



Munich Personal RePEc Archive

Entrepreneurship gender gap and entrepreneurial culture: Evidence from Google Trends

Gutiérrez, Antonio

Universidad de Zaragoza

January 2023

Online at <https://mpra.ub.uni-muenchen.de/115876/>
MPRA Paper No. 115876, posted 03 Jan 2023 16:35 UTC

La brecha de género en el emprendimiento y la cultura emprendedora: Evidencia con Google Trends

Antonio Gutiérrez

Enero, 2023

Resumen

Este trabajo analiza empíricamente la brecha de género en el emprendimiento y las variaciones geográficas en el interés para emprender en Estados Unidos. Para ello, empleamos las consultas realizadas en el buscador de Google. En primer lugar, elaboramos un índice compuesto mediante un análisis factorial sobre las búsquedas relativas al emprendimiento. En segundo lugar, exploramos el grado de sexismo local que existe a lo largo de la geografía estadounidense mediante un índice simple que representa las consultas machistas realizadas en el buscador. Los resultados señalan una correlación positiva y estadísticamente significativa entre el índice compuesto y el número de empresas asentadas en cada área de mercado. El índice de sexismo local no logra explicar la brecha de género en el emprendimiento, en su lugar, son las decisiones dentro del hogar y la proporción de individuos en cada grupo de edad las que presentan correlaciones estadísticamente significativas con la brecha de género en el emprendimiento.

Abstract

This paper empirically analyses the entrepreneurship gender gap and geographic variations of the entrepreneurship culture in the United States. To do so, we use Google search engine queries. First, we construct a composite index using factor analysis on searches related to entrepreneurship. Second, we explored the degree of local sexism that exists across the United States. Therefore, we use a simple index that represents sexist queries made on the search engine. The results indicate a positive and statistically significant correlation between the composite index and the number of companies based in each market area. The local sexism index fails to explain the gender gap in entrepreneurship. Conversely, it is intra-household decisions and the proportion of individuals in each age group that show statistically significant correlations with the entrepreneurship gender gap.

Keywords: entrepreneurship, gender gap, entrepreneurship culture, Google Trends, Big Data

JEL Classification: L26, J16, J71, O51

Introducción

El emprendimiento constituye una herramienta muy poderosa en la economía de los países como catalizador del crecimiento económico y el desarrollo tecnológico (Acs, 1992; Porter, 2011). El emprendimiento se relaciona con factores, no sólo académicos, sino institucionales y políticos. Normalmente, el análisis del emprendimiento se realiza bajo un enfoque macroeconómico, basándose en el apoyo gubernamental, institucional y hacendístico (Barrado et al., 2015). Sin embargo, esta metodología permite analizar el efecto agregado del nivel de emprendimiento, no los atributos individuales que determinan al emprendedor (Molina et al., 2016). El emprendimiento ha suscitado una amplia literatura nacional e internacional que compara las decisiones laborales de los asalariados con las de los emprendedores (Giménez et al., 2015; Campaña et al., 2016, 2020; Molina, 2020). Además, también se relacionan las decisiones emprendedoras con un amplio número de variables socio-demográficas que condicionan la actividad emprendedora de los individuos, como el género, la edad, la educación, la movilidad, las actividades dentro del hogar o sus finanzas (Giménez et al., 2012; Giménez et al., 2016; Campaña et al. 2017; Giménez et al., 2018; Velilla et al., 2018; Molina et al., 2016). Una fuente de información importante es el Global Entrepreneurship Monitor, el cual proporciona el indicador TEA habitualmente utilizado para identificar los emprendedores nacientes, emprendedores por necesidad y emprendedores por oportunidad. (Molina et al., 2016, 2017; Giménez et al., 2019; Velilla, 2018). En este trabajo, nos centraremos en el papel que juega el aspecto cultural sobre la propensión a emprender de los individuos. Concretamente, en el sexismo y el interés en el emprendimiento.

Para ello, utilizaremos los datos de Google Trends (GT), los cuales nos indican el grado de popularidad de una consulta en un momento y localización del tiempo determinado. En los últimos años, las consultas realizadas por los usuarios de Google han emergido como fuentes de datos relevantes y útiles para la investigación en ciencias sociales (Cebrián & Domenech, 2022). La utilización de estos datos, nos permite contribuir al estudio de las ciencias sociales a través de nuevas bases de datos de alta frecuencia. En otras palabras, nos permite contribuir a la aplicación del “Big data” en la investigación en economía, una línea con gran proyección (Einav & Levin, 2014). Por otro lado, utilizaremos los datos de American Community Survey de 2019 para obtener la información sociodemográfica de cada área de mercado de EEUU.

En este contexto, en este trabajo realiza un índice compuesto para captar el interés en el emprendimiento por área de mercado en EEUU y analiza la relación entre el sexismo local y la brecha en el emprendimiento medido a través del autoempleo. Este trabajo muestra una correlación positiva y estadísticamente significativa entre el índice compuesto de interés por el emprendimiento y el número de empresas localizadas en un área de mercado. Por otro lado, para el sexismo local observamos correlación significativa y positiva entre el índice y la brecha de género en el autoempleo. Además, contemplamos una correlación negativa y estadísticamente significativa entre el nivel educativo adquirido y la brecha de género en el autoempleo.

Este Trabajo contribuye a la literatura de distintas maneras. En primer lugar, demuestra la utilidad de los datos de alta frecuencia en la investigación en economía, concretamente con el objetivo de reflejar comportamientos discriminatorios. En segundo lugar,

documentamos una correlación estadísticamente significativa entre la brecha de género y el nivel de sexismo de una región. En tercer lugar, podemos explicar la variación en el número de empresas a través del interés que expresan los usuarios de Google en el emprendimiento. La estructura del trabajo sigue el siguiente orden. En la Sección 2 se revisa a literatura sobre el interés y la cultura en el emprendimiento, la brecha de género en el emprendimiento y el estudio con datos de alta frecuencia. La Sección tres expone los datos y variables empleadas. La Sección 4 formula la estrategia empírica y los resultados principales. La Sección 6 presenta las conclusiones fundamentales.

Interés en el emprendimiento, emprendimiento femenino y datos de alta frecuencia

El objetivo principal de este trabajo es contribuir a la investigación con datos de alta frecuencia probando la utilización de estas fuentes de datos para explicar los distintos fenómenos que se dan en el emprendimiento. En este contexto, tratamos de explicar las variaciones en el nivel de emprendimiento y la brecha de género existente en cada región. Con el análisis del interés en las consultas relacionadas con el emprendimiento, contribuimos a una de las corrientes de estudio sobre la cultura emprendedora la cual analiza la relación existente entre las creencias, cultura y valores sobre el emprendimiento entre las distintas regiones y países (Hayton et al., 2002; Hayton & Cacciotti, 2013).

Sin duda, la influencia y formación de la cultura emprendedora es un tema de gran relevancia en esta rama de investigación (McClelland, 1961). En este sentido, encontramos evidencia de que existe una relación positiva entre los aspectos culturales y el emprendimiento. En Davidsson (1995), observamos como los valores pueden afectar a los niveles de emprendimiento medidos por las tasas de creación de empresas en Suecia. Thomas & Mueller (2000) asocian la cultura con aspectos de personalidad vinculadas con comportamientos emprendedores. Morrison (2000) afirma que existe una relación entre ciertas especificidades culturales y el emprendimiento. En Suddle et al. (2010) encuentran una relación positiva entre el emprendimiento incipiente y la cultura del emprendimiento, incorporando como variables de control factores demográficos, institucionales y económicos. Por otro lado, existe evidencia de que la actitud hacia el emprendimiento puede variar para distintas regiones. Bosma & Schutjens (2011) atribuyen las diferencias en las actitudes hacia el emprendimiento en las distintas regiones europeas debido a aspectos demográficos, institucionales y demográficos. (Liñán et al., 2011) demuestran que las diferencias en el nivel de emprendimiento en España se pueden explicar a través de la valoración social que tienen los emprendedores. Beugelsdijk (2010) por otro lado, explica que la innovación de un país y su crecimiento económico pueden ser explicados a través de la cultura emprendedora en las distintas regiones europeas.

En cuanto al emprendimiento femenino, aunque las tasas de autoempleo femenino hayan aumentado en los últimos años, todavía existe cierta asimetría en el autoempleo por cuestiones de género (Koellinger et al., 2013). En la literatura encontramos numerosos trabajos que abordan este aspecto desde cuatro perspectivas diferentes: aversión al riesgo, segregación ocupacional por género, satisfacción laboral y comportamientos discriminatorios (Simoes et al., 2016). Por tanto, la discriminación por género puede ser

una de las causas detrás de la brecha en el emprendimiento. Sin embargo, no existe un consenso generalizado del efecto producido. En algunos trabajos, observamos que la discriminación provoca una reducción de oportunidades laborales de las mujeres en empleos por cuenta ajena, lo que actúa como factor de empuje de las mujeres hacia el autoempleo (Williams, 2012). Por otro lado, la discriminación por género puede suponer una barrera al emprendimiento femenino debido a que supone un freno a la financiación externa (Coleman & Robb, 2009, 2012; Gicheva & Link, 2013, 2015).

Para responder a las preguntas establecidas previamente, empleamos datos de alta frecuencia. Con este término, nos referimos a los distintos tipos de datos masivos que podemos encontrar y que permiten su uso para análisis económico e investigación (Einav & Levin, 2014; Cebrián & Domenech, 2022). Estos datos se diferencian de otros bancos de datos utilizados en la investigación en la recolección y procesamiento de estos. Así como la frecuencia, puntualidad y concreción de la publicación de los datos.

Uno de los trabajos pioneros en la utilización de estas fuentes de datos fue Choi & Varian (2012), dónde comprueban que las búsquedas realizadas en Google son útiles para comprender el comportamiento de los consumidores en distintos mercados como el automovilístico, el inmobiliario o el turístico. A partir de este trabajo, el uso de los datos de GT se aplicó a otros campos con el objetivo de mejorar las predicciones en la tasa de desempleo (Askatas & Zimmermann, 2009; Fondeur & Karamé, 2013; Baker & Fradkin, 2017; D'Amuri & Marcucci, 2017; Naccarato et al., 2018), el mercado de la vivienda (Webb, 2009; Wu & Brynjolfsson, 2009; Dietzel, 2016) o la confianza de los consumidores (Della Penna & Huang, 2009; Castelnuovo & Tran, 2017; Bontempi et al., 2019).

La utilización de datos de alta frecuencia en investigación en emprendimiento es relativamente escasa. Principalmente, la base de datos de alta frecuencia más utilizada es la que ofrece la red social Twitter. Mediante la utilización de estos datos, se ha contribuido en la investigación sobre la identidad digital de los emprendedores. Este concepto hace referencia a la auto-representación en redes del empresario que ayuda a comprender como los emprendedores son percibidas por las partes interesadas y su interacción con estas (Fischer & Rebecca, 2014). En Obschonka et al. (2017) estudian la personalidad de emprendedores con éxito a partir del análisis del lenguaje de sus tuits publicados. Para realizar este análisis utilizan herramientas conocidas como *language-based personality assessment tools*. En Fisch & Block (2021) analizan las variaciones en la identidad digital de los emprendedoras tras experimentar el fracaso en sus intentos de emprendimiento. Para ello, parten de una muestra de 760 emprendedores, los cuales, que previamente han errado en sus intentos de emprender. Observan que el tono emocional de los tuits de estos emprendedores disminuye y se advierten mayores niveles de angustia. Por otro lado, aumenta el contenido que denota mayor autoestima y reflexión. Obschonka et al. (2020) analiza a partir de una muestra de 1,5 billones de tuits la personalidad de los emprendedores de 1772 condados de EEUU. Waters et al. (2021) realizan un análisis de sentimiento a partir de los datos de Twitter con el fin explicar diferencias en las emociones que expresan los emprendedores y los no emprendedores.

Para los datos de GT, hasta dónde conocemos, únicamente se han utilizado en dos trabajos. En primer lugar, Gómez Martínez et al. (2015), documentan una relación

positiva entre la proporción de búsquedas para la palabra emprendimiento y el porcentaje de propietarios de un negocio en 5 países europeos. En segundo lugar, Drago (2022), realiza un índice compuesto para representar el interés de las mujeres para emprender en Italia durante el año 2021. frecuencia. Por otro lado, en el contexto de esta investigación, también se han empleado en distintos artículos para medir actitudes discriminatorias y aspectos culturales. En este plano, observamos el trabajo de Stephens-Davidowitz (2014), dónde a través de las consultas racistas logra explicar las diferencias en el apoyo electoral que recibe un candidato negro en el mercado político norteamericano. Utilizando una metodología similar, Corbi & Picchetti (2020), utilizan las consultas machistas para analizar el respaldo que recibe una mujer candidata en las elecciones de EEUU. Del mismo modo, desde la perspectiva de género, Owen & Wei (2021) estudian las variaciones de la brecha salarial en el continente norteamericano utilizando como variable explicativa un índice de sexismo elaborado a partir de las consultas de los usuarios de Google.

Datos

Para este trabajo utilizamos datos obtenidos de Google Trends (<https://trends.google.es>). Esta aplicación gratuita del buscador de Google permite obtener el grado de popularidad de las consultas de los usuarios utilizando el Search Volume Index (SVI). Este índice, representa la popularidad en una escala de 0 a 100 a lo largo de una serie temporal determinada. Para ello, GT emplea la cuota de consultas, el número de consultas para el término de búsqueda dividido por el número total de consultas en una región concreta. La incorporación de este tipo de datos presenta una serie de ventajas. En primer lugar, contribuye a la línea de investigación en economía con datos de alta frecuencia. En segundo lugar, estas fuentes de datos agilizan el proceso de investigación debido a que evitan el retraso producido por la recolección y procesamiento de los bancos de datos. En tercer lugar, estos datos reflejan las consultas que realizan los individuos en privado. De este modo podemos inferir pensamientos, opiniones e intereses de los individuos.

Desde otro punto de vista, debemos ser cuidadosos a la hora de emplear estos datos. En primer lugar, es preciso tener en cuenta que estos datos únicamente cubren las consultas realizadas en el buscador de Google lo que puede suponer problemas de cobertura (Cebrián & Domenech, 2022). En segundo lugar, el uso del lenguaje puede variar en función de cada región provocando errores en nuestras estimaciones (Artola & Galan, 2012). En tercer lugar, aunque el SVI constituye un índice normalizado a la unidad, lo que elimina el efecto sobre el volumen de consultas debido a un aumento en los usuarios de Google (Carrière-Swallow & Labbé, 2013), pueden surgir comportamientos tendenciales no relacionados con la búsqueda (Suhoy, 2009). Esto sucede cuando el SVI de una consulta disminuye a pesar de que el número bruto de consultas aumentó a causa de otras consultas con mayor popularidad (Google, 2016). Debido a esta problemática, surge un enfoque nuevo en el uso de estos datos, la perspectiva de sección cruzada. Siguiendo esta metodología, en lugar de estudiar la evolución de un índice simple (SVI) para una región y consulta determinadas estos estudios permiten la comparación de índices compuestos para regiones y consultas distintas (Silverstovs & Wochner, 2018) . Podemos encontrar trabajos que siguen este planteamiento para predecir las preferencias

de los votantes (Askitas, 2015a, 2015b) o la demanda turística (Siliverstovs & Wochner, 2018).

Para obtener una medida del grado de sexismo de una región de EEUU recurrimos a los datos de las consultas machistas recogidas por el índice de volumen de búsqueda en Google de Google Trends. En este sentido, los datos de Google Trends resultan muy atractivos, pues reflejan comportamientos privados de los usuarios que no estarían dispuestos a admitir en público (Stephens-Davidowitz, 2014; Corbi & Picchetti, 2020; Owen & Wei, 2021). En este trabajo seguimos los enfoques de Stephens-Davidowitz (2014) y Corbi & Picchetti (2020) y utilizaremos el volumen de consultas para una palabra [“Palabra 1”]¹ y su plural con el objetivo de capturar las diferencias geográficas de los comportamientos machistas.

Definimos el índice de sexismo de la siguiente manera:

$$\text{Índice de sexismo} = \frac{N. \text{ búsquedas en google para la Palabra 1 en el área } j}{N. \text{ de búsquedas en total en el área } j}$$

De esta manera, obtenemos un índice relativo del volumen de consultas para una zona geográfica y periodo concreto. En este caso, obtenemos el índice para 201 áreas de mercado² de Estados Unidos para el año 2019. Es preciso mencionar que este índice representa el sexismo de una zona geográfica, lo cual no refleja el nivel de sexismo de un individuo, y es probable que los individuos expresen estos comportamientos de distintas maneras. Para la medida del interés en el emprendimiento en cada área de mercado, realizamos un índice compuesto a través de las palabras relacionadas en el emprendimiento. Explicamos este índice en el apartado referido a la metodología empleada.

Por otro lado, tomamos datos de American Community Survey de 2019. Esta base de datos nos permite tener acceso a variables del hogar y sociodemográficas para 3200 condados EE. UU.. Obtenidos los datos, agregamos cada condado a su correspondiente área de mercado³. Desechamos los datos pertenecientes a Puerto Rico y Alaska. De esta manera recogemos datos demográficos para 206 áreas de mercado⁴. Podemos observar el conjunto de variables empleadas en este estudio en la Tabla 1.

Las variables dependientes principales son la brecha de género en el autoempleo⁵ y el número total de establecimientos en cada área de mercado. Obtenemos el número de establecimientos para cada condado de County Business Patterns, estadística de

¹ Este trabajo contiene términos de búsquedas en Google que pueden resultar ofensivas. Por este motivo, utilizamos lenguaje codificado para referirnos a estas palabras. Los términos empleados se pueden encontrar en el apéndice. Utilizamos la misma palabra utilizada por Corbi & Picchetti (2020).

² Google Trends permite dividir las consultas para 210 áreas de mercado en total. En este caso, no encontramos datos para las siguientes áreas: Alpena MI, Victoria TX, North Platte NE, Presque Isle ME, Zanesville OH, Juneau AK, Fairbanks AK y Glendive MT.

³ Asignamos cada condado al área de mercado correspondiente de acuerdo a la información obtenida a través del siguiente enlace: <https://www.kaggle.com/code/kerneler/starter-google-trends-county-dma-4a2dd71f-4/data>

⁴ No disponemos de datos demográficos para las siguientes áreas: Anchorage AK, Fairbanks AK, Juneau AK, Helena MT y Palm Springs CA.

⁵ Para obtener el número de autoempleados por condado, los datos de ACS contabiliza los individuos de autoempleados e individuos que trabajan en la empresa familiar.

American Census Bureau que indica el número de establecimientos con trabajadores asalariados en EEUU. Definimos la brecha de género en el emprendimiento como la diferencia entre la proporción de autoempleados que son varones y la proporción restante de mujeres. La variable explicativa principal es el índice de sexismo mencionado previamente. Al mismo tiempo disponemos de otras variables de control que reflejan la proporción de individuos por edades que componen el área de mercado. De esta forma, podemos observar el porcentaje de individuos que tienen entre 25-34 años, 35-44 años, 45-54 años y 55-64 años en cada zona geográfica. También, observamos variables relativas a las decisiones dentro del hogar. Estas son el porcentaje de mujeres mayores de 15 años casadas y el porcentaje de hogares constituidos por un matrimonio con hijos menores de 18 años. Por otro lado, disponemos de variables de control de cada área de mercado: porcentaje de población activa, porcentaje de población ocupada, porcentaje de población activa femenina y la brecha de género en la educación universitaria. Definimos esta última variable como la diferencia entre el porcentaje de hombres mayores de 25 años que cuentan con educación universitaria o superior y el porcentaje de mujeres mayores de 25 años que cuentan con educación universitaria o superior. Los datos disponibles de la ACS para 2019 también permiten obtener el peso relativo de cada sector y ocupación en cada área de mercado.

En la Tabla 1, se puede apreciar la gran variabilidad de establecimientos a lo largo de las distintas áreas de mercado analizadas. Respecto a la segunda variable dependiente, en la Figura 1, se puede observar la variación en la brecha de género para los distintos condados. Observamos que existe gran variabilidad de la brecha de género a lo largo de la geografía norteamericana. En este sentido, los resultados de la Tabla 1 -estadística descriptiva- confirman la variabilidad de la brecha de género y del índice de sexismo empleado para cada área de mercado. En el Gráfico 1, se observa una relación positiva entre la brecha de género en el emprendimiento y el índice de sexismo. El Gráfico 2 apunta a lo que podría considerarse como un predictor del índice de sexismo para las distintas áreas de mercado, el nivel adquirido de educación. Corbi & Picchetti (2020) señalan otras variables predictoras como la proporción por edades en cada área.

Metodología empírica

En este contexto, elaboramos un índice compuesto que refleja la popularidad de las consultas relacionadas con la creación de negocios con el objetivo de captar variaciones en la cultura emprendedora. Para ello, realizamos una lista de palabras clave de las cuales obtendremos el SVI de GT. Observamos en la Tabla 1 la lista de palabras clave utilizadas.

Tabla 1. Keywords

Business plan
Business ideas
Business proposal
Entrepreneurship
Self-employment
Entrepreneurial
Business finance

Business investment

Como se puede observar, son palabras relacionadas con el entorno del emprendimiento y el negocio.

Obtenida la proporción de consultas totales en cada área de mercado, procedemos con el índice compuesto. Para ello recurrimos al análisis factorial, una técnica de reducción de datos ampliamente utilizada en confección de índices compuestos porque permite la confección de índices con ponderaciones no arbitrarias mediante combinaciones lineales (Monge et al., 2022; Poza & Monge, 2020). De esta forma, nuestro índice de cultura emprendedora podemos expresarlo de la siguiente manera:

$$F_j = a_{j1} X_1 + a_{j2} X_2 + \dots + a_{j8} X_{j8}$$

Dónde a_{ji} son las cargas factoriales obtenidas tras el análisis factorial y X_i son las proporciones de consultas para cada término i en concreto. Es preciso mencionar que aplicamos un análisis factorial mediante componentes principales y con cargas factoriales rotadas mediante la rotación Varimax, rotación ortogonal. Aplicada esta técnica, observamos que el análisis factorial realizados es consistente de acuerdo con el criterio de adecuación de los datos KMO, con un valor de 0,816. Por otro lado, también rechazamos la hipótesis nula del test de Barlett. Esta técnica también permite identificar que variable aporta más información al índice, es decir, que término de búsqueda es el más relevante.

$$\begin{aligned} \text{Índice} = & 0.5607 * X_1 + 0.6868 * X_2 + 0.3868 * X_3 + 0.7102 * X_4 + \\ & 0.1021 * X_5 + 0.3843 * X_6 + 0.1631 * X_7 + 0.3352 * X_8 \end{aligned}$$

De la combinación lineal anterior, observamos que la variable más relevante, de acuerdo con las cargas factoriales, es la variable X_4 , es decir, las búsquedas del término *entrepreneurship*. Por otro lado, todas las cargas factoriales son positivas, lo que indica que a mayor número de búsquedas mayor es la cultura emprendedora.

En la Tabla 2 se presentan las correlaciones entre los distintos componentes sociodemográficos de cada área de mercado en EE. UU. y el índice compuesto de GT⁶. En el modelo incorporamos variables sociodemográficas. Observamos que las búsquedas relacionadas con el emprendimiento presentan una correlación positiva y estadísticamente significativa con la proporción de individuos que tienen entre 45-54 años. Por otro lado, observamos correlaciones estadísticamente significativas y negativas para el grupo de edad más avanzada (55-64 años), la proporción de matrimonios con hijos, la tasa de desempleo y la brecha en los estudios superiores⁷.

Utilizamos la estimación por mínimos cuadrados ordinarios para la estrategia empírica del trabajo. Para ello, establecemos las siguientes especificaciones:

⁶ El índice solo está disponible para 186 áreas de mercado, esto es debido a que en ciertas áreas las consultas no superan el umbral mínimo de búsquedas requerido para disponer de un SVI.

⁷ Definida como la diferencia entre el porcentaje de población que cuenta con educación secundaria y la proporción que cuenta con educación superior en cada área de mercado.

$$(1) \text{Log Establishments}_i = \alpha_0 + \alpha_1 * \text{Google Search Entrepreneurship Index} + \alpha_2 * \text{Household decisions}_i + \alpha_3 \text{Media market controls}_i + \mu_i,$$

$$(2) \text{Gender Gap}_i = \beta_0 + \beta_1 * \text{Google Search Sexism Index} + \beta_2 * \text{Household decisions}_i + \beta_3 \text{Media market controls}_i + \varepsilon_i,$$

En la Ecuación 1, partimos de la variable dependiente Log Establishments que refleja el logaritmo del número de establecimientos en cada área de mercado i . Regresamos la variable dependiente sobre la variable explicativa principal, en este caso, el índice compuesto que refleja el volumen de búsquedas en Google para consultar aspectos relacionados con el emprendimiento. Al mismo tiempo incluimos en esta especificación variables de control relacionadas con las decisiones en el hogar, *Household decisions*, para cada área de mercado i . Estas variables son: el porcentaje de mujeres mayores de 15 años casadas y el porcentaje de hogares compuestos por un matrimonio con hijos menores de 18 años a su cargo. Introducimos también variables sociodemográficas de control para cada área de mercado i , *Media market controls*. Concretamente: el porcentaje de individuos por grupo de edad, el porcentaje de individuos con empleo, el porcentaje de personas con educación universitaria, la proporción de población activa y el porcentaje de personas empleadas en cada tipo de ocupación e industria.

En esta primera especificación, el parámetro relevante es α_1 , pues es el que refleja la relación estadística existente entre las consultas en Google y número de establecimientos. De tal forma que un parámetro positivo y estadísticamente significativo significaría que las búsquedas en Google pueden utilizarse como termómetro del emprendimiento en un área de mercado determinada. De esta manera, un aumento unitario en el índice compuesto supondría un incremento $(e^{\alpha_1} - 1) \times 100$ puntos porcentuales en el número de establecimientos de un área de mercado.

En la Ecuación 2, establecemos como variable dependiente la brecha de género en el emprendimiento de cada área de mercado i . Esta variable ha sido definida previamente como la diferencia entre el porcentaje de autoempleados hombres y mujeres. Regresamos esta variable dependiente sobre las mismas variables de control empleadas en la Ecuación 1. Establecemos como variable explicativa principal el índice simple que refleja el volumen de consultas machistas en cada área de mercado i , *Google Search Sexism Index*. Este índice está normalizado de acuerdo con su valor z , lo que implica que el parámetro β_1 mide el impacto del incremento, en puntos porcentuales, de una desviación estándar en las consultas machistas sobre la brecha de género. Además, en la especificación incluimos variables de control con el objetivo de introducir el efecto de las decisiones dentro del hogar y las características socioeconómicas de cada área de mercado. Las variables que acompañan a β_2 son el porcentaje de mujeres mayores de 15 que están casadas, exceptuando aquellas que están separadas y el porcentaje de hogares compuestos por un matrimonio con hijos menores de 18 años. Las variables de control para las características socioeconómicas son el porcentaje de población activa, porcentaje de población ocupada, porcentaje de población activa femenina y la brecha de género en la educación universitaria. Por último, se

introduce el parámetro μ_i para captar la heterogeneidad inobservable de cada área de mercado.

Resultados

La Tabla 3 recoge los resultados para la Ecuación 1. Las estimaciones indican que existe una correlación positiva y estadísticamente significativa entre las consultas relacionadas con el emprendimiento y el número de establecimientos de un área de mercado. Esto sugiere, que las búsquedas en Google pueden actuar como termómetro de la actividad emprendedora de una región. Observamos que esta relación se cumple cuando incluimos todos los controles sociodemográficos. Concretamente, las estimaciones señalan que un incremento unitario en el índice compuesto de las búsquedas en Google se asocia con un incremento en torno al 2,4% en el número de establecimientos con trabajadores asalariados.

En la Columna 1, sin establecer variables de control por área de mercado, observamos que un incremento unitario del índice supone un incremento del número de establecimientos del 2,53%. En la Columna 2, incluyendo las variables de control sociodemográficas, el incremento unitario del índice se asocia a un incremento del número de establecimientos del 2,94%. En esta Columna, observamos una correlación estadísticamente significativa y positiva entre la brecha de educación secundaria y el número de establecimientos, así como la proporción de individuos entre 35 y 44 años. La Columna 3, incluye las variables sociodemográficas de control y también el peso de cada sector en el área de mercado. De nuevo la variable explicativa principal es significativa y positiva, señalando un efecto sobre la variable dependiente que supone un incremento del 2,22%. Con la inclusión del peso relativo de cada sector, las variables relativas al nivel de formación y la edad dejan de resultar estadísticamente significativas. En la cuarta Columna, controlamos mediante el peso relativo de cada ocupación en el área de mercado. La proporción de individuos entre 35-44 años y el índice compuesto resultan estadísticamente significativos y positivos. El efecto de la variable explicativa supone un incremento del 2,4% de la variable dependiente. En la última Columna, introducimos todas las variables de control, ocupaciones, sectores y sociodemográficas. De nuevo, el efecto de la variable explicativa supone un incremento del 2,4% de la variable dependiente. Por tanto, persiste la correlación estadísticamente significativa y positiva entre ambas variables. También persiste el mismo efecto sobre la proporción de individuos entre 35-44 años y el número de establecimientos.

En la Tabla 4, observamos los resultados para la Ecuación 2. El efecto del sexismo local expresado a través de las consultas de Google no está claro. En la Columna 1, especificación sin variables de control, un aumento de una desviación típica en las consultas machistas se asocia con un aumento en la brecha de género del 0,195%. En la Columna 2, controlando los aspectos sociodemográficos, el índice de sexismo mantiene su correlación estadísticamente significativa y positiva con la brecha de género. Concretamente, un incremento de una desviación típica de este índice se vincula a un incremento de la brecha de género del 0,211%. La brecha de formación universitaria entre hombres y mujeres, y los grupos de individuos que tiene 25-34, 35-44 y 55-64 también resulta estadísticamente significativa y negativa. Por el contrario, la variable que refleja

la proporción de individuos que tienen entre 45-54 años presenta una correlación estadísticamente significativa y positiva con la brecha de género en el emprendimiento. En la Columna 3, controlamos por el peso relativo que tiene cada sector en cada área de mercado. En esta Columna, la variable explicativa principal, el índice de sexismo local, ya no resulta estadísticamente significativa. En su lugar, las variables relativas a las decisiones dentro del hogar, porcentaje de mujeres casadas y porcentaje de hogares compuestos por un matrimonio con hijos, resultan estadísticamente significativas. La variable que refleja la proporción de mujeres mayores de 15 años que están casadas presenta una correlación negativa mientras que la proporción de hogares con matrimonio e hijos presenta una correlación positiva. En cuanto a los grupos de edad, la correlación estadísticamente significativa y negativa se mantiene para el grupo de entre 35-44 años y significativa y positiva para el grupo de 45-54 años.

En la Columna 4, controlando únicamente el peso relativo de cada ocupación, el índice de sexismo sí que resulta estadísticamente significativo y positivo. Un incremento unitario en este índice se asocia con un incremento en la brecha de género del 0,106%. La brecha de género en los estudios universitarios, el grupo de edad entre 35-44 años y el porcentaje de mujeres casadas resultan estadísticamente significativas y negativas. La Columna 5 incorpora las variables de control sociodemográficas y el peso relativo del sector y la ocupación. El índice de sexismo local no resulta estadísticamente significativo, por lo que no lo asociamos a efecto alguno sobre la brecha de género. Al contrario, sí que resultan significativas las variables relacionadas con la edad y las decisiones en el hogar. El grupo de edad de entre 35-44 años presenta una correlación negativa y estadísticamente significativa con la brecha de género, mientras que en el grupo de 45-54 años la correlación es estadísticamente significativa y positiva. Similar a lo que ocurría en la Columna 3, el porcentaje de mujeres casadas presenta una correlación estadísticamente significativa y negativa con la brecha de género. Al contrario, el porcentaje de hogares compuestos por un matrimonio con hijos menores de 18 años a su cargo presenta una correlación estadísticamente positiva y significativa sobre la brecha de género.

Conclusión

En este Trabajo se ha presentado evidencia empírica sobre la utilidad de las búsquedas en Google para estudiar distintos fenómenos en el emprendimiento para Estados Unidos. El objetivo principal es poner en valor las nuevas fuentes de datos masivos en la investigación en ciencias sociales. Para ello, se han utilizado los datos de Google Trends para captar la popularidad de distintos términos de búsqueda en el año 2019 en el país norteamericano. A partir de estos datos, se ha conformado un índice compuesto que refleja el interés de los individuos de cada área de mercado en el emprendimiento mediante la aplicación de un análisis factorial sobre las consultas relacionadas. Por otro lado, hemos obtenido un índice simple como medida del sexismo local de cada área de mercado.

Los resultados señalan una correlación estadísticamente significativa y positiva entre el número de establecimientos de un área de mercado y el índice compuesto de interés en la actividad emprendedora. Concretamente, un incremento unitario de este índice se asocia a un incremento del 2,4% en el número de establecimientos. En relación con el sexismo local, no encontramos una relación estadísticamente significativa entre el índice simple y

la brecha de género en el emprendimiento medida a través del autoempleo. Específicamente, esta variable deja de ser significativa cuando controlamos por el peso relativo de cada industria dentro de cada área de mercado. En su lugar, la brecha de género viene explicada por factores relacionados con las decisiones dentro del hogar, como la proporción de mujeres casadas o los matrimonios con hijos, o el peso relativo de cada grupo de edad. En particular, la brecha de género se correlaciona positivamente con el grupo de edad que tiene entre 45-54 años y negativamente con el grupo de entre 35-44 años.

En definitiva, este trabajo ha analizado empíricamente el efecto del interés en la actividad emprendedora y el sexismo local medido a través de las consultas en Google sobre el emprendimiento. Las implicaciones derivadas de este trabajo son relevantes puesto que logramos identificar aspectos culturales y comportamientos discriminatorios a través del comportamiento en la red. Las estimaciones sugieren que el comportamiento de los usuarios de Google puede funcionar como un termómetro de la actividad emprendedora en las distintas áreas de mercado en EEUU. Por otro lado, no podemos explicar la brecha de género a través del perfil sexista que un área de mercado pueda presentar. En su lugar, son las decisiones en el hogar y la edad, las que presentan una asociación significativa con la brecha de género en el emprendimiento.

Agradecimientos: Este trabajo se ha beneficiado de financiación procedente de la Beca de investigación de la Cátedra Emprender de la Universidad de Zaragoza. En otro orden, agradecer las reflexiones y consideraciones de los participantes del V Workshop online sobre Investigación en Emprendimiento organizado por la Cátedra Emprender de la Universidad de Zaragoza, en el contexto del Campus Iberus, el día 13 de diciembre de 2022.

Referencias

- Acs, Z. J. (1992). Small business economics: A global perspective. *Challenge*, 35(6), 38–44.
- Artola, C., & Galan, E. (2012). Las huellas del futuro están en la web: construcción de indicadores adelantados a partir de las búsquedas en Internet. *Banco de España Documentos Ocasionales N. 1203*.
- Askitas, N. (2015a). Calling the Greek Referendum on the nose with Google Trends. Available at SSRN 2633443.
- Askitas, N. (2015b). Predicting the Irish 'Gay Marriage' Referendum. Available at SSRN 2674243.
- Askitas, N., & Zimmermann, K. F. (2009). Google econometrics and unemployment forecasting. *Applied Economics Quarterly*, 55(2), 107–120.
- Baker, S. R., & Fradkin, A. (2017). The impact of unemployment insurance on job search: Evidence from google search data. *Review of Economics and Statistics*, 99(5), 756–768. https://doi.org/10.1162/REST_a_00674
- Barrado, B. and Molina, J.A. (2015). Factores macroeconómicos que estimulan el emprendimiento. Un análisis para los países desarrollados y no desarrollados. DTECONZ, 2005-06.
- Beugelsdijk, S. (2010). Entrepreneurial culture, regional innovativeness and economic growth. In *Entrepreneurship and culture* (pp. 129–154). Springer.
- Bontempi, M. E., Frigeri, M., Golinelli, R., & Squadrani, M. (2019). *Uncertainty, perception and the internet*.
- Bosma, N., & Schutjens, V. (2011). Understanding regional variation in entrepreneurial activity and entrepreneurial attitude in Europe. *The Annals of Regional Science*, 47(3), 711–742.
- Campaña, J.C., Giménez-Nadal, J.I. and Molina, J.A. (2016). Differences between self-employees and wage earners in time uses: Aragón vs. Spain. MPRA Paper 71463.
- Campaña, J.C., Giménez-Nadal, J.I. and Molina, J.A. (2017). Self-employment and educational childcare time: Evidence from Latin America.
- Campaña, J.C., Giménez-Nadal, J.I. and Molina, J.A. (2020). Self-employed and employed mothers in Latin American families: are there differences in paid-work, unpaid work and child care? *Journal of Family and Economic Issues*. DOI:10.1007/s10834-020-09660-5
- Carrière-Swallow, Y., & Labbé, F. (2013). Nowcasting with Google trends in an emerging market. *Journal of Forecasting*, 32(4), 289–298. <https://doi.org/10.1002/for.1252>
- Castelnuovo, E., & Tran, T. D. (2017). Google It Up! A Google Trends-based Uncertainty index for the United States and Australia. *Economics Letters*, 161(December 2008), 149–153. <https://doi.org/10.1016/j.econlet.2017.09.032>

- Cebrián, E., & Domenech, J. (2022). Is Google Trends a quality data source? *Applied Economics Letters*, 00(00), 1–5. <https://doi.org/10.1080/13504851.2021.2023088>
- Choi, H., & Varian, H. (2012). Predicting the Present with Google Trends. *Economic Record*, 88(SUPPL.1), 2–9. <https://doi.org/10.1111/j.1475-4932.2012.00809.x>
- Coleman, S., & Robb, A. (2009). A comparison of new firm financing by gender: evidence from the Kauffman Firm Survey data. *Small Business Economics*, 33(4), 397–411.
- Coleman, S., & Robb, A. (2012). Gender-based firm performance differences in the United States: Examining the roles of financial capital and motivations. In *Global Women's Entrepreneurship Research*. Edward Elgar Publishing.
- Corbi, R., & Picchetti, P. (2020). The cost of gendered attitudes on a female candidate: Evidence from Google Trends. *Economics Letters*, 196. <https://doi.org/10.1016/j.econlet.2020.109495>
- Cueto, B., Mayor, M., & Suárez, P. (2015). Entrepreneurship and unemployment in Spain: a regional analysis. *Applied Economics Letters*, 22(15), 1230–1235.
- D'Amuri, F., & Marcucci, J. (2017). The predictive power of Google searches in forecasting US unemployment. *International Journal of Forecasting*, 33(4), 801–816. <https://doi.org/10.1016/j.ijforecast.2017.03.004>
- Davidsson, P. (1995). Culture, structure and regional levels of entrepreneurship. *Entrepreneurship & Regional Development*, 7(1), 41–62.
- Della Penna, N., & Huang, H. (2009). *Constructing Consumer Sentiment Index for U.S. Using Google Searches* (Working Papers, Issues 2009–26). University of Alberta, Department of Economics. <https://doi.org/DOI:>
- Dietzel, M. A. (2016). Sentiment-based predictions of housing market turning points with Google trends. *International Journal of Housing Markets and Analysis*.
- Drago, C. (2022). A Quantile-Based Composite Indicators Approach on Woman's Entrepreneurship as measured by Google Search Activity. *Available at SSRN 4073363*.
- Einav, L., & Levin, J. (2014). Economics in the age of big data. *Science*, 346(6210). <https://doi.org/10.1126/science.1243089>
- Fisch, C., & Block, J. H. (2021). How does entrepreneurial failure change an entrepreneur's digital identity? Evidence from Twitter data. *Journal of Business Venturing*, 36(1). <https://doi.org/10.1016/j.jbusvent.2020.106015>
- Fischer, E., & Rebecca Reuber, A. (2014). Online entrepreneurial communication: Mitigating uncertainty and increasing differentiation via Twitter. *Journal of Business Venturing*, 29(4), 565–583. <https://doi.org/10.1016/j.jbusvent.2014.02.004>
- Fondeur, Y., & Karamé, F. (2013). Can Google data help predict French youth unemployment? *Economic Modelling*, 30, 117–125.

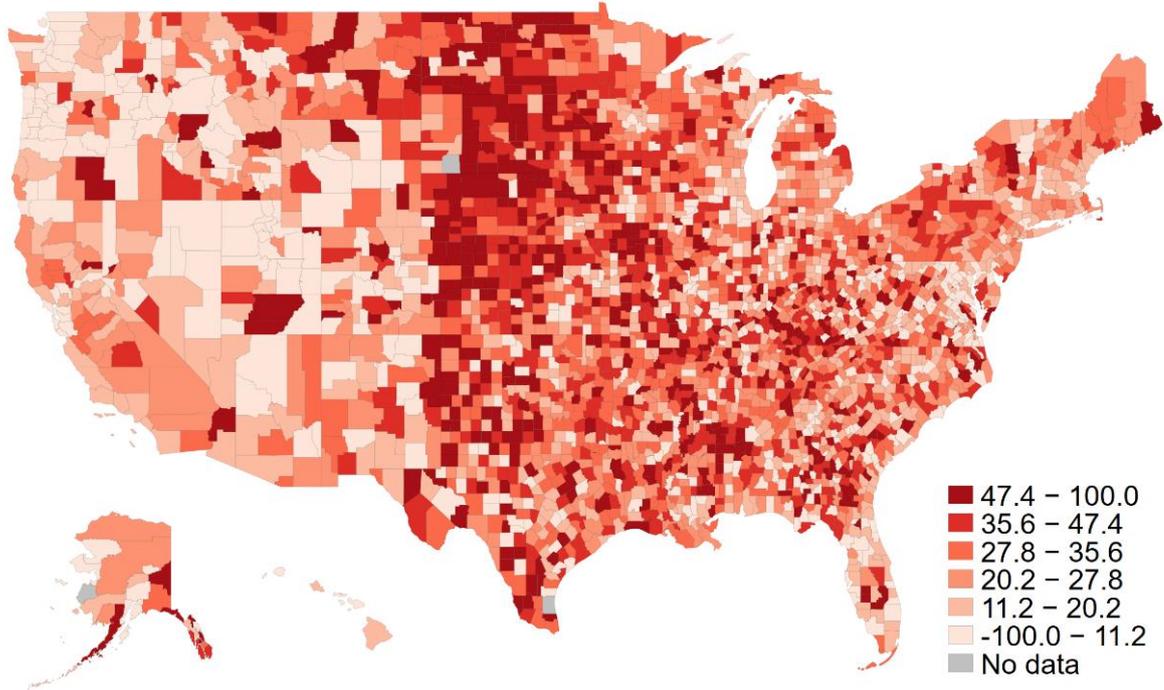
- Gicheva, D., & Link, A. N. (2013). Leveraging entrepreneurship through private investments: does gender matter? *Small Business Economics*, 40(2), 199–210.
- Gicheva, D., & Link, A. N. (2015). The gender gap in federal and private support for entrepreneurship. *Small Business Economics*, 45(4), 729–733.
- Gimenez-Nadal, J. I., Molina, J. A., & Ortega, R. (2012). Self-employed mothers and the work-family conflict. *Applied Economics*, 44(17), 2133–2147. <https://doi.org/10.1080/00036846.2011.558486>
- Giménez-Nadal, J.I. , Molina, J.A. and Velilla, J. (2015). Excess commuting in the US: Differences between the self-employed and employees. IZA DP N° 9425.
- Giménez-Nadal, J.I. , Molina, J.A. and Velilla, J. (2016). A wage- efficiency spatial model for US self-employed workers. IZA DP N° 9634.
- Giménez-Nadal, J.I. , Molina, J.A. and Velilla, J. (2018). The commuting behavior of workers in the United States: differences between the employed and the self-employed. *Journal of Transport Geography*, 66, 19-29. DOI: 10.1016/j.jtrangeo.2017.10.011.
- Giménez-Nadal, J.I. , Lafuente, M., Molina, J.A. and Velilla, J. (2019). Resampling and bootstrap algorithms to asses the relevance of variables: applications to cross-section entrepreneurship data. *Empirical Economics*, 56, 233-267. DOI:10.1007/s00181-017-1355-x.
- Giménez-Nadal, J. I., Molina, J. A., & Velilla, J. (2022). Intergenerational correlation of self-employment in Western Europe. *Economic Modelling*, 108. <https://doi.org/10.1016/j.econmod.2021.105741>
- Gómez Martínez, R., Prado Román, M., & Mercado Idoeta, C. (2015). Google search activity as entrepreneurship thermometer. In *New challenges in entrepreneurship and finance* (pp. 225–233). Springer.
- Hayton, J. C., & Cacciotti, G. (2013). Is there an entrepreneurial culture? A review of empirical research. In *Entrepreneurship and Regional Development* (Vol. 25, Issues 9–10, pp. 708–731). <https://doi.org/10.1080/08985626.2013.862962>
- Hayton, J. C., George, G., & Zahra, S. A. (2002). National culture and entrepreneurship: A review of behavioral research. *Entrepreneurship Theory and Practice*, 26(4), 33–52.
- Koellinger, P., Minniti, M., & Schade, C. (2013). Gender Differences in Entrepreneurial Propensity. *Oxford Bulletin of Economics and Statistics*, 75(2), 213–234. <https://doi.org/10.1111/j.1468-0084.2011.00689.x>
- Liñán, F., Urbano, D., & Guerrero, M. (2011). Regional variations in entrepreneurial cognitions: Start-up intentions of university students in Spain. *Entrepreneurship and Regional Development*, 23(3–4), 187–215.
- McClelland, D. C., & mac Clelland, D. C. (1961). *Achieving society* (Vol. 92051). Simon and Schuster.

- Molina, J.A. (2020). Family and entrepreneurship: New empirical and theoretical results. *Journal of Family and Economic Issues*. DOI:10.1007/s10834-020-09667-y.
- Molina, J.A., R. Ortega and J. Velilla (2016). Entrepreneurial activity in the OECD: Pooled and cross-country evidence.
- Molina, J.A., Velilla, J. and Ortega, R. (2016). The decision to become an entrepreneur in Spain: the role of household finances. *International Journal of Entrepreneurship*, 20(1), 57-73.
- Molina, J.A. and J. Velilla (2016). Innovation as a determinant of entrepreneurship. MPRA Papers 71471.
- Molina, J.A., R. Ortega and J. Velilla (2017). Feminization of entrepreneurship in developing countries.
- Monge, M., Poza, C., & Borgia, S. (2022). A proposal of a suspicion of tax fraud indicator based on Google trends to foresee Spanish tax revenues. *International Economics*, 169, 1–12. <https://doi.org/10.1016/j.inteco.2021.11.002>
- Morrison, A. (2000). Entrepreneurship: what triggers it? *International Journal of Entrepreneurial Behavior & Research*.
- Naccarato, A., Falorsi, S., Loriga, S., & Pierini, A. (2018). Combining official and Google Trends data to forecast the Italian youth unemployment rate. *Technological Forecasting and Social Change*, 130(November 2016), 114–122. <https://doi.org/10.1016/j.techfore.2017.11.022>
- Obschonka, M., Fisch, C., & Boyd, R. (2017). Using digital footprints in entrepreneurship research: A Twitter-based personality analysis of superstar entrepreneurs and managers. *Journal of Business Venturing Insights*, 8, 13–23. <https://doi.org/10.1016/j.jbvi.2017.05.005>
- Obschonka, M., Lee, N., Rodríguez-Pose, A., Eichstaedt, J. C., & Ebert, T. (2020). Big data methods, social media, and the psychology of entrepreneurial regions: capturing cross-county personality traits and their impact on entrepreneurship in the USA. *Small Business Economics*, 55(3), 567–588. <https://doi.org/10.1007/s11187-019-00204-2>
- Owen, A. L., & Wei, A. (2021). Sexism, household decisions, and the gender wage gap. *Labour Economics*, 72(September), 102062. <https://doi.org/10.1016/j.labeco.2021.102062>
- Porter, M. E. (2011). *Competitive advantage of nations: creating and sustaining superior performance*. simon and schuster.
- Poza, C., & Monge, M. (2020). A real time leading economic indicator based on text mining for the Spanish economy. Fractional cointegration VAR and Continuous Wavelet Transform analysis. *International Economics*, 163(February), 163–175. <https://doi.org/10.1016/j.inteco.2020.02.002>

- Román, C., Congregado, E., & Millán, J. M. (2013). Start-up incentives: Entrepreneurship policy or active labour market programme? *Journal of Business Venturing*, 28(1), 151–175.
- Silverstovs, B., & Wochner, D. S. (2018). Google Trends and reality: Do the proportions match?: Appraising the informational value of online search behavior: Evidence from Swiss tourism regions. *Journal of Economic Behavior and Organization*, 145, 1–23. <https://doi.org/10.1016/j.jebo.2017.10.011>
- Simoës, N., Crespo, N., & Moreira, S. B. (2016). INDIVIDUAL DETERMINANTS OF SELF-EMPLOYMENT ENTRY: WHAT DO WE REALLY KNOW? *Journal of Economic Surveys*, 30(4), 783–806. <https://doi.org/https://doi.org/10.1111/joes.12111>
- Stephens-Davidowitz, S. (2014). The cost of racial animus on a black candidate: Evidence using Google search data. *Journal of Public Economics*, 118, 26–40. <https://doi.org/10.1016/j.jpubeco.2014.04.010>
- Suddle, K., Beugelsdijk, S., & Wennekers, S. (2010). Entrepreneurial culture and its effect on the rate of nascent entrepreneurship. In *Entrepreneurship and culture* (pp. 227–244). Springer.
- Suhoy, T. (2009). *Query indices and a 2008 downturn: Israeli data*. Bank of Israel.
- Thomas, A. S., & Mueller, S. L. (2000). A case for comparative entrepreneurship: Assessing the relevance of culture. *Journal of International Business Studies*, 31(2), 287–301.
- Velilla, J., Molina, J. A., & Ortega, R. (2018). Why older workers become entrepreneurs? International evidence using fuzzy set methods. *Journal of the Economics of Ageing*, 12, 88–95. <https://doi.org/10.1016/j.jeoa.2018.03.004>
- Waters, J., Nicolaou, N., Stefanidis, D., Efstathiades, H., Pallis, G., & Dikaiakos, M. (2021). Exploring the sentiment of entrepreneurs on Twitter. *PLoS ONE*, 16(7 July). <https://doi.org/10.1371/journal.pone.0254337>
- Webb, G. K. (2009). Internet search statistics as a source of business intelligence: Searches on foreclosure as an estimate of actual home foreclosures. *Issues in Information Systems*, 10(2), 82.
- Williams, D. R. (2012). Gender discrimination and self-employment dynamics in Europe. *Journal of Socio-Economics*, 41(2), 153–158. <https://doi.org/10.1016/j.socec.2011.08.029>
- Wu, L., & Brynjolfsson, E. (2009). The future of prediction: how Google searches foreshadow housing prices and quantities. *ICIS 2009 Proceedings*, 147.

Figura 1. Mapa de calor de la brecha de género por condado de EEUU para 2019

Gender gap by County



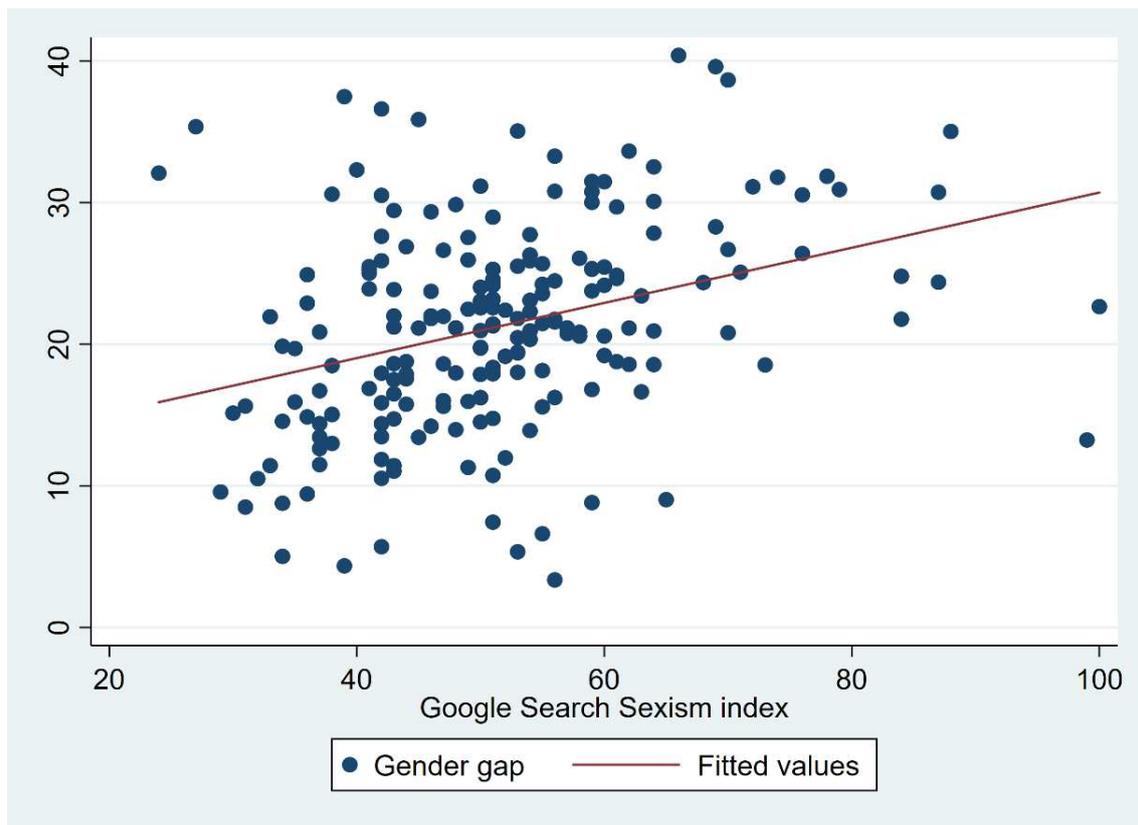
Nota: Elaboración propia a través de los datos disponibles en American Community Survey. Male self-employed (%) – Female selfemployed (%) = Gender gap

Tabla 1. Estadística descriptiva

VARIABLES	(1) mean	(2) sd	(3) max	(4) min
Unemployment rate	5.615	1.527	12.100	2.157
Google Search Sexism index	51.593	12.901	100.000	24.000
Entrepreneurial index	44.970	20.700	89.108	0.000
% Selfemployed	6.319	1.409	14.177	4.186
Selfemployed - Percentage of male	60.640	3.687	70.197	51.681
Selfemployed - Percentage of female	39.360	3.687	48.319	29.803
Gender gap	21.279	7.374	40.394	3.362
% Females, now married, except separated	47.382	4.087	56.297	22.596
Married-couple family, with own children under 18 years	17.823	2.900	29.721	6.001
College gender gap in media market	-1.168	1.049	1.392	-4.186
% Management, business, science, and arts occupations	35.299	4.420	50.119	24.589
% Service occupations	18.294	1.987	27.634	15.078
% Sales and office occupations	21.422	1.267	25.685	18.469
% Natural resources, construction, and maintenance occupations	10.139	2.273	20.797	6.301
% Production, transportation, and material moving occupations	14.846	3.520	25.604	7.103
% Agriculture	3.120	2.835	17.991	0.199
% Construction	6.778	1.171	11.890	4.513
% Manufacture	11.040	4.661	25.842	2.246
% Wholesale trade	2.405	0.462	3.753	0.979
% Retail trade	11.650	0.933	15.229	9.212
% Transportation and warehousing	5.198	1.146	13.969	2.996
% Information	1.573	0.448	4.061	0.525
% Finance and insurance, and real estate and rental and leasing	5.540	1.379	10.457	3.206
% Professional, scientific, and management, and administrative and waste management	9.140	2.603	19.621	4.913
% Educational services, and health care and social assistance	24.277	2.954	35.679	15.442
% Arts, entertainment, and recreation, and accommodation and food services	9.548	2.015	26.765	6.115
% Other services, except public administration	4.761	0.421	5.848	3.870
% Public administration	4.970	1.719	11.217	2.215
% people 25-34 years	13.133	1.245	16.888	10.274
% people 35-44 years	12.023	0.831	15.166	9.916
% people 45-54 years	12.502	0.873	14.344	10.269
% people 55-64 years	13.212	1.263	16.392	8.459
% Bachelor's degree or higher (25 years and older)	27.391	6.512	48.551	15.114
% High school graduate or higher	88.086	4.375	95.036	65.845
% of Employed	93.796	2.222	96.944	82.390
Labour force percentage	49.048	3.461	56.663	37.520
Female labour force percentage	47.159	1.520	53.414	41.696
Log number of establishments	10.024	1.215	12.766	5.403

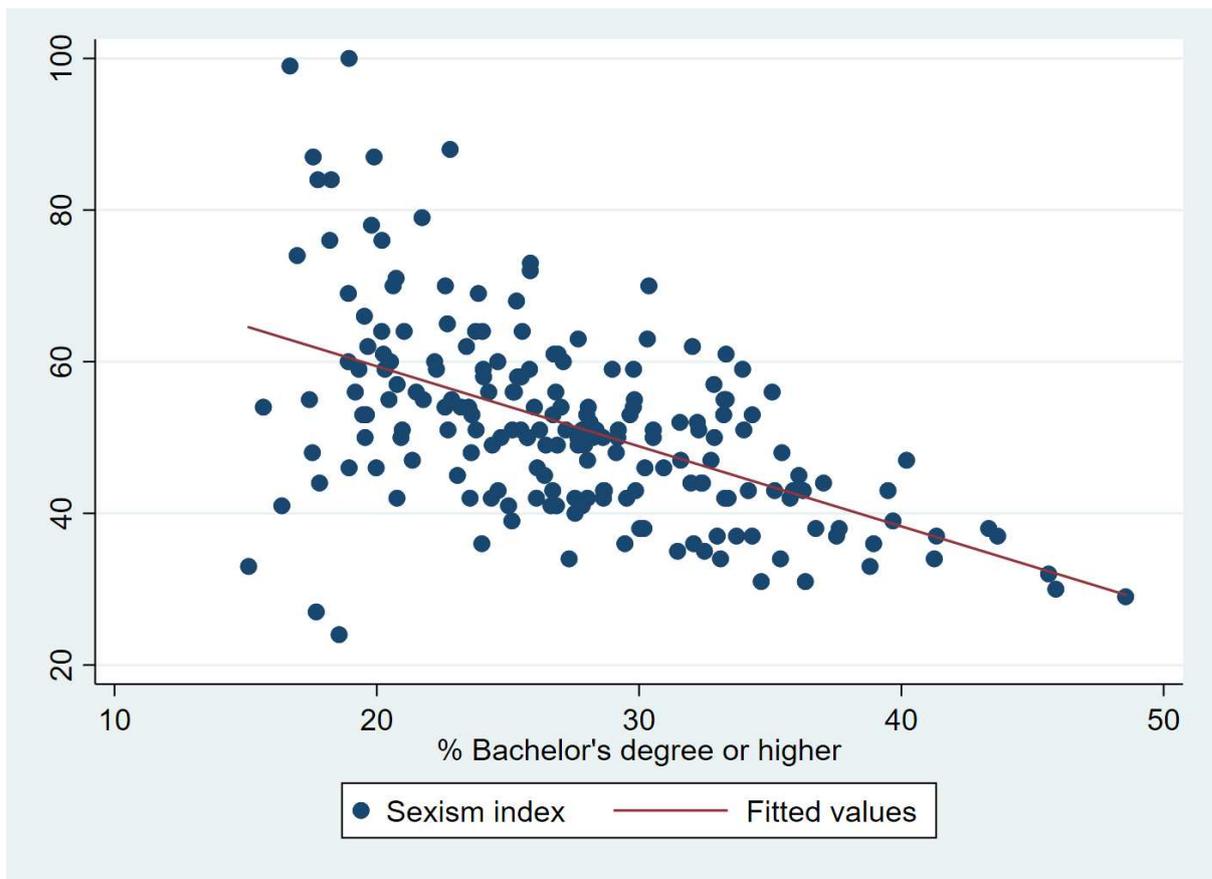
Nota: Estadística para las 199 áreas de mercado de las que disponemos de datos. El índice de sexismo está descrito en la Sección 2. Gender gap hace referencia al brecha de género en el emprendimiento entre hombres y mujeres (% hombres autónomos – % mujeres autónomas). College gap es la diferencia entre la proporción de hombres que cuentan con educación superior y la proporción de mujeres que cuentan con educación superior.

Gráfico 1. Brecha de género vs Índice de sexismo



Nota: El eje x representa el índice de sexismo explicado en la Sección 2 y el eje y la brecha de género en el emprendimiento para las 199 áreas de mercado de nuestra muestra.

Gráfico 2. Porcentaje de educación superior vs índice de sexismo



Nota: El eje x representa el índice de sexismo explicado en la Sección 2 y el eje y la proporción de población que cuenta con estudios universitarios o superiores para las 199 áreas de mercado de nuestra muestra.

Tabla 2. Predictores del índice compuesto del interés en el emprendimiento

VARIABLES	Index (1)
% people 25-34 years	-1.857 (1.469)
% people 35-44 years	-2.953 (2.921)
% people 45-54 years	5.421* (2.947)
% people 55-64 years	-7.618*** (2.425)
% Females, now married, except separated	-0.057 (0.553)
Married-couple family, with own children under 18 years	-2.321** (0.923)
Unemployment rate	-3.277*** (0.786)
College gap in media market	-1.360*** (0.316)
Constant	285.131*** (41.591)
Observations	186
R-squared	0.297

Robust standard errors in parentheses, clustered at the state level

*** p<0.01, ** p<0.05, * p<0.1

Tabla 3. Estimaciones para el índice compuesto de interés en el emprendimiento

VARIABLES	Log number of establishments				
	(1)	(2)	(3)	(4)	(5)
Entrepreneurial index	0.025*** (0.005)	0.029*** (0.004)	0.022*** (0.005)	0.024*** (0.005)	0.024*** (0.005)
College gap in media market		0.046** (0.022)	0.054 (0.037)	0.056 (0.037)	0.038 (0.037)
% people 25-34 years		-0.037 (0.143)	-0.153 (0.134)	-0.123 (0.119)	-0.171 (0.148)
% people 35-44 years		0.429* (0.216)	0.323 (0.216)	0.407** (0.201)	0.404* (0.224)
% people 45-54 years		0.146 (0.180)	0.099 (0.226)	0.099 (0.201)	0.047 (0.209)
% people 55-64 years		0.203 (0.191)	0.100 (0.212)	0.169 (0.199)	0.163 (0.214)
% Females, now married, except separated		-0.007 (0.044)	-0.020 (0.045)	-0.032 (0.042)	-0.007 (0.048)
Married-couple family, with own children under 18 years		0.100 (0.080)	0.063 (0.071)	0.095 (0.073)	0.043 (0.070)
% of Employed		0.028 (0.040)	-0.009 (0.046)	-0.014 (0.044)	-0.019 (0.048)
Constant	8.837*** (0.295)	-7.398* (4.300)	-0.041 (6.118)	2.816 (6.710)	6.297 (8.630)
Observations	186	186	186	186	186
R-squared	0.126	0.288	0.368	0.314	0.386
Occupation Controls	NO	NO	NO	YES	YES
Industry Controls	NO	NO	YES	NO	YES

Robust standard errors in parentheses, clustered at the state level

*** p<0.01, ** p<0.05, * p<0.1

Tabla 4. Estimaciones para el índice de sexismo local y la brecha de género

VARIABLES	Entrepreneurship Gender gap					
	(1)	(2)	(3)	(4)	(5)	
Google Search Sexism index	0.195*** (0.070)	0.211*** (0.056)	0.050 (0.041)	0.106** (0.044)	0.028 (0.042)	
% of Employed		0.129 (0.357)	0.145 (0.272)	0.073 (0.302)	0.145 (0.296)	
Labour force percentage			0.013 (0.197)	0.006 (0.187)	-0.155 (0.243)	
Female labour force percentage		-0.313 (0.667)	-0.608 (0.609)	-0.200 (0.637)	-0.868 (0.685)	
College gender gap in media market		-	1.875*** (0.593)	-0.674 (0.560)	-1.654** (0.637)	-0.659 (0.604)
% people 25-34 years		-1.495** (0.735)	-0.103 (0.728)	-0.467 (0.822)	-0.212 (0.759)	
% people 35-44 years		-	3.931*** (1.312)	-2.555** (1.257)	3.507*** (1.216)	-2.649* (1.398)
% people 45-54 years		2.353* (1.285)	2.480** (1.200)	1.469 (1.403)	2.544** (1.233)	
% people 55-64 years		-2.719* (1.353)	-0.895 (1.148)	-0.904 (1.386)	-0.668 (1.176)	
% Females, now married, except separated		-0.427 (0.296)	-0.744** (0.295)	-0.515** (0.244)	-0.788** (0.310)	
Married-couple family, with own children under 18 years		0.562 (0.592)	0.778* (0.405)	0.845 (0.503)	0.870* (0.435)	
Constant	11.230*** (3.904)	94.506** (38.600)	48.421 (36.123)	62.446 (49.769)	20.486 (47.626)	
Observations	199	199	199	199	199	
R-squared	0.116	0.349	0.533	0.442	0.545	
Occupation Controls	NO	NO	NO	YES	YES	
Industry Controls	NO	NO	YES	NO	YES	

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Apéndice 1: índice de sexismo local

Código de la palabra ofensiva	Palabra	Código
	whore, whores	Palabra 1
