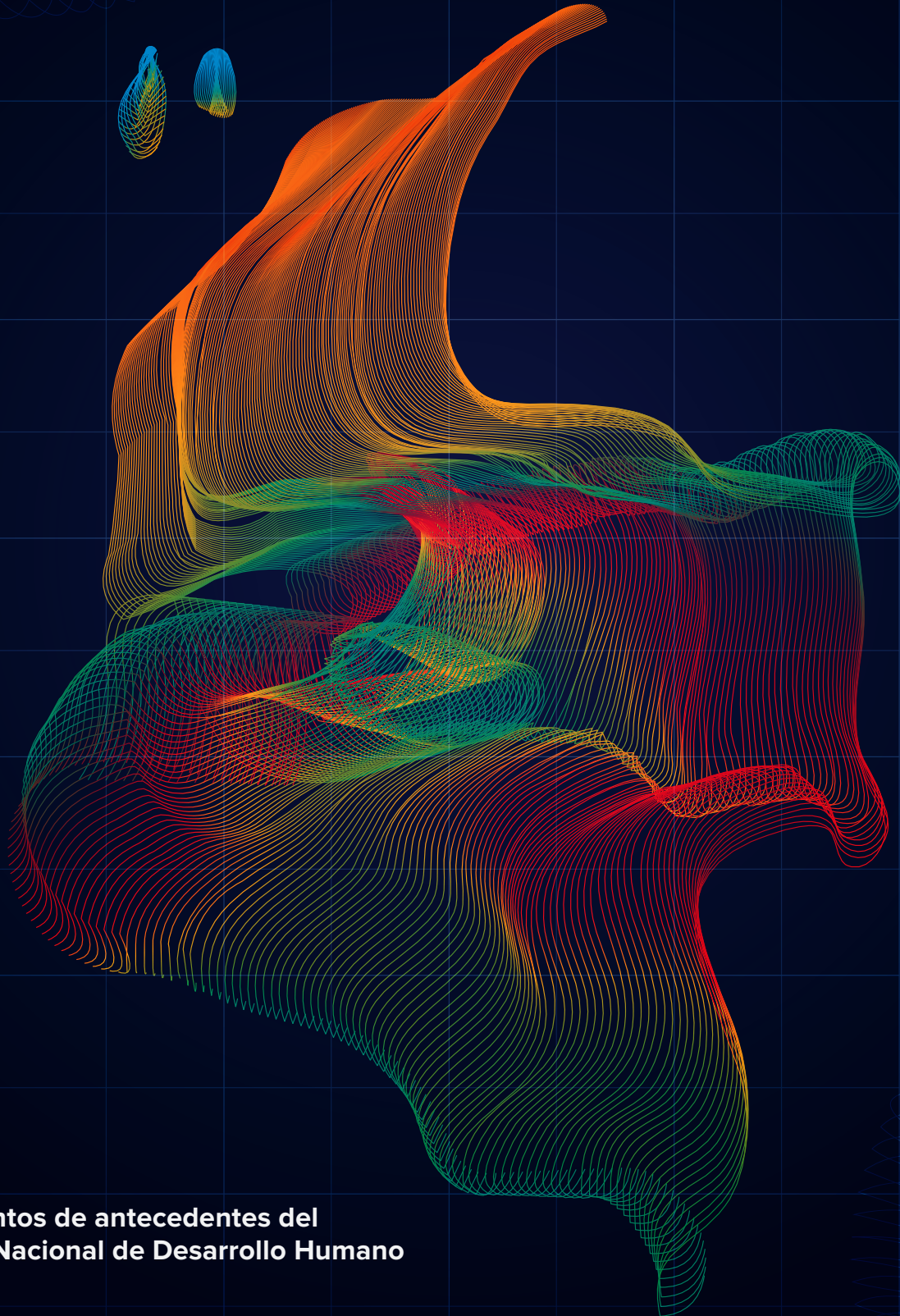

Informalidad, productividad e inequidad. Un análisis desde el punto de vista de las firmas y los trabajadores



07

Documentos de antecedentes del
Informe Nacional de Desarrollo Humano
2024

Informalidad, productividad e inequidad. Un análisis desde el punto de vista de las firmas y los trabajadores

Investigadora: Cristina Fernández¹
Asistente: Alexander Sarango

¹ Investigadora de Fedesarrollo. Se agradece el gran apoyo de Andrés García-Suaza en el desarrollo de este trabajo. María Angélica Arbeláez, Eduardo Lora, Daniel Gómez, Juan Miguel Gallego, Fernando Jaramillo y Alain Desdoigts también fueron fundamentales para desarrollar el trabajo. Marlon Salazar y Camilo Ríos contribuyeron en los pasos iniciales de este documento, y Mónica Ortiz proporcionó comentarios útiles a la base de datos. Este trabajo fue financiado por el Informe de Desarrollo Humano para Colombia de las Naciones Unidas. Se agradece también al Departamento Nacional de Planeación (DNP), que financió una versión anterior de este documento.

Como sucede en muchos otros países, en Colombia hay una carencia de información comprensiva y recurrente del tejido empresarial. El último censo empresarial se realizó en 1990. El presente trabajo propone subsanar esta carencia mediante el diseño de una metodología de construcción de una base de datos firma-empleado-empleador, a partir de encuestas periódicas del Dane, que permite realizar el primer análisis del universo de firmas en su conjunto con un énfasis especial en la informalidad, que es una característica común de los mercados emergentes.

El análisis identifica los siguientes hechos estilizados: 1. El tejido empresarial colombiano es granular tanto en términos de número de trabajadores como en valor agregado. 2. La informalidad empresarial es decreciente en el número de empleados y en el valor agregado por trabajador, este comportamiento se explica por la heterogeneidad del universo empresarial, donde algunas firmas son muy poco productivas para ser formales, otras prefieren ser informales y otras tienen la productividad suficiente para ser formales. 3. La informalidad empresarial también es decreciente en el número de trabajadores, este comportamiento se puede explicar por la combinación de la política tributaria y la de seguridad social que permite deducir la totalidad de costos laborales formales a las firmas grandes, los costos de seguridad social a las firmas del monotributo y ningún costo a las firmas que se encuentran por debajo del umbral del impuesto a la renta. 4. Esta situación afecta a los trabajadores más vulnerables. 5. Las unidades por cuenta propia abarcan una proporción muy alta de la población de firmas y de trabajadores en el país. Estas unidades tienen características diferentes al resto de firmas porque están ligadas a factores ocupacionales; pero también son decrecientes en la productividad, lo que se puede explicar por el sistema de seguridad social colombiano. Adicionalmente, estos trabajadores cuentan con limitaciones importantes para transitar al mercado formal.

JEL classification: J46, O17, L11, O47

Keywords: Informality, Firm informality, Informal labor market, Taxonomy of informality, Policy recommendations for informality, Size distribution of firms, Informality and taxes, Informality and productivity, Informality and welfare.

I. Introducción y resumen ejecutivo

Como sucede en muchos otros países, en Colombia hay una carencia de información comprensiva y recurrente del tejido empresarial. El último censo empresarial se realizó en 1990. El presente trabajo propone subsanar esta carencia mediante el diseño de una metodología de construcción de una base de datos firma-empleado-empleador que permite analizar el universo de firmas en su conjunto con un énfasis especial en la informalidad, que es una característica común de los mercados emergentes.

Para lograr este fin, se utiliza la encuesta de micronegocios (EMICRON) y las encuestas estructurales realizadas a los sectores de manufactura, comercio y servicios (EAC, EAS y EAM). Los faltantes de información se complementan con la encuesta de hogares (GEIH), que provee información representativa a nivel trimestral sobre los propietarios y algunas características de sus negocios. Bajo el supuesto de que un empleador representa a una firma, es posible asimilar la muestra de jefes y algunos trabajadores cuenta propia, como representativas del total de firmas en el país. Este ejercicio se realiza para el año 2019, teniendo en cuenta que el año 2020 no tiene información completa y la que se encuentra disponible está muy afectada por los efectos del COVID-19. Para el 2021 el ejercicio no puede ser replicado aún, porque la información de las encuestas estructurales (EAC, EAM

Y EAS) no está disponible. La base de datos resultante no incluye sector agrícola y el financiero, donde la representatividad era baja. También se excluyen el gobierno y el servicio doméstico, que en realidad no pertenecen al universo empresarial.

Este esfuerzo es importante por tres razones. La primera es que, en línea con las recomendaciones del documento CONPES (DNP, 2019), la base de datos generada permite entender el comportamiento de todo el conjunto de empresas colombianas, incluyendo las más pequeñas y a las informales. Esto es fundamental tanto para las descripciones analíticas como para la elaboración de modelos de carácter más teórico. En particular, opera como complemento del Directorio Estadístico Empresarial (DEE) y de los registros administrativos de seguridad social (PILA); cuya distribución es restringida y su disponibilidad de variables, limitada. Adicionalmente, contribuye a sentar las bases para la realización de un eventual nuevo censo económico, porque permite tener un estimativo de los resultados esperados y de los faltantes de información más relevantes. La segunda es que, al estar en su mayoría unida a la encuesta de hogares, que es la base de cálculo de los indicadores de distribución del ingreso, permite un análisis conjunto de los aspectos económicos y sociales del sector productivo; y por lo tanto, de las políticas de productividad, informalidad y bienestar. La tercera, es que la metodología utilizada puede ser aplicada periódicamente, no sólo en el caso colombiano; sino también en el de muchos países en desarrollo que no cuentan con un censo económico o una encuesta empresarial representativa periódica.

La base de datos compilada EEG (Emicron, Estructurales, GEIH) no está exenta de limitaciones, siendo la más importante el que los pesos poblacionales de los jefes de las firmas no fueron diseñados con el propósito de describir el universo de firmas. No obstante, estos pesos obedecen principalmente a la disponibilidad de información, que se adapta a las necesidades de crear una demografía de las empresas. De todas maneras, para limitar el problema de representatividad se realizaron algunas comprobaciones de consistencia con otras fuentes de información.

La segunda limitación es que es necesario imputar algunas de las variables que no son observadas en segmentos determinados de la nueva base de datos, como las características sociodemográficas de los trabajadores. Para realizar estas imputaciones, se utilizan métodos de inteligencia artificial que se alimentan de diferentes bases de entrenamiento como la parte de la encuesta de hogares que hace preguntas a los trabajadores (GEIH, 2019); la EMICRON (2019) y la encuesta de hogares 2021 (GEIH, 2021, marco 2018), que a pesar de proveer información postpandemia, es útil para identificar relaciones estructurales entre variables porque tiene una batería más extensa de preguntas sobre la firma realizadas al trabajador. Otra variable que es necesario calcular es el valor agregado de las firmas que no se encuentran reportadas en la EMICRON, ni en las encuestas estructurales. Esta estimación se realiza a partir de la definición del valor agregado como suma de la remuneración a factores.

La conjunción de esta información permite, por primera vez, la identificación de los principales hechos estilizados de la totalidad del tejido empresarial en Colombia, y su relación con la informalidad, la productividad y la equidad. El primero de estos hechos es su inmensa granularidad; incluyendo una alta incidencia del trabajo cuenta propia. Los cálculos hechos en este trabajo estiman que las firmas de menos de 10 trabajadores constituyen el 99% del total de firmas, y el 96% de las firmas con trabajadores. Esta estadística, y la alta tasa de empleo por cuenta propia nos sitúan en el extremo alto de la distribución de países de la OECD. La causa de la proliferación de firmas pequeñas es aún un tema por explorar, pero la evidencia sugiere que puede estar relacionada con la existencia de umbrales exentos del pago de impuestos, la disponibilidad de capital físico y humano, el monitoreo y control que crece con el tamaño de las firmas y la informalidad.

En el caso del trabajo cuenta propia, interfieren factores más propios de una decisión ocupacional, como la razón por la cual tiene ese trabajo (idea de negocios, subsistencia o flexibilidad).

El segundo hecho estilizado identificado para Colombia es una informalidad empresarial alta y decreciente no sólo en el número de trabajadores en la firma sino también en su productividad. Una versión simplificada del modelo de Ulyssea (2018) permite explicar este comportamiento a partir de tres tipos de firmas: 1) las firmas muy pequeñas y poco productivas que no pueden operar como empresas formales, por barreras de entrada al mercado formal; (2) las firmas con una productividad mayor que podrían operar como formales pero no lo hacen porque no les resulta rentable; y (3) las firmas de productividad más alta, para las cuáles es más rentable operar formalmente.

Este trabajo realiza la estimación contable de esta taxonomía utilizando como variable de productividad el valor agregado por trabajador y como aproximación al valor de la firma, sus ganancias. El ejercicio no replica bien la taxonomía para el trabajo cuenta propia, pero sí para las firmas que tienen trabajadores, para quienes encuentra que en el primer segmento de productividad no operan firmas formales; en el segundo coexisten firmas formales e informales, pero predominan las informales, y en el tercero, aunque existen algunas firmas informales, predominan las formales. No obstante, la variable del valor agregado tiene un problema de endogeneidad, para poder probar una relación causal es necesaria la estimación completa del modelo, como lo hace Fernández (2022) para el caso colombiano.

El tercer hecho estilizado es una informalidad laboral alta pero fuertemente decreciente en el número de trabajadores de la firma. En este caso, una taxonomía de la informalidad laboral también contribuye a explicar su carácter decreciente en la productividad de las firmas. Para las firmas con trabajadores remunerados que pagan impuestos, hace sentido contratar a los trabajadores formalmente porque los costos de contratación formal son deducibles de impuestos. Por el contrario, si la firma genera ingresos inferiores al umbral de contratación, el costo relativo de contratar a un trabajador formalmente frente a hacerlo informalmente es cercano al 47%. Este mecanismo se refuerza con el mayor monitoreo y control que ejercen las autoridades sobre las firmas de mayor tamaño y que hacen para las firmas más grandes más riesgosas las contrataciones informales.

En el caso de las firmas de un trabajador, la decisión de informalidad laboral depende también de la estructura de las cotizaciones a seguridad social, que es regresivo y no lineal. En Colombia el ingreso base mínimo de cotización para trabajadores independientes es de un salario mínimo mensual, lo que implica que los trabajadores con menos ingresos pueden realizar menos deducciones de costos antes de aplicar el costo de cotización del 33%. El caso extremo es el de los trabajadores independientes que ganan exactamente un salario mínimo y no pueden realizar ninguna deducción; o el de los trabajadores que ganan menos de un salario mínimo y no tienen probabilidad de ser formales a un costo razonable. Mientras tanto, los trabajadores que ganan 2.2 salarios mínimos o más pueden deducir hasta el 60% de sus ingresos laborales. Esta inequidad se ve reforzada por el hecho de que la probabilidad de pensionarse y el tamaño de los subsidios son decrecientes en el salario y por lo tanto, en la productividad de los trabajadores. En otras palabras, los trabajadores de menores ingresos tienen menor probabilidad de pensionarse y si lo hacen reciben menos subsidios, y por lo tanto, tienden a percibir las cotizaciones a seguridad social como un impuesto.

Los incentivos y desincentivos a la formalidad descritos anteriormente afectan de manera particular a los trabajadores más vulnerables. En efecto, las firmas pequeñas y las menos productivas tienen más probabilidad de tener trabajadores en condición de pobreza que

el resto de las firmas, como indican en Eslava, Meléndez y Urdaneta (2021). Así mismo, en las firmas pequeñas existe una probabilidad mayor de tener trabajadores migrantes, jóvenes y poco calificados. Desde el punto de vista de las cotizaciones de trabajadores independientes, la relación es más clara: los que reciben menores ingresos, que puede ser una medida aproximada de productividad, enfrentan mayores tasas de cotización.

En los análisis realizados hasta el momento parece existir una correspondencia entre firmas formales muy productivas que contratan trabajadores formales y más calificados y firmas informales poco productivas que contratan trabajadores informales y poco calificados; así como niveles de productividad y sectores donde sólo operan firmas formales, y niveles de productividad y sectores donde sólo operan firmas informales. Sin embargo, estas tendencias no permiten afirmar Colombia es una economía dual. Ejercicios realizados en este capítulo indican que, en el país, las firmas formales e informales comparten amplios rangos de sectores y de productividad. Así mismo, una estimación econométrica sugiere que las diferencias de productividad de los trabajadores formales e informales contratados por la misma firma sólo se explica marginalmente por la formalidad laboral y empresarial. De todas maneras, todos los mecanismos que faciliten esta transición del empleo informal y/o del autoempleo a el empleo formal, como certificaciones de trabajos por horas, son bienvenidas.

De este análisis se derivan algunas recomendaciones de política de carácter regulatorio para incrementar la formalidad laboral. Como indica la Misión de Empleo (Levy y Maldonado, 2022), el gobierno debe buscar una manera de pasar a impuestos generales la mayoría de la carga prestacional, que no incluye aspectos como el seguro de desempleo o los riesgos profesionales. Contrario a lo que se piensa, este cambio no implicaría un costo demasiado oneroso, porque hoy en día el estado financia la seguridad social de todos los trabajadores de empresas que pagan impuestos a partir de deducciones laborales. Un primer paso hacia esta transición sería la creación de una pensión básica universal o dirigida a los trabajadores de bajos ingresos que no pueden jubilarse. Así mismo, debe considerarse establecer una tarifa de cotización progresiva o por lo menos plana, para todos los tipos de firmas y niveles de ingresos de trabajadores independientes. Arbeláez, Fernández y Hernández (2020) diseñaron un plan para plataformas digitales que podría obrar como una caja de arena en esta dirección.

Algunas recomendaciones más operativas, pero menos eficaces y horizontales, incluyen la generación de vouchers de deducciones tributarias de los costos formales de contratación que los agentes puedan deducir cuando crezcan o cuando ganen más ingresos; la inclusión de la seguridad social en el monotributo, y/o el establecimiento de la tarifa del monotributo sobre ingresos netos de costos salariales y de seguridad social. Estas disposiciones harían más atractiva la contratación formal de trabajadores y daría más sentido a los esfuerzos de formalización empresarial. Sin embargo, es necesario diseñar mecanismos adicionales, que por ejemplo prevengan el enanismo de empresas en el esquema de monotributo.

Otras recomendaciones, implican aumentar la productividad de las empresas pequeñas. En el enfoque tradicional, se ha entendido la importancia de aumentar la productividad de las empresas grandes para aumentar el crecimiento y incrementar la contratación de trabajadores formales. Sin embargo, las conclusiones con respecto al aumento de la productividad de las firmas pequeñas son menos claras. Hasta hace poco existía la percepción de que las empresas pequeñas no crecían y sobrevivían largos años con bajos niveles de productividad. Estudios recientes de Angulo (2022) y Confecámaras (2022) señalan no sólo que existen algunas firmas pequeñas que crecen; sino que es posible identificar algunos factores comunes de estas firmas como el acceso a crédito, a tecnología y la formalidad (independientemente de la mutua causalidad entre esta variable

y la productividad). Dada la relación que existe entre el tamaño de las firmas, la productividad y la incidencia de pobreza, las ayudas del gobierno para el crecimiento de estas firmas podrían constituir en sí mismas una política social. Sin embargo, es necesario disponer de mayores estudios y bases de datos para poder diseñar lo que podría denominarse una política social de productividad. En particular, resulta primordial que el Dane realice un nuevo censo para el país o una encuesta representativa que incluya a las firmas informales.

El trabajo se estructura de la siguiente manera: La sección II presenta la descripción de la metodología utilizada en la creación de la base de datos; la sección III, los principales hechos estilizados derivados del análisis del conjunto de firmas; las secciones IV y V, la relación entre el tamaño de firma y la informalidad empresarial y laboral, respectivamente; la sección VI, la relación entre la informalidad, la productividad y la equidad; la sección VII, un análisis sobre el dualismo en Colombia y la sección IV, las conclusiones y recomendaciones de política. Estos ejercicios constituyen un ejemplo de las posibilidades de análisis que ofrece de la base de datos construida, pero en realidad su potencialidad va más allá del alcance que se le da en el presente trabajo.

II. Metodología

Perry et al (2007), en el informe líder del Banco Mundial, realizan uno de los primeros intentos por comprender la relación de la dinámica empresarial y la informalidad a nivel latinoamericano. Según sus hallazgos, las empresas informales engloban no solo pequeñas empresas de subsistencia sino también empresas de mayor tamaño que no cumplen con las regulaciones. También sugieren que se pueden obtener considerables ganancias de eficiencia mediante el traslado de recursos de empresas de baja productividad a empresas de alta productividad. En cuanto a las recomendaciones de política, argumentan que algunas empresas pueden beneficiarse de reducir los costos de la informalidad y reaccionar ante un aumento en los costos de ser informales, pero también entienden que la mejor política para las pequeñas empresas requiere políticas como acceso al crédito formal, capacitación, y servicios de desarrollo empresarial.

De otra parte, uno de los ejercicios más exhaustivos para describir la taxonomía de las empresas en el contexto de la informalidad lo realizó Levy (2018) para el caso de México. Según Levy, la principal causa de la baja productividad en México es la mala asignación de recursos, muy ligada a la informalidad y al trabajo no asalariado. Para llegar a esta conclusión, realizó un análisis exhaustivo a nivel de empresas, que concluye lo siguiente sobre las empresas informales; 1. Absorben una cantidad significativa de capital y mano de obra. 2. Se pueden encontrar en todos los sectores de la economía y en todo el territorio. 3. No son necesariamente ilegales porque la mayoría contratan trabajadores a través de trabajo no asalariado (eso es legal en México), y no necesariamente completamente informales (hay una escala de informalidad). 4. Son muy pequeñas (menos de 5 trabajadores), en su mayoría, pero no todas las empresas informales son pequeñas. La proliferación de pequeñas empresas genera una proliferación de empresarios sobre trabajadores, que no siempre es óptima. 4. Es creciente. La composición de la actividad económica con el tiempo se ha desplazado hacia el sector informal. 5. Son poco productivas. Las empresas que contratan trabajadores formales y/o1 son formales son más productivas, pero no todas las empresas formales son más productivas que todas las empresas informales.

Este análisis lleva a Levy a la premisa de que, las empresas de baja productividad absorben más capital y mano de obra de lo que deberían; mientras que las más productivas no reciben suficientes recursos; y este proceso se ve reforzado por la

longevidad de esas pequeñas empresas improductivas. Además, las pequeñas empresas improductivas se comportan de manera socialmente ineficiente. Por ejemplo, eligen ser informales cuando pueden ser formales. Al ser informales permanecen lejos del control de las autoridades y por lo tanto, no aceptan efectivo, no crecen, no hacen uso de tecnología, y no contratan trabajadores formales. En algunos casos estas firmas tienen altas tasas de entrada y bajas tasas de supervivencia, por lo que generan puestos de trabajo de corta duración. Este análisis fue resumido posteriormente por Álvarez y Ruane (2019) para estimar Ulyssea (2018) para el caso de México.

Más condensada e igualmente rica es la descripción de la informalidad empresarial realizada en Ulyssea (2018), Ulyssea (2019) y Ulyssea (2020). Los principales hallazgos de estos artículos son los siguientes: 1) Existen niveles decrecientes del margen intensivo y extensivo de informalidad sobre el tamaño de las empresas. 2) Las empresas informales son en promedio más pequeñas, están dirigidas por personas con menos educación, pagan salarios más bajos y son menos productivas que las empresas formales, pero a pesar de estas diferencias, las empresas formales e informales en Brasil coexisten en todos los sectores y niveles de productividad, y también argumentan que no existe evidencia de que falten empresas medianas características de una economía dualista. 3) La brecha salarial entre trabajadores formales e informales -característica de las economías segmentadas ya que significa que los trabajadores formales e informales realizan diferentes tareas en la economía- desaparece cuando se controla por las características propias de las empresas (efectos fijos). 4) el proceso de selección dinámica tiene lugar tanto en el sector formal como en el informal, pero es más débil en el segundo y, por lo tanto, la heterogeneidad ex-ante más que ex-post es el determinante de la dinámica de la empresa.

Con base en todos estos hechos estilizados, Ulyssea desarrolló un modelo en el que las empresas menos productivas ni siquiera pueden pagar los costos fijos de ser formales; la empresa de mediana productividad podría pagar los costos fijos de ser formal pero encontrar que es más rentable ser informal, y las empresas más productivas, que son una minoría pero también las más grandes, son las únicas que encuentran rentable ser formal. Esta perspectiva tiene la ventaja de integrar en un mismo modelo dos puntos de vista de la informalidad que hasta ahora se consideraban excluyentes: la perspectiva dual del mercado (Lewis, 1954 y Harris y Todaro, 1970); y el margen de De Soto según el cual la informalidad se debe a los altos costos de entrada (De Soto et al., 1989 y De Soto 2000). Otra ventaja de Ulyssea (2018) es que es un modelo de equilibrio general que tiene en cuenta tanto el margen intensivo como el extensivo de la informalidad.

Para el caso de Colombia, Eslava, Haltiwanger y Pinzon (2019) analizaron toda los establecimientos no micro manufactureros en Colombia y encontraron que la distribución del tamaño de las empresas manufactureras formales en Colombia exhibe una alta concentración de pequeñas empresas antiguas, en términos de número y empleo, lo que apunta a una escasez de espíritu empresarial de alto crecimiento y una probabilidad relativamente alta de supervivencia a largo plazo para las empresas pequeñas y probablemente improductivas. Los autores consideran que ésta es una característica clave de las bajas tasas de crecimiento de la productividad en Colombia. Más enfocado en la informalidad, Fernández (2019) hizo un primer intento por comprender la dinámica de la informalidad empresarial, pero con base en la Encuesta de Microestablecimientos (2013-2016), que en ese entonces cubría solo microempresas y no era una representativa.

Si bien estos documentos, realizados para entender el caso colombiano, hicieron contribuciones importantes a la comprensión de la dinámica empresarial y la informalidad, se basaron en conjuntos parciales de información y, por lo tanto, no pueden abordar

algunas de las características de informalidad empresarial identificadas para México o Brasil. Este trabajo busca enmendar esta situación.

III. Metodología

Esta sección presenta cinco aspectos metodológicos que se tuvieron en cuenta para la creación de la base de datos y su análisis respectivo. La primera parte presenta las definiciones básicas utilizadas en el resto del documento, la segunda contiene las directrices que se adoptaron para construir la base de datos y homologar variables, la tercera explica la metodología para completar la base de datos con imputaciones, desde el punto de vista de las características de los trabajadores, la cuarta presenta las estimaciones de la productividad que fue necesario en los segmentos de la nueva base de datos que no disponían de esta información, y la quinta y última parte presenta la metodología para medir las ganancias de los empresarios.

II.A. Definiciones:

Antes de explicar en detalle la metodología de compilación de datos, vale la pena establecer algunas definiciones de las variables que van a ser utilizadas en el artículo.

- Firma: Unidad económica que desarrolla una actividad productiva de bienes o servicios, con el objeto de obtener un ingreso, actuando en calidad de propietario o arrendatario de los medios de producción². Puede o no incluir trabajo por cuenta propia y firmas de un trabajador
- Tamaño de firma: Número de trabajadores. Incluye trabajadores directos, contratistas directos, trabajadores no remunerados, socios y jefes.
- Firma formal: La definición más utilizada en este trabajo es estricta: firmas registradas recientemente, con contabilidad formal y que pagan impuestos o están exentas. Otras definiciones utilizadas incluyen: i) registradas (recientemente o no) y con contabilidad formal ii) registradas (recientemente o no) y iii) registrada (recientemente o no) o con contabilidad formal
- Trabajador formal: contribuye a salud y pensiones
- Trabajador calificado: trabajador con educación secundaria.

III.B. Construcción del universo empresarial

El universo empresarial en este trabajo se construye de la siguiente manera: las empresas con menos de 10 trabajadores están representadas por EMICRON; las empresas con más de 10 trabajadores (11+) están representadas por las encuestas estructurales (EAM, EAS y EAC)³, y por la GEIH si son informales o pertenecen a un sector o a un tamaño no cubierto por las encuestas estructurales. En efecto, bajo el supuesto de que cada jefe representa a una firma, y dado que las estadísticas de jefes son representativas del total de jefes del país en una base trimestral, es posible afirmar que la muestra de jefes de empresas es representativa del total de firmas en Colombia⁴. Se asume también que las empresas referenciadas en las encuestas estructurales son formales, sus jefes cotizan a pensiones

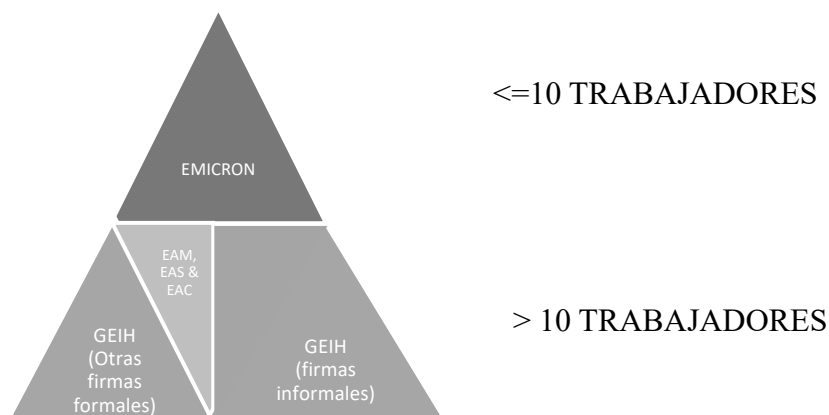
² Elste criterio se adapta a la definición del Dane de micronegocios en su metodología general encuesta de micronegocios – Emicron.

³ Se descartan las empresas con menos de 10 trabajadores en esas encuestas

⁴ . Los trabajos de los segundos empresarios no se consideran en el ejercicio, porque las variables que describen el trabajo son diferentes y generalmente no están disponibles.

o a salud, y contratan trabajadores formales. La nueva base de datos se conoce como EEG (EMICRON-Estructurales-GEIH). El Diagrama 1 muestra las fuentes de información del EEG, cuyas fuentes principales se detallan a continuación.

Diagrama 1. Fuentes de información del EEG (2019)



Las bases de datos utilizadas para construir la EEG (2019) son las siguientes:

- **EMICRON (Encuesta de Micronegocios).** Esta encuesta empleado-empleador de corte transversal es un módulo de la Encuesta de Hogares y por lo tanto, representativa a nivel de microempresarios. Los pesos poblacionales son ajustados al universo de microempresarios y a las pérdidas de información en el proceso de recolección de datos. Su cobertura excluye negocios gubernamentales, transporte aéreo, sector financiero y de seguros, y algunas áreas departamentales (Dane, 2020d).⁵
- **Gran Encuesta Integrada de Hogares (GEIH, dic 2018 – nov 2019).** Esta encuesta de carácter transversal es representativa mensualmente para los grandes agregados y a nivel trimestral, para niveles más desagregados. La parte de la encuesta de hogares que se utiliza más activamente en este ejercicio es el conjunto de preguntas realizadas sobre trabajadores independientes (Dane, 2013)⁶. Sin embargo, la sección de preguntas a los trabajadores independientes se utiliza para hacer inferencias. Para hacer esta encuesta compatible con la EMICRON se ajustaron los meses de análisis, teniendo en cuenta que la EMICRON se realiza un mes después que la de hogares y se excluyeron los sectores no cubiertos por EMICRON. La parte de la encuesta de hogares dirigida a los trabajadores también se utiliza para completar la información de los trabajadores.

⁵ Incluye microempresas en 24 ciudades y sus respectivas áreas metropolitanas, pero no incluye Arauca (81), Casanare (85), Vichada, Guainía (94), Guaviare (95), Vaupés (97), Putumayo (86) y Amazonas (91). Se excluyen los datos de San Andrés (88) ya que no están incluidos en GEIH (alternativamente podemos incluirlos en GEIH). El sector agrícola no está excluido, como en Ulyseas. La base de datos original de Emicron 2019 tiene 101 681 observaciones de trabajadores que el uso de pesos se expande a 6.894 711 trabajadores y 86 969 empresas de información que se amplían a 5.874 177 empresas. Excluyendo San Andrés (1434 observaciones) hay 100247 observaciones de los trabajadores y 85756 observaciones de la empresa, de las cuales 75.122 son de trabajo por cuenta propia.

⁶ Inicialmente el GEIH cuenta con 337.507 observaciones, lo que representa 22'303.304 trabajadores (promedio para todo el año, pesos/12). Sin embargo, es necesario excluir algunos sectores para que sea comparable a EMICRON: Se excluyeron los siguientes sectores: gobierno (12743), servicios públicos (3447), transporte aéreo (230), sectores financieros y de seguros (4815) y actividades domésticas que actúan como empleadores (12002). Adicionalmente, fusionamos ambas bases de datos para obtener los siguientes datos que faltaban en EMICRON: nivel de habilidad imputado de los trabajadores en EMICRON, utilizando técnicas de análisis discriminario (detalladas en la siguiente sección), nivel escolar y horas trabajadas de emprendedores y nivel CIU de cuatro dígitos (según la respuesta del emprendedor).

- **Encuesta Anual Manufacturera (EAM, 2019).** Se trata de una encuesta estructural y exhaustiva que considera todos los establecimientos manufactureros con 10 o más de 10 personas empleadas a nivel nacional. La información corresponde a 2019 pero fue recopilada en 2020. Durante 2019, 7.631 establecimientos de los 6.742 tienen 11 o más trabajadores, involucrando a 706.563 ocupados (Dane, 2020b).
- **Encuesta Anual de Comercio (EAC, 2019).** Se trata de una encuesta estructural y exhaustiva que considera todas las unidades económicas minoristas con 10 o más de 10 personas empleadas a nivel nacional, eximiendo a los minoristas de bienes usados (se incluyen las reventas de automóviles) (Dane, 2020a). La información corresponde a 2019 pero fue recopilada en 2020. Durante 2019, 9.859 unidades económicas de las cuales 8.033 tienen 11 o más trabajadores, involucrando a 632.270 ocupados (trabajadores remunerados y no remunerados).
- **Encuesta Anual de Servicios (EAS, 2019).** Se trata de una encuesta estructural y exhaustiva que considera todas las unidades económicas formales de servicio a nivel nacional de bienes no usados. A diferencia de EAC y EAM, el nivel de corte de cubrimiento de la encuesta difiere con los subsectores.⁷ Para 2019 se entrevistaron 6.398 unidades económicas, de las cuales 5.985 tienen información sobre el número de trabajadores y menos de 10 trabajadores que involucran a 2'054.164 ocupados (trabajadores remunerados y no remunerados).
- **Gran Encuesta Integrada de Hogares (2021), marco 2018.** Esta base de datos no se utiliza directamente para crear el universo de empresas como tal, pero sí para estimar variables como el margen intensivo de la formalidad, o la probabilidad de que una firma formal o informal contrate trabajadores vulnerables. En efecto, a pesar de que esta información se recolectó después de la pandemia, cuenta con la ventaja de preguntar a cada empleado si la firma donde trabaja es formal o informal⁸.

La unidad de observación utilizada en esta base de datos es la “*unidad económica*”, en línea con lo sugerido por el manual de la OCDE (2007). El tamaño de la empresa en términos de trabajadores incluye trabajadores directos, contratistas directos e indirectos, pasantes, trabajadores no remunerados y socios y empresarios⁹. Además de los sectores no incluidos en EMICRON, como el financiero, el gobierno y los servicios domésticos, este análisis excluye el sector agrícola, en aras de representatividad sectorial. En efecto, el número de trabajadores en este sector en la EEG (2019) era tan sólo el 30% de lo que se reporta en la GEIH en firmas mayores de 10 trabajadores.

La base de datos se construye para el 2019, que es el primer año para el cual se tiene información completa de las encuestas utilizadas. Esta información también se encuentra disponible para el 2020, pero tiene limitaciones en la recolección de información y está alterada por el efecto del COVID. Cuando se dispongan los datos para el 2021, el ejercicio podrá completarse con los avances de la nueva GEIH, orientados a tener una base de datos empleado/empleador. En el caso de que se realice el Censo Económico se podrán ajustar los pesos poblacionales utilizados para asegurar la representatividad de la encuesta. Toda la información disponible utiliza los nuevos factores de expansión, provistos por el DANE con el nuevo marco de encuesta que tiene proyecciones de la población con base en el último censo 2018.

⁷ Una lista detallada de los niveles de corte por sectores se puede encontrar en Dane (2020c) <https://www.Dane.gov.co/index.php/estadisticas-por-tema/servicios/encuesta-annual-de-servicios-eas>

⁸ En general las características de las personas que trabajan en determinado tipo de firma no cambian significativamente en el tiempo y se pueden controlar utilizando una variable dummy.

⁹ Esta unidad se refiere en lo general al establecimiento, lo que es consistente con la EMICRON, la EAS y la EAC. Sin embargo, la información del número de trabajadores en la encuesta de hogares y en la EAM se refiere a la unidad que incluye sucursales etc.

En el Cuadro 1 se resumen los datos utilizados en este documento. Se utilizan 96.174 observaciones de las cuales 76.130 provienen de EMICRON, 13.420 de las encuestas estructurales (EAM, EAC y EAS) y 458 de las preguntas que se hacen a trabajadores independientes en la GEIH. Estas observaciones representan a 4'399 709 empresas y cerca de 10 millones de trabajadores. Excluyendo las firmas de un trabajador, representan un poco más de 1 millón de empresas y 6.4 millones de trabajadores; y excluyendo a todo el trabajo cuenta propio se tienen 575 mil firmas y 5.4 millones de trabajadores¹⁰. Una vez construido el universo empresarial, se procedió a homologar las variables utilizadas¹¹:

Cuadro 1. Composición de la base de datos EEG

	EMICRON			EAM, EAS y EAC	GEIH	TOTAL		
	Total	> 1 trabajador	No cuenta propia	> 10 trabajadores	> 10 trabajadores	Total	> 1 trabajador	No cuenta propia
# Firmas (Observaciones)	76 130	16 350	9 256	13 420	458	96.174	36.394	29 283
# Firmas (muestra expandida)	4'355 827	970 814	532 549	13 420	24 307	4'399 709	1'014 573	575 262
# Trabajadores (Observaciones)	76.130	16.350	9 256	13.285	458	96.039	36.259	29 148
# Trabajadores (muestra expandida)	6'001392	2'610 843	1'645639	3'168 969	603 450	9'773 812	6'338 810	5'397 366

Fuente: EEG. Pesos poblacionales marco 2018.

III.C. Comparación del número de firmas cubierto con otras fuentes de información

Una primera aproximación para establecer la representatividad de la EEG (2019), es segmentar el análisis para cada uno de los subcomponentes de esta base de datos. El subcomponente de la EMICRON no cuenta con un marco de representatividad en el universo empresarial, porque las encuestas y los censos a este nivel son costosos e inefectivos. Sin embargo, esta encuesta se estima como un módulo de la encuesta de hogares, lo que se considera, en la literatura como la mejor práctica para contabilizar negocios pequeños. Así mismo, el carácter censal de las encuestas estructurales (EAC, EAM y EAS) despeja las dudas existentes sobre su representatividad¹².

Queda entonces por definir la representatividad de las firmas de más de 10 trabajadores que no se encuentran representadas en las encuestas estructurales. Una manera de proceder es comparar diferentes agregados en términos del número de trabajadores dependientes de la encuesta de hogares (GEIH, 2019) con la EEG (2019) extendida a través de unos pesos poblacionales de firma multiplicados por el número de trabajadores de cada firma. Adicionalmente, la EEG (2019) se ajusta a la cobertura de la EEG (2019) descrita en la sección anterior. Adicionalmente, como en la parte de la GEIH dirigida a los dependientes no se puede identificar la formalidad de la firma, es necesario analizar la representatividad conjunta de la EEG (2019) para el segmento de más de 10 trabajadores.

¹⁰ De acuerdo con el Dane y la OIT las unidades por cuenta propia son aquellas que no tienen trabajadores remunerados. De manera tal, que pueden existir unidades por cuenta propia con más de un trabajador.

¹¹ Los resultados de esta homologación están disponibles bajo demanda

¹² No obstante, algunos ejercicios nos demuestran que en los rangos de firmas entre 10 y 50 trabajadores es posible que no se capturen todos los negocios existentes.

Como es posible observar en el Cuadro 2, el número de trabajadores en la EEG (2019) es un 82% de los trabajadores que reportan estar en el sector servicios de la GEIH. Esta estadística es del 72% en comercio y del 48% en industria y construcción. Por tamaño de firma, se observa una buena representatividad en las firmas de más de 50 trabajadores, pero en las firmas ente 11 y 50 trabajadores la representatividad es baja (40%). Como se mencionó anteriormente, esto puede estar relacionado con fallas en la recolección de las encuestas estructurales (EAC, EAM y EAS) en los segmentos de firmas relativamente pequeñas. Es importante anotar que en este ejercicio no es posible diferenciar las fallas en representatividad de las firmas de la GEIH y las estructurales, porque no se tiene información en la GEIH sobre si los trabajadores de la GEIH se encuentran en firmas formales o informales.

Cuadro 2. Representatividad de la EEG en firmas de más de 10 trabajadores.

	Miles de trabajadores (> 10 trabajadores)		
	EEG	GEIH trabajadores	EEG/GEIH
Servicios	2378	2901	82%
Comercio	637	880	72%
Industria y construcción	756	1564	48%
de 11 a 50	665	1,654	40%
de 51 a 100	326	466	70%
101+	2781	3226	86%
Total	3,772	5,422	69%

Fuente: EEG y GEIH.

La comparación de la EEG (2019) con otras fuentes de información empresarial es difícil por las diferentes delimitaciones de la cobertura en términos de los sectores y de la definición de firma utilizada. Por ejemplo, el Directorio de firmas (DEST), restringe su cobertura a la existencia de algún registro administrativo y bajo este supuesto encuentra un total de 5 millones de firmas¹³. Sin embargo, la cobertura sectorial es menos clara, en el segmento de trabajadores cuenta propia. Por esta razón se restringe el universo de comparación a aquellas firmas con algún tipo de registro y por lo menos un empleado. Los resultados, que se presentan en el Cuadro 3, indican una cobertura cercana al 67%. En efecto, la EEG contabiliza 548 mil unidades, y la DEST 826 mil unidades. Esta similitud se logra a pesar de que los procedimientos de recolección de información y las fuentes son totalmente distintas.

Cuadro 3. Comparación con otras fuentes de información.

	# firmas, > 1 trabajador			
	EEG	EEG registrado	DEST	EEG registrado /DEST
Comercio	373	209	409	51%
Servicios	363	194	229	85%
Industria y construcción	279	144	188	77%
Total	1015	548	826	67%

Fuente: DEST, GEIH y EEG

¹³ El pre-censo del censo que se refiere a las unidades económicas que tienen establecimiento es de 2.5 millones de firmas. Sin embargo, los datos no son comparables por divergencias en la definición de firma.

III.D. Imputación de variables internas a la firma (margen intensivo)

La base de datos resultante tiene una buena batería de variables de análisis, en contraste con otros ejercicios de estimación del número de empresas que sólo cuentan con unas pocas variables. El Cuadro 4 presenta la información observada (O), asumida (A), imputada (I) y no disponible (ND) en la EEG (2019) y en las bases de datos, que se utilizan para completar los faltantes de información como: 1) la encuesta de hogares GEIH (2019) en su segmento dirigido a los trabajadores, para aquellas firmas con más de 4 trabajadores; 2) la base de datos de EMICRON (2019) a nivel de trabajadores y 3) la encuesta de hogares GEIH (2021, marco 2018). Nótese en el cuadro que la ventaja de estas dos últimas bases de datos es que contienen información simultánea sobre las características de los trabajadores y la formalidad e informalidad de las firmas¹⁴.

Para realizar las imputaciones de las características sociodemográficas de los trabajadores se utilizó el método de inteligencia artificial, y más específicamente, el “bosque aleatorio” o “random forest”, cuya metodología y comparación con otros métodos de estimación alternativos se encuentran descritos en el Anexo 1. Este método estima las probabilidades de que un trabajador tenga determinada característica a partir de árboles de decisión cuyos nodos corresponden a las características de las firmas como el total de empleados de la firma donde trabaja, el área geográfica y la rama sectorial donde se desempeña, si el trabajador es remunerado, no remunerado o socio, la formalidad de la firma y si el dato es de postpandemia, en el caso de que se incluye la encuesta GEIH 2021 entre las bases de entrenamiento. Una ventaja del bosque aleatorio es que no es posible tener datos y variables omitidas en algunas de las bases de entrenamiento, lo que es una clara diferencia con una regresión tradicional, cuando se quiere obtener diferentes variables de diferentes bases de entrenamiento.

¹⁴ La EMICRON contiene información acerca de la informalidad empresarial porque es una encuesta empleado-empresario, pero se restringe a firmas de menos de 10 trabajadores; lo que resulta un impedimento cuando se trata de estimar lo que sucede en las firmas de mayor tamaño.

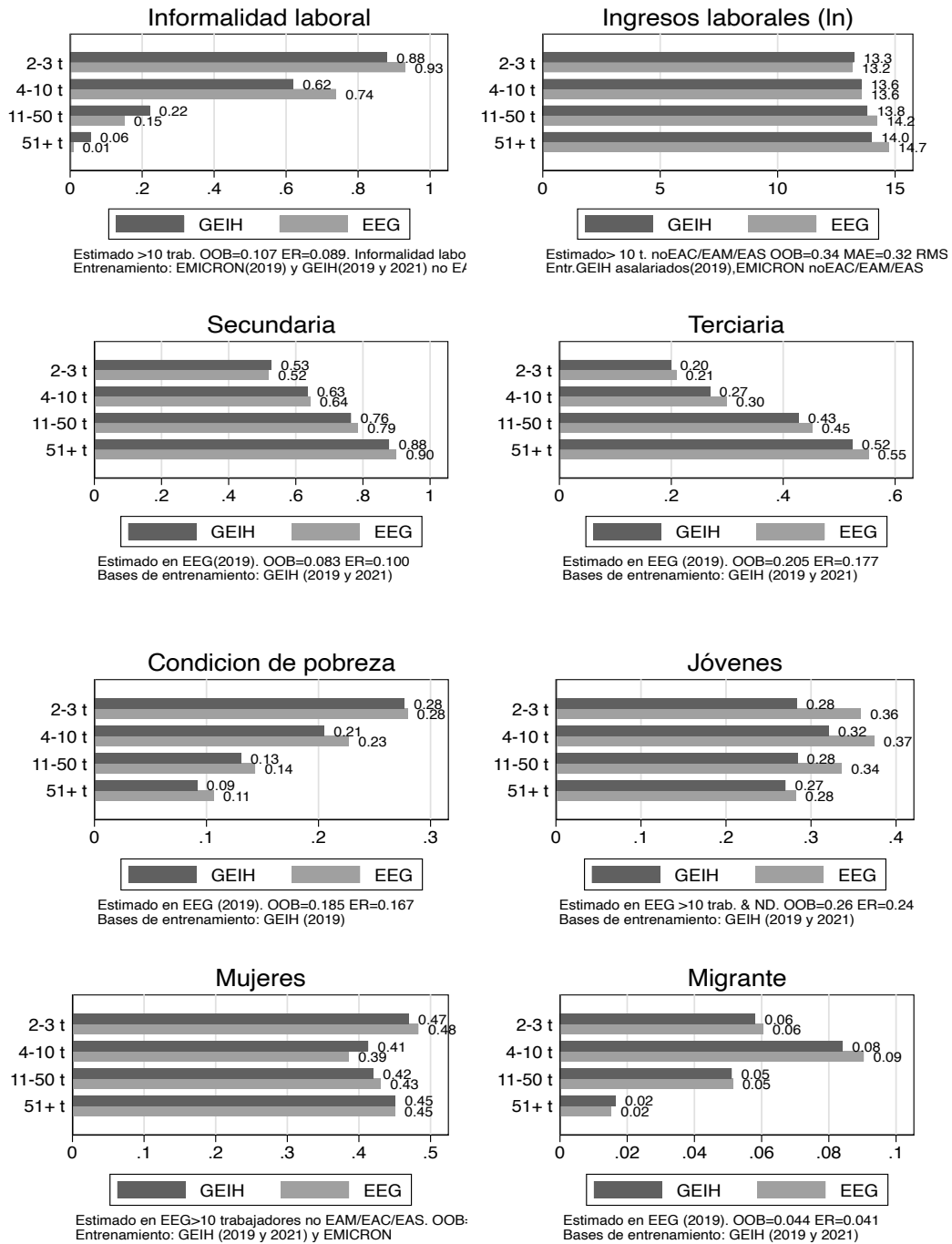
Cuadro 4. Disponibilidad de información para estimar la productividad a nivel de firma

			EEG (2019)					Base de datos de entrenamiento		
			EMI-CRON Firmas	EAM	EAC	EAS	GEIH firmas	GEIH Trabajadores (2019)	GEIH Trabajadores (2021, Marco18)	EMICRON Trabajadores (2019)
Características	Trabajadores	Informalidad laboral trabajadores	O	A	A	A	I	O	O	O
		Mujeres %	O	O	O	O	I	O	O	O
		Ingreso laboral (%)	O	I	I	I	I	O	O	O
		Calificación trabajadores (%)	I	I	I	I	I	O	O	I
		Jóvenes % (<=28)	O	I	I	I	I	O	O	I
		Migrantes (%)	I	I	I	I	I	O	O	I
		Pobreza	I	I	I	I	I	O	ND	O
		Valor agregado	O	O	O	O	I	O	O	O
	Consumo intermedio	O	O	O	O	I	O	O	O	
	Independientes	Calificación jefe	O	A	A	A	O	ND	ND	O
		Informalidad laboral jefes	O	A	A	A	I	O	O	O
		Género jefe	O	A	A	A	O	ND	ND	O
		Pobreza jefe	O	A	A	A	O	O	O	O
		Razón para crear el negocio	O	A	A	A	O	ND	ND	O
Características de las firmas	CIU2d	O	O	O	O	O	O	O	O	
	Area	O		O	O	O	O	O	O	
	Rural	O	A	A	A	O	O	O	O	
	Asalariado	O	O	O	O	O	O	O	O	
	Socios	O	O	O	O	O	O	O	O	
	No remunerados	O	I	I	I	O	O	O	O	
	# trabajadores	O	O	O	O	O	ND	ND	O	
	Informalidad empresarial	O	A	A	A	O	ND	O	O	
# worker (rangos)	O	O	O	O	O	O	O	O		

Fuente: EEG (2019), GEIH (2019), EMICRON (2019) y GEIH (2021, Marco 2018)

Estas estimaciones asumen que las probabilidades de ocurrencia de las variables dependientes son las mismas a nivel de trabajadores y de firmas, condicionadas a ciertas características de las firmas. La concordancia entre las variables estimadas para la EEG, usando pesos poblacionales a nivel de trabajadores (pesos poblacionales de las firmas multiplicados por el número de trabajadores) y las observadas en la GEIH se presentan en la Gráfica 1.

Gráfica 1. GEIH y EEG (estimado para algunos segmentos) por tamaño de firma.



Fuente: EEG (2019, pesos poblacionales de trabajadores), GEIH (2019). Elaboración propia con base en EMICRON (2019) GEIH (2019) y GEIH (2021). T: número de trabajadores en la firma. Variables de estimación: tamaño por rangos, área (cuando la estimación no involucra a las encuestas estructurales), sector 2 dígitos, trabajador remunerado o socio, dummy postpandemia, formalidad de firma flexible (no cotiza a salud y pensiones) y estricta(cotiza a salud y pensiones), cotizaciones del jefe de la firma, en la estimación de informalidad empresarial y educación del jefe en estimación del ingreso. Método de estimación: random forest. ER: tasa de error. OOB: error fuera de canasta. Incluye trabajo por cuenta propia.

III.E. Estimación contable del valor agregado de las firmas asociadas a la GEIH

El Dane calcula el valor agregado de las firmas en la EMICRON y en las encuestas estructurales como la producción bruta menos el consumo intermedio (eq 1). Sin embargo, esta información no se le pregunta a los empresarios en la GEIH. Para calcularlo, en este trabajo se utiliza la estimación del valor agregado desde el punto de vista de la remuneración a factores, según la cual la producción bruta es igual al ingreso disponible, que a su vez es igual a la remuneración a trabajadores, los impuestos netos sobre la producción, el ingreso mixto y el ingreso bruto de explotación (eq 2). Formalmente, el ejercicio se puede explicar de la siguiente manera:

$$VAB_i = PB_i - C_i + Tnetos_i \quad (eq 1)$$

$$VAB_i = Yf_i + Tprod_i$$

$$VAB_i = Re_i + EBE_i + IngMixto_i + Tprod_i \quad (eq 2)$$

$$VAB_i = (1 + t_w)w_{if} l_{if} + w_{ij} l_{ij} + t_{vat} VA_i + \Pi_i + ebe_{ciiu} VA_i$$

$$VAB_i(1 - t_{vat} - ebe_{ciiu}) = ((1 + t_w)w_{if} l_{if} + w_{ij} l_{ij} + \Pi_i)$$

$$VAB_i = ((1 + t_w)w_{if} l_{if} + w_{ij} l_{ij} + \Pi_i) / (1 - t_{vat} - ebe_{ciiu})$$

$$VAB_i = ((1 + t_w)w_{iL^s}(1 - P^i) + w_{iL^s}(1 - P^i) + \Pi_i) / (1 - t_{vat} - ebe_{ciiu}) \quad (eq 3)$$

Donde:

VAB: Valor Agregado Bruto

Pb: Producción

Ci: Consumo intermedio

Tprod: Otros impuestos netos sobre la producción $t_{vat} VA_i$, $t_{vat}=0$ si los ingresos anuales < 100m

Yf: Ingreso de los factores

Re: remuneración a los trabajadores $(1 + t_w)w_{if} l_{if} + w_{ij} l_{ij}$

EBE: Excedente bruto de explotación $EBE=0$ $t_{vat}=0$ si los ingresos anuales < 100m

ebe_{ciiu}: Excedente bruto de explotación/valor agregado de cuentas nacionales, a nivel CIIU 4d

IngMixto: Ingreso Mixto, Π_i se asume que sólo el socio principal gana este salario y que los no remunerados no ganan ingresos

L^s : total de trabajadores asalariados

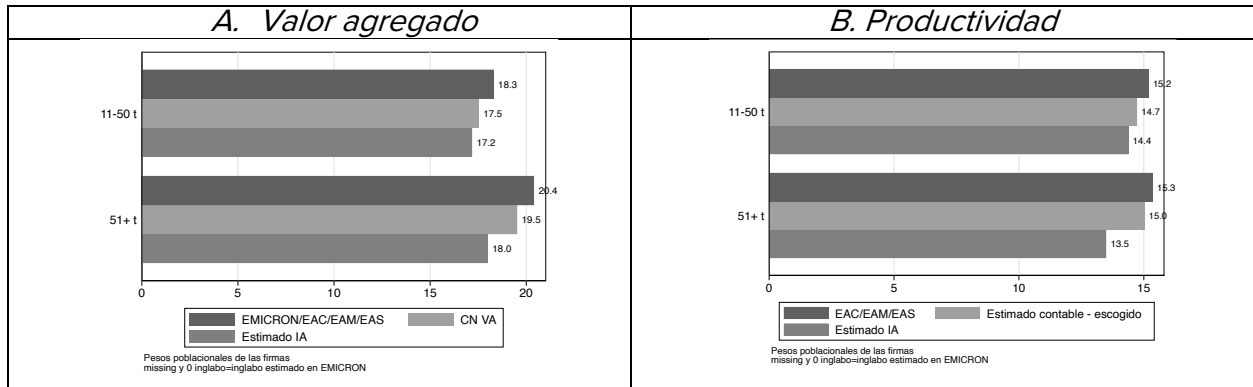
P^i : margen intensivo, probabilidad de un trabajador sea informal en la firma (estimado).

En la EEG (2019) se disponen de todas las variables necesarias para estimar la ecuación (3), menos el ingreso bruto de explotación (EBE), que representa la renta sobre el capital que no se recoge en el ingreso mixto. Para estimarlo se hace uso de la matriz de ingreso-utilización del Dane, que tiene 61 agrupaciones de actividades, que pueden correlacionarse con las categorías de CIIU a cuatro dígitos¹⁵. El EBE se asume cero en las firmas cuyos jefes ganan menos de 100 millones de pesos al mes.

La gráficas 2A y 2B presentan la comparación de la estimación en logaritmos del valor agregado y la productividad en firmas de más de 10 trabajadores. La barra superior se refiere a los datos de productividad laboral observados en las encuestas estructurales (EAC, EAM Y EAS). La barra del medio hace referencia al cálculo realizado de manera contable y la barra de abajo a la estimación de la productividad laboral con el método de inteligencia artificial, descrito anteriormente.

Gráfica 2. Estimación de logaritmo del valor agregado con dos métodos alternativos

¹⁵ Se asume que todas las actividades CIIU que pertenecen a una agrupación contribuyen al excedente bruto de explotación de manera equitativa, razón por la cual a cada CIIU se le imputará con el promedio de la agrupación de cuentas nacionales repartido entre todos los CIIU que pertenezcan a cada una de las 61 categorías de actividades de Cuentas Nacionales.



Fuente: EEG (2019) y cálculos propios con base en EMICRON, EAC, EAM y EAS. No incluye trabajo por cuenta propia

III.F. Estimación contable de las ganancias de las firmas

Las ganancias brutas de las firmas pueden establecerse como la suma del valor agregado, el pago del IVA, el impuesto de renta neto de la deducción del pago de nómina formal (como es el caso en Colombia) y el pago de la nómina formal e informal, como se ilustra en la Ecuación (4)¹⁶.

$$p_{f,i} = (1-t_y) \{ (1-t_{vat}) VAB_i - (1-t_w) w_i l_i^f \} - w_i l_i^i \quad (Eq 4)$$

$$p_{f,i} = (1-t_y) \{ (1-t_{vat}) VAB_i - (1-t_w) w_i L^s (1-P^i) \} - w_i L^s P^i$$

Donde

t_y : Impuesto de renta (28%), 0% si la firma gana menos de 46' anuales o la firma es informal

t_{vat} : IVA (19%), 0% si la firma gana menos de 100' anuales o la firma es informal

t_w : cotizaciones a seguridad social (47%)

w_i : salario promedio de la firma a los trabajadores asalariados

l_i^f : trabajadores formales

L^s : total de trabajadores asalariados

P^i : margen intensivo, probabilidad de un trabajador sea informal en la firma (estimado en EMICRON y GEIH y asumido en estructurales).

La ecuación 5 (Eq 5) ilustra la determinación de las ganancias de las firmas netas de los costos que implican las actividades de monitoreo y control por parte de las autoridades (Ulyssea, 2018)¹⁷. Nótese que, a diferencia de Ulyssea(2018), se permite que las firmas informales contraten trabajadores formales. Lo anterior es particularmente importante, si el criterio de informalidad es estricto, o incluye el pago de impuestos. En otras palabras la firma puede contratar informalmente si está registrada en la Cámara de Comercio así no pague impuestos o esté exento de los mismos.

$$p_i^f = va_i - (1-t_w) w_i L^s (1-P^i) - (1 - L^s (P^i P_c) / b_c) w_i L^s (P^i P_c) - L^s (P^i (1-P_c)) / b_{nc} w_i L^s (P^i (1-P_c)) \quad (Eq 5)$$

$$p_i^i = (1-t_y) \{ (1-t_{vat}) va_i - (1-t_w) w_i L^s (1-P^i) \} - (1 - L^s (P^i P_c) / b_i) w_i L^s (P^i P_c) - L^s (P^i (1-P_c)) / b_i w_i L^s (P^i (1-P_c))$$

p_i^f : ganancias de las firmas formales

p_i^i : ganancias de las firmas informales

b_c =coeficiente de costo del monitoreo y control para la firma, trabajadores calificados, mayor=menor costo

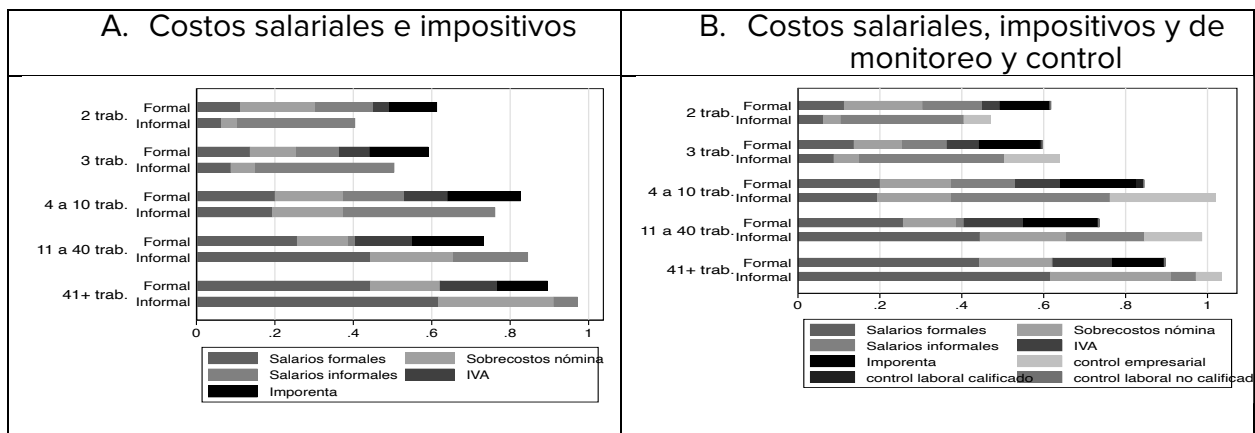
¹⁶ Se supone que las ganancias de los socios son parte de las ganancias y que los trabajadores no remunerados no reciben ingresos.

¹⁷ Para su cálculo se utilizaron los parámetros encontrados por Fernández (2022) para el caso de Colombia.

b_{nc} =coeficiente de costo del monitoreo y control para la firma, trabajadores no calificados, mayor=menor costo
 b_i =coeficiente de costo del monitoreo y control para la firma para todos los trabajadores, firmas informales
 P_c =probabilidad de tener a un trabajador calificado

Las gráficas 3A y 3B presentan la estructura de costos salariales, regulatorios e impositivos relativos de las firmas formales e informales sin incluir e incluyendo los costos que trae para la firma el monitoreo y control de las autoridades. Como es posible observar en los resultados, en las firmas más pequeñas los costos de ser formal son mayores, mientras que en las más grandes los costos de ser informal son mayores; pero el cambio ocurre en firmas de mayor tamaño relativo si no se consideran los costos de monitoreo y control.

Gráfica 3. Estimación de costos diferentes a insumos de firmas formales e informales.



Fuente: EEG (2019). Excluye empleo por cuenta propia

Recuadro 1 Metodología de anualización y reponderación de factores de expansión

Una de las críticas de la construcción de indicadores anualizados, especialmente de mercado laboral a partir de cortes transversales repetidos como es el caso de la GEIH, que se recolecta mensualmente, es que un promedio directo anual entre todos los meses no explica realmente la evolución mes a mes que existe en el mercado laboral.

De esta manera, al calcular un promedio directo de los ocupados de cada mes se estaría dejando por fuera a aquellos ocupados que, si bien es cierto, pasaron por alguna transición, ya sea por el desempleo o la inactividad, al menos estuvieron ocupados en algún momento dentro del año de referencia y estos últimos estarían quedando por fuera del cálculo y por lo tanto se estaría subestimando el número de ocupados si no se toman en cuenta a estos casos particulares.

Por lo tanto, se sigue la metodología de Lora, Benitez y Gutiérrez (2021) para la adaptación de los factores de expansión que brinda la GEIH para que, respetando la composición mes a mes de ocupados propuesta por el DANE, el promedio anual sea igual a la anualización que proponen los autores (ver Cuadro R1).

Cuadro R1. Identificación de los ocupados anuales con datos mensuales de las encuestas de hogares (diciembre 2019)

Clasificación	Pregunta	Criterios	Número de individuos
Ocupados plenamente		Trabajó todos los últimos 12 meses	18.863.272
No ocupados plenamente		No trabajó todos los últimos 12 meses, pero trabajó el último mes ("trabajó al menos una hora en la última semana" o "reportó tener trabajo en el último mes")	3.897.458
Inactivo pero previamente ocupado	¿cuánto hace que ... Trabajó por última vez? (P7440)	Menos de un año	1.504.447
Desempleado pero previamente ocupado	¿cuántas semanas hace que ... Dejó de trabajar por última vez? (P7320)	Menos de 53 semanas	1.682.915
Total de ocupados en el año (ocupados anuales)			25.948.092

Fuente: Lora, Benitez y Gutiérrez (2021)

Por lo tanto, como se puede notar en el Cuadro 1R, el promedio anual del factor de expansión modificado coincide con el total de ocupados que estiman Lora, Benitez y Gutiérrez (2021), 25.948.092 ocupados durante el 2019 que incluyen no solamente los trabajadores catalogados como ocupados sino también los desempleados que han pasado menos de 53 semanas en esa posición y aquellos inactivos que se han mantenido así por menos de un año. Además, para mantener la evolución mensual del empleo en el mercado laboral se observa que, tanto el factor de expansión del DANE como el modificado respetan las mismas tasas de variación mes a mes para conservar en lo máximo posible el ciclo del mercado de trabajo.

Este ajuste no se implementa sobre el análisis para no afectar la comparabilidad con las cifras tradicionales del mercado laboral

Cuadro 1R. Comparación de los factores de expansión DANE y factores de expansión modificados

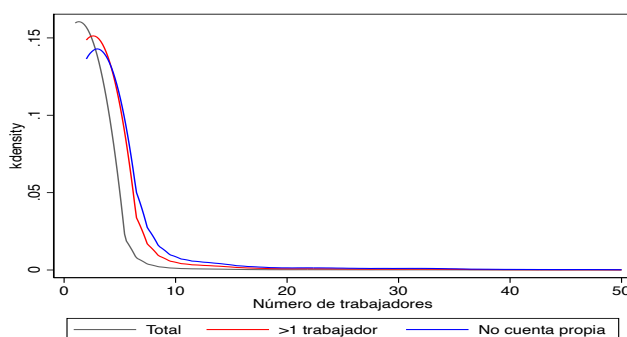
	Factor de expansión DANE	Factor de expansión modificado
Enero	21.649.861	25.205.974
Febrero	22.071.182	25.696.499
Marzo	22.114.055	25.746.414
Abril	21.896.028	25.492.575
Mayo	22.164.214	25.804.812
Junio	22.618.084	26.333.233
Julio	22.139.997	25.776.617
Agosto	22.115.973	25.748.647
Septiembre	22.229.542	25.880.870
Octubre	22.813.247	26.560.452
Noviembre	22.874.449	26.631.707
Diciembre	22.760.730	26.499.309
Promedio anual	22.287.280	25.948.092

IV. Características del universo empresarial

a. Características de la distribución por tamaño de firmas

Una de las características más propias de las economías emergentes, es la granularidad del tejido empresarial. De acuerdo con la EEG (2019) éste también es el caso de Colombia: las firmas de 10 o menos trabajadores constituyen el 99% del total de firmas del universo empresarial y el 61% del empleo, lo que implica un promedio de 2.2 trabajadores por firma. Cuando se restringe el universo a las firmas con empleados, estos micronegocios cubren el 96% de las firmas y el 41% del empleo (6.2 empleados por firma) y cuando se consideran únicamente las firmas con empleados pagos, la participación de los micronegocios se reduce al 93% y 30% de las firmas y del empleo, respectivamente (9.3 empleados por firma)¹⁸. La Gráfica 4 presenta la distribución de firmas en Colombia para los tres agregados considerados.

Gráfica 4. Distribución de firmas por número de trabajadores



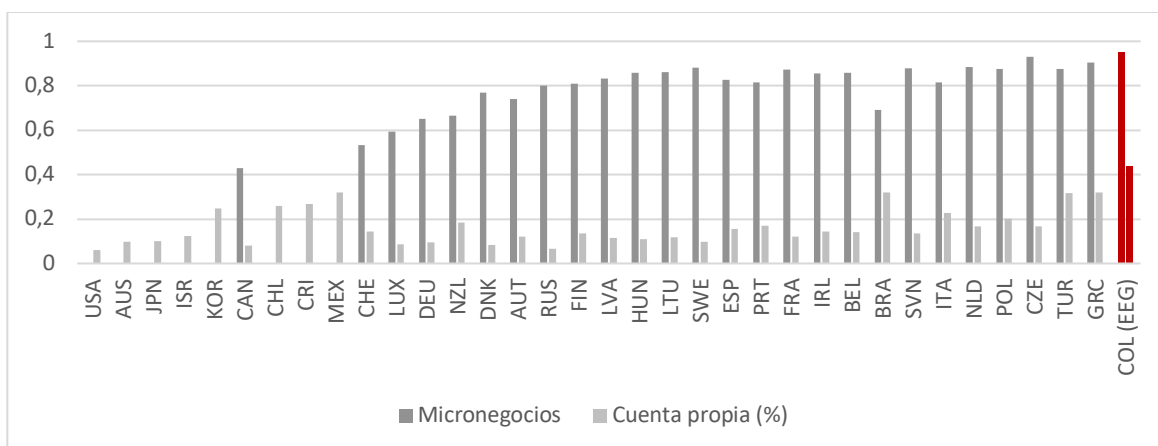
Fuente: EEG (2019)

Así mismo, existe una alta proporción de trabajo por cuenta propia en Colombia¹⁹. De acuerdo, con la encuesta de hogares de 2021, en Colombia hay 20 millones de trabajadores; de los cuáles el 44% (8,9 millones) son trabajadores cuenta propia. Esta tasa es muy superior a la observada en otros países de Latinoamérica, como Perú (38%), Brasil (24%), México (13%), Argentina (25%) y Chile (23%), conforme con las estadísticas del BID (SIMS, 2019). La Gráfica 4, que presenta el porcentaje de micronegocios y auto empleo en los países de la OECD, sitúa a Colombia en el extremo alto de la distribución en ambas estadísticas.

¹⁸ Estas estadísticas se restringen al universo de firmas considerado en este trabajo que por ejemplo excluye al sector agrícola. A manera de confirmación de las estadísticas presentadas, en la GEIH (2019) el promedio de empleadores de firmas con menos de 10 trabajadores también es de 95%, cuando se ajusta la muestra al universo considerado en este trabajo.

¹⁹ Se entiende por trabajo cuenta propia aquel que no tiene a cargo trabajadores pagos

Gráfica 4. Firmas por cuenta propia y con menos de 10 trabajadores (% OECD, 2019)



Fuente: OECD (2022), *Self-employed with and without employees, employment in small firms (indicator)*. doi: 10.1787/b7bf59b6-en (Accesed on 27 September 2022 y EEG (2019). El dato de empleo por cuenta propia presentado por al OECD para Colombia es del 50%, ligeramente superior al 44% de la encuesta de hogares.

Sin embargo, no todo el trabajo por cuenta propia constituye una firma. De acuerdo con la OIT, sólo aquellos trabajadores que tienen la propiedad o uso de los medios de producción pueden ser considerados como firmas. Bajo este criterio, que es el utilizado en el resto del trabajo, en la EEG (2019) hay 3,8 millones de firmas por cuenta propia (87% del total) que emplean a 4.4 millones de trabajadores (45% de los trabajadores); y 3.3 millones de firmas/trabajadores de un trabajador/jefe, que constituyen el 77% del total de firmas y emplean al 35% de los trabajadores²⁰.

b. Características de la distribución por valor agregado

En la distribución de firmas por valor agregado también existe una granularidad importante, pero con largas colas hacia la derecha de la distribución; lo que significa la existencia de unas pocas firmas responsables de un porcentaje importante del PIB en Colombia. Es por esta razón que, mientras la participación de los micronegocios en el total de firmas y en el empleo es de 99% y 66%, respectivamente; la participación en el valor agregado reportado en las encuestas es de tan sólo 16% (7%, si sólo se tiene en cuenta las firmas con trabajadores pagos)^{21 22}.

Las gráficas 6A y 6B presentan la distribución del valor agregado, productividad, ganancias y ganancias netas para firmas totales sin cuenta propia y cuenta propia, corregidas por las diferencias sectoriales²³. En estas gráficas, la distribución de las ganancias y las ganancias netas replican el comportamiento del valor agregado se replica

²⁰ Estos cálculos difieren de Eslava y Meléndez (2020) porque no incluyen sector agrícola, financiero, gobierno y servicio doméstico; ni a todo el trabajo cuenta propia, sino únicamente el que puede considerarse como firma de acuerdo a la OIT.

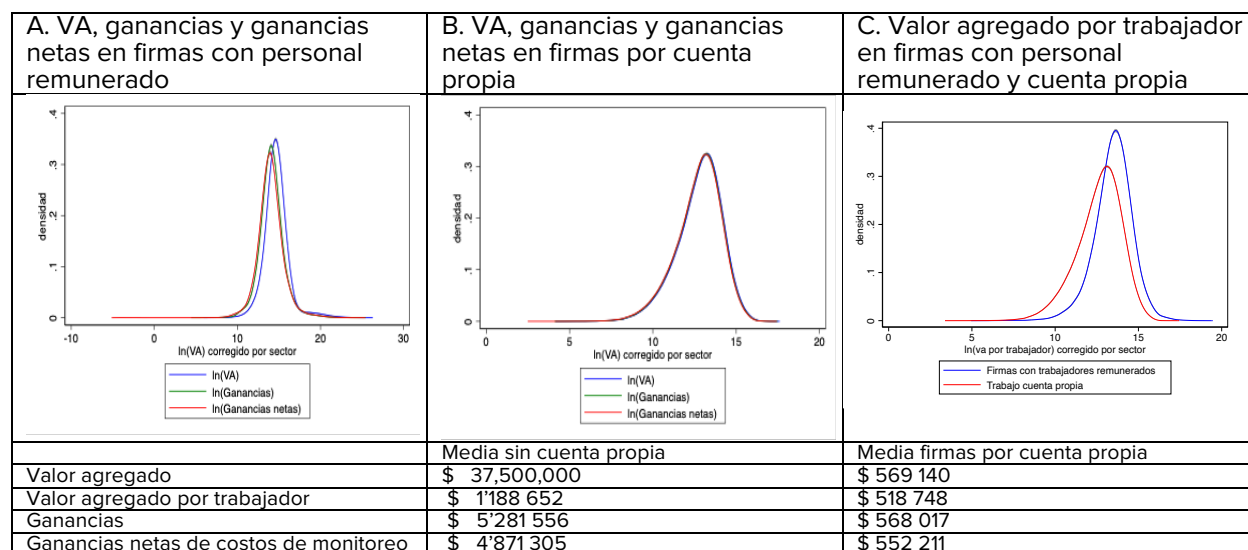
²¹ Consecuentemente la media de ingresos mensuales de las firmas con trabajadores pagos es de COP\$37 millones (cerca de US\$11 300) y la de las firmas por cuenta propia es tan sólo COP\$ 567 mil pesos (cerca de US\$172).

²² La relación directa entre estas dos variables ya había sido identificada anteriormente por Levy (2018) y Eslava, Meléndez y Urdaneta (2020).

²³ Las ganancias se calculan como el valor agregado neto de los costos salariales y tributarios, y las ganancias netas, como las ganancias netas de los costos de operar informalmente ante el control de las autoridades. La media de ganancias mensuales de las firmas con empleo pago es cercana US\$1500. Para controlar las diferencias sectoriales se obtienen los errores de predicción de una regresión entre logaritmo de la variable de interés y los 3 sectores económicos y se les añade la constante de la regresión. Para evitar el impacto de los valores extremos a cada base de datos involucrada se le quita el 1% de la cola derecha de la distribución y los valores negativos, puesto que se trabaja con logaritmos.

pero con menos datos extremos. En el panel A de esta gráfica se observa que en las firmas con trabajadores la distribución se corre hacia la izquierda como resultado del componente impositivo y laboral, y un poco más a la izquierda como consecuencia de los costos de monitoreo y control²⁴. Esto no ocurre con el trabajo cuenta propia (panel B); probablemente porque la mayoría de estas firmas tiene un valor agregado inferior al umbral de pago de impuestos en Colombia (\$ 5.6 millones mensuales)²⁵. El panel C se observa la comparación del valor agregado por trabajador para el empleo sin trabajadores y para el empleo por cuenta propia, que en promedio es menos productivo.

Gráfica 6. Distribución de firmas por valor agregado y productividad, ganancias y ganancias netas de costos de monitoreo y control



Fuente: EEG (2019) valores de las gráficas corregidos por diferencias sectoriales.

c. Racionalidad de la distribución del universo empresarial

Las razones de la alta participación de pequeñas firmas en el país, es un tema aún por explorar en la literatura, pero puede estar relacionada con el umbral exento de impuestos de renta y de IVA para firmas de bajos ingresos, la falta de capital físico y humano para crecer, el menor monitoreo y control que ejercen las autoridades sobre las firmas pequeñas y la alta incidencia de informalidad.

Así mismo, las preferencias por flexibilidad laboral y el espíritu emprendedor juegan un papel en la demografía de firmas en Colombia; especialmente en el caso de las firmas por cuenta propia, donde las decisiones de bienestar se mezclan con las decisiones de productividad. En efecto, como se ilustra en la Gráfica 7A y en menor medida en la gráfica 7B, las firmas que se crean porque se tiene una idea de negocio, se quiere ejercer una profesión o continuar con una tradición, tienen un mayor valor agregado por trabajador que las creadas por razones de subsistencia o de flexibilidad²⁶. Así mismo, en las firmas por cuenta propia se espera que las características propias del jefe, como horas dedicadas

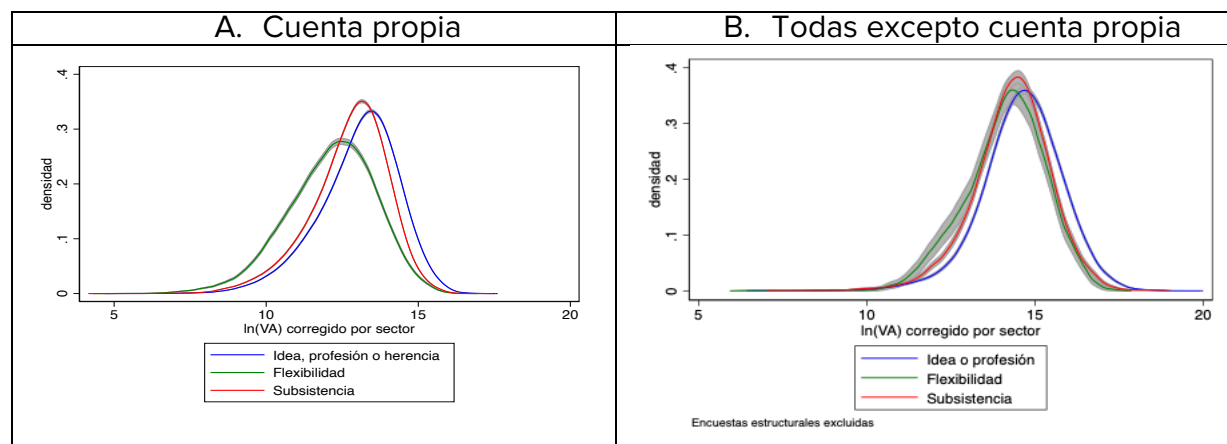
²⁴ Las diferencias entre medias de valor agregado, ganancias y ganancias netas, resultan significativamente diferentes de cero en un test de diferencia de medias (t-test).

²⁵ Sin embargo, como se muestra en la sección V, las ganancias de los trabajadores cuenta propia sí se afectan con las inequidades en pagos a seguridad social

²⁶ Estas diferencias resultan significativas en regresiones para estimar el logaritmo del valor agregado por trabajador y en pruebas de diferencias de medias que están disponibles para el lector pero no se presentan en este trabajo.

al negocio o el nivel de calificación, afecten más directamente la productividad del negocio.

Gráfica 7. Razones para empezar el negocio y valor agregado por trabajador promedio.



	Cuenta propia		Total de firmas (sin cuenta propia)		
	Composición	Va por trabajador promedio	Composición	Va por trabajador promedio	Tamaño promedio
Empieza negocio, mayor ingreso	41%	\$644 681	61%	\$1'344 439	12.8
Negocio familiar	6%	\$527 339	9%	\$1'132 018	3.8
Flexibilidad	13%	\$301 864	7%	\$1'094 054	8.2
Subsistencia	40%	\$457 338	23%	\$827 245	3.1

Fuente: EEG (2019). Las distribuciones están corregidas por sector e incluyen intervalos de confianza.

d. Resumen de las particularidades del universo de firmas

En resumen, en el país existe una alta proporción de firmas pequeñas y en particular, firmas de un trabajador. Esta granularidad puede estar relacionada con los márgenes de impuestos, acceso a capital físico y humano, y por una tolerancia de las autoridades a la informalidad en las firmas pequeñas. En el caso del trabajo por cuenta propia, las preferencias por flexibilidad laboral de los trabajadores por cuenta propia y jefes, podrían estar jugando un papel importante.

V. Margen extensivo de la informalidad o informalidad empresarial.

a. Características generales de la informalidad empresarial

Uno de los problemas para analizar la informalidad empresarial es que no existe uno, sino múltiples criterios para su identificación; y esta abundancia de criterios la convierten en una variable continua, que identifica el cumplimiento de un porcentaje mayor o menor de la normativa empresarial. Esta situación es muy diferente en países como Brasil donde el monotributo, que considera el cumplimiento de una amplia normativa empresarial, incluyendo la laboral, tiene aplicación generalizada.

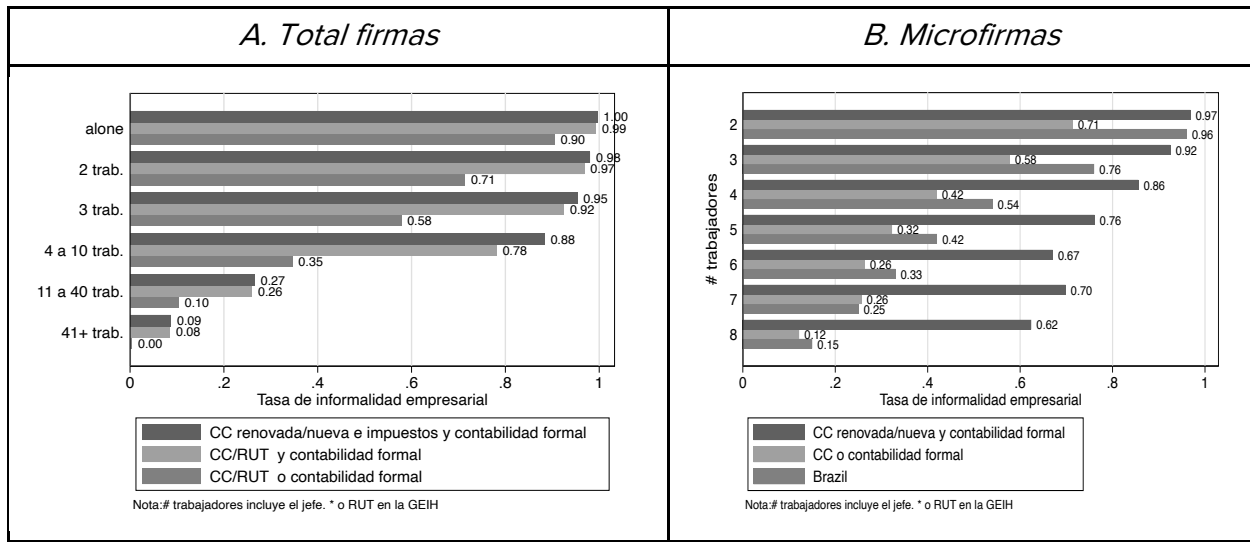
Para enfrentar la variedad de criterios en el caso colombiano, el Dane está creando una metodología de medición de la informalidad empresarial que considera 4 componentes: entrada, impuestos, insumos y producto. Dado que esta metodología aún está en proceso, en este trabajo se utiliza la medida más estricta de informalidad empresarial (paga impuestos o está eximido de pagarlos, tiene registro en Cámara de Comercio renovado y lleva contabilidad formal) que además permite asumir que las firmas formales no evaden impuestos. Sin embargo, cuando los resultados son sensibles a la definición de informalidad utilizada se utilizan criterios adicionales como el de informalidad flexible, que considera como formales a las firmas que tienen el registro renovado o alternativamente, a aquellas que llevan una contabilidad formal.

La primera conclusión que puede obtenerse de la informalidad empresarial es que tiene un carácter generalizado. Tan sólo el 25% de las firmas tienen registro en cámara de comercio renovado y el 2% tienen registro renovado, llevan contabilidad formal y pagan impuestos o están eximidos de hacerlo. En estas condiciones, dependiendo del criterio de formalidad que se utilice (estricto o flexible), la informalidad empresarial cubre a entre el 31% y el 88% de las firmas, entre el 10% y el 31% del personal y entre el 2% y el 8% del valor agregado reportado, en las empresas que tienen trabajadores remunerados. En las unidades por cuenta propia, la informalidad representa casi la totalidad de las firmas, personal y valor agregado reportado.

La Gráfica 10A muestra las tasas de informalidad de acuerdo con las definiciones estricta y flexibles de informalidad. Como se muestra en el gráfico, la informalidad empresarial se reduce monótonicamente con el tamaño de las firmas; no obstante, el porcentaje de informalidad empresarial en las firmas entre 10 y 50 trabajadores se encuentra entre 30% y 10% dependiendo del concepto utilizado de informalidad empresarial. Incluso hay algunas empresas con más de 50 trabajadores que son informales bajo la definición estricta de informalidad.

La Gráfica 10B muestra la informalidad empresarial para los micronegocios de Colombia y Brasil. Según este gráfico, las pequeñas empresas brasileñas muestran niveles de informalidad similares a los escenarios de informalidad más estrictos de Colombia, pero la informalidad de las empresas en Brasil decrece más rápido en el tamaño de la empresa. La información derivada de las pendientes de las variables en este gráfico permite suponer que las empresas mayores de 10 trabajadores son formales en el caso de Brasil, pero no en el colombiano.

Gráfica 10. Margen extensivo de informalidad por criterios



Fuente: EGG y Ulyssea (2018)

b. Racionalidad del comportamiento decreciente de la informalidad empresarial

Ulyssea (2018) explica la relación decreciente entre la productividad y la informalidad empresarial mediante una taxonomía que, en una versión simplificada, involucra tres tipos de firmas: 1) las firmas de subsistencia que son muy pequeñas y poco productivas y no pueden operar formalmente por la existencia de barreras de entrada al mercado formal, como los costos de registro y contabilidad; (2) las firmas de informalidad inducida que, aunque su mayor nivel de productividad les permitiría operar formalmente, les es más rentable hacerlo informalmente; y (3) las firmas de productividad alta, para las cuáles es más rentable operar como formales.²⁷

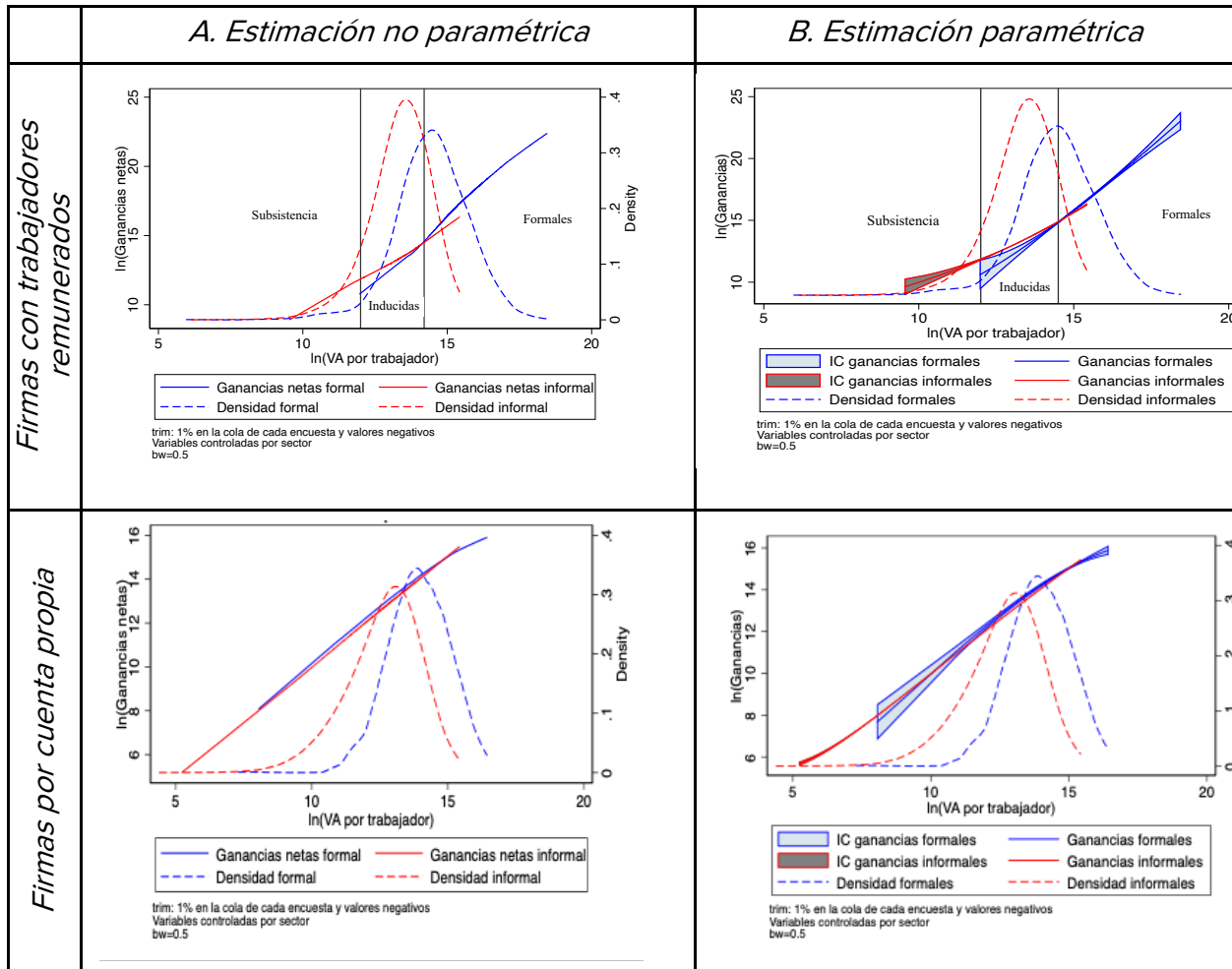
La Gráfica 11 replica esta taxonomía de manera contable para el caso colombiano, pero aproxima la productividad con el valor agregado por trabajador, y el valor de la firma, con las ganancias netas.²⁸ Las convenciones utilizadas en esta gráfica son las siguientes: 1) las líneas sólidas representan los estimativos de la rentabilidad promedio de las empresas, y las punteadas, se refieren a la distribución de las firmas; 2) los datos en azul se refieren a las firmas formales y los rojos a las informales; 3) las gráficas de la izquierda muestran las estimaciones no paramétricas (estimador de Lowess)²⁹ y las de la derecha, las paramétricas, que incluyen intervalos de confianza. para las firmas colombianas con trabajadores pagos, y 4) el panel superior realiza la estimación para las firmas con trabajadores remunerados y el panel inferior para las firmas por cuenta propia.

Gráfica 11. Ganancias netas, valor agregado por trabajador e informalidad en firmas sin trabajadores pagos

²⁷ En realidad, Ulyssea (2018) distingue dos tipos de firmas de productividad inducida, las que de muy baja productividad que identifica como parásitas y las que están más cerca del punto en el cual es rentable ser formal, que identifica como las firmas de De Soto, por cuanto la relajación de costos excesivos de entrada a la formalidad las llevarían a operar formalmente.

²⁸ La utilización de estas aproximaciones a las variables, reduce el carácter exógeno de las estimaciones.

²⁹ Esta estimación no acepta pesos poblacionales, ni genera intervalos de confianza.



Fuente: EEG(2019). Firmas sin trabajadores remunerados pagos

En las gráficas superiores se observa que hay un rango de productividad muy baja donde no operan firmas formales (Subsistencia, 52% de las firmas)³⁰; otro rango de productividad mayor, donde las ganancias de las firmas que operan formalmente son más altas que las del promedio de las firmas informales (Inducida, 41% de las firmas); y un tercer segmento, de productividad alta, donde las ganancias de las firmas que operan formalmente son mayores (Formales, 7% de las firmas). Como indican las curvas de densidad, en el primer segmento no operan firmas formales, en el segundo coexisten firmas formales e informales, pero predominan las informales, y en el tercero, aunque existen algunas firmas informales, predominan las formales³¹.

Los resultados anteriores se formalizan en el Cuadro 5, que ilustra una regresión entre las utilidades netas, el valor agregado y la informalidad. De acuerdo con la primera regresión (1), existe una relación positiva entre los beneficios netos y el valor añadido y entre las utilidades netas y la informalidad, lo que significa que las utilidades crecen con la

³⁰ Las firmas con ganancias netas negativas, que no aparecen en el ejercicio, se incluyen en este primer grupo.

³¹ El punto a partir del cual empieza la coexistencia de firmas formales e informales es un valor por trabajador de apenas de COP\$150 mil y el punto a partir del cual se hace relativamente más productivo ser formal que informal es de (COP\$2.2 millones). Este último punto es cercano al punto de cruce de las distribuciones de firmas informales y formales.

productividad y que las empresas informales tienden en promedio a ser más rentables. Sin embargo, la mayor rentabilidad relativa de la informalidad disminuye a medida que la empresa tiende a ser más productiva, hasta convertirse en una menor rentabilidad relativa. Las columnas (2), (3) y (4) indican que los resultados son robustos a la inclusión de otras variables de control, y a la inclusión de conglomerados en los errores por informalidad y por encuesta.

Cuadro 5. Estimación de ganancias netas en firmas con trabajadores pagos

	(1)	(2)	(3)	(4)
	Ganancias	Ganancias	Ganancias	Ganancias
Valor agregado por trabajador	1.671*** [27.88]	1.622*** [27.47]	1.622** [518.25]	1.622** [7.05]
Firma informal	10.38*** [11.41]	9.399*** [10.52]	9.399** [210.72]	9.399* [2.99]
Valor agregado por trabajador* firma informal	-0.728*** [-12.10]	-0.652*** [-11.01]	-0.652** [-201.81]	-0.652* [-2.84]
Errores robustos	Si	Si	Si	Si
Controles	No	Si	Si	Si
Cluster errores informalidad	No	No	Si	Si
Cluster informalidad y encuesta	No	No	No	Si
Observations	25570	25524	25524	25524
R-squared	0.848	0.862	0.862	0.862

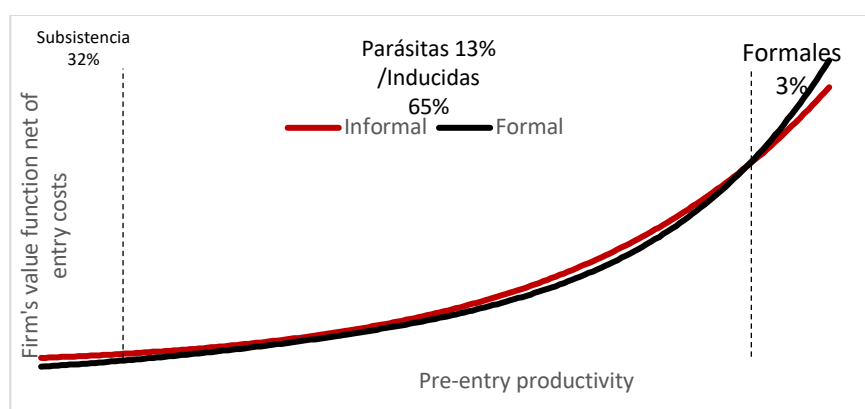
*Fuente EEG (2019) t estadísticas entre paréntesis, * p<0.05, ** p<0.01, *** p<0.001. Se utilizan pesos poblacionales. El valor añadido y los beneficios netos están en logaritmos. Los datos recortados se encuentran recortados al 1% en cada encuesta y no se consideran los valores negativos del VA.*

Los ejercicios anteriores, sugieren que la taxonomía diseñada por Ulyssea (2018) para entender el comportamiento de firmas con trabajadores remunerados es aplicable para el vaso colombiano³²; pero como se observa en la parte inferior de la Gráfica 11, este no es el caso de las firmas por cuenta propia, donde intervienen factores de carácter ocupacional. La ventaja del modelo estructural, que se ilustra en la Gráfica 12 es que permite establecer el impacto causal de la productividad inherente al empresario como factor explicativo de la decisión de informalidad³³.

³² Este resultado es obtenido a pesar de que los costos de monitoreo y control de Fernández (2022) son muy bajos, porque los incentivos tributarios crecientes en el tamaño de las firmas son considerables.

³³ Este modelo permite reproducir las distribuciones de la demografía empresarial presentadas en este trabajo, sin necesidad de suponer que existen choques de productividad diferenciados para firmas formales e informales. No obstante, estos resultados no son suficientes para excluir el impacto que puede tener la informalidad sobre la proliferación de firmas poco productivas.

Gráfica 12. Ganancias, productividad exógena e informalidad



Fuente: Fernández (2022)

c. Resumen de las particularidades de la informalidad empresarial

En resumen, es posible concluir que la informalidad de la empresa en Colombia es mayor y más extendida entre el número de trabajadores de la firma, que en el caso brasileño. También es una condición de carácter menos binario que en el caso de Brasil, donde hay más prevalencia del esquema de impuesto único, que incluye varios tipos de responsabilidades de las firmas, incluyendo las laborales. Otro hallazgo de este capítulo es que la informalidad empresarial en Colombia decrece con el tamaño de las firmas, pero este decrecimiento no es tan pronunciado como en el caso de Brasil; razón por la cual no es posible asumir que todas las empresas de más de 10 trabajadores son formales. Esta evidencia enfatiza la importancia de que el Dane extienda la encuesta EMICRON, o desarrolle una fuente de información periódica para firmas informales de mayor tamaño.

Este capítulo también provee una explicación a esta relación inversa entre el tamaño de firmas y la informalidad empresarial, según la cual las empresas de baja productividad y pequeñas son informales porque no pueden permitirse ser formales, o porque encuentran menos rentable ser formales; mientras que las más productivas encuentran más rentable ser formales.

VI. Margen intensivo (informalidad laboral)

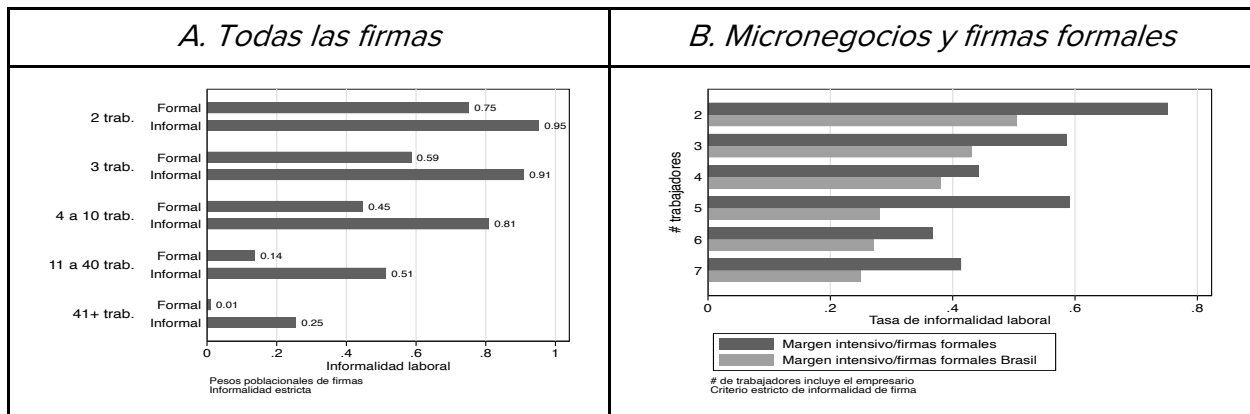
a. Características generales de la informalidad laboral (margen intensivo)

Aunque la informalidad laboral cuenta con criterios más claros de identificación que la empresarial, la divergencia de definiciones entre países es grande y depende de la legislación interna de cada país (Fernández, Villar y Gómez, 2013). En el caso colombiano existen tres criterios básicos para medir la informalidad. El criterio de formalidad que se utiliza en este trabajo es el de cotizaciones a salud y pensiones, que puede tener alguna variación que incluye cotizaciones a salud o a pensiones, o a alguna de ellas por separado. La ventaja de este criterio es que es internacional y que no se confunde con otros tópicos como el tamaño de las firmas o la informalidad empresarial, como sucede con los criterios que utiliza el Dane; la desventaja, es que identifica de como informales a los trabajadores que no están obligados a cotizar (informales legales de acuerdo con Levy, 2018).

La informalidad laboral también cuenta con mejores fuentes de información porque puede medirse directamente de la encuesta de hogares. Sin embargo, para analizar el margen intensivo, que considera si la contratación la realiza una firma formal o informal, se requieren estimaciones adicionales cuyos resultados se presentan en la Gráfica 13³⁴. La primera observación que puede hacerse de esta gráfica es que no hay firmas formales contratando a trabajadores formales y firmas informales contratando a trabajadores informales, sino que los criterios de informalidad laboral y empresarial se entrelazan entre sí³⁵.

La contratación formal por parte de las firmas informales se explica porque para contratar formalmente a sus trabajadores una firma no necesariamente tiene que cumplir con todos los criterios de formalidad empresarial. En las Gráfica 13 también se corrobora el comportamiento decreciente de la informalidad laboral en el número de trabajadores. La pendiente de la relación no permite suponer que todos los trabajadores de las firmas formales de más de 10 trabajadores son formales, pero el supuesto es plausible a partir de 40 trabajadores.

Gráfica 13. Informalidad laboral y empresarial



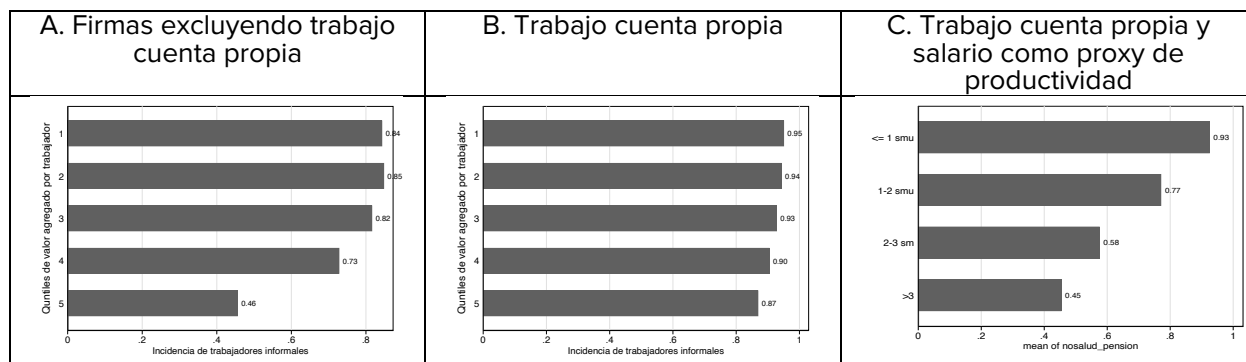
Fuente: EEG (2019) y Ulysea (2018).

La informalidad laboral también es decreciente en el nivel de productividad de las firmas tanto en el caso del trabajo por cuenta propia, como en total de las firmas. Las gráficas 14A y 14B muestran la distribución de la tasa de informalidad laboral por nivel de productividad, para el caso de los jefes y trabajadores dependientes y para el trabajo por cuenta propia. La Gráfica 14C, muestra la misma estadística para los trabajadores independientes, pero utiliza el salario de los trabajadores como proxy de productividad.

³⁴ La metodología de imputaciones utilizadas en este análisis se encuentra en el Capítulo 2 de este trabajo. La encuesta 2021 también permite establecer el margen intensivo más directamente a partir de preguntas a los trabajadores sobre el nivel de la formalidad de la firma, pero tiene la limitación de ser información postpandémica.

³⁵ Esta situación atípica también se observa en los datos observados de la GEIH (2021)

Gráfica 14. Informalidad laboral en trabajadores independientes por nivel de ingreso laboral



Fuente: EEG (2019) y GEIH (2019)

b. Racionalidad del comportamiento decreciente de la informalidad empresarial

Cómo sucede en el caso de la informalidad empresarial, el comportamiento decreciente de la informalidad en el número de trabajadores y la productividad de la firma también puede explicarse por las decisiones racionales de los agentes, que se ilustra en el Cuadro 6. Para las firmas, de mayor tamaño relativo que pagan impuestos, hace sentido contratar a los trabajadores formalmente, porque los costos de contratación formal son deducibles de impuestos, y el monto de ahorros tributarios es equivalente al costo de seguridad social a cargo del empresario. Mientras tanto, si la firma paga sus impuestos a través del monotributo, la deducción tributaria sólo se hace sobre el pago de seguridad social, y por lo tanto el incentivo a contratar informalmente es menor (31%). Por último, si la firma tiene ingresos inferiores al umbral de tributación (46 millones anuales en 2019 y 53 millones en 2021) y está registrada como persona natural, no puede hacer ninguna deducción; y por lo tanto, enfrenta un costo relativo de contratar formalmente da un trabajador de 47% del salario.

Cuadro 6. Costo relativo de contratar a un trabajador formal e informalmente

Ingresos brutos anuales (millones)	Tasa de impuesto de renta	Costo de seguridad social	Deducción de impuestos	Diferencia entre el costo de contratar trabajador formal e informal
< 53.	0%	47%	0	47%
> 53 monotributo		47%	47%*35%=16%	47%- 16%= 31%
> 53	35%	47%	1.47*35%=51%	47% - 51%=-4%

Fuente: cálculos propios, tasas impositivas del 2021

Esta asimetría en el costo relativo de contratar trabajadores formales no es algo marginal y afecta a la mayoría de las firmas. De acuerdo con la EEG (2019) y el Cuadro 7, el 86% de las firmas se encuentra por debajo del umbral de ingresos. De estas firmas el 33% está registrada en Cámara de Comercio y el 92% lo hace bajo la figura de persona natural. Este mecanismo se refuerza con el mayor monitoreo y control que ejercen las autoridades sobre las firmas de mayor tamaño.

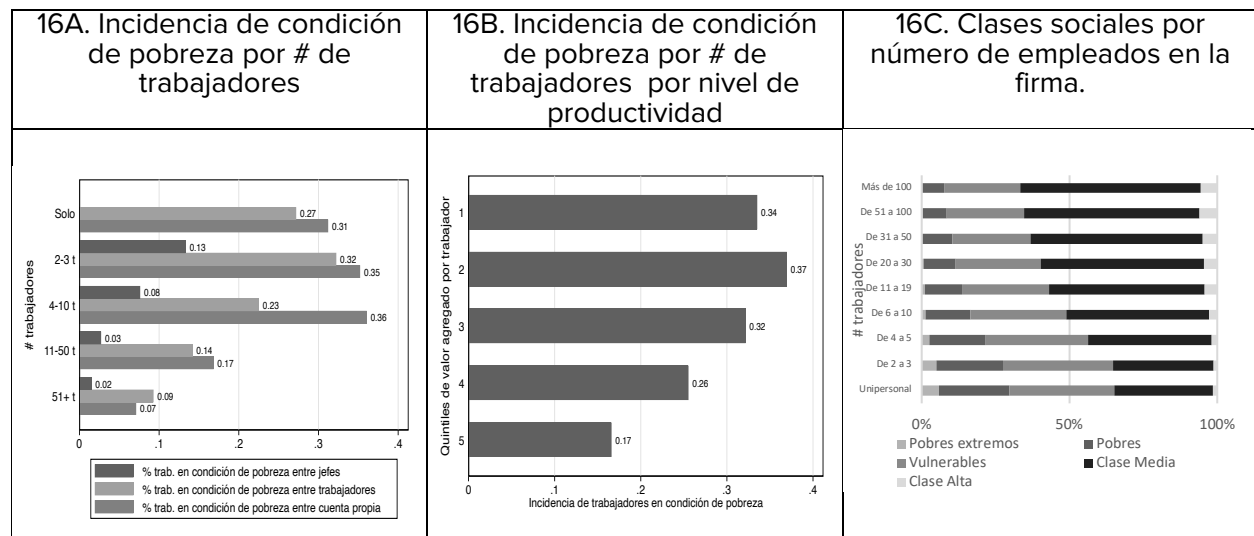
c. Resumen de las particularidades de la informalidad empresarial

El margen intensivo es de más fácil análisis porque existen buenas bases de información y menos diversidad de criterios. Como en el caso de la informalidad empresarial, se corrobora una tendencia decreciente de este tipo de informalidad en el tamaño de las firmas y su productividad. Este comportamiento se explica en gran parte por la regulación que dicta tarifas más altas para los trabajadores independientes de menos ingresos, y asalariados de empresas pequeñas.

VII. Equidad

Cómo se explicó en la introducción, la gran ventaja de la base de datos utilizada es que permite establecer una conexión el bienestar, la formalidad y el tamaño de las firmas. En la Gráfica 16 se observa el porcentaje de los jefes y trabajadores que viven en condición de pobreza, según el tamaño de la firma (Gráfica 16A) y según la distribución por deciles del valor agregado por trabajador (Gráfica 16B). De acuerdo con las gráficas, existe mayor incidencia de trabajadores en condición de pobreza en las empresas pequeñas y poco productivas, y existe mayor incidencia de condición de pobreza entre los trabajadores y trabajadores por cuenta propia en las firmas de tamaño pequeño. La Gráfica 16C, muestra la distribución de las clases sociales por tamaño de la firma y corrobora la existencia de un porcentaje mayor de trabajadores en condición de pobreza y en condición de pobreza extrema en las firmas pequeñas y mayor incidencia de la clase media y alta en las firmas de mayor tamaño relativo.

Gráfica 16. Porcentaje de trabajadores en condición de pobreza monetaria por tamaño de firma

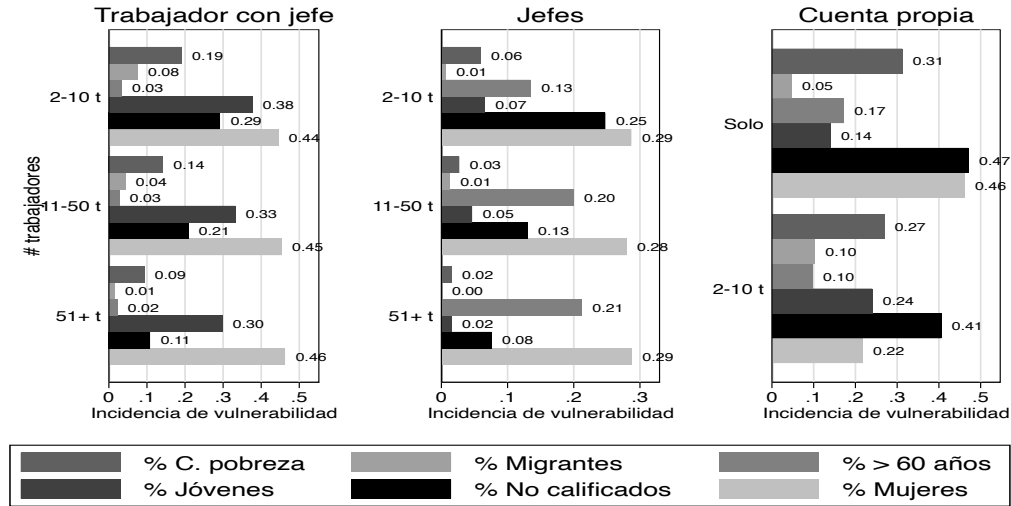


Fuente: EEG (2019).

Así mismo, como ilustra la Gráfica 17, las firmas pequeñas tienen mayor probabilidad de tener población vulnerable que el resto de firmas. Existe también mayor incidencia población vulnerable entre trabajadores y trabajadores por cuenta propia, que entre jefes. En particular, los jóvenes, las mujeres y los migrantes, tienen poca propensión a ser jefes

y participan más como trabajadoras normales o como trabajadores por cuenta propia³⁷. En esta última ocupación, las mujeres tienden a trabajar solas, y los migrantes, en negocios de más de un trabajador, sin ganar remuneración fija. Los mayores de 60 años y los no calificados también tienen una mayor participación entre los trabajadores cuenta propia. Estas divergencias son útiles a la hora de generar políticas de apoyo a los trabajadores vulnerables.

Gráfica 17. Incidencia de trabajadores vulnerables por tamaño de firma

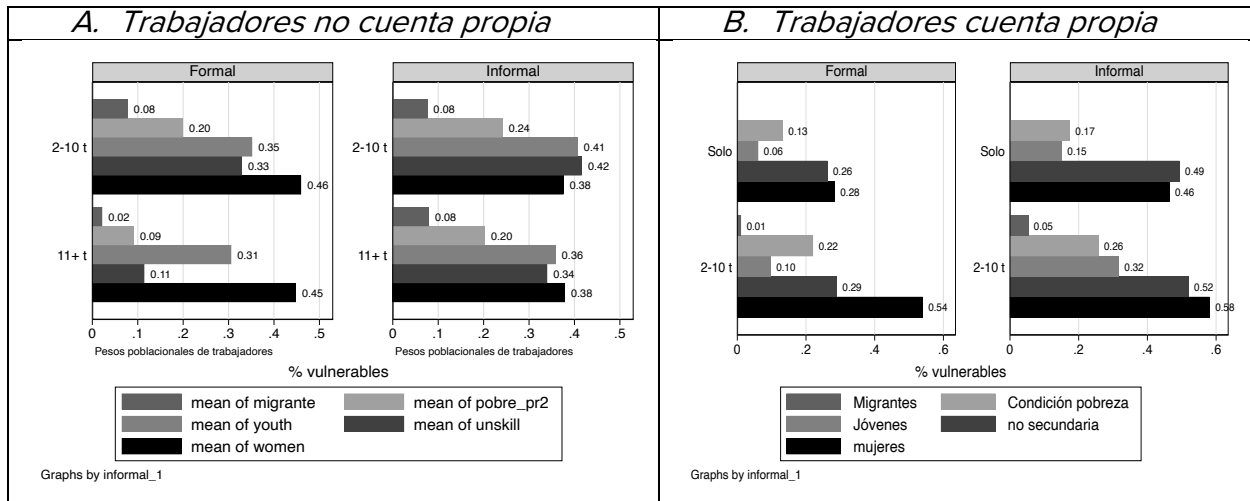


Fuente GEIH (2019)

Este porcentaje de trabajadores vulnerables es mayor en las firmas informales que en las formales; con excepción del caso de las mujeres que tienden a estar más o igualmente representadas en las firmas formales (Gráfica 17). Así mismo, dentro de las firmas que no incluyen trabajo por cuenta propia, la incidencia de vulnerabilidad es mayor en las firmas pequeñas que en las grandes. Esto no ocurre en las firmas cuenta propia, donde con frecuencia, los trabajadores vulnerables son contratados sin remuneración fija.

³⁷ De acuerdo con la OECD (2022), la baja participación de la población vulnerable puede estar relacionada con los “missing entrepreneurs” o trabajadores que no pueden iniciar su negocios debido a barreras de entrada en términos de financiamiento o capacitación empresarial.

Gráfica 17. Porcentaje de trabajadores en condiciones de vulnerabilidad



Fuente: EEG (2019).

Con respecto a la informalidad laboral, y de acuerdo con Fernández y Villar (2013) existen diferencias significativas entre las características observables de los trabajadores formales e informales, asalariados e independientes. En particular los asalariados formales tienen mayor probabilidad de ser hombres en edad productiva (28-50), jefes de hogar, con altos niveles de educación, y residentes de las grandes ciudades; mientras que los informales tienen una mayor probabilidad de ser migrantes, mayores, mujeres y vivir en áreas rurales. Lo anterior, y las mejores condiciones laborales y de satisfacción laboral de los asalariados formales, indica la existencia de brechas de acceso a este mercado.

Adicionalmente, las inequidades en las tarifas de seguridad social y en los beneficios del sistema pensional, explicados en capítulos anteriores, afectan de manera particular a los trabajadores independientes de menores ingresos y a los trabajadores y jefes de las firmas pequeñas, que tienen mayor incidencia y vulnerabilidad. En particular, la no existencia de un esquema de afiliación formal a seguridad social para trabajadores que ganan menos de un salario mínimo explica por sí sola una parte importante de la informalidad en Colombia. De acuerdo con la encuesta de hogares de 2019, en Colombia hay más de 7 millones de trabajadores independientes que ganan menos del salario mínimo y por lo tanto no pueden ser formales, a tasas de cotización razonables. Adicionalmente, 1,5 millones de trabajadores que ganan entre 1 y 2.5 salarios mínimos, se ven afectados por las tarifas regresivas para trabajadores independientes.

En resumen, las firmas pequeñas e informales y el trabajo por cuenta propia tienden a mostrar mayor incidencia de pobreza y vulnerabilidad, que se exagera en las firmas y trabajadores informales. El esquema de tributación, cotizaciones a seguridad social y beneficios pensionales es regresivo y por lo tanto, agrava esta situación; precisamente en uno de los rubros que más le cuesta al estado, y que es más importante para los trabajadores desde el punto de vista de crecimiento personal y bienestar para la vejez. Solucionar estas inequidades es positivo para los trabajadores de menos ingresos pero también para los que están excluidos del mercado laboral.

VIII. Dualismo

Del análisis que se ha realizado en los capítulos anteriores se desprende que las firmas formales son más productivas y grandes; y los trabajadores formales son más calificados

y ganan mayores ingresos que sus pares informales. Lo mismo ocurre con las unidades económicas por cuenta propia, que tienden a ser más informales y menos productivas que las firmas que contratan trabajadores remunerados. Pero lo anterior no necesariamente implica que existen dos economías inherentemente diferentes con pocos vasos comunicantes entre sí.

La identificación del grado de dualismo en la economía es importante porque las recomendaciones de política que se formulen dependen en gran medida de este indicador. Por ejemplo, de acuerdo con la teoría de la economía dual, la formalización de las firmas y los trabajadores mejoraría por sí sola la productividad, ingresos y bienestar de los trabajadores; lo cual no necesariamente es válido, si las razones por las cuáles habían decidido ser informales continúan vigentes. Es decir la causalidad entre la informalidad y la productividad se encuentra en el centro de esta discusión.

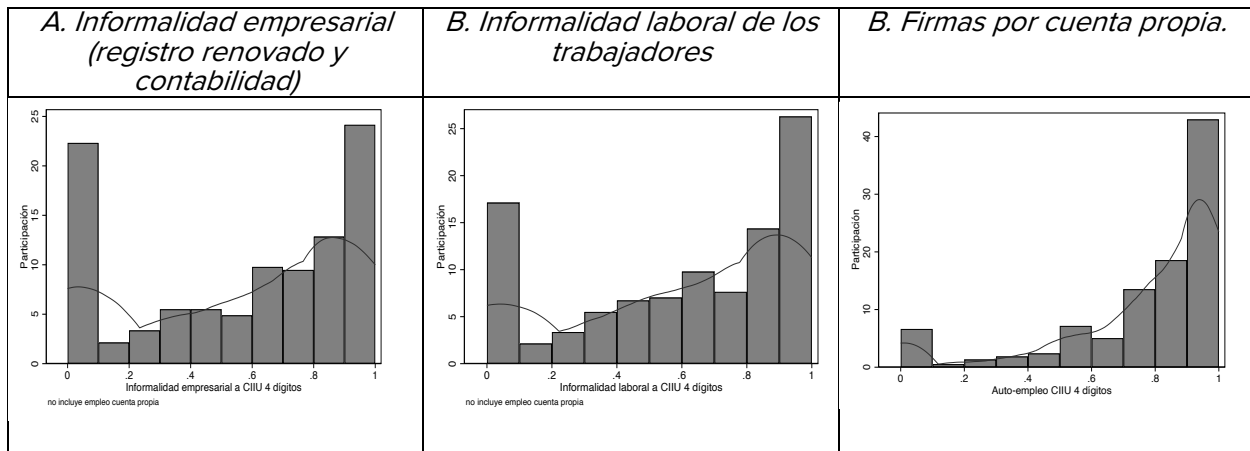
La literatura ha diseñado varias herramientas para determinar el grado de segmentación de la economía. Tal vez la más utilizada es la identificación de las transiciones entre la formalidad y la informalidad, ampliamente utilizada por Maloney (2004) para establecer la poca segmentación de los mercados de Brasil y México. Desafortunadamente, en Colombia no es posible replicar el ejercicio por que la encuesta de hogares en Colombia no contiene información de panel y las encuestas que lo hacen no son representativas.

Sin embargo, Lasso (2020) utiliza información de la encuesta de hogares para hacer un pseudo panel con cadenas de Markov que permite identificar las transiciones entre el trabajo dependiente e independiente, el desempleo y la inactividad. De acuerdo con este ejercicio, los trabajadores por cuenta propia tienen una persistencia importante (75% continúan en este sector) y transitan más a la inactividad (13%) que al trabajo asalariado (5%) y al desempleo (5%) del total de la población considerada). Lo anterior hace sentido, si se tiene en cuenta que para un trabajador por cuenta propia, que no tiene referencias, ni hoja de vida, le es complicado aplicar a un trabajo formal. Esta misma razón explica porque estos trabajadores tienden a resguardarse en la inactividad más que en el desempleo.

Otro ejercicio recurrente es la identificación de sectores y áreas geográficas predominantemente formales o informales, utilizada por Ulysea (2018) para identificar la poca segmentación causada por la informalidad empresarial en el caso de Brasil. La Gráfica 18 replica este ejercicio para el caso colombiano. De acuerdo con la Gráfica 18A, la proporción de sectores predominantemente formales (tasa de informalidad inferior al 10%) es de tan sólo el 22%, y la de sectores predominantemente informales (tasa de informalidad inferior al 10%) del 23%; el 55% restante se distribuye uniformemente en diferente los otros 8 niveles de informalidad³⁸. En el caso de la informalidad laboral la segmentación es un poco menor: el 16% es predominantemente formal y el 26% predominantemente informal y por el contrario, en el caso de las firmas por cuenta propia resulta un poco mayor: el 42% de las firmas están en sectores donde prevalece el trabajo por cuenta propia; aunque existen pocos sectores donde no haya trabajadores con estas características. Lo anterior podría reforzar los hallazgos de Lasso (2018).

³⁸ Se utiliza un criterio flexible de informalidad y una desagregación a CIU 4 dígitos (376 sectores), Cuando se utiliza el criterio más estricto, como es natural, el porcentaje de sectores informales se incrementa a 60%, pero el porcentaje de sectores formales es menor al 10%.

Gráfica 14. Informalidad empresarial y subsectores económicos



Fuente: EEG (2019). Cálculos propios. Los paneles A y B excluyen empleo por cuenta propia.

Así mismo, como se ha ilustrado repetidamente en este trabajo, las firmas formales son en promedio más productivas y tienen mayores ganancias que las informales; y hay segmentos donde sólo operan firmas formales o informales; y lo mismo ocurre con las firmas por cuenta propia y el resto de firmas pero las distribuciones distan mucho de ser consideradas dicótomas porque existe un rango amplio de productividad donde coexisten los dos tipos de firmas.

Finalmente, se presenta un ejercicio utilizado por Perry et al (2007) y por Ulyssea (2018) para identificar segmentación que busca identificar si las diferencias en productividad de los trabajadores se deben a la informalidad de sus trabajadores. En la implementación de este ejercicio no sólo se identifica si las diferencias de productividad están asociadas con la productividad laboral, sino también la relación con la informalidad empresarial y el auto empleo. Las fuentes de datos principales son la encuesta de hogares (2019) y la EMICRON³⁹

Para comprobar lo anterior, se realiza un ejercicio de diferencia entre los salarios de los trabajadores formales e informales que se ilustra en el Cuadro 8. La columna (1) de este cuadro muestra el ejercicio realizado con la encuesta de hogares donde se puede controlar por características de los trabajadores, como el nivel de calificación, y de las firmas, pero no se puede incluir el efecto de la formalidad de las firmas y ni efectos fijos de firma. Los resultados indican que la informalidad laboral tiene un impacto del 21% en los salarios. La columna (2) realiza la misma estimación pero sin controlar por la educación del trabajador. Esta omisión le añade 2 puntos porcentuales al coeficiente de la informalidad laboral. La columna (3) realiza los mismos estimativos con la Emicron, y obtiene un coeficiente 4 puntos mayor. Las columnas (4), (5) y (6) incluyen algunos efectos fijos de firmas como la informalidad de la empresa, la informalidad laboral del jefe y su nivel educativo, ocasionando una reducción significativa en el coeficiente de informalidad laboral (de 0.28 a 0.21).

Este resultado no es tan tajante como el encontrado por Ulyssea (2018) para el caso brasilero, donde la significancia del coeficiente desaparece del todo con los efectos fijos de firma, lo que permite concluir que no existe un mercado dual; pero de todas maneras el valor del coeficiente es reducido y es posible que se reduzca aún más al incluir la

³⁹ No se utiliza el resto de la EEG para no hacer inferencias a partir de datos estimados.

variable de calificación; lo que indica que existe cierto nivel de dualismo, pero moderado en el mercado laboral.

Cuadro 8. Determinantes de las diferencias salariales, como indicador de dualismo en el mercado laboral

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	GEIH	GEIH	EMICRON	EMICRON	EMICRON	EMICRON	EMICRON
Informalidad laboral	-0.2163***	-0.2356***	-0.2787***	-0.2669***	-0.2608***	-0.2140***	-0.1388***
Calificación	0.1426***						
Mujer	-0.1487***	-0.1363***	-0.1455***	-0.1466***	-0.1467***	-0.1503***	-0.0537***
Edad	0.0270***	0.0262***	0.0281***	0.0279***	0.0284***	0.0274***	0.0121***
Edad2	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0001**
Experiencia	0.0108***	0.0105***	0.0094***	0.0093***	0.0098***	0.0096***	0.0126***
Informalidad empresa				-0.0858***	-0.0794***	-0.0554***	
Jefe con secundaria +						0.0734***	
Formalidad laboral jefe						-0.1422***	
Número de empleados	34,776	34,776	17,023	17,023	17,023	17,023	17,023
Observaciones	0.269	0.257	0.312	0.314	0.316	0.322	0.939
R-2	No	No	No	No	No	No	Yes
Efectos de firma	All	All	All	All	All	All	All

Fuente: GEIH (2019) y EMICRON (2019)

Los ejercicios realizados permiten establecer que la economía colombiana no es una economía dual, aunque es importante señalar que los resultados podrían cambiar si se incluyera el sector agrícola en los resultados (no significativamente de acuerdo con ejercicios realizados y no presentados). Sin embargo, sí existen algunas firmas y trabajadores de subsistencia, la mayoría de los cuáles opera como trabajadores por cuenta propia, para los cuáles es muy difícil transitar al mercado formal. Los mecanismos que se implementen para remediar esta situación pueden tomar la forma de capacitaciones y registros formales de las actividades que realicen.

IX. Reflexión

Antes de generar las recomendaciones para reducir la informalidad vale la pena preguntarnos para qué se quieren reducir, porque la informalidad no puede ser un objetivo en sí mismo. Las razones más importantes para reducir la informalidad empresarial son: fiscales; control a la competencia desleal; cumplimiento de otros estándares, incluyendo los laborales; productividad y, equidad.

Realizar una reducción de la informalidad por razones fiscales carece de mucho sentido, porque es posible que los costos del monitoreo sobre las firmas pequeñas sobrepasen los beneficios en términos de ingresos. Si la razón es equidad en el tratamiento de las firmas (control a la competencia desleal), debe tenerse en cuenta que la regulación actual ha creado también inequidades que tienden a favorecer a las firmas más grandes y productivas y lo ideal sería establecer un tratamiento horizontal a las firmas.

Respecto al cumplimiento de otros estándares, es importante revisar los incentivos que existen para el cumplimiento de las diferentes normativas. En el caso laboral por ejemplo, es importante considerar que un aumento en la formalidad empresarial no necesariamente implica un aumento en la formalidad laboral, si las firmas que se van a formalizar son pequeñas; porque para estas firmas es muy costoso contratar a un trabajador formalmente. Con respecto al control de la formalidad con un objetivo de

productividad, debe tenerse en cuenta que los aumentos de productividad de las firmas pequeñas no necesariamente implican aumentos en productividad de la economía en general, porque si bien la participación en el número de firmas y el de trabajadores de estas firmas es marginal, su participación en el valor agregado es limitada; y la efectividad de esta medida es limitada.

Finalmente, la reducción de la informalidad por temas de equidad requiere un replanteamiento. El aumento en la formalidad empresarial por sí sola no genera aumentos en bienestar, por esta razón debe venir acompañada de incentivos para cumplir con otra normativa como la ambiental o la laboral. Sin embargo, en este punto también debe considerarse que un mayor pago de seguridad social por parte de los trabajadores no necesariamente implica un aumento en el bienestar, porque los trabajadores de bajos ingresos tienen bajas posibilidades de lograr una pensión.

De este análisis se derivan algunas recomendaciones de política de carácter regulatorio para incrementar la formalidad laboral. Como indica la Misión de Empleo (Levy y Maldonado, 2022), el aparato institucional debería pasar la mayor parte de la carga prestacional, que no incluye aspectos como el seguro de desempleo o los riesgos profesionales, a los impuestos generales. Este cambio no implicaría un costo demasiado alto, porque hoy en día el estado financia la seguridad social de todos los trabajadores de empresas que pagan impuestos a partir de deducciones laborales. En efecto, una deducción de 30% sobre unos costos laborales que incluyen salarios de 150%, equivalen a un subsidio de la carga prestacional equivalente a todos los costos de seguridad social.

Un primer paso hacia esta transición sería el establecimiento de una pensión básica universal para los trabajadores de bajos ingresos que no pueden jubilarse⁴⁰, para lo cual no es necesario cambiar el régimen de aseguramiento privado. Para los trabajadores que ganan más de un salario mínimo, podría establecerse una tarifa de cotización progresiva (Lora y Mejía, 2019) o por lo menos plana, para todos los tipos de firmas y niveles de ingresos de trabajadores independientes. Arbeláez, Fernández y Hernández (2020) proponen para las plataformas digitales, un fondo de aportes alimentado por las empresas que corrija las inequidades en las tarifas que enfrentan los trabajadores independientes de menores ingresos, donde se encuentran la mayoría de los colaboradores de este sector. Esta propuesta podría obrar como una caja de arena regulatoria en el objetivo de lograr un sistema más igualitario.

Algunas recomendaciones más operativas, incluyen la generación de unos vouchers de deducciones de costos formales que las firmas que se encuentran por debajo del umbral de ingresos puedan deducir cuando crezcan o cuando ganen más ingresos. La ventaja de esta medida es que incentiva el crecimiento de las empresas, que como se observó anteriormente es un problema particular del caso colombiano.

También existen algunas recomendaciones relativas al monotributo. El monotributo ha sido asociado en la literatura con la reducción de la informalidad en casos como el de Brasil o el de Uruguay. Sin embargo, es importante tener en cuenta los grandes costos fiscales y los posibles incentivos al enanismo de las empresas. Adicionalmente, debe tenerse presente que poco se logra con el monotributo en términos de bienestar, si las condiciones para cumplir con otra normativa como la laboral no están dadas, o si los trabajadores de menores ingresos no logran jubilarse. En este sentido la inclusión de la seguridad social en el monotributo, y/o el establecimiento de la tarifa del monotributo sobre ingresos netos de costos salariales y de seguridad social, pueden ser unas soluciones efectivas. Así mismo, es necesario diseñar mecanismos adicionales, que

⁴⁰ El financiamiento de esta pensión no requeriría disponer de los ingresos del RAIS.

incentiven el crecimiento de las empresas del monotributo, como la graduación de empresas o los beneficios tributarios decrecientes.

Otro tipo de políticas que vale la pena analizar son los esfuerzos para aumentar la productividad con fines de equidad. Hasta hace poco existía la percepción de que las empresas pequeñas no crecían y sobrevivían largos años con bajos niveles de productividad. Estudios recientes de Angulo (2022) y Confecámaras (2022) señalan no sólo que existen algunas firmas pequeñas que crecen; sino que es posible identificar algunos factores comunes de estas firmas como el acceso a crédito, a tecnología y la formalidad (independientemente de la mutua causalidad entre esta variable y la productividad). Dada la relación que existe entre el tamaño de las firmas, la productividad y la incidencia de pobreza, las ayudas del gobierno para el crecimiento de estas firmas podrían constituir en sí mismas una política social. Sin embargo, es necesario disponer de mayores estudios y bases de datos para poder diseñar lo que podría denominarse una política social de productividad.

Bibliografía

- Álvarez, J., & Ruane, C. (2019). Informalidad y productividad agregada: el caso de México.
- Dane (2013) Metodología general Gran Encuesta Integrada de Hogares - EAC
- Dane (2020a) Metodología general Encuesta Anual de Comercio - EAC
- Dane (2020b) metodología general Encuesta Anual Manufacturera -EAM
- Dane (2020c) Metodología general Encuesta Micro-establecimientos - EMICRON
- Dane (2020d) Metodología general Encuesta Anual de Servicios - EAS
- Danés (2021) Censo económico de Colombia. Documento metodológico censo económico, 2021
- DNP (2019). CONPES 3956 Formalización Empresarial. <https://colaboracion.dnp.gov.co/CDT/Conpes/Economicos/3956.pdf>
- Eslava, Meléndez y Urdaneta (2021). Market Concentration, Market Fragmentation, and Inequality in Latin America. UNDP Working Papers # 11.
- Fernández, C. (2020). Informalidad empresarial en Colombia. Coyuntura Económica: Investigación Económica y Social. 50, 133-168, diciembre.
- Fernández, C. & Mejía, L. F. (2021). "Rigideces del mercado laboral en Colombia: tendencias, perspectivas y recomendaciones". En Fedesarrollo, *Descifrar el futuro: la economía colombiana en los próximos diez años*, capítulo 4, 261-319, Penguin Random House.
- Fernández, Villar y Gómez (2017). "Taxonomía de la informalidad en América Latina," Coyuntura Económica, Fedesarrollo, vol. 47(1 y 2), pages 137-167, December
- Levy Algazi, Santiago, 2018. "Esfuerzos insuficientemente recompensados: la búsqueda elusiva de la prosperidad en México", Publicaciones del BID (Libros), Banco Interamericano de Desarrollo, número 8971
- Lasso-Valderrama, Francisco 2020. "LA DINÁMICA DEL MERCADO LABORAL COLOMBIANO: Cuál es el rol de los flujos de trabajadores?," Borradores de Economía 1120, Banco de la Republica de Colombia.
- Maloney, William F. "Informality revisited." World development 32.7 (2004): 1159-1178
- OCDE (2007) Eurostat-OCDE Manual sobre Demografía Empresarial, Statistics Comisión Europea ISBN 978-92-79-04726-8, Cat. Número: KS-RA-07-010-ES-N-N, ISSN 1977-0375

- OECD/European Commission (2021), *The Missing Entrepreneurs 2021: Policies for Inclusive Entrepreneurship and Self-Employment*, OECD Publishing, Paris, <https://doi.org/10.1787/71b7a9bb-en>.
- Ulyssea, G. (2018). Empresas, informalidad y desarrollo: Teoría y evidencia de Brasil. *American Economic Review*, 108(8), 2015-47.
- Ulyssea G. 2019. Dinámicas de empresa formales e informales. Manuscrito inédito, Univ. Oxford, Oxford, Reino Unido

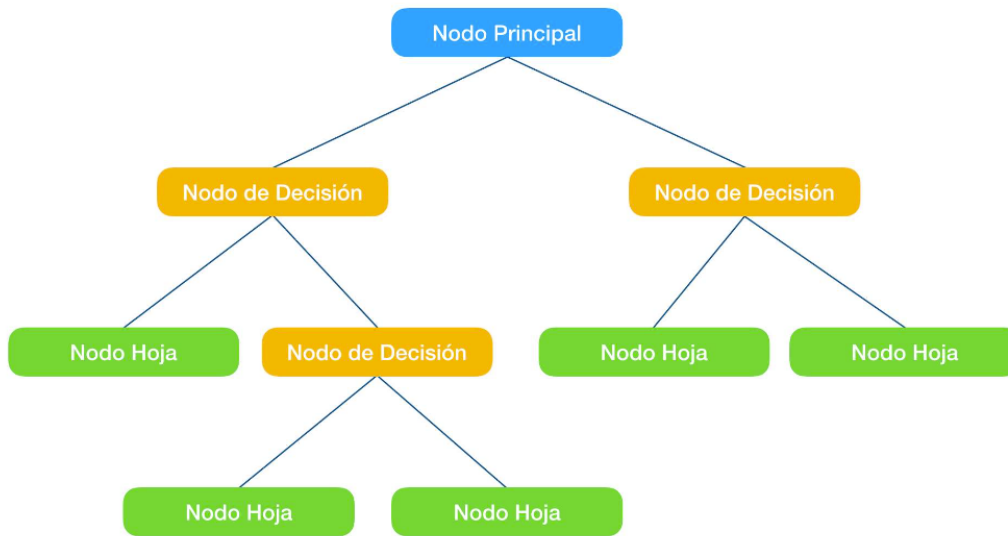
Anexo 1. Métodos de imputación

En la actualidad, con el crecimiento del flujo de los datos y la evolución computacional, existen algunos algoritmos de clasificación que pueden ser ejecutados para evaluar su desempeño y contrastar cuál de estos tiene un mejor comportamiento con los datos disponibles. Para el presente análisis, se utiliza como base de entrenamiento a la base de trabajadores con las siguientes variables: total de empleados de a firma donde trabaja, área geográfica, rama sectorial donde se desempeña y si el trabajador es remunerado, no remunerado o socio. La base de prueba, donde se evalúa el desempeño del modelo fuera de muestra, es la base de firmas tanto de la GEIH como de las encuestas estructurales. Se utilizan 4 métodos de clasificación de *machine learning* para evaluar la clasificación de formalidad: árboles de decisión, bosques aleatorios, modelo logístico y redes neuronales.

Árboles de decisión

Los árboles de decisión utilizan una variedad de algoritmos para decidir la división de un nodo en dos o más subnodos como se nota en la [Diagrama 2](#), la forma de hacer subnodos amplía la homogeneidad de los subnodos consecuentes; toma todas las variables disponibles y luego selecciona la división que resulta en los subnodos más homogéneos. Los bosques aleatorios se componen básicamente de múltiples árboles de decisión, que son los clasificadores básicos que conforman los bosques (Li et al., 2019; Kumar et al., 2016).

Diagrama 2. Estructura de un árbol de decisión



Luego de ejecutar este modelo en la base de entrenamiento y evaluarlo se obtiene la matriz de confusión que se encuentra en la Cuadro 4, donde se nota las frecuencias de aquellas observaciones que fueron clasificadas correctamente, es decir, el modelo clasificó como formal a 139.980 de los 167.598 trabajadores formales, a los 27.618 los clasificó incorrectamente como informales. Con esto en mente, la precisión del modelo de árbol de decisión tiene una precisión del 83.8%.

Cuadro 4. Matriz de confusión del árbol de decisión (base GEIH (trabajadores) y EMICRON)

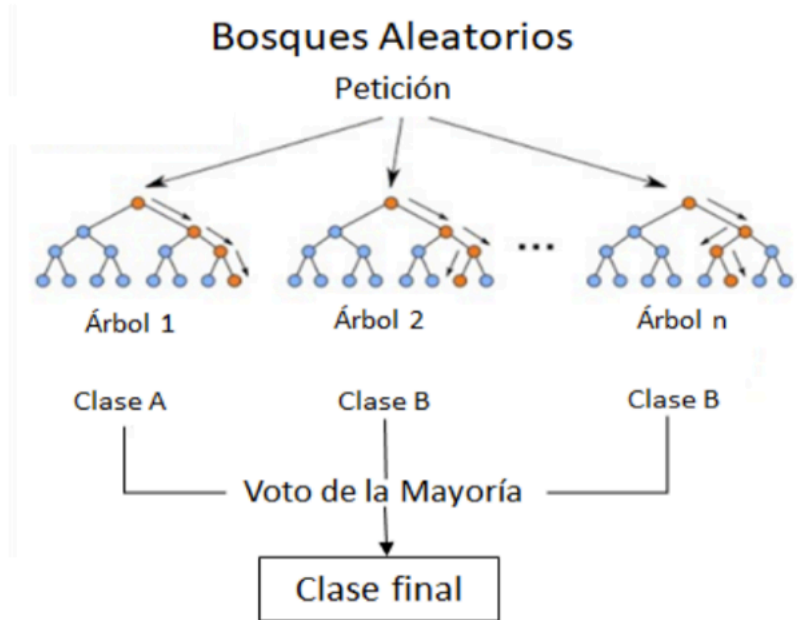
		Predicción		
		Formal	Informal	Total
Observado	Formal	151 148	16 450	167 568
	Informal	19 130	85 352	104 482
Total		170 278	101 802	86.9%

Fuente: EEG y GEIH, cálculos del autor

Bosque aleatorio

El modelo de bosque aleatorio es un clasificador que consiste en un conjunto de clasificadores estructurados en forma de árbol con vectores aleatorios independientes idénticamente distribuidos, como se observa en la Diagrama 3 cada árbol emite un voto unitario en la “petición” para la clase más popular (Reis et al., 2018). Luego, se genera un vector aleatorio independiente de los vectores aleatorios anteriores de la misma distribución y se genera un árbol utilizando la prueba de entrenamiento, se extrae un límite superior para los bosques aleatorios para obtener el error de generalización en términos de dos parámetros exactitud e interdependencia para obtener el voto de la mayoría y con este la clase final (Ozgođe Yigin et al., 2020).

Diagrama 3. Estructura de un bosque aleatorio



De esta forma, se desarrolla un algoritmo de agrupación de aprendizaje que combinará una serie de modelos más débiles para crear otro más robusto. La principal ventaja de este método es que normalmente no necesitaremos podar el bosque aleatorio (ya que el modelo es muy resistente al ruido). Sin embargo, es mucho menos interpretable que los árboles de decisión.

Luego de ejecutar el modelo para la base de trabajadores se obtuvo una precisión del 87.7%, la matriz de confusión se muestra en Cuadro 5.

Cuadro 5. Matriz de confusión del bosque aleatorio

		Predicción		
		Formal	Informal	Total
Observado	Formal	152 034	15 564	167 568
	Informal	17 930	86 552	104 482
Total		169 964	102 116	87.7%

Fuente: EEG y GEIH, cálculos del autor

Regresión logística

La regresión logística es un algoritmo de clasificación que se utiliza cuando el valor de la variable objetivo es de naturaleza categórica y es usada con mayor frecuencia cuando los datos en cuestión tienen una salida binaria, o son un 0 o un 1.

La función logística (ecuación 1) es una función que se asemeja a una curva en forma de "S" cuando se representa en un gráfico. Toma valores entre 0 y 1 y los jalona hacia los márgenes de la parte superior e inferior, etiquetándolos como 0 o 1.

$$\text{logit}(x_j) = \beta_0 + \sum_{k=2}^{k=11} \beta_k \text{rangos}_{jk} + \sum_{k=12}^{k=111} \beta_k \text{CIIU2}d_{jk} + \varepsilon_j$$

Tras ejecutar este modelo en para la base de trabajadores se obtuvo una precisión del 86.6%, la matriz de confusión se muestra en la Cuadro 6.

Cuadro 6. Matriz de confusión del modelo logístico

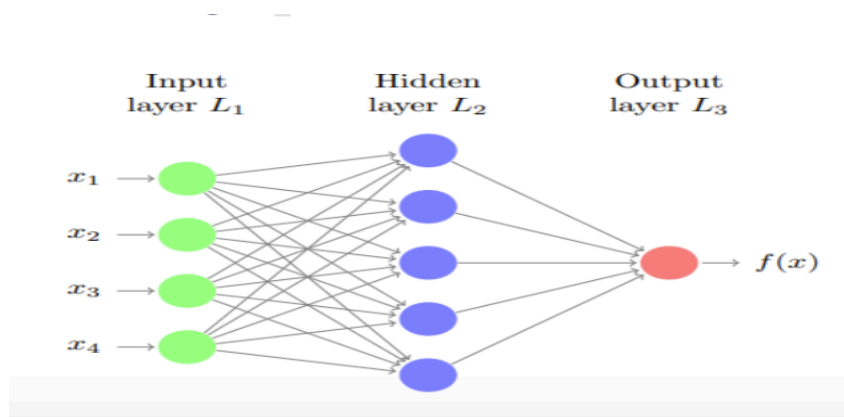
		Predicción		
		Formal	Informal	Total
Observado	Formal	149 862	17 730	168 598
	Informal	18 649	85 833	104 482
Total		168 511	103 569	86.6%

Fuente: EEG y GEIH, cálculos del autor

Red neuronal

Finalmente, una red neuronal es un modelo altamente parametrizado, inspirado en la arquitectura del cerebro humano, una máquina que con suficientes datos podría aprender cualquier relación predictiva fluida.

Diagrama 4. Estructura de una red neuronal



El Diagrama 4 muestra un ejemplo simple de un diagrama de red neuronal. Hay cuatro predictores o entradas (círculos verdes) y cinco unidades ocultas (círculos violetas). Estas unidades de memoria o neuronas aprenden automáticamente nuevas características de los datos a través de un proceso llamado aprendizaje supervisado. Cada neurona está conectada a la capa de entrada a través de un vector de parámetros o pesos se refiere a la primera capa. La idea era que cada neurona aprendería un binario simple en la función logística. La capa final o de salida también tiene pesos y una función de salida h . Para la regresión cuantitativa, h suele ser la función de identidad y, para una respuesta binaria, es

una vez más la función logística. Por último, las redes neuronales se ajustan por máxima verosimilitud (Efron y Hastie, 2021).

Luego de ejecutar el modelo para la base de trabajadores se obtuvo una precisión del 86.9%, la matriz de confusión se muestra en la Cuadro 7

Cuadro 7. Matriz de confusión del bosque aleatorio

		Predicción		
		Formal	Informal	Total
Observado	Formal	151.148	16.450	167.598
	Informal	19.130	85.352	104.482
Total		170.278	101.802	86.9%

Fuente: EEG y GEIH, cálculos del autor

Escogencia de modelo de clasificación.

Después de evaluar estos 4 modelos de clasificación, se concluye que el más eficiente es el bosque aleatorio con una precisión de 87,7%. Estas estimaciones asumen que las probabilidades de ocurrencia de estas variables dependientes son las mismas a nivel de trabajadores y de firmas.



Documentos de antecedentes del
Informe Nacional de Desarrollo Humano
2024



Suecia
Sverige



AÑOS - COLOMBIA & SUECIA