chi-cuadrado)	101609***	-	73432,90***	-
Conclusión	Rechaza H0	-	Rechaza H0	-
Autocorrelación de Wooldridge (F Snedecor)	280,97***	72,72***	171,06***	79,09***
Conclusión	Rechaza H0	Rechaza H0	Rechaza H0	Rechaza H0
Individuos	17	17	17	17
Periodos	15-18	15-18	15-18	15-18
Observaciones	278	278	276	276

Notas: i) Los * se refieren al nivel de significación de los contrastes individuales, *** 1%, ** 5% y *1%, ii) Debajo de cada contraste figura la conclusión, rechazo o no rechazo de la hipótesis nula, iii) El estadístico en el c. conjunto de regresores en efectos aleatorios es una Chi-cuadrado.

Tabla 5.2.1: Modelos en diferencias.

Fuente: elaboración propia.

5.3 MODELOS CON RETARDOS

Los datos que utiliza el GEM para confeccionar los índices referentes al proceso emprendedor se recogen mediante la encuesta APS, la cual se realiza generalmente en los meses centrales del año. Por ejemplo, la muestra utilizada en el informe GEM España 2021-2022 es de 31.785 individuos y se recogió entre los meses de julio y septiembre de 2021, Laviada, A. F., y otros (2022). Además, la *TEP* (emprendedores potenciales) se calcula con la proporción de adultos (18 a 64 años) que han expresado la intención de emprender *en los próximos 3 años*. Con el fin de capturar los posibles rezagos en las variables *Emprendedores* y *EmprendedoresPot*, se ha decidido aplicar un retardo en ellas y analizar los resultados. Los modelos planteados son:

- **Modelo 1.3:** $NuevasSoc_{it} = \beta_1 + NuevosEmp_{it-1} * \beta_2 + Pob18_{64} * \beta_3 + u_{it}$
- **Modelo 2.3:** $NuevosSSAut_{it} = \beta_1 + NuevosEmp_{it-1} * \beta_2 + Pob18_{64} * \beta_3 + u_{it}$
- **Modelo 3.3:** $NuevasSoc_{it} = \beta_1 + NuevosEmpPot_{it-1} * \beta_2 + Pob18_{64} * \beta_3 + u_{it}$



- **Modelo 4.3:** $NuevosSSAut_{it} = \beta_1 + NuevosEmpPot_{it-1} * \beta_2 + Pobl18_{64} * \beta_3 + u_{it}$

Los retardos se han introducido sobre las variables en diferencias, *NuevosEmp* y *NuevosEmpPot*, que ofrecen una mejor interpretación y estadísticamente es más adecuado expresar las variables explicativas en los mismos términos que las variables endógenas. Los resultados se muestran en la tabla 5.3.1.

Variable	Modelo 1.3	Modelo 2.3	Modelo 3.3	Modelo 4.3
const	37180,50*** (12608,80)	26563,30*** (8811,62)	40508,20*** (11843,50)	147837** (54716,70)
Emprendedores	0,007** (0,003)	0,03 (0,03)	-	-
EmprendedoresPot	-	-	0,002 (0,001)	-0,03*** (0,009)
Poblacion18_64	-0,02** (0,007)	0,09*** (0,005)	-0,02** (0,01)	0,03 (0,03)
R2 MCVF	0,96	-	0,96	0,99
R2 "intra"	0,30	-	0,30	0,03
Breusch-Pagan (Chi-cuadrado)	391,95***	1428,39***	380,67***	1450,44***
Conclusión	Rechaza H0	Rechaza H0	Rechaza H0	Rechaza H0
Hausman (Chi-cuadrado)	12,98***	6,22*	13,18***	6,81**
Conclusión	Rechaza H0	No rechaza H0	Rechaza H0	Rechaza H0
Conjunto de regresores (F Snedecor)	55,33***	263,84***	5,45**	18,10***
Conclusión	Rechaza H0	Rechaza H0	Rechaza H0	Rechaza H0
Diferentes interceptos por grupos (F Snedecor)	53,75***	-	63,90***	215,57***
Conclusión	Rechaza H0	-	Rechaza H0	Rechaza H0
Heterocedasticidad de Wald (Chi-cuadrado)	209608***	-	343413***	731781***
Conclusión	Rechaza H0	-	Rechaza H0	Rechaza H0
Autocorrelación de Wooldridge (F Snedecor)	267,05***	71,46***	77,95***	62,77***
Conclusión	Rechaza H0	Rechaza H0	Rechaza H0	Rechaza H0
Individuos	17	17	17	17
Periodos	14-17	14-17	14-17	14-17
Observaciones	261	261	259	259
Notas: i) Los * se refieren al nivel de significación de los contrastes individuales, *** 1%, ** 5% y *1%, ii) Debajo de cada contraste figura la conclusión, rechazo o no rechazo de la hipótesis nula, iii) El estadístico en el c. conjunto de regresores en efectos aleatorios es una Chi-cuadrado.				

Tabla 5.3.1: Modelos con retardos.

Fuente: elaboración propia



Los modelos 1.3, 3.3 y 4.3, a la vista de los resultados de los *tests* de Breusch-Pagan y Hausman, se han estimado por efectos fijos, mientras que el 2.3 tiene un nivel crítico lo suficientemente grande como para rechazar la hipótesis nula en Hausman y concluir que es preferible estimar por efectos aleatorios. Los resultados no difieren en términos generales de los planteados previamente.

Centrándonos en el modelo 1.3, que *a priori* parece el más apropiado tal y como hemos comentado anteriormente, vemos que las variables son relevantes en el modelo. El efecto de *NuevosEmp_1* sobre *NuevasSoc* es de 0,007, algo menos que en los anteriores modelos. En este modelo, como en los anteriores, volvemos a obtener un coeficiente negativo de la variable de control, *Pob18_64*. El coeficiente de determinación lineal R^2 "intra" es menor que en los anteriores, 0,30, por lo que las variables *NuevosEmp_1* y *Pob18_64* explican conjuntamente menos variaciones de *NuevasSoc* que los modelos 1.1 y 1.2. El R^2 MCVF sigue en niveles muy cercanos, 0,96, lo que se explica, principalmente, porque considera las variables ficticias de los individuos.

Los indicadores habituales, *test* de Wald y Woolridge, detectan la presencia de heterocedasticidad y autocorrelación de primer orden.

Cabe añadir, que también se ha probado a introducir un segundo y tercer retardo en las variables sin obtener los resultados esperados.

5.4 MODELOS EN LOGARITMOS

Por último, se van a transformar las variables a ambos lados de la ecuación en logaritmos. Esto se fundamenta en algunas ventajas que ofrece esta transformación: una es la interpretación, que coincide con el de la elasticidad, es el aumento porcentual en la variable dependiente cuando la independiente aumenta en un punto porcentual, *ceteris paribus* el resto de variables. Otra, es la mejora en la satisfacción de los supuestos del modelo lineal clásico (MLC), entre los que se encuentra la homocedasticidad, en los modelos en los que se aplica logaritmo en la variable endógena, según señala Wooldridge, J. M. (2006). Las variables transformadas en logaritmo llevarán el prefijo "l_", los modelos a estimar:

- Modelo 1.4: $l_NuevasSoc_{it} = \beta_1 + l_Emprendedores_{it} * \beta_2 + Pob18_{64} * \beta_3 + u_{it}$
- Modelo 2.4: $l_NuevosSSAut_{it} = \beta_1 + l_Emprendedores_{it} * \beta_2 + Pob18_{64} * \beta_3 + u_{it}$
- Modelo 3.4: $l_NuevasSoc_{it} = \beta_1 + l_EmprendedoresPot_{it} * \beta_2 + Pob18_{64} * \beta_3 + u_{it}$
- Modelo 4.4: $l_NuevosSSAut_{it} = \beta_1 + l_EmprendedoresPot_{it} * \beta_2 + Pob18_{64} * \beta_3 + u_{it}$



Variable	Modelo 1.4	Modelo 2.4	Modelo 3.4	Modelo 4.4
const	5,22*** (0,67)	11,72*** (0,26)	10,40*** (0,56)	11,58*** (0,22)
Emprendedores	0,41*** (0,05)	-0,02 (0,02)	-	-
EmprendedoresPot	-	-	-0,07 (0,05)	-0,02 (0,02)
Poblacion18_64	-8,69*e-7*** (2,63*e7)	2,01*e-7** (7,76*e-8)	-7,88*e-7*** (2,62*e-7)	2,24*e-7** (7,83*e-8)
R2 MCVF	0,97	0,99	0,96	0,99
R2 "intra"	0,29	0,08	0,08	0,08
Breusch-Pagan (Chi-cuadrado)	416,85***	806,07***	241,06***	1192,70***
Conclusión	Rechaza H0	Rechaza H0	Rechaza H0	Rechaza H0
Hausman (Chi-cuadrado)	38,65***	103,75***	130,65***	1192,70***
Conclusión	Rechaza H0	Rechaza H0	Rechaza H0	Rechaza H0
Conjunto de regresores (F Snedecor)	42,54***	3,94***	8,90***	4,23**
Conclusión	Rechaza H0	Rechaza H0	Rechaza H0	Rechaza H0
Diferentes interceptos por grupos (F Snedecor)	13,34***	27,32**	8,94***	43,39***
Conclusión	Rechaza H0	Rechaza H0	Rechaza H0	Rechaza H0
Heterocedasticidad de Wald (Chi-cuadrado)	324,39***	260,45***	145,37***	310,96***
Conclusión	Rechaza H0	Rechaza H0	Rechaza H0	Rechaza H0
Autocorrelación de Wooldridge (F Snedecor)	55,21***	191,22***	132,46***	158,89***
Conclusión	Rechaza H0	Rechaza H0	Rechaza H0	Rechaza H0
Individuos	17	17	17	17
Periodos	16-19	16-19	16-19	16-19
Observaciones	295	295	293	293
Notas: i) Los * se refieren al nivel de significación de los contrastes individuales, *** 1%, ** 5% y *1%, ii) Debajo de cada contraste figura la conclusión, rechazo o no rechazo de la hipótesis nula, iii) El estadístico en el c. conjunto de regresores en efectos aleatorios es una Chi-cuadrado.				

Tabla 5.4.1: Modelos en logaritmos.

Fuente: elaboración propia

Los resultados de las estimaciones, que vemos en la tabla 5.4.1, muestran la necesidad de estimar los 4 modelos por efectos fijos (en todos los tests de Breusch-Pagan y Hausman proporcionan niveles críticos muy bajos, con lo que se rechazan las hipótesis nulas). Sin embargo, nuevamente el único modelo cuya variable explicativa es estadísticamente significativa y con efecto positivo es el 1.4, que relaciona el logaritmo de *Emprendedores* con el de *NuevasSoc*. El efecto es de 0,41, lo que se traduce en que el aumento de un 1% en emprendedores estimados provoca un crecimiento del 0,41% en las nuevas sociedades constituidas, *ceteris paribus* el resto de variables. Aunque con un efecto insignificante, la variable *Pob18_64* tiene en el modelo también un coeficiente negativo. Todas las variables y el modelo en son conjunto son significativos. El R^2 "intra" es casi el mismo que en el modelo 1.3, de 0,29 en este caso y el R^2 MCVF se mantiene en 0,97. Al igual que en el resto de modelos, se detecta heterocedasticidad y autocorrelación de primer orden.



6. ANÁLISIS DE RESULTADOS

En el capítulo 5 se han planteado, estimado y analizado una serie de modelos, inicialmente se plantearon 4 que posteriormente se han ido transformando y valorando los resultados con el fin de encontrar una relación clara entre las variables *Emprendedores* y *EmprendedoresPot* con *NuevasSoc* y *NuevosSSAut*. Finalmente, se han validado 4 modelos, todos ellos relacionan las nuevas sociedades mercantiles constituidas, *NuevasSoc*, con el número de emprendedores estimados por el GEM, *Emprendedores*, y las diferentes transformaciones que se han aplicado sobre estas variables. No se encuentran relaciones significativas con la afiliación al epígrafe RETA, *NuevosSSAut* y tampoco entre los emprendedores potenciales, *EmprendedoresPot*, o alguna de sus transformaciones nuevos autónomos o sociedades, *NuevosSSAut* y *NuevasSoc*.

Los modelos estadísticamente significativos y coherentes con nuestra hipótesis son los 4 siguientes:

- Modelo 1.1: $NuevasSoc_{it} = \beta_1 + Emprendedores_{it} * \beta_2 + Pob18_{64} * \beta_3 + u_{it}$
- Modelo 1.2: $NuevasSoc_{it} = \beta_1 + NuevosEmp_{it} * \beta_2 + Pob18_{64} * \beta_3 + u_{it}$
- Modelo 1.3: $NuevasSoc_{it} = \beta_1 + NuevosEmp_{it-1} * \beta_2 + Pob18_{64} * \beta_3 + u_{it}$
- Modelo 1.4: $l_NuevasSoc_{it} = \beta_1 + l_Emprendedores_{it} * \beta_2 + Pob18_{64} * \beta_3 + u_{it}$

Que nos permiten afirmar que existe una relación positiva entre los niveles de emprendimiento estimados y la creación de sociedades mercantiles. Aunque los coeficientes son pequeños, podemos interpretar el del modelo 1.1, 0,02, como el número de nuevas sociedades creadas por cada emprendedor más estimado. El del modelo 2.1, 0,01, como el número de nuevas sociedades por cada nuevo emprendedor estimado, de igual forma el 3.1, que con un retardo nos da un efecto algo menor, 0,007. Y el 4.1, que se interpreta como la elasticidad, nos dice que por cada 1% más de emprendedores estimados hay un 0,41% más de sociedades.

A continuación, se estimarán y analizarán los efectos individuales y temporales que contienen estos modelos.

6.1 EFECTOS INDIVIDUALES

En la tabla 6.1.1 se muestran los resultados de efectos fijos mostrando los efectos individuales de cada comunidad. Estos se han calculado estimando por MCO el modelo incluyendo 17 variables *dummy*, una por comunidad, excluyendo la Comunidad de Madrid, cuyo efecto queda contenido en la constante y a partir de él se interpretan los efectos del resto de comunidades, los positivos están por encima de la de referencia y los negativos por debajo. El efecto neto se obtendría sumando el efecto individual de la comunidad con la constante.



	Modelo 1.1	Modelo 1.2	Modelo 1.3	Modelo 2.1
const	84170,20*** (11093,80)	91219,50*** (19165)	92265,30*** (30008,20)	8,57*** (1,25)
Emprendedores	0,02*** (0,01)	-	-	0,41*** (0,05)
NuevosEmp	-	0,01*** (0,001)	-	-
NuevosEmp_1	-	-	0,007** (0,003)	-
l_Emprendedores	-	-	-	-
Pob_18_64	-0,016*** (0,002)	-0,02*** (0,004)	-0,02** (0,007)	-8,69*e-7*** (2,71*e-7)
PAÍS VASCO	-59146,50*** (7460,82)	-64534,80*** (12869,60)	-65350,60*** (20190)	-3,68*** (0,77)
CATALUÑA	6051,11*** (1531,03)	7997,02*** (2560,89)	8184,71* (4042,97)	0,31* (0,15)
GALICIA	-53194,20*** (6571,14)	-57809*** (11356)	-58551,30*** (17836,10)	-3,27*** (0,98)
ANDALUCÍA	13025,60*** (2898,13)	14760,90*** (5051,23)	15118,30* (7964,87)	0,68** (0,30)
ASTURIAS	-72292,40*** (9343,30)	-78697,80*** (16128)	-79670,40*** (25284,40)	-4,86*** (0,97)
CANTABRIA	-77657,90*** (10106,90)	-84162,70*** (17462,20)	-85150,90*** (27351,20)	-5,61*** (1,05)
LA RIOJA	-80612,50*** (10567,60)	-87402,70*** (18253,70)	-88426,10*** (28586)	-5,99*** (1,10)
R. DE MURCIA	-66934,00*** (8630,39)	-72716,80*** (14894,10)	-73616,10*** (23323,90)	-4,20*** (0,89)
C. VALENCIANA	-23323,00*** (2699,18)	-25834,10*** (4582,72)	-26266*** (7184,29)	-1,27*** (0,28)
ARAGÓN	-69260,80*** (8915,79)	-75206,80*** (15392,60)	-76166,90*** (24120)	-4,51*** (0,93)
CAST. LA MANCHA	-61310,40*** (7690,76)	-66391,20*** (13270,10)	-67240,30*** (20775,20)	-3,95*** (0,80)
I. CANARIAS	-58613,40*** (7368,01)	-63659,20*** (12690,30)	-64374*** (19862,70)	-3,65*** (0,77)
NAVARRA	-76985,20*** (10040)	-83498,10*** (17336,40)	-84475,90*** (27147,40)	-5,37*** (1,04)
EXTREMADURA	-72521,60*** (9306,20)	-78478,10*** (16075,70)	-79760,60*** (25176,80)	-5,09*** (0,97)
I. BALEARES	-69943,30*** (9148,17)	-75863*** (15791,70)	-76760,60*** (24715,90)	-4,27*** (0,95)
CAST. Y LEON	-57316,70*** (7051,61)	-62209,90*** (12183,10)	-62984,20*** (19120,10)	-3,73*** (0,73)
R2/R2CORR	0,97/0,96	0,96/0,96	0,96/0,95	0,99
Número individuos	17	17	17	17
Número años	16-19	15-18	14-17	16-19
Observaciones	295	278	261	295

Notas: i) Los * se refieren al nivel de significación de los contrastes individuales, *** 1%, ** 5% y *1%, ii) Debajo de cada contraste figura la conclusión, rechazo o no rechazo de la hipótesis nula.

Tabla 6.1.1: Efectos individuales.

Fuente: elaboración propia

A la vista de los resultados se ve que todas las comunidades tienen un efecto significativo para explicar las variables dependientes, que son las nuevas sociedades. A esto encontramos la excepción de Cataluña para el modelo 1.4, cuyo nivel crítico



asociado al coeficiente está por encima de 0,05, por lo que el efecto individual de esta comunidad en ese modelo lo consideraremos estadísticamente no significativo.

Se sospecha de la presencia de correlación entre las variables de comunidad y la variable de población, *Pob18_64*. Es decir, suponemos que, a parte de las políticas de apoyo al emprendimiento u otros factores que expliquen las diferencias entre comunidades, los efectos individuales también contienen la población. Para verificarlo se van a crear 4 variables nuevas que contengan los efectos individuales de cada modelo y otra que contenga la media de población en los 19 años para cada comunidad:

- ***EF_Mod1.1***: Contiene los efectos individuales de cada comunidad, exceptuando Madrid que es la de referencia, del modelo 1.1.
- ***EF_Mod1.2***: Contiene los efectos individuales de cada comunidad, exceptuando Madrid que es la de referencia, del modelo 2.1.
- ***EF_Mod1.3***: Contiene los efectos individuales de cada comunidad, exceptuando Madrid que es la de referencia, del modelo 3.1.
- ***EF_Mod1.4***: Contiene los efectos individuales de cada comunidad, exceptuando Madrid que es la de referencia, del modelo 4.1.
- ***POB***: Media de población de 18 a 64 años de cada comunidad en los 19 periodos.

Una vez tenemos las variables, se ha estimado en Gretl una matriz de correlación con estas 5 variables. La correlación es perfecta entre los efectos individuales y la población, por lo que se confirma la presencia de la población dentro de los efectos individuales. Para solucionar este sesgo en la estimación de efectos individuales se propone una regresión auxiliar que controle por la población los efectos individuales y cuyos residuos representen las diferencias entre comunidades en cuanto a efectos de políticas de apoyo al emprendimiento. Por lo que se proponen las siguientes regresiones MCO, cuyos resultados se muestran en la tabla 6.1.2 y sus residuos en la 6.1.3:

- $EF_Mod1.1_{it} = \beta_1 + POB_{it} + u_{it}$
- $EF_Mod2.1_{it} = \beta_1 + POB_{it} + u_{it}$
- $EF_Mod3.1_{it} = \beta_1 + POB_{it} + u_{it}$
- $EF_Mod4.1_{it} = \beta_1 + POB_{it} + u_{it}$

	EF_Mod1.1	EF_Mod2.1	EF_Mod3.1	EF_Mod4.1
cons	-84715,90*** (311,824)	-92117*** (397,91)	-93248,6*** (402,311)	-5,67*** (0,13)
POB	0,018*** (0,00026)	0,02*** (0,0003)	0,02*** (0,0003)	1,24*e-6*** (5,76*e-8)
R2/R2CORR	0,99/0,99	0,99/0,99	0,99/0,99	0,98/0,97

Notas: i) Los * se refieren al nivel de significación de los contrastes individuales, *** 1%, ** 5% y *1%.

Tabla 6.1.2: Resultados de la regresión auxiliar.

Fuente: elaboración propia.



RESIDUOS	EF_Mod1.1	EF_Mod2.1	EF_Mod3.1	EF_Mod4.1
PAÍS VASCO	-443,70	-879,27	-932,52	0,22
CATALUÑA	1536,82	2485,40	2538,99	-0,07
GALICIA	-730,10	-979,36	-1047,60	0,21
ANDALUCÍA	-1690,30	-1912,47	-1833,87	-0,39
ASTURIAS	-229,50	-424,70	-445,25	-0,05
CANTABRIA	66,90	305,22	349,37	-0,42
LA RIOJA	376,00	636,14	691,38	-0,58
R. DE MURCIA	504,00	496,15	483,30	0,30
C. VALENCIANA	1754,10	1030,82	884,53	0,36
ARAGÓN	28,60	31,81	-15,58	0,12
CAST. LA MANCHA	-344,30	-259,38	-313,83	0,11
I. CANARIAS	-343,40	-477,24	-435,62	0,22
NAVARRA	253,50	437,98	485,64	-0,21
EXTREMADURA	-523,00	-275,31	-183,07	-0,29
I. BALEARES	1210,20	1415,09	1456,64	0,48
CAST. Y LEON	-1425,80	-1630,88	-1682,51	-0,02
OBSERVACIONES	1-17	1-17	1-17	1-17
POBLACIÓN	16	16	16	16

Tabla 6.1.3: Residuos de la regresión auxiliar.

Fuente: elaboración propia.

Estos residuos, por tanto, se interpretan como las diferencias en cuanto a políticas de emprendimiento tomando la Comunidad de Madrid como referencia. Se observa en los resultados que los modelos 1.1, 2.1 y 3.1 son similares en cuanto a las comunidades que tienen más peso y las que menos, con ligeras diferencias. Estas diferencias, aumentan lógicamente cuando asumimos relaciones no lineales (modelo 1.4). Por simplicidad comentaremos los de estos modelos. No obstante, hay que tener en cuenta que la metodología econométrica utilizada en la investigación tiene limitaciones y estas clasificaciones se deben interpretar con cautela. Por lo tanto, se va a intentar explicar estas diferencias haciendo uso de distintos indicadores socioeconómicos que puedan estar relacionados con el fenómeno emprendedor y nos permitan contrastar los resultados y darles un sentido.¹

Entre las que tienen un mayor efecto están las comunidades del Levante español: Cataluña, Comunidad Valenciana, Islas Baleares y Región de Murcia, junto con La Rioja, dependiendo del modelo.

Cataluña es una región tradicionalmente industrial y emprendedora. La ciudad de Barcelona es un núcleo emprendedor a nivel internacional, con instituciones como la “Fira de Barcelona”, la feria internacional más grande de España. También la Universidad de Barcelona y la Autónoma de Barcelona, polos de emprendimiento. También, dentro de la Comunidad Valenciana, la ciudad de Valencia es un polo de emprendimiento que en 2021 albergó 1.012 “start-ups”, un 15% más respecto al año anterior, según el Observatorio Startup de la Comunidad Valenciana.

Estas dos comunidades tienen también en común un peso relativo, junto con la Región de Murcia y La Rioja, del sector industrial, que, aunque ha experimentado cierto declive en las últimas décadas, sigue teniendo en estas comunidades un peso sobre el empleo en 2021 del 13% al 15%, salvo La Rioja, que llega al 25%, lo que podría explicar su posición. El sector industrial es normalmente potenciador del emprendimiento, tiene unas externalidades positivas como la transferencia de tecnología o las inversiones en

¹ Todos los datos referentes a PIB, PIB per cápita, desempleo, creación de empresas, peso del empleo por sectores económicos y densidad de población citados en este apartado son extraídos de la EPA que publica el INE.



Universidad de Oviedo

investigación y desarrollo. Por otro lado, el sector servicios, eje del emprendimiento en España, donde de las 314.033 empresas creadas en 2021, el 61% pertenecían a “otros servicios” y el 21,3% a actividades comerciales. Las comunidades del Levante concentran la mayoría de su actividad laboral en el sector servicios, siendo el nivel más alto de España en Baleares, un 75,5% y el menor de las 4 la Región de Murcia, un 62,9%. Hay que añadir, que la Región de Murcia, una comunidad menos desarrollada históricamente, tuvo una tasa de crecimiento compuesta del PIB entre 2003 y 2021 del 2,55%, la segunda mayor de España.

En la mitad de la tabla encontramos comunidades como Aragón, Cantabria, Asturias, Islas Canarias o Extremadura. También la Comunidad de Madrid o el País Vasco, que está en los últimos puestos, resultados que llaman la atención y son difíciles de explicar.

Por último, en los 3 últimos puestos en cuanto a fomento del emprendimiento tenemos a Galicia, Castilla y León, y Andalucía. La explicación para estas comunidades podría darse, por un lado, por la despoblación. El fenómeno de la “España Vacía” puede en parte explicar esto, la densidad de población en las dos Castillas y Extremadura están por debajo de 50 habitantes por kilómetro cuadrado (hab/km²), mientras que en zonas del Levante; provincias de Sevilla, Málaga y Cádiz; y zonas de la Cornisa Cantábrica se encuentra entre 200 y 500 hab/km² y las áreas metropolitanas de Madrid y Barcelona por encima de 500 hab/km². De estas tres comunidades, la más afectada por esta despoblación es Castilla y León, según datos del Ministerio de Política Territorial y Función Pública, es la segunda comunidad con una mayor proporción de municipios que pierden población, el 86,1%, entre 2001 y 2018. En ese sentido, Galicia ocupa el 5º puesto en la estadística, con el 76,7% de sus municipios que pierden población. Andalucía presenta una explicación más compleja y diversa, tiene cierta concentración de población en las provincias mencionadas, pero al igual que la Región de Murcia, es una comunidad menos desarrollada históricamente. Con los mayores niveles de paro del país y de la Unión Europea, el 20,18% en el cuarto trimestre de 2021. Además, la tasa de paro para menores de 25 años es del 39,15%, significativamente alta, lo que puede explicar que la gente joven busque oportunidades de empleo en otras comunidades o en el extranjero.



6.2 EFECTOS TEMPORALES

	Modelo 1.1	Modelo 1.2	Modelo 1.3	Modelo 1.4
const	29040*** (3821,26)	30095,80*** (5828,90)	137933*** (15692,90)	8,30*** (0,45)
Emprendedores	0,02*** (0,005)	-	29612*** (6598,44)	-
NuevosEmp	-	0,01*** (0,003)	-	-
NuevosEmp_1	-	-	0,002 (0,002)	-
l_Emprendedores	-	-	-	-0,02 (0,03)
Pob18_64	-0,01*** (0,002)	-0,01*** (0,003)	-0,01*** (0,003)	1,89*e-7 (1,65*e-7)
2004	1792,61*** (558,52)	1328,89 (1071,65)	-	0,08** (0,04)
2005	2691,68** (1021,72)	2342,81*** (647,07)	3041,90*** (1040,33)	0,13** (0,05)
2006	2561,47** (1115,85)	3336,16*** (682,89)	3954,91*** (706,66)	0,25*** (0,08)
2007	2268,09* (1078)	2236,63*** (442,85)	3243,37*** (596,89)	0,22** (0,09)
2008	624,46 (881,08)	587,809 (464,28)	431,98 (443,97)	-0,11 (0,09)
2009	-39,38 (800,64)	-410,418 (572,73)	-783,41 (613,75)	-0,45*** (0,08)
2010	393,06 (824,62)	-419,077 (493,22)	-600,68 (574,96)	-0,43*** (0,08)
2011	155,04 (885,45)	-577,708 (461,50)	-364,22 (460,72)	-0,39*** (0,07)
2012	307,20 (868,77)	-201,492 (407,46)	-402,49 (411,14)	-0,36*** (0,07)
2013	585,63 (908,49)	-12,4602 (356,96)	-135,84 (368,78)	-0,31*** (0,06)
2014	232,08 (833,19)	-364,636 (404,03)	-384,14 (396,11)	-0,30*** (0,06)
2015	-0,45 (874,77)	-527,832 (403,63)	-590,62 (411,51)	-0,31*** (0,06)
2016	379,15 (894,74)	-188,418 (399,67)	-329,01 (402,94)	-0,22*** (0,06)
2017	-338,02 (795,93)	-850,957* (428,69)	-763,21* (427,92)	-0,31*** (0,06)
2018	-311,58 (791,42)	-666,144* (349,19)	-736,48** (321,39)	-0,32*** (0,06)
2019	-183,67 (910,49)	-544,373** (216,41)	-649,20** (248,70)	-0,35*** (0,06)
2020	-592,52 (905,88)	-1105,80*** (367,81)	-1299,16*** (412)	-0,52*** (0,06)
2021	598,20 (1108,41)	-	-	-0,29*** (0,06)
R2	0,98/0,72	0,908/0,71	0,98/0,71	0,99/0,87
Individuos	17	17	17	17
Número años	16-19	15-18	14-17	16-19
Observaciones	295	278	261	295

Notas: i) Los * se refieren al nivel de significación de los contrastes individuales, *** 1%, ** 5% y *1%, ii) Debajo de cada contraste figura la conclusión, rechazo o no rechazo de la hipótesis nula.

Tabla 6.2.1: Efectos temporales.

Fuente: elaboración propia.



Al igual que en el apartado anterior, en la tabla 5.2 se muestran, en este caso, las estimaciones para nuestros cuatro modelos incluyendo variables ficticias anuales que nos permitan aislar el efecto temporal de cada año. Se ha tomado el primero, 2003, como año de referencia, y en este caso, se ha estimado por efectos fijos, ya que Gretl permite en este tipo de estimaciones añadir variables ficticias temporales.

Encontramos ciertas tendencias en estos resultados, hay que señalar que muchos de los coeficientes de estas variables ficticias temporales no son estadísticamente significativos, situación común en estas estimaciones, ya que se tienen que detectar tendencias significativas por año. Como venimos señalando, los resultados hay que tratarlos con cierta cautela. Además de los años que pueden no ser significativos para explicar los modelos, la variable asociada a 2021 ha sido omitida en el 2.1 y 3.1 y la asociada a 2004 también en este último, se debe a que se ha detectado colinealidad perfecta en estas variables.

Si nos fijamos en los años que tienen un efecto mayor son los previos a la crisis de 2008, que en España comenzó a mediados de ese mismo año. 2004, 2005, 2006 y 2007 son los años, en general, con más peso. Son años de gran crecimiento económico en España, con tasas del 3% y 4% de crecimiento interanual del PIB, datos provenientes del INE. Con los niveles de desempleo bajos para España, un país tradicionalmente con niveles altos, se registraban tasas por debajo del 10% en 2005, 2006 y 2007, niveles en mínimos históricos. Por otro lado, años como 2020, donde se produjo la emergencia por la pandemia de la COVID19 y el confinamiento, están en los últimos puestos de la tabla. Por lo que podríamos decir que antes de la crisis de 2008 el emprendimiento experimentó su mayor auge y parece no haber recuperado esos niveles.

7. CONCLUSIONES

A lo largo de esta investigación se ha intentado responder a la cuestión de si los índices elaborados por el Observatorio GEM España y sus equipos regionales son un fiel reflejo del fenómeno emprendedor. Para esto se ha recopilado una base de datos con los propios índices, TEA, emprendedores y TEP, emprendedores potenciales; y la representación del fenómeno emprendedor mediante la creación de nuevas sociedades mercantiles, provenientes de bases de datos del INE y la afiliación al epígrafe RETA, los autónomos, datos publicados por la Seguridad Social. Con estos datos se ha elaborado un panel y se ha expuesto una metodología básica para estimar modelos. Los propuestos, suponen una relación entre el número de emprendedores y emprendedores potenciales con el mayor número de sociedades o de afiliados al RETA, además se han propuesto diferentes variaciones de estos modelos con el fin de capturar todos los posibles efectos, introduciendo diferencias, retardos y logaritmos.

Las relaciones reflejadas en los modelos 1.1, 2.1, 3.1 y 4.1 son coherentes con nuestra hipótesis de partida, detectando una relación positiva entre emprendedores estimados y nuevos emprendedores estimados y la creación de sociedades. De ellos, la capacidad explicativa, nos fijamos en el coeficiente de determinación "intra", oscila entre el 0,29 y el 0,5, niveles bajos. En el último modelo se aplicaron logaritmos a ambos lados de la ecuación, lo que nos permite interpretar los resultados como una elasticidad. El coeficiente asociado al logaritmo de emprendedores nos dice que por cada 1% que aumentan estos, las nuevas sociedades crecen 0,41%.



Universidad de Oviedo

Sin embargo, no se encuentra evidencia sólida de correlación de estos emprendedores con la afiliación a autónomos. Tampoco entre emprendedores potenciales y afiliación o creación de sociedades. Lo que nos lleva a dudar de la capacidad del índice de emprendedores potenciales para explicar las variaciones en el emprendimiento.

Sin embargo, sí tenemos evidencia para afirmar que existe una relación positiva y significativa entre la tasa total de emprendedores (TEA) estimada por el GEM y la creación de sociedades mercantiles en España. Esto es un hecho a destacar, puesto que este índice, como se menciona, el más importante de los confeccionados por el proyecto, es una referencia en términos de medición del fenómeno emprendedor y sirve a los diferentes agentes económicos y sociales como referencia y orientación en la toma de decisiones.

También, gracias a las ventajas que nos ofrecen las estimaciones de paneles, hemos podido aproximar las comunidades autónomas que más peso tienen en el emprendimiento, mediante los efectos individuales, donde parece tener cierta ventaja la zona del Levante frente al centro y sur del país. En el mismo sentido, aislando los efectos temporales, encontramos que los años con más efecto son los previos a la crisis de 2008, donde se alcanzaron los niveles de emprendimiento más altos, que no se han recuperado hasta ahora. Insistimos en que estos resultados deben ser interpretados con mucha precaución por lo rudimentario de los métodos de estimación.

Hay que señalar que nuestros modelos no están exentos de limitaciones y problemas, como se ha visto, todos ellos presentan problemas de autocorrelación y heterocedasticidad. Quedaría como propuesta para futuras investigaciones profundizar en este tema, aplicando métodos econométricos más sofisticados, como son los modelos dinámicos de panel, de gran utilidad cuando las observaciones actuales dependen de las pasadas, como puede ser este caso, ya que permite medir este efecto incluyendo retardos tanto en las variables dependientes como independientes. Otra ventaja sería la posibilidad que ofrecen de solventar la presencia de heterocedasticidad y autocorrelación.



8. BIBLIOGRAFÍA

REFERENCIAS BIBLIOGRÁFICAS:

- **Acs, Z. J., C. Armington and A. Robb, (1999):** 'Measures of Job Flow Dynamics in the U.S. Economy. Upper Marlboro, Maryland', U.S. Bureau of the Census, Center for Economic Studies, Discussion Paper.
- **Baumol, W. J., (1968):** 'Entrepreneurship in Economic Theory', American Economic Review 58 (2), 64–7.
- **Birch, D. L., (1979):** 'The Job Generation Process.' Unpublished report. Massachusetts Institute of Technology Program on Neighbourhood and Regional Change for the Economic Development Administration, U.S. Department of Commerce, Washington, DC.
- **Cantillon, R., & Jevons, W. S. (1950):** Ensayo sobre la naturaleza del comercio en general (No. HB151. C36 1950.). México: Fondo de cultura económica.
- **Hill, S., Ionescu-Somers, A., Coduras, A., Guerrero, M., Roomi, M. A., Bosma, N., ... & Shay, J. (2022, February):** Global entrepreneurship monitor 2021/2022 global report: opportunity amid disruption. In Expo 2020 Dubai. Disponible en: <https://www.gemconsortium.org/file/open?fileId=50900>
- **Laviada, A. F., Calvo, N., Samsami, M., Neira, I., Atrio, Y., & Barros, E. (2022):** Global Entrepreneurship Monitor. Informe GEM España 2021-2022 (Vol. 262). Ed. Universidad de Cantabria. Disponible en: <http://gemgalicia.org/wp-content/uploads/2022/11/Informe-GEM-Espana-2021-2022.pdf>
- **Navarro, J. R., & Garrido, J. A. M. (2007):** Global Entrepreneurship Monitor. 2005: Informe Ejecutivo 2005. Andalucía. Servicio Publicaciones UCA.
- **Pérez, R., & López, A. J. (2019):** Econometría aplicada con Gretl. Oviedo: Departamento de Economía Aplicada. Disponible en: [https://www.researchgate.net/publication/334771581 Econometria Aplicada c on Gretl](https://www.researchgate.net/publication/334771581_Econometria_Aplicada_c_on_Gretl)
- **Reynolds, P., Bosma, N., Autio, E., Hunt, S., De Bono, N., Servais, I., ... & Chin, N. (2005):** Global entrepreneurship monitor: Data collection design and implementation 1998–2003. Small business economics, 24, 205-231. Disponible en: [https://www.researchgate.net/publication/5158277 Global Entrepreneurship Monitor Data Collection Design and Implementation 1998-2003](https://www.researchgate.net/publication/5158277_Global_Entrepreneurship_Monitor_Data_Collection_Design_and_Implementation_1998-2003)
- **Schumpeter, J. A. (1976):** Teoría del desenvolvimiento económico: una investigación sobre ganancias, capital, crédito, interés y ciclo económico. (No Title).
- **Smith, A., Quintana, E. F., & Blas, L. P. (1996):** Investigación de la naturaleza y causas de la riqueza de las naciones. Valladolid: Junta de Castilla y León, Consejería de Educación y Cultura.
- **Wooldridge, J. M. (2006):** Introducción a la econometría. Un enfoque moderno: un enfoque moderno. Ediciones Paraninfo, SA. Disponible en: <http://hdl.handle.net/10498/25685>



BASES DE DATOS:

- **Diáñez-González, J.P., Franco-Leal, N., Bornay-Barrachina, M.M., Guerrero-Villegas, J., Segundo-Gallardo, J. (2022):** Global Entrepreneurship Monitor. Comunidad Autónoma de Andalucía. Informe 2021-2022. Disponible en: <https://www.gem-spain.com/wp-content/uploads/Informes-Regionales-Andalucia/Informe%20GEM%20Andalucia%202021-2022.pdf>
- **González, M. D., Rodríguez V. S., Barcala, M. F., Bayón, S. L., López, e I. M., Fernández, M. M. Cátedra de Emprendimiento Caja Rural de Asturias-Universidad de Oviedo (2022):** Informe GEM Asturias 2021-2022. Global Entrepreneurship Monitor. Servicio de publicaciones de la Universidad de Oviedo. Disponible en: <https://www.gem-spain.com/wp-content/uploads/Informes-Regionales-Asturias/Informe%20GEM%20Asturias%202021-2022.pdf>
- **Guemes, J. J., Coduras, A., Rachida, C. C., & Pampillon, R. (2010):** Global entrepreneurship monitor. *Informe Ejecutivo 2010*. Disponible en: <https://www.gem-spain.com/wp-content/uploads/2015/03/Informes%20antiguos/GEM2010.pdf>
- **Instituto Nacional de Estadística** (21 de septiembre, 2022). Sociedades mercantiles constituidas. Disponible en: <https://www.ine.es/jaxiT3/Tabla.htm?t=13913>
- **Instituto Nacional de Estadística.** Principales series de población desde 1998. Disponible en: <https://www.ine.es/jaxi/Tabla.htm?path=/t20/e245/p08/l0/&file=02003.px&L=0>
- **Laviada, A. F., Calvo, N., Samsami, M., Neira, I., Atrio, Y., & Barros, E. (2022):** Global Entrepreneurship Monitor. Informe GEM España 2021-2022 (Vol. 262). Ed. Universidad de Cantabria. Disponible en: <http://gemgalicia.org/wp-content/uploads/2022/11/Informe-GEM-Espana-2021-2022.pdf>
- **Ministerio de inclusión, Seguridad Social y migraciones.** Histórico de estadísticas de afiliación publicadas desde 1999. Disponible en: <https://www.seg-social.es/wps/portal/wss/internet/EstadisticasPresupuestosEstudios/Estadisticas/EST8/EST167>
- **Navarro, J. R., Medina, J. A., Lorenzo, J. D. y Ramos, A. R. (2010):** Global Entrepreneurship Monitor. Informe Ejecutivo 2010 Andalucía. Disponible en: <https://rodin.uca.es/handle/10498/25696>

MATERIAL DE CONSULTA:

- **de España, G. (2019):** Diagnóstico Estrategica Nacional frente al Reto Demográfico. Eje Despoblación. Disponible en: https://www.miteco.gob.es/content/dam/miteco/es/reto-demografico/temas/analisis-cartografia/diagnostico_eje_despoblacion_tcm30-517769.pdf
- **Encuesta de Población Activa (EPA) elaborada por el INE.** Disponible en: https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176918&menu=ultiDatos&idp=1254735976595
- **Observatorio startup Valencia:** <https://startupvalencia.org/es/>