

# **Efectos del COVID-19 en búsquedas de empleo en Colombia**

Diego Alejandro Contreras López

Tesis de Maestría

Maestría en Economía

Universidad Nacional de La Plata

Director de tesis: Mariana Viollaz

Co-director de tesis: Mariana Marchionni

Noviembre, 2022

Códigos JEL: J22, J64, N36..

# Efectos del COVID-19 en búsquedas de empleo en Colombia.<sup>†</sup>

Diego Alejandro Contreras López<sup>††</sup>

## Resumen

En este documento estudio el efecto de la pandemia del COVID-19 en búsquedas de empleo usando datos de Google Trends para 24 departamentos de Colombia. Utilizando los novedosos métodos de Sun y Abraham, 2021 y Gardner, 2021, además de Estudio de Eventos (ES), estudio dos tratamientos escalonados: la fecha de reporte del primer caso de COVID-19 y la fecha de la primera muerte para cada departamento. Los *outcomes* son las búsquedas en Google de palabras clave relacionadas con páginas web de servicios de búsqueda de empleo y el promedio de estas búsquedas. Existe una heterogeneidad en los resultados, dependiendo de la palabra clave usada como *outcome*. Encuentro aumentos entre 50% y 75% promedio de búsquedas en la palabra “computrabajo”, cuyo efecto se mantiene hasta 30 semanas después de los tratamientos. Estos efectos son mayores y más precisos en departamentos con mayor calidad de vida y en departamentos que reportaron el primer caso de COVID-19 antes de la cuarentena. Adicionalmente, aplique un modelo autorregresivo de las búsquedas de “computrabajo” sobre tasas de desempleo de varios grupos demográficos y encuentro que, en promedio y en pandemia, las búsquedas estarían asociadas con una caída en 0.7 puntos porcentuales del desempleo femenino, una caída de la tasa de desempleo juvenil de 0.5 p.p y la baja de 0.67 p.p de la tasa de desempleo masculino.

**Códigos JEL:** J22, J64, N36.

**Palabras clave:** coronavirus, Colombia, búsqueda de empleo, Google Trends, tratamiento escalonado.

---

<sup>†</sup>Este trabajo constituye la tesis de Maestría en Economía de la Universidad Nacional de La Plata, realizada bajo la dirección de Mariana Viollaz y Mariana Marchionni.

<sup>††</sup>Economista, Universidad Nacional de Colombia. Estudiante de Maestría en Economía, Universidad Nacional de La Plata. E-Mail: diacontreraslo@unal.edu.co. Cualquier error u omisión es de exclusiva responsabilidad del autor.

# 1. Introducción

La pandemia causada por el virus COVID-19 ocasionó una crisis sanitaria, económica y social en todo el planeta. Las cuarentenas promulgadas por la mayoría de los gobiernos, con el objetivo de reducir la movilidad de la gente y mitigar el contagio del virus, provocaron una caída en la actividad económica y el subsiguiente aumento del desempleo, además de un cambio en los sitios de trabajo, ya que muchos trabajadores tuvieron que pasar al trabajo en casa. El día 23 de marzo de 2020, el Gobierno Nacional de Colombia decretó una cuarentena a nivel nacional, llamado “Aislamiento Preventivo Obligatorio”, donde prohibió la circulación de personas excepto en 34 casos<sup>1</sup>. Originalmente, esta medida estaría vigente por 19 días, pero fue extendida varias veces hasta el 30 de agosto, con aperturas graduales según las afectaciones en los municipios. Según el Departamento Administrativo Nacional de Estadística (DANE), en el segundo trimestre de 2020 el PIB decreció un 15.7% con respecto al mismo trimestre del año anterior<sup>2</sup>. En abril, la economía tuvo su peor desempeño, con una caída del 20.1% anual y las actividades más afectadas fueron las actividades artísticas, de entretenimiento y recreación y otras actividades de servicios (-37,1%), comercio al por mayor y al por menor (-34,3%) y construcción (-31,7%), aunque hubo una recuperación lenta de los sectores conforme se flexibilizaron las medidas por parte del Gobierno (Urrea-Ríos & Piraján, 2020).

En este trabajo propongo evaluar los efectos de la pandemia causada por el COVID-19 en las búsquedas de empleo en Colombia. Para esto, utilizo datos de Google Trends de 24 departamentos y la ciudad de Bogotá en palabras claves de búsqueda de empleo como “computrabajo”, “elempleo” o “linkedin”. Las búsquedas son por semana, desde el primero de enero de 2017 hasta el 31 de diciembre de 2020. La pandemia y las medidas de contención ordenadas por los gobiernos menguaron la producción en muchos sectores, lo que llevaría a los trabajadores a buscar un nuevo empleo, a cambiar el empleo actual o a cambiar la estrategia de búsqueda de empleo de forma presencial a virtual por no poder salir o por querer evitar el contagio.

Este trabajo es relevante en el sentido que es un complemento a la información oficial sobre el empleo. El DANE recolecta datos sobre el mercado laboral y la tasa de desempleo, pero no detalla con precisión cómo es la búsqueda de empleo de los individuos. Adicionalmente, la posibilidad de hacer búsquedas anónimas en Google eliminaría posibles sesgos de respuesta que tengan individuos al responder encuestas, a cambio de no tener información demográfica de los individuos.

Debido a la pandemia y sus respectivas cuarentenas ordenadas por los gobiernos, fue más

---

<sup>1</sup><https://id.presidencia.gov.co/Paginas/prensa/2020/Para-garantizar-derecho-vida-salud-supervivenciadurante-Aislamiento-Preventivo-Obligatorio-Decreto-457-permite-dere-200323.aspx>

<sup>2</sup>[https://www.dane.gov.co/files/investigaciones/boletines/pib/bol\\_PIB\\_Itrim20\\_produccion\\_y\\_gasto.pdf](https://www.dane.gov.co/files/investigaciones/boletines/pib/bol_PIB_Itrim20_produccion_y_gasto.pdf)

complicada la recolección de datos por parte de los institutos de estadística en más de 150 países. Esto llevó a suspensiones de encuestas o cambios en la estrategia de recolección de datos de forma presencial a virtual, telefónica o mixta, o utilizando datos administrativos, lo que provocaría rupturas en series y cambios en procesos de estimación<sup>3</sup>. En el caso colombiano, el DANE hizo cambios en el mecanismo de recolección de datos de la Gran Encuesta Integrada de Hogares (GEIH), donde redujeron el cuestionario de 200 a 39 preguntas, haciendo encuestas presenciales en zonas rurales y telefónicas en áreas urbanas, además de evaluar la posibilidad de datos tipo panel<sup>4</sup>. Esta situación justifica el uso de datos novedosos para entender los efectos de la pandemia en variables económicas y sociales y para complementar la información oficial. Por otra parte, a mi conocimiento, no existen estudios como este en Colombia, por lo que le da un carácter novedoso a este estudio.

El trabajo se organiza de la siguiente manera. En la sección 2 se revisan algunos trabajos anteriores sobre búsquedas en Google Trends, el COVID-19 y su efecto en diferentes variables. La sección 3 explica los datos utilizados y la estrategia de identificación. La sección 4 detalla los resultados del estudio y la sección 5 discute resultados y concluye.

## 2. Revisión literaria

La pandemia del COVID-19 dio origen a varios estudios que usaron datos de internet, ya sea por los cambios en recolección y estimación de datos oficiales, causada por las medidas de confinamiento y su consiguiente dificultad para recopilar estadísticas o por la facilidad de adquirir datos en internet gracias al paso a la virtualidad. En Brodeur et al., 2021 evalúan el efecto del COVID-19 en la salud mental de las personas. Utilizan datos de Google Trends de palabras clave de bienestar como “aburrimiento, soledad, irritabilidad y suicidio” para diferentes países. Como metodología, utilizan diferencias en diferencias (DiD) y regresión discontinua (RD), basándose en el supuesto que, ante la ausencia de COVID-19, el comportamiento de la búsqueda de estas palabras clave hubiese sido similar al año anterior. Realizan un proceso de escalamiento para poder medir adecuadamente la serie de cada variable. Encuentran un aumento significativo en la búsqueda de palabras como “aburrimiento”, “soledad”, “preocupación” y “tristeza” en Estados Unidos y Europa.

Para saber las reacciones de las personas, Bento et al., 2020 realizan un estudio de eventos en Estados Unidos con datos de Google para evaluar las búsquedas de la gente después del anuncio oficial de los primeros infectados por COVID-19. Los autores buscan palabras claves

---

<sup>3</sup><https://ilostat.ilo.org/topics/covid-19/covid-19-impact-on-labour-market-statistics/>

<sup>4</sup>[https://www.dane.gov.co/files/investigaciones/boletines/ech/ech/CP\\_ajustesmetodologicos\\_covid19\\_feb\\_20.pdf](https://www.dane.gov.co/files/investigaciones/boletines/ech/ech/CP_ajustesmetodologicos_covid19_feb_20.pdf)

relacionadas con el virus, sus síntomas y desinfectantes de manos. Búsquedas por “coronavirus” aumentaron en un 36%. Además, los autores afirman que los reportes oficiales de primeros casos no afectan las búsquedas de palabras clave relacionadas con la idea de si la pandemia puede ser un montaje. Chin y Wang, 2020 comparan las búsquedas en Google Trends de “face masks” en Taiwán, Singapur, Filipinas, Italia y España, con el número de casos de infectados con COVID-19 en estos países. No realizan un estudio de evaluación de impacto o regresiones. Afirman que Google Trends puede ser usado para políticas de salud por su tiempo real. En esta misma línea de implementos sanitarios, Chakraborti y Roberts, 2020 estudian el efecto de las leyes anti-especulación en las búsquedas de “papel higiénico”, “limpiador de manos” y “loción de manos” como placebo en Google Trends, entre febrero y marzo de 2020 en estados de Estados Unidos. Encuentran un aumento de hasta 40% comparado con estados sin leyes en contra de la especulación.

Existen trabajos en Latinoamérica y el Caribe (LAC) que utilizan Google Trends como fuente de datos. Por ejemplo, Silverio-Murillo et al., 2021 buscan el efecto de las cuarentenas en las búsquedas asociadas a la salud mental, con las palabras insomnio, estrés, ansiedad, tristeza, depresión y suicidio en 11 países latinoamericanos (Argentina, Bolivia, Chile, Colombia, Ecuador, Guatemala, Honduras, México, Paraguay, Perú y Uruguay). Usando estudio de eventos y DiD, encuentran que las búsquedas por insomnio aumentan, pero luego caen a lo largo del tiempo. También hallan que las búsquedas por estrés, ansiedad y tristeza aumentan y se mantienen constantes durante la cuarentena y no hay cambios significativos en las búsquedas de depresión o suicidio.

Berniell y Facchini, 2021 estudian el impacto de la pandemia en la violencia doméstica en 11 países: Estados Unidos, Argentina, Brasil, Chile, Colombia, México, Francia, Alemania, Italia, España y Reino Unido. Usan palabras relacionadas con “violencia doméstica” y líneas de atención a la mujer en cada país. Realizan un estudio de eventos y un modelo DiD, donde encuentra un aumento de búsquedas de entre 25% a 85% y el efecto se mantiene durante las primeras 10 semanas de cuarentena. Repiten el estudio, pero separando entre países desarrollados y en desarrollo, donde encuentran que el efecto de las cuarentenas en las búsquedas es más del doble en países en desarrollo comparado con países desarrollados, lo que puede ser dado por un menor cumplimiento de las medidas o por una mayor respuesta de los violentadores a las medidas de cuarentena en países desarrollados.

En economía laboral también crece el uso de Google Trends, en especial como fuente de datos para predecir la tasa de desempleo. Un artículo seminal de *nowcasting* es el de Choi y Varian, 2012, donde proponen un método utilizando datos de Google Trends para “predecir”

datos oficiales que se publican con rezago. Para esto, usan un modelo autorregresivo controlando por datos de palabras clave de Google Trends y muestran varios ejemplos en precios de vivienda, turismo y ventas de autos.

De Choi y Varian, 2012 surgen trabajos como el de Smith, 2016, donde realiza una predicción de la tasa del desempleo en el Reino Unido utilizando Google Trends con palabras como “redundancy”, un seguro de desempleo inglés, utilizando el ARIMA propuesto por Choi y Varian, 2012 y, como novedad, un modelo MIDAS. Encuentra menores errores de estimación en el modelo MIDAS y propone métodos para lidiar con la temporalidad de las series de Google Trends contra las series mensuales oficiales. O González-Fernández y González-Velasco, 2018, donde sigue el trabajo de Choi y Varian, 2012 con Google Trends con la palabra “desempleo” en España. Cardona Rojas y Rojas, 2017 utilizan datos de Colombia para un periodo pre-pandemia. Emplean modelos de regresión lineal simple, ARIMA y ARIMAX para predecir la tasa de desempleo en Colombia con diferentes términos relacionados al empleo de Google Trends hasta el 2017. Encuentran que los mejores predictores son las búsquedas de “Busco trabajo” y “Trabajo”.

Lenaerts et al., 2016 realizan una revisión de páginas de internet como potenciales buscadores de trabajo, fuentes de datos y trabajos realizados alrededor de estas páginas. Las páginas son Google Trends, LinkedIn, Facebook, Twitter y Glassdoor. Resalta el alto uso de LinkedIn como plataforma de búsqueda de empleo y como fuente de datos para artículos, mientras que Facebook es poco usado para conseguir empleo y Twitter sirve en buena parte como complemento para los reclutadores para publicar ofertas laborales. En cambio, Google Trends se usa para mostrar tendencias y hacer predicciones de tasa de desempleo, como se observa en la literatura existente.

Aaronson et al., 2020 realizan un estudio de eventos con estados afectados por huracanes para obtener un *nowcasting* de la búsqueda en Google Trends sobre las solicitudes de seguro de desempleo. Además, emparejan al grupo tratamiento con estados no afectados por huracanes.

Con datos de Google Trends, Dilmaghani, 2019a busca predecir el desempleo de jóvenes entre 16 y 24 años en Estados Unidos, dividiendo blancos, negros y latinos. Controlando por búsquedas de actividades de ocio, el autor encuentra que las predicciones de la tasa de desempleo son mayores para los blancos, comparados con negros y latinos, pero que las actividades de ocio son significativas en los latinos, mejorando la predicción para este grupo. Dilmaghani, 2019b realiza una metodología similar a Dilmaghani, 2019a, donde busca predecir el desempleo de Canadá usando palabras clave de páginas de búsqueda de empleo, controlando por búsquedas de sitios de ocio. Encuentra que los modelos que controlan por ocio tienen un mejor desempeño. Adicional, muestra sugerencias para manejar datos de Google Trends.

### 3. Datos y metodología

#### 3.1. Datos

Google Trends es una herramienta diseñada por Google<sup>5</sup>, donde se analiza un porcentaje de las búsquedas en la web para determinar cuántas búsquedas se han hecho durante un periodo de tiempo dado. Se determina un indicador de una palabra clave en una región por un periodo que elige el usuario, generando un índice. Acorde con Google, cada punto del índice es el número de búsquedas de ese día dividido por el número de búsquedas más alto del rango de la fecha dada, este índice está normalizado en una escala de 0 a 100, donde 100 es el día donde más búsquedas se han hecho sobre dicha palabra clave. Cuando se hacen búsquedas de varias palabras clave al mismo tiempo, el índice cambia, siendo ahora normalizado por el número de búsquedas más alto de la palabra más buscada. A modo de ejemplo: si se buscan “manzana”, “naranja”, “pera”, “mango” y “guayaba” y la palabra más buscada fuese “manzana”, entonces el índice de “naranja” sería el número de búsquedas de “naranja” sobre el número de búsquedas máximo de “manzana”. Por esta razón, es posible que una palabra clave no tome un valor de 100, comparada con otras. Google Trends excluye búsquedas realizadas por muy pocas personas, donde asigna un valor de 0. También elimina búsquedas duplicadas y con caracteres especiales.

Para este estudio, utilizo las siguientes palabras claves asociadas a la búsqueda de empleo: “computrabajo”, “empleo”, “linkedin”, “indeed”, “empleo sin experiencia” y “servicio de empleo”. Algunas de estas palabras corresponden a páginas web que buscan hacer el matching entre empresas y buscadores de empleo, mientras que otras páginas pertenecen a sitios donde se prestan servicios a los buscadores de empleo<sup>6</sup>. Con el paquete de R *GtrendsR*<sup>7</sup> descargo los datos de Google Trends desde el 01 de enero de 2017 hasta el 31 de diciembre del 2020 de 24 regiones de Colombia más la ciudad de Bogotá. El programa descarga los índices de búsquedas de hasta 5 palabras a la vez, en el departamento o la ciudad que se le especifique, sin embargo, para este estudio, las búsquedas se descargan de forma separada para cada palabra clave. Los datos se descargan por año. La frecuencia de los datos es semanal, lo que permite formar una base de datos en panel para el estudio. Elijo estas 24 regiones porque son un proxy de las regiones que identifica el Departamento Administrativo Nacional de Estadística (DANE) como cabeceras urbanas en la recolección de datos. Aquellas semanas cuya búsqueda es muy pequeña, Google Trends asigna un valor de  $> 1$ , lo que genera *missing values*. A estos *missing values* les asigno

---

<sup>5</sup>Disponible en <https://trends.google.es/trends/>

<sup>6</sup>Según el sitio [similarweb.com](http://similarweb.com), el cual monitorea el tráfico de paginas web en el mundo, “computrabajo”, “indeed” y “empleo” son las páginas de empleo con mayor tráfico en Colombia. La tabla A1 muestra una descripción de las palabras usadas.

<sup>7</sup>Disponible en <https://github.com/PMassicotte/gtrendsR>

un valor de 0. Para obtener una perspectiva general de las búsquedas de empleo en Colombia, genero un promedio con estas 6 palabras de búsqueda, similar a Bacher-Hicks et al., 2021. Con esto, se pretende dar una visión más ampliada del comportamiento de las búsquedas y mitigar posibles búsquedas nulas en algunos departamentos de determinadas palabras clave.

La tabla 1 muestra las estadísticas descriptivas de las palabras claves de búsqueda para las 24 regiones y del promedio de las palabras clave, llamado “meanJobsGS”. Las palabras “computrabajo”, “servicio de empleo” y “linkedin” tienen un mayor número de búsquedas en promedio. En cambio, “empleo sin experiencia” e “indeed” tienen el promedio más bajo entre las palabras consideradas. En el caso del promedio general de las palabras (meanJobsGS), el promedio es de 29.93, mientras que el p(25) es de 17.16 y el p(75) es de 40. Todas las palabras tienen un valor mínimo de 0 y un valor máximo de 100, excepto por meanJobsGS (97.5). Las palabras clave con mayor variabilidad son linkedin, elempleo y servicio de empleo, con una desviación estándar de 30.6, 28.66 y 28.05 respectivamente.

**Tabla 1:** Estadísticas descriptivas de palabras clave de búsqueda.

Palabra Clave	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
meanJobsGS	5,016	29.932	17.150	0	17.167	40.333	97.500
computrabajo	5,016	50.076	20.418	0	38	62	100
elempleo	5,016	25.802	28.662	0	0	46	100
linkedin	5,016	27.844	30.685	0	0	51	100
servicio de empleo	5,016	38.746	28.057	0	17	60	100
empleo sin experiencia	5,016	18.192	27.347	0	0	36	100
indeed	5,016	18.934	27.052	0	0	36	100

Fuente: Google Trends, cálculos propios.

La figura 1 muestra el comportamiento semanal del promedio de las palabras claves en los años 2017 a 2020. La serie morada muestra el año 2020 y la línea naranja representa las búsquedas en el año 2019. Los años 2017 y 2018 están reflejados en líneas grises. La línea roja vertical es de la semana del 22 de marzo de 2020, fecha donde el gobierno declaró la cuarentena general en Colombia. La fecha de la cuarentena fue la misma para todos los departamentos del país. Las líneas azules verticales son las semanas de la menor y mayor fecha de los tratamientos que se aplicarán en este estudio; la fecha del primer caso reportado y de la primera muerte reportada en cada departamento. Se observa un descenso generalizado de las búsquedas relacionadas con el empleo durante las 10 semanas previas a la cuarentena, seguramente dado por un efecto estacional<sup>8</sup>. El promedio de las palabras tiene una caída en las búsquedas durante las primeras

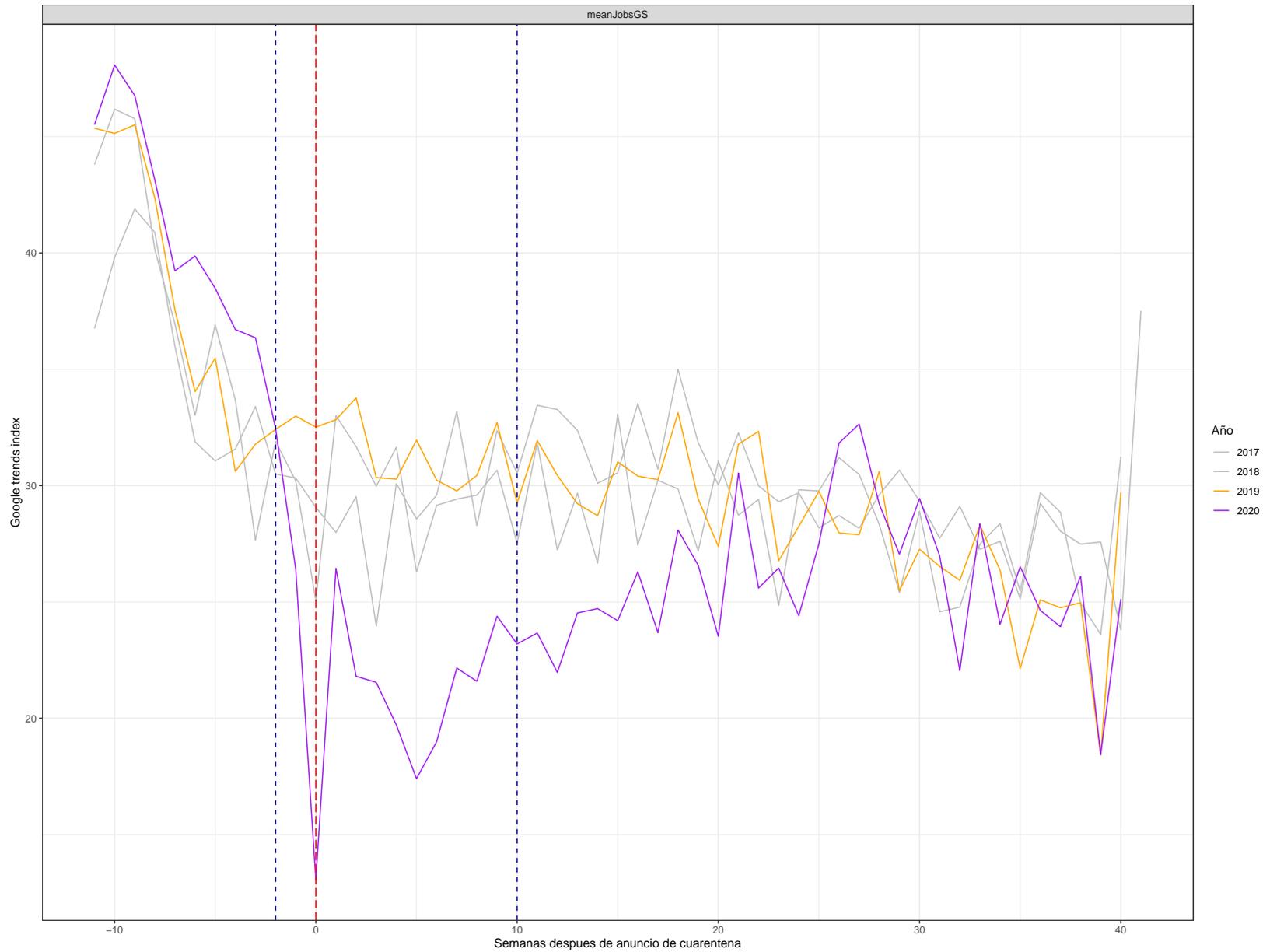
<sup>8</sup>Véase Arango Thomas y Posada, 2001 y Henao y Rojas, 1999 para estudios sobre tasa de desempleo en Colombia.

10 semanas en años anteriores, pero la caída es más pronunciada en el año 2020. Posterior a la cuarentena, se observa una senda de recuperación en el año 2020. Las búsquedas de 2020 parecen retomar valores similares a años anteriores a partir de la semana 20, donde hay un patrón estable.

Una versión desagregada del comportamiento de las palabras clave de búsqueda está en el gráfico [A1](#). Computrabajo es la palabra clave con la caída más pronunciada de búsquedas con respecto a años anteriores, retornando a su senda hasta la semana 25. Empleo y empleo sin experiencia presentan caídas en su índice de búsqueda, pero no tan notorias como Computrabajo. Para el caso de Indeed y LinkedIn, se observa un comportamiento de la serie en el año 2020 similar a años anteriores. Servicio de empleo muestra una tendencia similar a Computrabajo, excepto por un pico una semana después de la cuarentena y un retorno a la senda cercano a la semana 30.

Estos comportamientos en las búsquedas estarían dados posiblemente por la cuarentena, ya que las personas pudieron estar más enfocados en mantenerse en cuarentena para no exponerse al virus y preservar su salud que en buscar empleo. También es posible que la gente que no pudo trabajar desde casa ni pudo salir se hayan resignado y decidieron no buscar empleo.

Existen problemas con el uso de Google Trends como fuente de datos. La información demográfica de los individuos que buscan en Google no está disponible. No es posible saber su nivel de ingreso, sexo, edad u otras variables por las cuales se pueda controlar y que pueden ser relevantes a la hora de buscar empleo. Tampoco es posible saber si la persona que busca empleo en la web está empleada o no. Adicionalmente, la fiabilidad de los datos depende de la cobertura de internet, donde es mayor en regiones urbanas relativo a las zonas rurales y que existe un sesgo del uso de internet hacia las personas más jóvenes, por lo que se podría subestimar a individuos de mayor edad que carezcan de afinidad y conocimientos de ofimática, además de computadores o celulares para buscar empleo.



**Figura 1:** Búsquedas promedio de palabras clave de empleo de Google Index, 2017 - 2020. El gráfico muestra el promedio de búsquedas de las palabras clave relacionadas con búsquedas de empleo para 24 departamentos por semana para el año 2020 (en púrpura), 2019 (en naranja) y los años anteriores en gris. La línea roja punteada vertical representa la semana que dictaron la cuarentena en Colombia. Las líneas azules punteadas representan el intervalo de semanas donde ocurren los tratamientos de este estudio en todos los departamentos. Fuente: Google Trends, cálculos propios.

### 3.2. Estrategia de identificación

La pandemia y las medidas de contención ordenadas por los gobiernos detuvo o mermó la producción en muchos sectores, aumentando el desempleo, lo que llevaría a los trabajadores a buscar un nuevo empleo. Sin embargo, existe la posibilidad que las personas por querer cuidar su salud y evitar contagiarse, se hayan enfocado en mantenerse en cuarentena y hayan dejado de buscar empleo. Por lo tanto, no es claro cual pueda ser el efecto esperado en las búsquedas de empleo. Para hallar los posibles efectos, siguiendo a Bacher-Hicks et al., 2021<sup>9</sup>, empleo el método de Estudio de Eventos (ES, por sus siglas en inglés), donde el modelo toma la siguiente forma:

$$\text{Log}(Y_{i,t}) = \alpha_i + \lambda_t + \sum_{\tau=-5}^{-2} \gamma_{\tau} D_{i,t}^{\tau} + \sum_{\tau=0}^{40} \gamma_{\tau} D_{i,t}^{\tau} + \beta X_{i,t} + \epsilon_{i,t} \quad (1)$$

Donde  $Y_{i,t}$  es el índice de Google Trends para cada una de las palabras clave buscadas en el departamento  $i$  y la semana  $t$ ,  $D_{i,t}$  es un indicador para la semana  $t$  que toma valor de uno si la semana  $t$  se ubica antes o después de la semana del tratamiento,  $\alpha_i$  son dummies de efectos fijos por departamento y  $\lambda_t$  son dummies de efecto fijo por semana. Asigno dos tratamientos. Primero, el día que fue detectado el primer caso por COVID-19 en cada uno de los departamentos y segundo, el día que fue reportada la primera muerte por este virus en cada uno de los departamentos<sup>10</sup>.

Excluyo  $\tau = -1$  como categoría omitida, lo que permite comparar todos los  $\gamma_{\tau}$  con respecto a la semana anterior de cada tratamiento y los cambios semanales en la intensidad de búsqueda durante 2020 con respecto a los tres años anteriores. Incluyo también un vector de controles  $X_{i,t}$ , donde están dummies por cada año y una dummy que toma valor de 1 cuando inició la cuarentena en Colombia, es decir, el 22 de marzo de 2020.

Aunque la fecha de cuarentena sería el tratamiento más adecuado ya que es exógena para cada departamento, no es posible utilizarla por la unicidad de este tratamiento, ya que esta ocurrió en todo el territorio nacional, lo que imposibilita diferenciarlas por departamentos. Es decir, impide comparar los que han sido tratados de los que no, al no tener un grupo de control. Por esta razón elijo estos dos tratamientos, por que dan la posibilidad de comparar acorde con la semana en la que fueron tratados los departamentos. Existe la posibilidad que los dos tratamientos sean endógenos. Por ejemplo, es posible que, en los departamentos con mejor conectividad, a mayores búsquedas en Google lleven a retardar las muertes en estos departamentos. Espero mitigar el posible problema de endogeneidad con los ejercicios de heterogeneidades que se presentan más

<sup>9</sup>Bento et al., 2020; Berniell y Facchini, 2021 realizan trabajos similares con algunas variantes.

<sup>10</sup>Las fechas de los tratamientos de cada departamento se encuentran en la tabla A2 del apéndice.

adelante en este trabajo.

El gráfico A2 del apéndice muestra la semana de cada tratamiento por departamento. La línea roja vertical representa la semana de inicio de la cuarentena nacional. Los primeros departamentos en reportar el primer caso de coronavirus fueron Antioquia, Bogotá y Valle del Cauca, dos semanas antes de la cuarentena en tanto que los primeros departamentos que reportaron la primera muerte fueron Bolívar, Huila y Valle, quienes lo hicieron la misma semana en que inició la cuarentena.

Es conocido por la literatura actual que, con un tratamiento escalonado en los individuos, los coeficientes estimados por el modelo ES pueden no mostrar un promedio ponderado de los efectos del tratamiento. ES estima comparaciones “limpias” entre tratados y no tratados, pero también hace comparaciones “prohibidas” con unidades que ya han sido tratadas. Si el efecto del tratamiento es heterogéneo a través de las unidades o el tiempo, estas comparaciones “prohibidas” pueden afectar los coeficientes ES, inclusive hasta posibles cambios de signo, debido a problemas de ponderación negativa (Gardner, 2021; Roth et al., 2022; Sun & Abraham, 2021).

Para solucionar este problema con los estimadores ES y siguiendo a Roth et al., 2022, realicé estimaciones con dos métodos para tratamientos escalonados. Primero, utilicé el método de Gardner, 2021, donde propone una estimación por dos etapas, bajo el supuesto de tendencias paralelas y no anticipación. La primera etapa consiste en la estimación de este modelo:

$$Y_{gpit} = \lambda_g + \gamma_t + \varepsilon_{gpit} \quad (2)$$

Para la unidad  $i$  del grupo  $g$  en el tiempo  $t$  del periodo  $p$ , con la muestra de observaciones no tratadas o aun no tratadas, es decir  $D_{gp} = 0$ , reteniendo las estimaciones de los efectos del grupo y del tiempo ( $\hat{\lambda}_g$  y  $\hat{\gamma}_t$ ) para generar los outcomes ajustados  $\tilde{Y}_{igt} \equiv Y_{igt} - \hat{\lambda}_g - \hat{\gamma}_t$ . La segunda etapa consiste en hacer una regresión de los outcomes ajustados  $\tilde{Y}_{gpit}$  en el tratamiento  $D_{gp}$  para estimar los efectos del tratamiento<sup>11</sup>.

El segundo método es el de Sun y Abraham, 2021, quienes presentan una descomposición del estimador ES en un ATT por cohortes, llamado por las autoras CATT, estimado de la siguiente forma:

$$CATT_{e,l} = E[Y_{i,e+l} - Y_{i,e+l}^{\infty} | E_i = e] \quad (3)$$

En donde cada  $CATT_{e,l}$  corresponde al efecto promedio del tratamiento (ATE, por sus siglas

---

<sup>11</sup>A modo de ilustración y por facilidad de implementación, selecciono estos dos métodos sobre otras opciones recomendadas por Roth et al., 2022. Para el método de Gardner, 2021 utilicé el programa *did2s* de Butts, 2021 en R. Para los métodos de ES y Sun y Abraham, 2021, uso *fixest* de Bergé, 2018 en R.

en inglés) de  $l$  periodos desde el tratamiento inicial para la cohorte de unidades  $i$  tratadas primero en el tiempo  $e$  (Sun & Abraham, 2021). Los supuestos que usan son el de tendencias paralelas en todos los grupos, no anticipación relevante del tratamiento y homogeneidad en el efecto del tratamiento, es decir que cada grupo tiene el mismo perfil de tratamiento. Luego, proponen un estimador de interacciones (IW) ponderadas, donde las ponderaciones son participaciones de las cohortes que pasan por al menos  $l$  periodos relativos al tratamiento, normalizados por el tamaño del grupo  $g$ . El método permite el uso de covariables. El estimador se calcula de la siguiente forma:

1. Estiman una regresión con ES usando nunca tratados o la última cohorte tratada como controles, excluyendo indicadores de cohortes de un conjunto  $C$ . La regresión toma la siguiente forma:

$$Y_{i,t} = \alpha_i + \lambda_t + \sum_{e \notin C} \sum_{l \neq -1} \delta_{e,l} (1\{E_i = e\} \cdot D_{i,t}^l) + \epsilon_{i,t}$$

2. Estimar los ponderadores  $Pr(E = e)$  con muestreos de cada cohorte en cada periodo.

3. Estimar IW tomando el promedio ponderado de todos los estimadores para CATT estimados del paso 1, multiplicados por los ponderadores estimados del paso 2.

$$\hat{v}_g = \frac{1}{|g|} \sum_{l \in g} \sum_e \hat{\delta}_{e,l} \hat{Pr}(E_i = e | E_i \in [-l, T - l])$$

### 3.3. Heterogeneidades

Se entiende por brecha digital a las diferencias en el uso y manejo de las Tecnologías de la Información y las Comunicaciones (TIC) entre los diferentes departamentos del país. Esta brecha puede estar dada por condiciones preexistentes en las regiones o los ciudadanos; nivel de ingreso, nivel educativo, si el departamento es urbano o rural, entre otros (MinTIC, 2021). Colombia ha tenido una leve reducción de la brecha digital durante los últimos años, sin embargo, el Índice de Brecha Digital del 2021 fue de 0.41, un valor alto<sup>12</sup>.

Habida cuenta de la brecha digital en Colombia, resulta interesante evaluar el efecto del coronavirus en las búsquedas relacionadas con el empleo en departamentos con un alto acceso a las TIC y en aquellos con un nivel bajo de acceso, ya que los resultados podrían diferir gracias a la brecha digital. Para esto, divido la muestra de departamentos en dos y, siguiendo a Bacher-Hicks et al., 2021, utilizo un indicador que me permite diferenciar los departamentos con un

<sup>12</sup>Estimado por el Ministerio de las Tecnologías de la Información y las Comunicaciones (MinTIC), el índice estima la brecha digital en un rango de 0 a 1, donde valores más cercanos a cero implican una menor brecha digital (MinTIC, 2021).

nivel de vida alto de los que tienen un nivel de vida bajo. Este indicador lo estimo con el método de Análisis de Componentes Principales (ACP), donde, con datos de la GEIH de 2016, incluyo las variables de ingreso promedio, ingreso mediano, años promedio de educación, tasa de acceso a la educación superior, tasa de acceso a internet y tasa de acceso a computador de cada departamento. Con las muestras separadas por nivel de vida, realizo estimaciones de los dos tratamientos con los tres métodos, pero sólo para el promedio de todas las palabras clave.

### 3.4. Modelo de desempleo

Es conocido por la literatura que los datos de Google Trends pueden ayudar a predecir diferentes actividades económicas (Cardona Rojas & Rojas, 2017; Choi & Varian, 2009; Tuhkuri et al., 2015; Vázquez & López-Araiza B, 2020). Aunque no estimo un efecto causal, hipotetizo que las búsquedas de “computrabajo” en Google Trends podrían estar correlacionadas con la tasa de desempleo de algunos grupos poblacionales<sup>13</sup>. Esto puede ser relevante como una aplicación de Google Trends como indicador, para una ejecución más eficiente de políticas públicas focalizadas a los grupos poblacionales más afectados por el desempleo. Al tener un indicador en tiempo real, podría ayudar a los *policymakers* a tomar medidas más rápido para reducir el desempleo.

Con el fin de poder entender la relación de las palabras de búsqueda de empleo y la tasa de desempleo, propongo comparar los siguientes modelos:

$$y_t = \alpha + \beta_1 y_{t-1} + \beta_2 y_{t-12} + \beta_3 Covid + \epsilon_t \quad (4)$$

$$y_t = \alpha + \beta_1 y_{t-1} + \beta_2 y_{t-12} + \beta_3 Covid + \beta_4 Computrabajo + \beta_5 Covid * Computrabajo + \epsilon_t \quad (5)$$

En donde  $y_t$  son las series de tiempo de cuatro tasas de desempleo mensuales de los años 2010 a 2020, a saber: Tasa de desempleo total, tasa de desempleo de hombres, tasa de desempleo de mujeres y tasa de desempleo de los jóvenes<sup>14</sup>. El modelo base está representado en la ecuación 4. La literatura sugiere un proceso autorregresivo (AR) con un *lag* de 1 periodo y otro de 12 periodos, para controlar por la tendencia temporal y la estacionalidad respectivamente. En el modelo alternativo (Eq. 5), la variable Computrabajo representa la serie de Google Trends de la palabra “computrabajo” para el mismo periodo de tiempo. Al ser buscada en un periodo tan largo, Google calcula el índice por un periodo mensual. En los dos modelos incluyo la variable dummy Covid para controlar por un posible efecto de la pandemia en el desempleo y en el

<sup>13</sup>Elijo “computrabajo” para este ejercicio por que es la palabra con resultados más relevantes del ejercicio anterior y por ser la página mas buscada de empleo en Colombia, según similarweb.com.

<sup>14</sup>La gráfica A7 muestra el comportamiento de estas series. Resalta la estacionalidad de computrabajo y las tasas de desempleo con picos a inicio de cada año. También llama la atención que los grupos poblacionales con mayor desempleo son los jóvenes y las mujeres.

modelo alternativo, una interacción entre Covid y Computrabajo para ver posibles relaciones de las búsquedas en pandemia.

## 4. Resultados

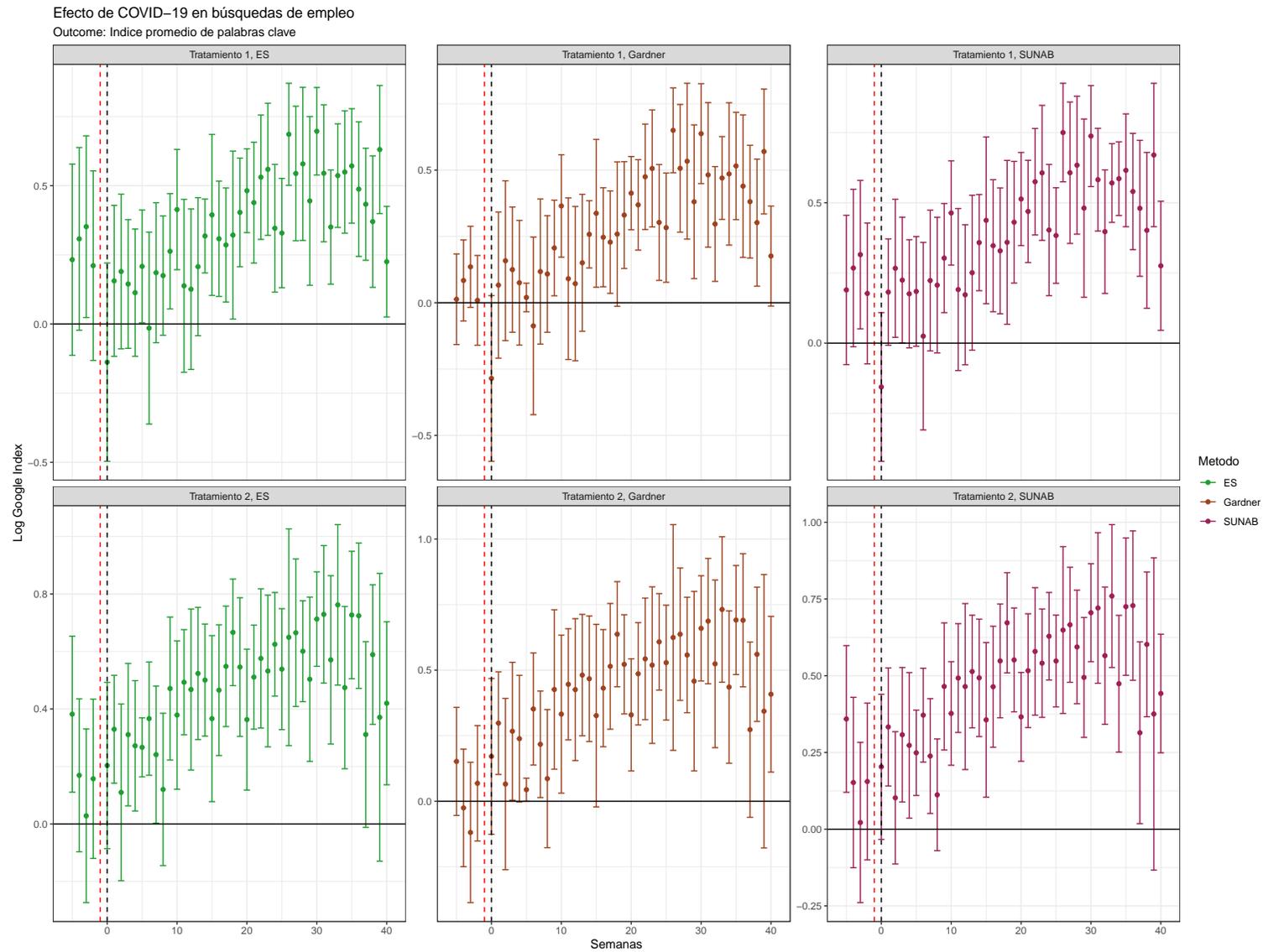
Para tener una idea general del efecto de la pandemia en las búsquedas de empleo, ejecuto las tres metodologías con los dos tratamientos en un promedio de los índices de búsqueda de las palabras de empleo en Google. Los resultados están en el gráfico 2. Las filas representan el tratamiento y las columnas, el método. La línea vertical punteada negra representa la semana 0 o de inicio del tratamiento y la línea punteada roja es la semana -1, es decir, la semana anterior al tratamiento (la semana de referencia)<sup>15</sup>.

Con el primer tratamiento, el primer caso reportado por cada departamento, se observa un efecto creciente a partir de la semana 9 desde el reporte del primer caso de COVID-19 con algunas semanas no significativas. Las búsquedas en Google de palabras relacionadas a búsquedas de empleo aumentaron de manera significativa en un rango de entre 25 % y 60 % desde la semana 14 hasta la última semana. Este efecto se mantiene constante con las 3 metodologías.

Para el segundo tratamiento, la primera muerte reportada, el efecto en el promedio de las búsquedas muestra un comportamiento creciente, positivo y significativo hasta la semana 35, con algunas semanas no significativas al inicio, donde llega a un valor máximo de más del 75 % en los 3 métodos. Luego, decrece el efecto, pero se mantiene positivo hasta la semana 40, donde llega a valores cercanos al 50 % con las tres estimaciones. Hay una semana con efectos positivos antes de la semana de tratamiento con los estimadores de EfS y Sun y Abraham, 2021, un indicador de posible violación al supuesto de tendencias paralelas previas.

---

<sup>15</sup>En las subsecciones siguientes analizo los resultados para cada palabra por separado.



**Figura 2:** Estudio de eventos. Outcome: Promedio de búsquedas de palabras clave de empleo. Notas: Las filas son los tratamientos. Tratamiento 1: Primer caso reportado. Tratamiento 2: Primera muerte reportada. Las columnas representan las metodologías utilizadas: Estudio de Eventos, Gardner, 2021 y Sun y Abraham, 2021. Las líneas verticales para cada coeficiente muestran intervalos de confianza del 95 %, con errores estándar clusterizados por departamento. La semana omitida es la semana anterior al tratamiento en cada departamento, representada en la línea roja punteada.

#### 4.1. Tratamiento 1: Día del primer caso reportado.

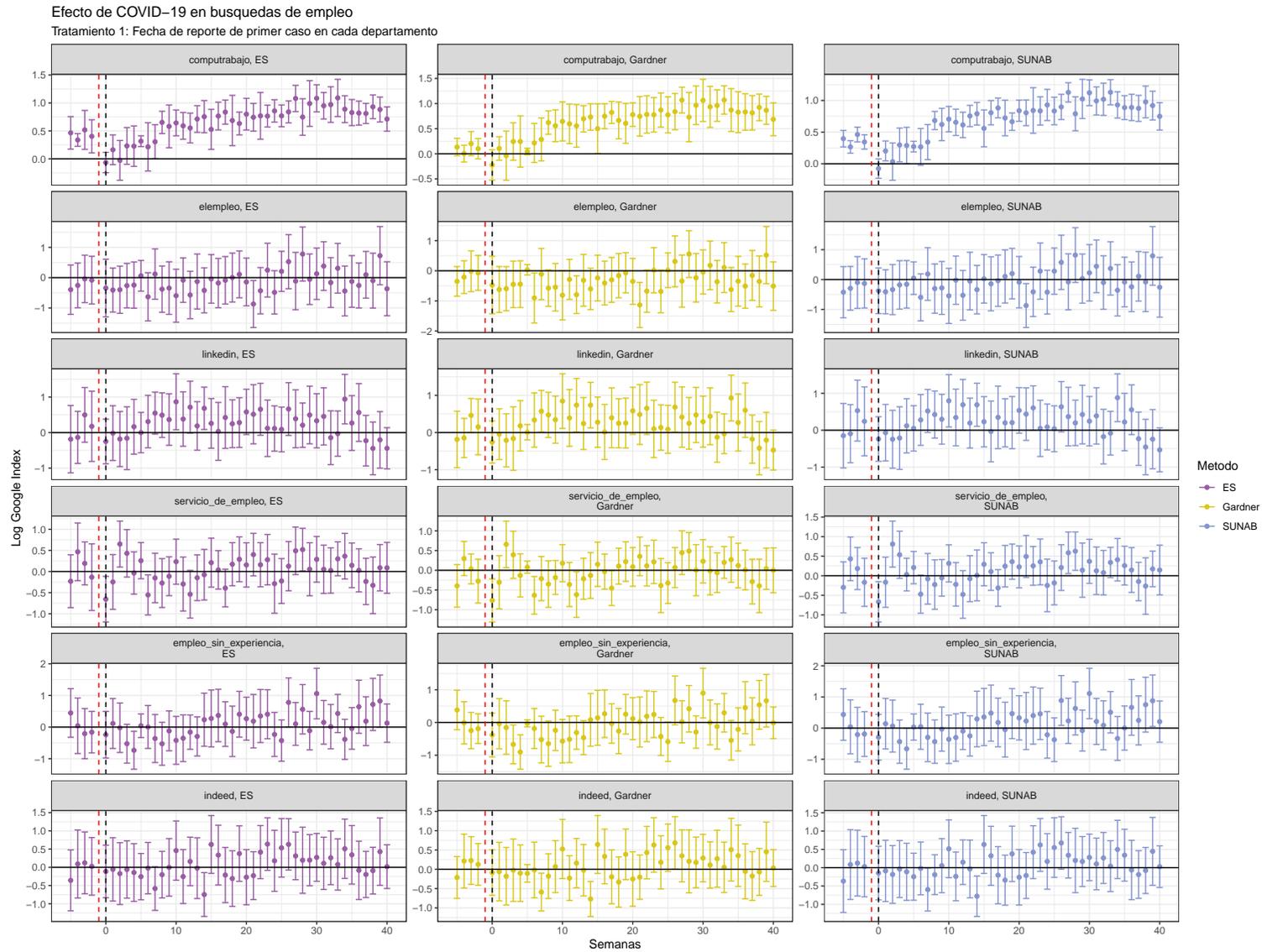
La figura 3 muestra las estimaciones de los efectos dinámicos del primer tratamiento, la fecha del primer caso reportado, con las tres metodologías para todas las palabras relacionadas con el empleo por separado. Las columnas muestran cada metodología y en las filas se encuentran ubicadas las palabras buscadas. La línea vertical punteada roja representa la semana -1, es decir, la semana anterior al tratamiento y la línea punteada negra es la semana 0, la semana de inicio del tratamiento. El comportamiento del efecto de Computrabajo es muy similar al del promedio de las palabras de empleo, lo que sugiere una correlación alta.

Para Computrabajo, después de la semana 6 el efecto crece de forma significativa hasta la semana 40, manteniendo un valor estable de alrededor del 100 % a partir de la semana 20 con los tres estimadores. Las metodologías de ES y Sun y Abraham, 2021 arrojan efectos positivos antes del tratamiento.

Las siguientes palabras clave de búsqueda presentan un comportamiento similar en los resultados. No tienen efectos significativos después del tratamiento y para algunas palabras con ciertos métodos hay efectos significativos, pero de muy corta duración. También se cumple el supuesto de tendencias paralelas previas en todas las palabras.

En el caso de Elemplo, no se observan efectos significativos, excepto por una caída en las búsquedas de 100 % en la semana 21, retomando rápidamente a un valor estadísticamente cero. LinkedIn presenta un efecto no significativo hasta la semana 34, con un efecto positivo del 100 %, que pierde significatividad en la semana siguiente. Según ES, hay algunas semanas con un efecto positivo rápido e intermitente, que toma valor estadísticamente cero ágilmente.

Con servicio de empleo se observa un efecto de significancia nula, excepto por un rápido aumento en la segunda semana de 50 % y una breve intermitencia significativa durante las primeras 10 semanas. Luego, el efecto no es significativo hasta la semana 40. Empleo sin experiencia muestra un efecto no significativo durante las 40 semanas, con algunas semanas con efectos significativos intermitentes. En Gardner, 2021 se observa una caída significativa de hasta 100 % durante las primeras 4 semanas, que se pierde rápidamente. Indeed no presenta efectos significativos a lo largo de las 40 semanas con las tres metodologías.



**Figura 3:** Estudio de eventos: Tratamiento 1 - Primer caso reportado en cada departamento. Notas: Las filas son el *outcome* de las palabras clave. Las columnas representan las metodologías utilizadas: Estudio de Eventos, Gardner, 2021 y Sun y Abraham, 2021. Las líneas verticales para cada coeficiente muestran intervalos de confianza del 95 %, con errores estándar clusterizados por departamento. La semana omitida es la semana anterior al tratamiento en cada departamento, representada en la línea roja punteada.

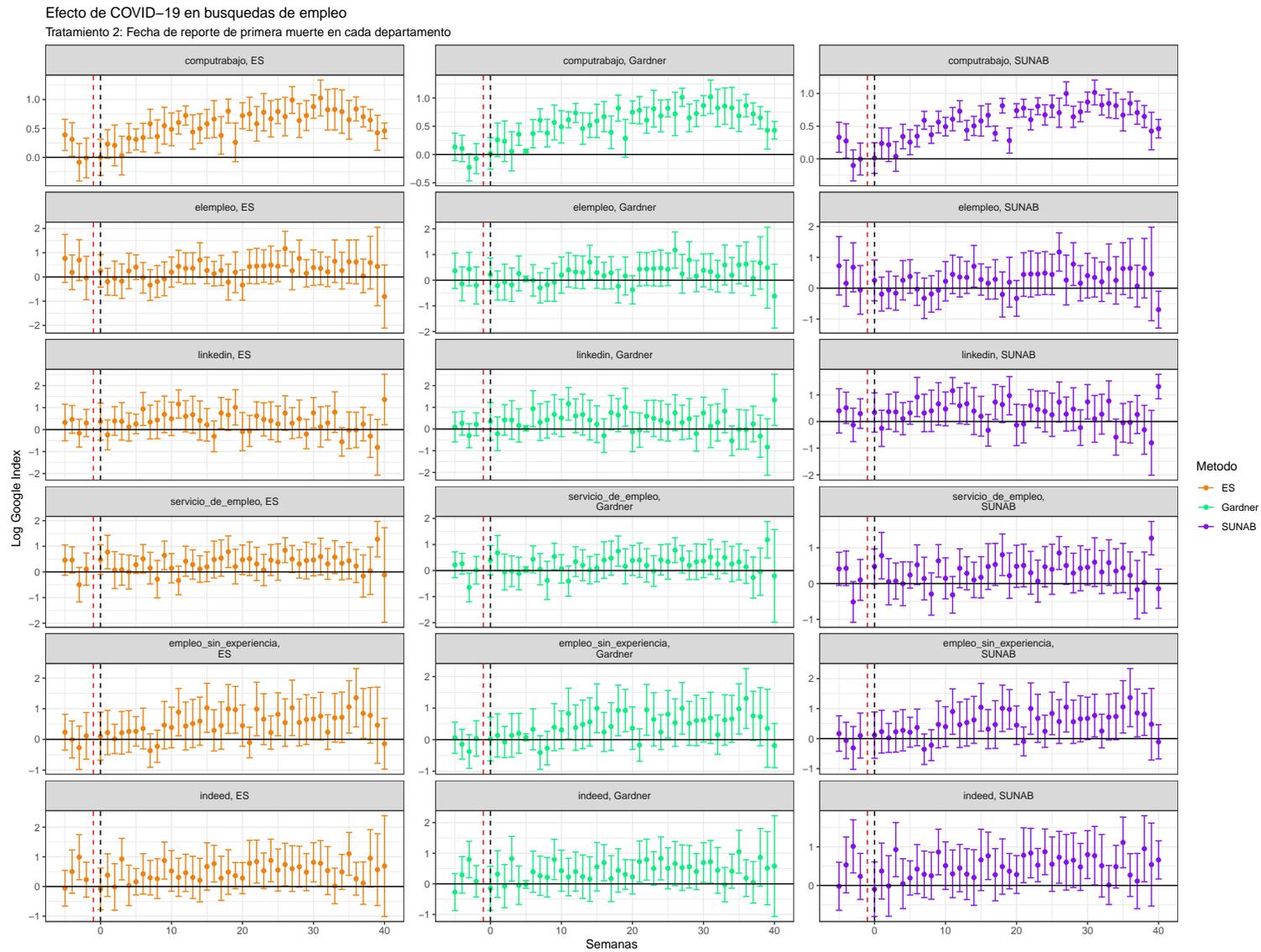
## 4.2. Tratamiento 2: Día de la primera muerte reportada.

La figura 4 muestra el efecto del segundo tratamiento, la fecha de reporte de la primera muerte de cada departamento, en las búsquedas de empleo. Para Computrabajo hay un efecto positivo y significativo a partir de la cuarta semana con picos máximos de 100 % en las cercanías de la semana 25 y con los tres métodos. Este efecto decrece, pero se mantiene positivo hasta la semana 40.

De forma similar al primer tratamiento; estas palabras clave de búsqueda también presentan un comportamiento similar en los resultados. No tienen efectos significativos después del tratamiento y para algunas palabras con ciertos métodos hay efectos significativos, pero de muy corta duración. Con este tratamiento, se cumple el supuesto de tendencias paralelas previas en casi todas las palabras.

No se observa una tendencia clara del efecto de la primera muerte reportada en las búsquedas de empleo. Hay algunas semanas con efectos positivos y significativos, pero son intermitentes. Resalta sólo un aumento del 100 % de las búsquedas en la semana 26, según los tres métodos. LinkedIn no muestra cambios significativos en las búsquedas, aparte algunos efectos intermitentes positivos que rápidamente vuelven a ser estadísticamente cero. Resalta el efecto positivo de 100 % encontrado con los tres métodos en la semana 40.

Con respecto al servicio de empleo, los tres estimadores no reportan una tendencia clara. A lo largo de las 40 semanas, el efecto no es significativo, aunque se presentan algunas semanas positivas que decaen y vuelven a ser estadísticamente cero. Empleo sin experiencia se observa una tendencia positiva del efecto con aumento búsquedas de hasta 100 % a partir de la semana 10, aunque con intermitencia de semanas no significativas hasta la semana 40. Indeed muestra un comportamiento similar con los tres estimadores. No se ve una tendencia clara y el efecto no es significativo durante las 40 semanas, excepto por intermitencias de algunas semanas con efecto positivo que decaen rápidamente.



**Figura 4:** Estudio de eventos: Tratamiento 2 - Primera muerte reportada en cada departamento. Notas: Las filas son el *outcome* de las palabras clave. Las columnas representan las metodologías utilizadas: Estudio de Eventos, Gardner, 2021 y Sun y Abraham, 2021. Las líneas verticales para cada coeficiente muestran intervalos de confianza del 95 %, con errores estándar clusterizados por departamento. La semana omitida es la semana anterior al tratamiento en cada departamento, representada en la línea roja punteada.

### 4.3. Efectos heterogéneos y pruebas de robustez

#### 4.3.1. Heterogeneidad por nivel socioeconómico de los departamentos

A fin de encontrar posibles heterogeneidades por nivel socioeconómico de los departamentos, realice la separación de los departamentos por estrato y estimo con los tres métodos para el promedio de las palabras de búsqueda. Las figuras de los estimadores con efectos heterogéneos se encuentran en las gráficas [A3](#) y [A4](#) en el apéndice. Similar a las estimaciones anteriores, la línea roja punteada representa la semana -1 y la línea negra punteada es la semana de inicio del tratamiento o la semana 0.

Para el promedio de palabras, con el primer tratamiento (Figura [A3](#)), según el método de Gardner, 2021, se observa un efecto positivo más temprano en los departamentos altos, ya que inicia el efecto significativo en la semana 2, mientras que en los departamentos bajos presentan el efecto positivo hasta después de la semana 10 y el efecto es mayor en los departamentos altos, con un aumento de las búsquedas promedio de 100% a lo largo de las 40 semanas. Esta diferencia no se presenta con la metodología de ES ni con Sun y Abraham, 2021. En el segundo tratamiento (Figura [A4](#)) no se refleja heterogeneidad alguna, ya que no hay una diferencia notable en la magnitud de los efectos ni en la velocidad en la que aparecen. En casi todos los estimadores se cumple el supuesto de tendencias paralelas.

Con los dos tratamientos y los tres métodos, el comportamiento tendencial es similar, pero en los departamentos con un nivel de vida alto, el efecto es ligeramente más preciso, con un intervalo de confianza un poco menor. Estas diferencias podrían ser explicadas por la brecha digital; baja conectividad, acceso a computadores y bajo nivel educativo que tienen estos departamentos comparado con aquellos de mayores ingresos y nivel de vida.

#### 4.3.2. Pruebas de robustez

Como prueba de robustez, repito el estudio usando datos de 2017 a 2019, como si la pandemia hubiese ocurrido en 2019. Asigno las semanas de los tres tratamientos al año 2019 y aplico los tres tratamientos para todas las palabras de búsqueda. Los gráficos [A5](#) y [A6](#) muestran que no hay una significancia estadística en las búsquedas de empleo con los dos tratamientos.

### 4.4. Modelo de desempleo

La tabla [A3](#) resume los resultados de los modelos para las cuatro tasas de desempleo. Cada par de columnas representa el resultado de los modelos para los respectivos outcomes. Las columnas impares son del modelo base y las columnas pares reflejan el modelo alternativo. Los rezagos de 1 periodo y 12 periodos son significativos en todos los modelos, lo que indica

un proceso autorregresivo de las tasas de desempleo, un resultado consistente con la literatura actual.

Acorde con los modelos, durante los meses del Covid, la tasa de desempleo fue más alta, pero este efecto aumenta en los modelos alternos. En la tasa de desempleo general, el efecto promedio es de 6 puntos porcentuales (p.p.), mientras que, con controles, el efecto del covid aumenta a 16 p.p. El efecto es similar para la tasa de desempleo masculina, con un aumento de 4 p.p. y con controles adicionales, aumenta a casi 16 p.p. Las mujeres tuvieron un mayor aumento de la tasa de desempleo, ya que el efecto del Covid sobre la tasa de desempleo femenina fue de 8 p.p. y con controles de 21 puntos porcentuales, efecto mayor que en las demás tasas. Para la tasa de desempleo juvenil, el efecto sigue siendo positivo y significativo, pero similar a la tasa general y la tasa masculina, con valores de 7.5 p.p. en el modelo base y 15.5 p.p en el modelo alternativo. Estos resultados indicarían posibles efectos heterogéneos del Covid en el mercado laboral por sexo.

Previo a la pandemia, las búsquedas de Computrabajo en Google no parecen tener una asociación significativa con las tasas de desempleo. Sin embargo, parece haber una relación negativa y significativa sobre las diferentes tasas de desempleo en tiempos de pandemia, aunque leve. Los resultados sugieren que una unidad adicional del índice de búsquedas de Computrabajo se asocia a una tasa de desempleo general 0.66 p.p más baja durante pandemia. El resultado es similar para la tasa de desempleo masculina, asociada con una tasa de desempleo 0.67 p.p más baja en pandemia. Resalta el resultado en la tasa de empleo femenina. Una unidad adicional del índice de búsquedas en Google de Computrabajo estaría aunada con una baja de la tasa de empleo femenina de 0.77 p.p. promedio en pandemia, siendo una magnitud mayor con respecto a las otras tasas. La caída de la tasa de desempleo juvenil de 0.5 p.p parece estar vinculada al aumento de las búsquedas de Computrabajo en pandemia. Aunque no es causal, estos resultados sugerirían que las búsquedas de empleo por internet podrían ser una herramienta útil para mitigar el desempleo en tiempos de crisis, en especial para las mujeres, por lo que este tipo de modelos serían útiles para posibles implementaciones de política pública.

## 5. Discusión y conclusiones

En este trabajo estudié el efecto de la pandemia de COVID-19 en las búsquedas de empleo en Google en 24 departamentos de Colombia, usando la herramienta de Google Trends. En los resultados más relevantes destaca que, con el estimador de Gardner, [2021](#), el promedio de las palabras clave tuvo un aumento significativo y creciente de hasta 50% a lo largo de las 40 semanas para los dos tratamientos. El efecto es similar con Computrabajo, donde hay un

aumento de las búsquedas sostenido de hasta 100% a partir de la semana 5 hasta la semana 40 con el estimador de Gardner, 2021. El COVID-19 no afectó significativamente las búsquedas en Google de las otras palabras clave relacionadas con búsqueda de empleo. La metodología de Gardner, 2021 es la más confiable en este estudio, ya que sus estimadores no violan el supuesto de tendencias paralelas previas, mientras que las metodologías de Event Studies y Sun y Abraham, 2021 presentan efectos significativos antes del tratamiento, lo que implicaría una violación a dicho supuesto. Estos resultados podrían ser explicados por el hecho que la pandemia y la cuarentena ordenada por el Gobierno Nacional trajo una caída de la producción y del empleo, lo que llevó a un aumento del desempleo y de búsquedas de empleo por parte de los nuevos desempleados.

Separando los departamentos por calidad de vida, se encuentran crecimientos de búsquedas del promedio de las palabras clave más robustos y precisos en los departamentos con mayor nivel socioeconómico, con aumentos de hasta 100% en los departamentos de nivel alto, con una diferencia significativa con los de nivel bajo y estimados con Gardner, 2021. Esto podría ser explicado por la brecha digital; una mayor cobertura y acceso a internet de los departamentos de mejor calidad de vida llevaría a resultados más precisos, lo que significaría que la calidad de la información disponible aumentaría si hubiese una mejor cobertura de internet en sitios apartados, combinado con mayores ingresos y políticas de mayor acceso y calidad educativa.

Con respecto a los modelos de desempleo, resalta el resultado para la tasa de empleo femenina, donde el aumento del índice de búsquedas en Google de Computrabajo estaría asociado con la reducción de 0.77 puntos porcentuales promedio la tasa de empleo femenina en pandemia. Aunque hace falta profundizar en estudios de este tipo de modelos, es una señal de diagnóstico para generar políticas públicas enfocadas a las mujeres y a los jóvenes y mejorar las existentes. Es posible que un mayor acceso a las TIC ayudaría a reducir brechas de género en el mercado laboral, aunque se necesitarían más estudios.

Este estudio es una muestra de la utilidad de los nuevos estimadores de estudios de eventos con tratamientos escalonados con datos en tiempo real, en especial la metodología de Gardner, 2021 que mostró resultados fiables. Asimismo, la fortaleza de los datos radica en su fácil disponibilidad y el sesgo reducido, si no nulo, de las búsquedas en Google, ya que son anónimas. Pero al ser agregadas, no es posible identificar variables demográficas clave, que permitirían controlar por sexo o edad. Si existiera la posibilidad, sería notable ver dichos efectos desagregados por sexo, edad o nivel de estudio.

Similar a Berniell y Facchini, 2021, estos resultados están limitados a la gente que tiene acceso a internet, por lo que no tiene en cuenta a aquellos que utilizan otros métodos de búsqueda de empleo, como el *networking* u ofertas de empleo en periódicos y estos resultados no podrían

ser aplicados al empleo informal, que es casi la mitad de la fuerza laboral del país, ya que las dinámicas de búsqueda de empleo en el sector informal son diferentes. Este estudio también resulta útil como primer paso para futuros estudios con Google Trends u otras herramientas de datos en tiempo real. Sería interesante ver estudios similares en patrones de consumo, salud o educación no limitados a países de primer mundo.

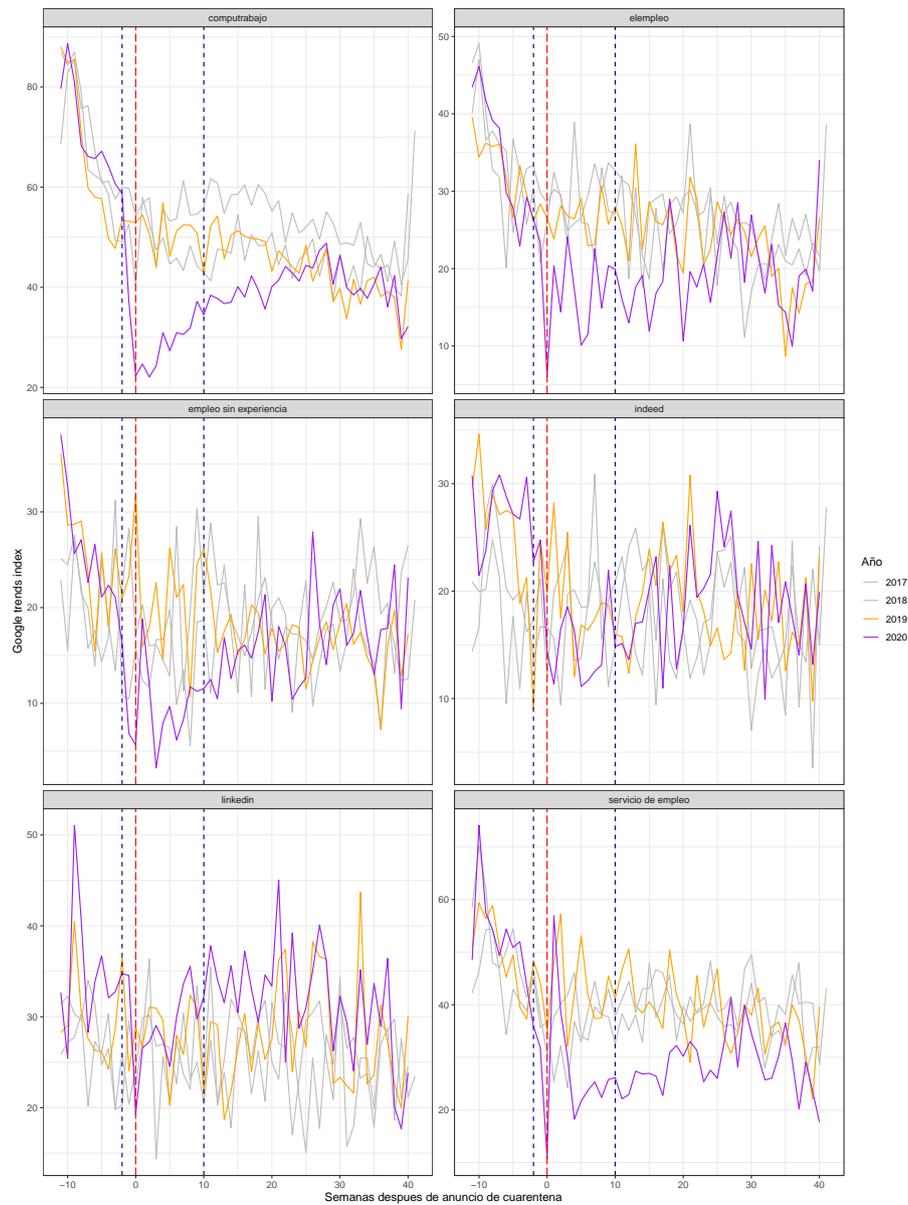
## Bibliografía

- Aaronson, D., Brave, S. A., Butters, R. A., Sacks, D. & Seo, B. (2020). *Using the Eye of the Storm to Predict the Wave of Covid-19 UI Claims* (inf. téc.). Federal Reserve Bank of Chicago. <https://doi.org/10.21033/wp-2020-10>
- Arango Thomas, L. & Posada, C. (2001). *El desempleo en Colombia* (inf. téc.). Banco de la Republica de Colombia.
- Bacher-Hicks, A., Goodman, J. & Mulhern, C. (2021). Inequality in household adaptation to schooling shocks: Covid-induced online learning engagement in real time. *Journal of Public Economics*, 193, 104345.
- Bento, A. I., Nguyen, T., Wing, C., Lozano-Rojas, F., Ahn, Y.-Y. & Simon, K. (2020). Evidence from internet search data shows information-seeking responses to news of local COVID-19 cases. *Proceedings of the National Academy of Sciences*, 117(21), 11220-11222. <https://doi.org/10.1073/pnas.2005335117>
- Bergé, L. (2018). Efficient estimation of maximum likelihood models with multiple fixed-effects: the R package FENmlm. *CREA Discussion Papers*, (13).
- Berniell, I. & Facchini, G. (2021). COVID-19 Lockdown and Domestic Violence: Evidence from Internet-Search Behavior in 11 Countries. *European Economic Review*, 103775.
- Brodeur, A., Clark, A. E., Fleche, S. & Powdthavee, N. (2021). COVID-19, lockdowns and well-being: Evidence from Google Trends. *Journal of Public Economics*, 193, 104346. <https://doi.org/10.1016/j.jpubeco.2020.104346>
- Butts, K. (2021). *did2s: Two-Stage Difference-in-Differences Following Gardner (2021)*. <https://www.github.com/kylebutts/did2s/>
- Cardona Rojas, L. F. & Rojas, J. (2017). *Pronósticos para la tasa de desempleo en Colombia a partir de Google Trends* (Archivos de Economía). Departamento Nacional de Planeación. <https://EconPapers.repec.org/RePEc:col:000118:016050>
- Chakraborti, R. & Roberts, G. (2020). Anti-Price Gouging Laws, Shortages, and COVID-19: Big Data Insights from Consumer Searches. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3613726>
- Chin, C.-Y. & Wang, C.-L. (2020). Effectiveness of COVID-19 Pandemic Prevention: A cross-country Comparison of Digital Footprint of Google Search Data. *Advances in Management and Applied Economics*, 10(4), 23-34.
- Choi, H. & Varian, H. (2009). Predicting initial claims for unemployment benefits. *Google Inc*, 1, 1-5.
- Choi, H. & Varian, H. (2012). Predicting the Present with Google Trends, 23.

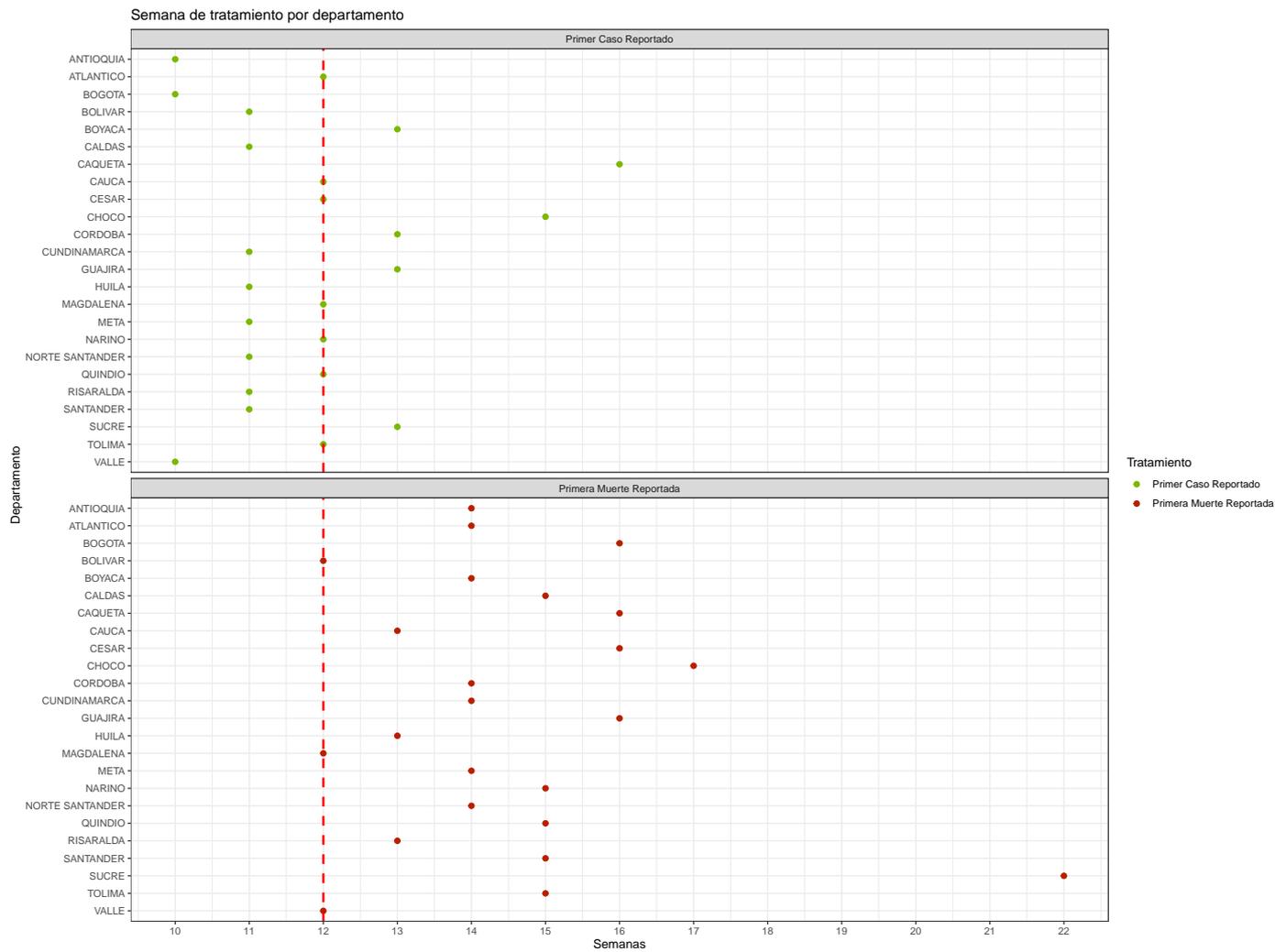
- Dilmaghani, M. (2019a). The racial ‘digital divide’ in the predictive power of Google trends data for forecasting the unemployment rate. *Journal of Economic and Social Measurement*, 43(3-4), 119-142. <https://doi.org/10.3233/JEM-180458>
- Dilmaghani, M. (2019b). Workopolis or The Pirate Bay: what does Google Trends say about the unemployment rate? *Journal of Economic Studies*, 46(2), 422-445. <https://doi.org/10.1108/JES-11-2017-0346>
- Gardner, J. (2021). Two-stage differences in differences.
- González-Fernández, M. & González-Velasco, C. (2018). Can Google econometrics predict unemployment? Evidence from Spain. *Economics Letters*, 170, 42-45.
- Henao, M. L. & Rojas, N. (1999). La tasa natural de desempleo en Colombia.
- Lenaerts, K., Beblavý, M. & Fabo, B. (2016). Prospects for utilisation of non-vacancy Internet data in labour market analysis—an overview. *IZA Journal of Labor Economics*, 5(1), 1. <https://doi.org/10.1186/s40172-016-0042-z>
- MinTIC. (2021). *Índice de brecha digital regional* (inf. téc.). Ministerio de las Tecnologías de la Información y las Comunicaciones. [https://colombiatic.mintic.gov.co/679/articles-238353\\_recurso.1.pdf](https://colombiatic.mintic.gov.co/679/articles-238353_recurso.1.pdf)
- Roth, J., Sant’Anna, P. H., Bilinski, A. & Poe, J. (2022). What’s Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature. *arXiv preprint arXiv:2201.01194*.
- Silverio-Murillo, A., Hoehn-Velasco, L., Tirado, A. R. & de la Miyar, J. R. B. (2021). COVID-19 Blues: Lockdowns and Mental Health-Related Google Searches in Latin America. *Social Science & Medicine*, 114040.
- Smith, P. (2016). Google’s MIDAS Touch: Predicting UK Unemployment with Internet Search Data: Google’s MIDAS Touch. *Journal of Forecasting*, 35(3), 263-284. <https://doi.org/10.1002/for.2391>
- Sun, L. & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2), 175-199.
- Tuhkuri, J. et al. (2015). Big Data: Do Google Searches Predict Unemployment?
- Urrea-Ríos, I. L. & Piraján, J. (2020). Impacto de la pandemia covid-19 sobre la economía colombiana. Una pandemia temporal con efectos permanentes (Impact of the COVID-19 Pandemic on the Colombian Economy. A Temporary Pandemic with Permanent Effects). *Una pandemia temporal con efectos permanentes (Impact of the COVID-19 Pandemic on the Colombian Economy. A Temporary Pandemic with Permanent Effects)*(August 20, 2020).

Vázquez, R. M. C. & López-Araiza B, S. E. (2020). Grandes datos, Google y desempleo. *Estudios Económicos*, 35(1), 125-15.

## A. Apéndice



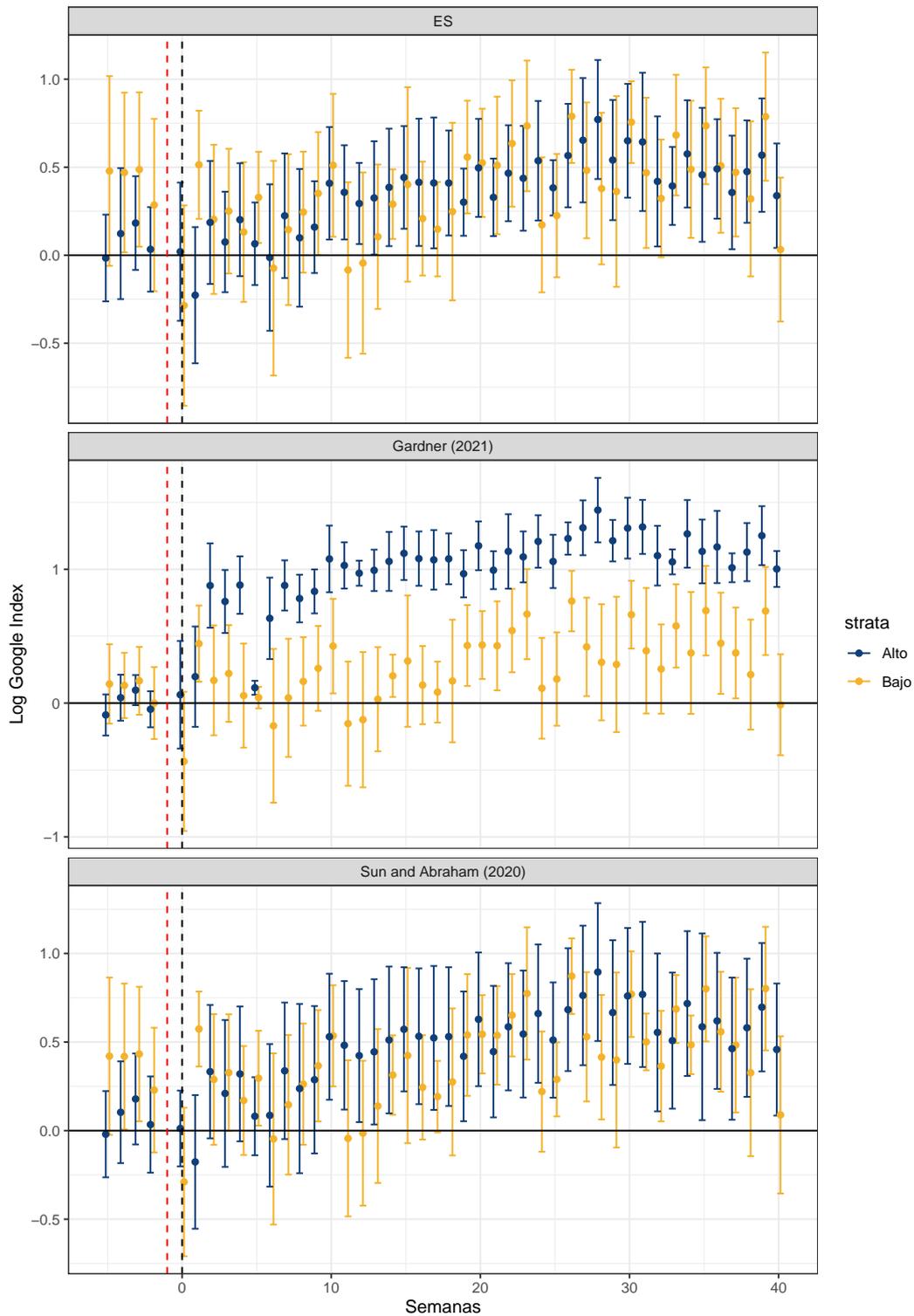
**Figura A1:** Búsquedas promedio nacional de Google Index por palabra clave, 2017 - 2020. El gráfico muestra el promedio de búsquedas de las palabras clave relacionadas con búsquedas de empleo para 24 departamentos por semana para el año 2020 (en púrpura), 2019 (en naranja) y los años anteriores en gris. La línea roja punteada vertical representa la semana que dictaron la cuarentena en Colombia. Las líneas azules punteadas representan el intervalo de semanas donde ocurren los tratamientos de este estudio en todos los departamentos. Fuente: Google Trends, cálculos propios.



**Figura A2:** Semanas de tratamiento por departamento. El primer panel muestra las semanas del año 2020 donde ocurrió el primer caso reportado en cada departamento (punto verde). El segundo panel muestra las semanas donde ocurrió la primera muerte reportada en cada departamento (punto rojo). La línea vertical roja muestra la semana donde fue dictada la cuarentena.

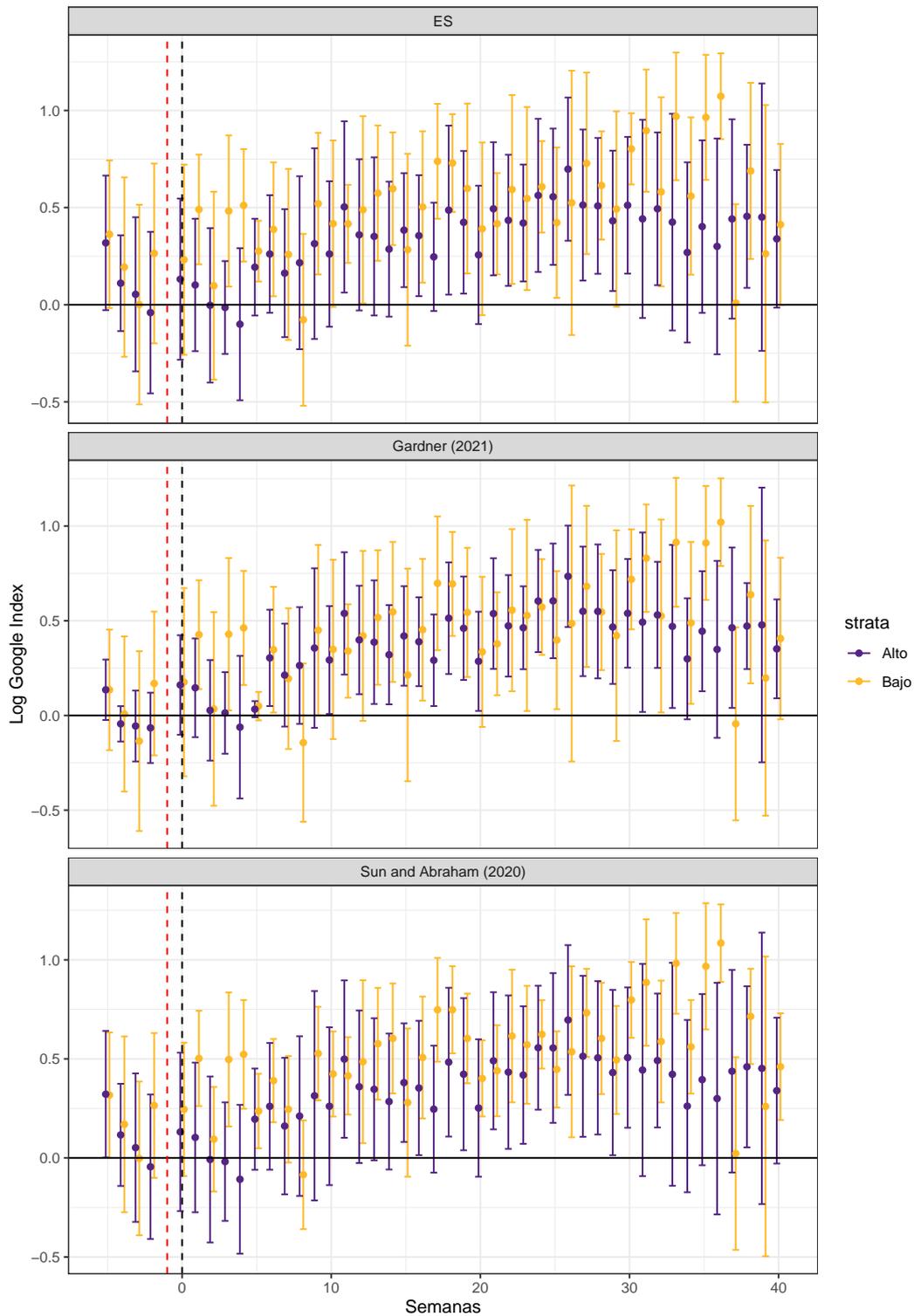
### Efecto heterogéneo de COVID-19 en búsquedas: meanJobsGS

Tratamiento 1: Fecha de reporte de primer caso en cada departamento

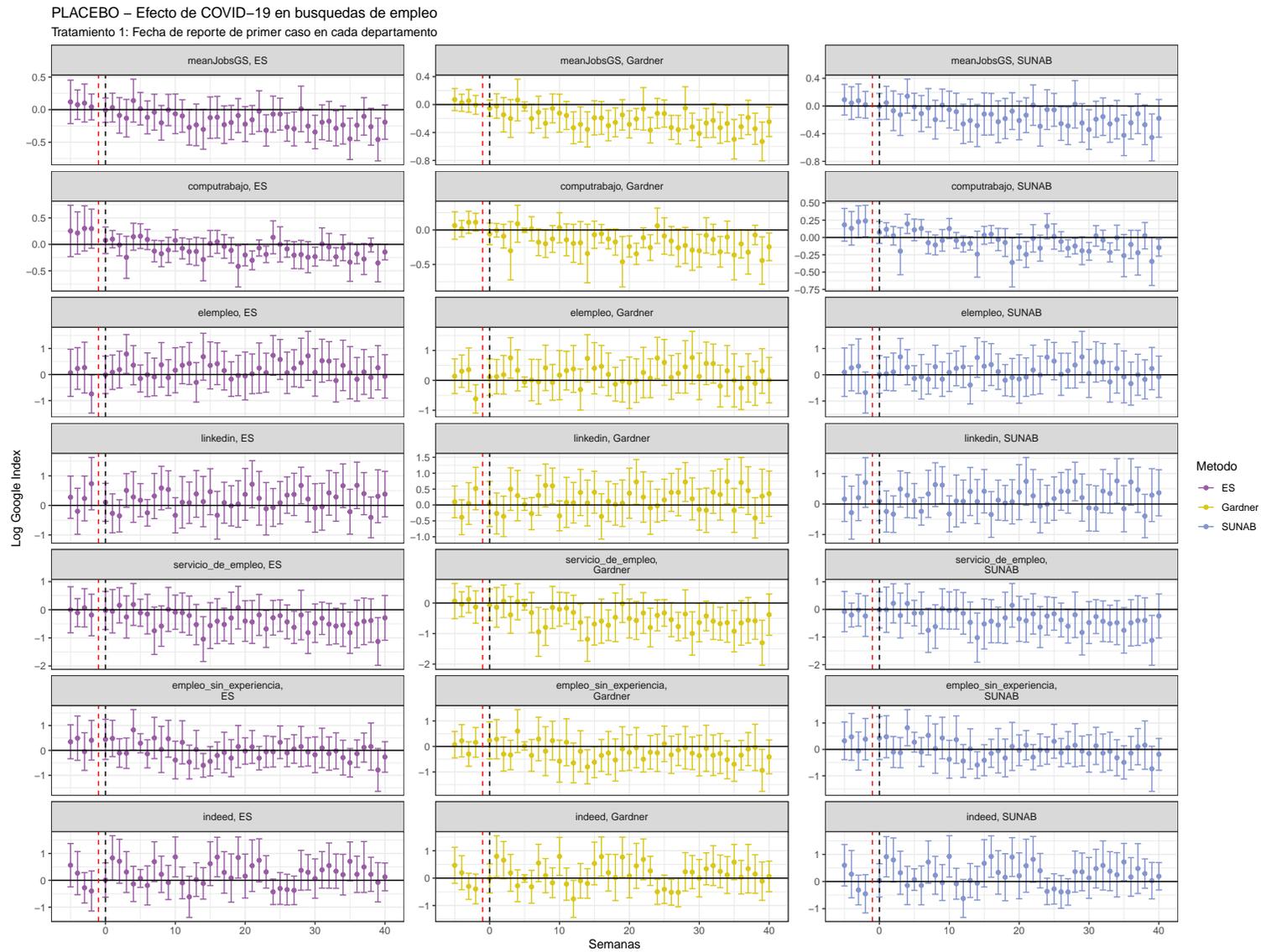


**Figura A3:** Estudio de eventos por departamentos de nivel alto y bajo para el promedio de búsquedas de empleo. Notas: Tratamiento 1: Primer caso reportado. Cada panel representa la metodología utilizada: Estudio de Eventos, Gardner, 2021 y Sun y Abraham, 2021. Los coeficientes en azul son la estimación de los departamentos de nivel de vida alto y los coeficientes en amarillo son la estimación de los departamentos de nivel de vida bajo. Las líneas verticales para cada coeficiente muestran intervalos de confianza del 95 %, con errores estándar clusterizados por departamento. La semana omitida es la semana anterior al tratamiento en cada departamento, representada en la línea roja punteada.

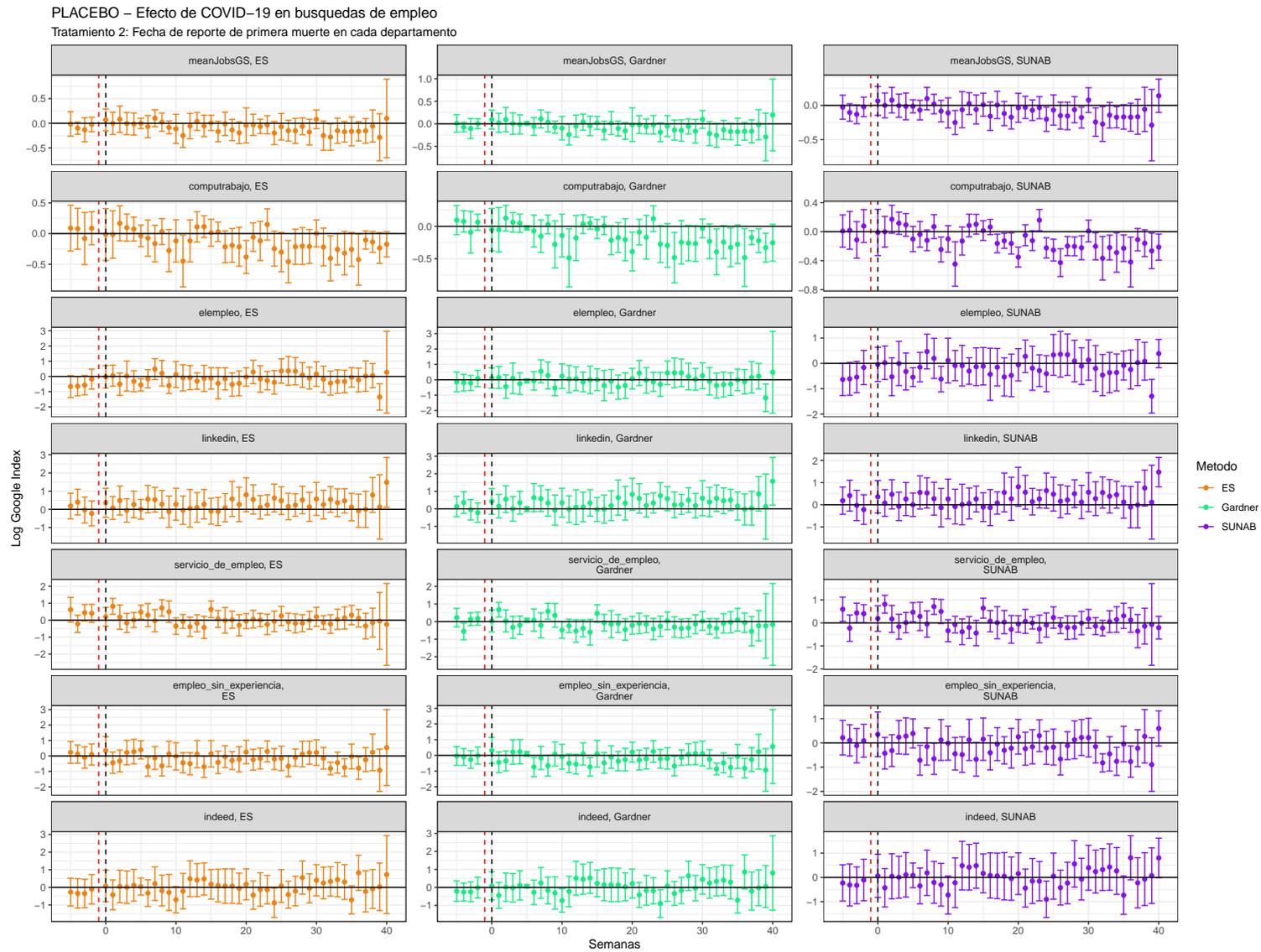
Efecto heterogéneo de COVID-19 en búsquedas: meanJobsGS  
 Tratamiento 2: Fecha de reporte de primera muerte en cada departamento



**Figura A4:** Estudio de eventos por departamentos de nivel alto y bajo para el promedio de búsquedas de empleo. Notas: Tratamiento 2: Primera muerte reportada. Cada panel representa la metodología utilizada: Estudio de Eventos, Gardner, 2021 y Sun y Abraham, 2021. Los coeficientes en morado son la estimación de los departamentos de nivel de vida alto y los coeficientes en naranja son la estimación de los departamentos de nivel de vida bajo. Las líneas verticales para cada coeficiente muestran intervalos de confianza del 95 %, con errores estándar clusterizados por departamento. La semana omitida es la semana anterior al tratamiento en cada departamento, representada en la línea roja punteada.



**Figura A5:** Estudio de eventos: Prueba placebo usando 2019 como año de tratamiento. Tratamiento 1: Primer caso reportado. Notas: El gráfico muestra los coeficientes de una prueba placebo donde uso 2019 como año de tratamiento en vez de 2020. Las filas son el *outcome* de las palabras clave y el promedio de las mismas. Las columnas representan las metodologías utilizadas: Estudio de Eventos, Gardner, 2021 y Sun y Abraham, 2021. Las líneas verticales para cada coeficiente muestran intervalos de confianza del 95 %, con errores estándar clusterizados por departamento. La semana omitida es la semana anterior al tratamiento en cada departamento, representada en la línea roja punteada.



**Figura A6:** Estudio de eventos: Prueba placebo usando 2019 como año de tratamiento. Tratamiento 2: Primera muerte reportada. Notas: El gráfico muestra los coeficientes de una prueba placebo donde uso 2019 como año de tratamiento en vez de 2020. Las filas son el *outcome* de las palabras clave y el promedio de las mismas. Las columnas representan las metodologías utilizadas: Estudio de Eventos, Gardner, 2021 y Sun y Abraham, 2021. Las líneas verticales para cada coeficiente muestran intervalos de confianza del 95%, con errores estándar clusterizados por departamento. La semana omitida es la semana anterior al tratamiento en cada departamento, representada en la línea roja punteada.



**Figura A7:** Búsquedas de Computrabajo y tasas de desempleo: 2010 - 2021. Notas: El panel superior muestra el comportamiento de las búsquedas de Computrabajo en Colombia para el periodo 2010 a 2021 (serie azul oscuro). El panel inferior muestra el comportamiento de cuatro tasas de desempleo de Colombia: General (rojo), hombres (verde oliva), mujeres (morado), jóvenes (aguamarina). Fuente: GEIH y Google Trends, cálculos propios.

Palabra Clave	Descripción
Computrabajo	Fundada en 1999 en Escocia, es una de las páginas de búsqueda de empleo con más visitas en Latinoamérica. Afirman que “Computrabajo ayuda a las personas a encontrar un trabajo mejor y a crecer profesionalmente y a las empresas a encontrar al profesional que mejor encaje con sus necesidades. Con este objetivo ofrece también soluciones innovadoras para la gestión del talento y la digitalización del reclutamiento.”
Elemplo	Fue creada en el año 2000 por la Casa Editorial El Tiempo como una plataforma de noticias del mercado laboral y ofertas de empleo. Ahora prestan servicios tanto a candidatos como a empresas en Colombia y Costa Rica.
LinkedIn	Empresa americana fundada en 2003 como red social profesional para networking y desarrollo de carreras, donde también se encuentran ofertas laborales. En 2016 fue adquirida por Microsoft y cuenta con más de 830 millones de usuarios registrados de todo el mundo.
Servicio de empleo	El Servicio Público de Empleo, es una iniciativa del Gobierno creada en 2013 con el objetivo de ayudar a los trabajadores a encontrar un empleo adecuado a su perfil y a los empleadores a contratar trabajadores de acuerdo a sus necesidades. Prestan servicios de orientación laboral, capacitación y buscadores de empleo de forma gratuita al público.
Indeed	Empresa americana fundada en 2004 por Paul Forster y Rony Kahan. Es una de las páginas más visitadas de búsquedas de empleo en Estados Unidos. Los servicios de Indeed incluyen búsqueda de empleo, tendencias del mercado de empleo, almacenamiento y búsqueda de curriculum, tendencias de industrias, búsqueda de salario, índices de competencia laboral y páginas de empresas. Desde 2012 hace parte del conglomerado japonés Recruit Co.

**Tabla A1:** Descripción de palabras clave.

Departamento	Primer Caso Reportado	Primera Muerte Reportada
ANTIOQUIA	2020-03-09	2020-04-06
ATLANTICO	2020-03-23	2020-04-03
BOGOTA	2020-03-06	2020-04-20
BOLIVAR	2020-03-11	2020-03-21
BOYACA	2020-03-25	2020-04-04
CALDAS	2020-03-15	2020-04-11
CAQUETA	2020-04-18	2020-04-18
CAUCA	2020-03-19	2020-03-25
CESAR	2020-03-22	2020-04-21
CHOCO	2020-04-11	2020-04-26
CORDOBA	2020-03-29	2020-04-04
CUNDINAMARCA	2020-03-15	2020-04-01
GUAJIRA	2020-03-31	2020-04-20
HUILA	2020-03-13	2020-03-30
MAGDALENA	2020-03-20	2020-03-23
META	2020-03-13	2020-04-06
NARINO	2020-03-24	2020-04-09
NORTE SANTANDER	2020-03-15	2020-04-05
QUINDIO	2020-03-18	2020-04-08
RISARALDA	2020-03-15	2020-03-31
SANTANDER	2020-03-13	2020-04-08
SUCRE	2020-03-29	2020-05-30
TOLIMA	2020-03-18	2020-04-12
VALLE	2020-03-09	2020-03-22

**Tabla A2:** Fechas de inicio de los tratamientos en cada departamento.

**Tabla A3:** Resumen de modelos de tasa de desempleo.

	<i>Variable dependiente</i>							
	TD		TD_hombre		TD_mujer		TD_joven	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
lag(TD, 1)	0.394*** (0.059)	0.419*** (0.055)						
lag(TD, 12)	0.637*** (0.092)	0.713*** (0.119)						
lag(TD_hombre, 1)			0.425*** (0.061)	0.449*** (0.056)				
lag(TD_hombre, 12)			0.562*** (0.094)	0.580*** (0.109)				
lag(TD_mujer, 1)					0.372*** (0.061)	0.398*** (0.059)		
lag(TD_mujer, 12)					0.615*** (0.096)	0.657*** (0.130)		
lag(TD_joven, 1)							0.340*** (0.063)	0.356*** (0.063)
lag(TD_joven, 12)							0.597*** (0.081)	0.636*** (0.106)
dcovid	5.896*** (0.632)	16.582*** (2.730)	4.521*** (0.532)	15.642*** (2.257)	8.096*** (0.875)	20.589*** (3.919)	7.573*** (0.849)	15.583*** (3.941)
computrabajo		-0.010 (0.008)		-0.004 (0.006)		-0.009 (0.012)		-0.010 (0.011)
dcovid:computrabajo		-0.663*** (0.159)		-0.674*** (0.131)		-0.772*** (0.228)		-0.502** (0.230)
Constant	-0.430 (1.096)	-1.040 (1.126)	0.024 (0.888)	-0.159 (0.841)	0.057 (1.482)	-0.465 (1.585)	1.058 (1.779)	0.498 (1.878)
Observations	120	120	120	120	120	120	120	120
R <sup>2</sup>	0.842	0.866	0.826	0.860	0.827	0.845	0.817	0.827
Akaike Inf. Crit.	395.273	379.577	357.116	335.208	474.282	465.500	467.914	465.419
Bayesian Inf. Crit.	409.210	399.089	371.053	354.720	488.219	485.012	481.852	484.932

Notas: Los errores estándar robustos por heteroscedasticidad están en paréntesis (\*p<0.1; \*\*p<0.05; \*\*\*p<0.01). Las observaciones son a nivel nacional y mensual desde enero de 2010 hasta diciembre de 2021. Las columnas impares son las estimaciones del modelo 4 y las columnas pares son del modelo 5. Columnas (1) y (2) son resultados para el *outcome* de tasa de desempleo general para Colombia. Columnas (3) y (4), la tasa de desempleo masculina. Columnas (5) y (6), tasa de desempleo femenina y columnas (7) y (8) el *outcome* es la tasa de desempleo juvenil.