

Una guía técnica para la GEIH

José Manuel Bastidas Burbano
Wilman Arturo Gómez

Este trabajo hace parte de los compromisos del proyecto CODI 2024-67651
Definiciones Internacionales de Informalidad: un análisis comparativo para el
caso colombiano



**UNIVERSIDAD
DE ANTIOQUIA**

1 8 0 3

**FACULTAD DE CIENCIAS
ECONÓMICAS**

**DEPARTAMENTO DE
ECONOMÍA**

Medellín - Colombia

Comité editorial:

Carlos Andrés Vasco Correo M.Sc
Ramon Javier Mesa Callejas



© Facultad de Ciencias Económicas. Universidad de Antioquia. 2026

Jair Albeiro Osorio Agudelo
Decano de Facultad

Ramon Javier Mesa Callejas
Jefe de Departamento de Economía

Carlos Andrés Vasco Correa
Director Revista Lecturas de Economía.

Una guía técnica para la GEIH

José Manuel Bastidas Burbano¹

Wilman Arturo Gómez²

Introducción. – I. Resumen y objetivo. – II. Un recorrido por el cambio en el enfoque conceptual y metodológico de la Gran Encuesta Integrada de Hogares – GEIH. – III. Trilogía de datos maestros: GEIH-M05, GEIH-M18 y GEIH-EMP. – IV. Marco conceptual: definiciones operativas – V. Construcción del diccionario maestro – VI. Procedimiento operativo para la construcción de la Serie Trimestral del Mercado Laboral Colombiano: El camino a la unificación total, una serie única y consistente – VII. Conclusiones. – Referencias bibliográficas.

Resumen

En este trabajo se propone una metodología que no solo se limita a unir, de forma mecánica, las series provenientes de distintos marcos estadísticos, sino que ofrece una ruta de armonización capaz de distinguir entre cambios los componentes muestrales y conceptuales, y además valores atípicos como los observados en pandemia. La aplicación de esta metodología mostró que cada variable exige un análisis concreto, esto es, en algunas el ajuste muestra es suficiente, mientras que el componente conceptual permitió una mejor clasificación de los formales e informales. Adicionalmente propone un método de imputación para el recálculo de los trabajadores formales e informales más consistente con la dinámica del mercado laboral durante los primeros tres trimestres de 2020.

Palabras clave: Metodología, Gran Encuesta Integrada de Hogares, Empleo, Informalidad, Modelos SARIMA, Marco conceptual.

¹ Estudiante - Joven Investigador del proyecto CODI 2024-67651, Facultad de Ciencias Económicas, Universidad de Antioquia, Grupo de Macroeconomía Aplicada. josem.bastidas@udea.edu.co. ORCID: <https://orcid.org/0009-0005-8154-0897> -

² Profesor Departamento de Economía, Facultad de Ciencias Económicas, coordinador Grupo de Macroeconomía Aplicada, Universidad de Antioquia. Wilman.gomez@udea.edu.co. ORCID: <https://orcid.org/0000-0002-0194-7189> -

Abstract

This paper proposes a methodology that is not merely limited to mechanically linking series from different statistical frameworks but also offers a harmonization strategy capable of distinguishing between changes in sample and conceptual components, as well as outliers such as those observed during the pandemic. The application of this methodology showed that each variable requires a specific analysis; that is, for some variables, the sample adjustment is sufficient, whereas for others the conceptual component allowed for a better classification of formal and informal workers. Additionally, it proposes an imputation method for recalculating formal and informal workers that are more consistent with labor market dynamics during the first three quarters of 2020.

Key words: Methodology, Great Integrated Household Survey (GEIH), Employment, Informal Employment, SARIMA Models, Conceptual Framework.

Clasificación JEL: O17, N36, J82, J46, J31, J21, J01, E24

Introducción

La Gran Encuesta Integrada de Hogares (GEIH), es una encuesta muestral desarrollada mensualmente por el Departamento Nacional de Estadística (DANE) de Colombia. Su principal propósito es proporcionar información básica sobre el tamaño y la estructura de la fuerza de trabajo de la población del país. En este contexto, es de suma importancia para la toma de decisiones de política pública. Por lo tanto, su comprensión y manejo claro, permite que tanto investigadores como *policy makers*, puedan tomar decisiones, posturas y análisis importantes para mejorar la calidad de vida de los colombianos.

Según DIMPE & DANE (2016), desde finales de la década de 1960 se ha desarrollado e implementado, de forma transversal, encuestas a hogares con el fin de recolectar información de la fuerza de trabajo, la calidad de vida y los ingresos y gastos. En 2006 se amplía la cobertura de la Encuesta Continua de Hogares (ECH) de 13 ciudades (con sus áreas metropolitanas) a 24 ciudades, al resto de cabeceras y al resto rural. En el mismo año, se empieza con la recolección de la GEIH con su módulo central de mercado laboral e ingresos de hogares.

Ahora, es importante tener en cuenta que la clasificación de la población según el objetivo del estudio se realiza con base a los conceptos y definiciones de la fuerza de trabajo que establecen las Conferencias Internacionales de Estadísticos del Trabajo (CIET) de la Organización Internacional del Trabajo (OIT). Surgen, por lo tanto, varios cuestionamientos: ¿cuáles son las metodologías que utiliza el DANE con base a la GEIH para clasificar la población? ¿son comparables entre sí? ¿qué aspectos se deben tener en cuenta para poder establecer uniformidad en los datos proporcionados a la comunidad?

El objetivo principal es ofrecer a la comunidad una guía para lograr uniformidad entre los cambios metodológicos: como las definiciones y la adopción de nuevos marcos muestrales (en concreto: CNPV05 a CNPV 18). Y de esta forma, lograr series largas que puedan tener comparabilidad según el objetivo de análisis. Para ello, el foco se encuentra en los valores absolutos de la población que integra el mercado laboral, las tasas que derivan de ellos y los ingresos laborales promedio³ en frecuencia trimestral entre el 2007 y el 2024. Aquellos son: población ocupada, población ocupada formal, población ocupada informal, población desocupada, población económicamente activa (fuerza de trabajo), población en edad de trabajar y población total.

I. Resumen y objetivo

El documento presenta una guía técnica para unificar tres series de tiempo básicas procedentes de la *Gran Encuesta Integrada de Hogares (GEIH)* para distintos grupos poblacionales del mercado laboral. Esta surge de la necesidad de tener una sola serie

³ Los ingresos laborales promedio solo se calculan para la población ocupada, la población ocupada formal y la población ocupada informal.

consistente ante los cambios conceptuales y metodológicos que se dieron a lo largo de los años y generan rupturas de comparabilidad. El objetivo es proponer un *método* replicable para unificar y construir series consistentes a lo largo del tiempo. En este caso, de forma trimestral, el alcance temporal cubre desde el primer trimestre del 2007 y el cuarto trimestre del 2025; geográficamente, aborda el total nacional de los valores absolutos de los grupos poblacionales. Lo anterior, por supuesto, se logra al utilizar microdatos oficiales proporcionados por el *Departamento Administrativo Nacional de Estadística (DANE)*. La metodología combina los siguientes elementos: estandarización de diccionarios, reglas de codificación y validaciones de consistencia a través de la comparación con cifras oficiales. Por último, como resultado, se entrega un paso a paso para la unificación de variables y la base de datos final con las series de tiempo unificadas y consistentes para el mercado laboral colombiano.

II. Un recorrido por el cambio en el enfoque conceptual y metodológico de la Gran Encuesta Integrada de Hogares – GEIH

1. Contextualización histórica y descriptiva de la GEIH y el Marco 2005

La GEIH es consecuencia de operaciones estadísticas históricas. Por ejemplo, en los años setenta, se inició el desarrollo e implementación de la Encuesta Nacional de Hogares (ENH). Esta encuesta fue insumo básico para producir estadísticas demográficas, sociales y económicas de la población colombiana. Conforme pasaron los años, se incluyeron municipios circundantes a las siete ciudades principales y otras ciudades. De forma que, en los años noventa, se presentaron resultados estadísticamente significativos para las áreas metropolitanas del país. Para 1996, se amplió la muestra, hito que permitió presentar resultados a nivel de 24 departamentos del país. Ahora, en concordancia con el desarrollo del plan de trabajo del DANE, en 1999 se inicia como prueba piloto, en Bogotá y luego en Medellín, Cali, Barranquilla, Bucaramanga, Manizales, y Pasto, la aplicación de la Encuesta Continua de Hogares (ECH). Esta encuesta formaría parte del Sistema de Encuestas a Hogares.

Sin embargo, en el 2005, con el propósito de extender el alcance temático y reducir costos, se integran las encuestas más importantes a nivel de hogares del DANE: la Encuesta Continua de Hogares (ECH), la Encuesta Nacional de Ingresos y Gastos (ENIG) y la Encuesta de Calidad de Vida (ECV). La integración de estas encuestas no fue del todo efectiva. Para 2006, se recibió un apoyo financiero del Banco Mundial para el Mejoramiento de las Encuestas y la Medición de las Condiciones de Vida en América Latina y el Caribe (MECOVI). Este apoyo permitió el desarrollo inicial de la Gran Encuesta Integrada de Hogares (GEIH). En 2008, la encuesta se aplica de manera paralela en la muestra diseñada alrededor del marco muestral del Censo Nacional de Población y Vivienda de 2005 (CNPV 2005). De esta forma, surge esta encuesta como resultado de la construcción de un marco estructural a lo largo de diez años y, en este mismo proceso, se define que su eje central debía ser el mercado laboral (DIMPE & DANE, 2016).

Llegados a este punto, es importante realizar una descripción de la GEIH. Su misión fundamental es proporcionar información sobre el tamaño y la estructura de la fuerza de trabajo de la población del país, al igual que características sociodemográficas como sexo, edad, parentesco, nivel educativo, afiliación al sistema de seguridad social, entre otros. La encuesta se realiza mensualmente por muestreo. En cuanto al alcance, los resultados permiten una desagregación por total nacional, total de las cabeceras municipales, total de los centros poblados y rural disperso. Además, cada una de las 23 ciudades capitales y áreas metropolitanas, San Andrés, Buenaventura, Barrancabermeja, Soacha, Tumaco, Rionegro y las 8 ciudades capitales de la Amazonía y la Orinoquía.

En cuanto al marco normativo de las estadísticas del trabajo, para la GEIH Marco 2005 se implementan las siguientes resoluciones o directrices: la *Resolución II* acerca de las normas de la población económicamente activa, ocupados, desocupados e inactivos de la CIET 13 (International Labour Office, 1982); la *Resolución III* sobre la ocupación en el sector informal (International Labour Office, 1993); y la *Resolución I* que aborda la medición del subempleo y las situaciones de empleo inadecuado (International Labour Office, 1998). Es decir, antes del rediseño conceptual y metodológico que desembocaría en la GEIH Marco 2018, se implementaron las definiciones proporcionadas por la Organización Internacional de Trabajo hasta 1998. Es entonces que, dados los cambios a lo largo de los años, y a partir de la actualización del marco geoestadístico a través del Censo Nacional de Población y Vivienda del 2018 (CNPV 2018), es necesario realizar un rediseño a la GEIH.

2. Actualización conceptual y metodológica: GEIH – Marco 2018

A partir de 2018, el DANE comienza con el rediseño de la GEIH. El objetivo principal fue “actualizar los diseños muestrales, operativos y conceptuales de la GEIH, con la información del CNPV 2018 y los actuales lineamientos de mercado laboral y pobreza monetaria” (DANE, 2022b). A continuación, se presenta una síntesis de los principales hitos del rediseño de la GEIH. El propósito es contextualizar los cambios metodológicos y conceptuales del marco 2018. Esta revisión permite comprender el origen del objetivo que se plantea en este estudio: ofrecer una serie trimestral unificada para el mercado laboral colombiano. El proceso de actualización de la GEIH se desarrolla en siete momentos entre 2018 y 2022.

- Primer y segundo momentos (2018 – 2019): En 2018 aparece la necesidad de actualizar el *Marco* con el CNPV 2018, debido a los cambios que son inherentes al paso del tiempo: surgimiento de nuevos barrios edificaciones urbanas, como también la densificación de centros poblados rurales (Hermida & Pulido-Mahecha, 2022). En 2019 se contratan dos expertos —uno conceptual-temático y otro estadístico— con el fin realizar una revisión de los fundamentos teóricos y técnicos de la encuesta. En este segundo momento se realiza una revisión conceptual y metodológica de la GEIH. Se revisan aspectos como los objetivos y el alcance, además, se incluyen las recomendaciones de la *Conferencia Internacional de Estadísticos del Trabajo (CIET)* diecisieteava a la veinteava; se ajustan las preguntas relacionadas con los módulos de migración, ruralidad, no ocupados, características generales, salud, educación, ingresos e impuestos; y por supuesto, se toma

en cuenta el *Plan Nacional de Desarrollo (PND)* 2018-2022 para mejorar la visibilidad estadística de grupos poblacionales como el campesinado, la comunidad LGBT y las personas con discapacidad.

- Tercer y cuarto momentos (2020): En 2020 se llevan a cabo las fases de diseño temático, estadístico y operativo. Estas incluyen la elaboración de manuales, el plan de pruebas, la definición y la estratificación de la muestra, y el diseño del piloto experimental. Simultáneamente se trabaja en la construcción del marco estadístico, la selección de la muestra y el desarrollo de los instrumentos de recolección y pruebas del sistema de producción. El tercer y cuarto momento, permiten consolidar los procedimientos que servirían de base para el levantamiento de información para el *Marco 2018*.
- Quinto y sexto momentos (2021): En 2021 comienza la recolección paralela a través de dos operativos independientes para los marcos 2005 y 2018. La recolección de la GEIH marco 2005 fue continua con un seguimiento operativo para obtener los indicadores del mercado laboral durante 2021 (DANE, 2022b). Mientras que la recolección de la GEIH marco 2018 se realiza con el objetivo de monitorear semanalmente, los posibles cambios operativos y de medición relacionados con la transición. En esta etapa, se detectan inconsistencias en la información, y por supuesto, se ejecutan ajustes conceptuales y operativos con el personal de campo. Este proceso permite evaluar los indicadores que surgen de ambos marcos, e igualmente, detectar diferencias entre los diseños muestrales y los nuevos criterios conceptuales.
- Séptimo momento (2022): Por último, en 2022 la GEIH adopta plenamente el *Marco 2018* y con ello, se consolida la actualización temática, conceptual y muestral iniciada en 2018.

Adicionalmente, en cuanto a las actualizaciones del diseño muestral de la encuesta, se incorpora la desagregación de siete ciudades sin área metropolitana en mediciones trimestrales: Medellín, Barranquilla, Pereira, Cúcuta, Manizales, Bucaramanga y Cali. También, en las actualizaciones conceptuales se tienen los siguientes cambios en el marco normativo de las estadísticas de trabajo, en concordancia con la OIT (ILO): se adopta la directriz sobre la definición estadística de la ocupación informal de la CIET 17(ILCS) (International Labour Office, 2003); la *Resolución I* sobre una definición más amplia del trabajo que incorpora cambios en la medición de la fuerza de trabajo (fuerza de trabajo potencial) de la CIET 19 (International Labour Office, 2013) y la *Resolución I* relativa a las estadísticas sobre las relaciones laborales, las directrices adoptadas sobre las estadísticas de migración laboral internacional y las directrices relativas a la medición de las cualificaciones y la inadecuación de las competencias de las personas en la ocupación de la CIET 20 (International Labour Office, 2018). Los anteriores cambios permiten un marco de medición para todas las formas de trabajo y proporcionan la posibilidad de desarrollar nuevos indicadores que contribuyan al análisis de la dinámica del mercado laboral colombiano.

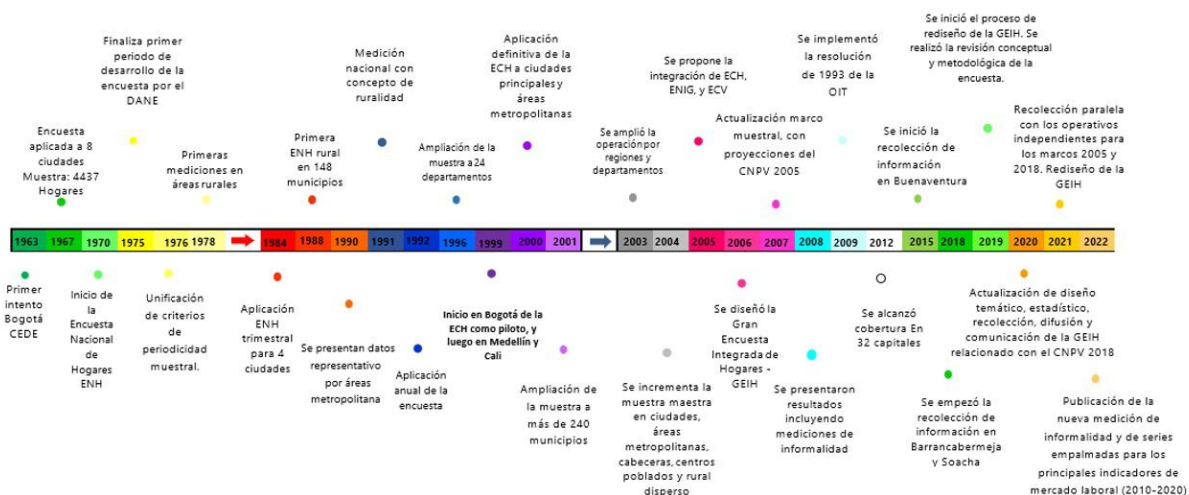
Otro cambio, no menos importante, surge a partir del convenio 138 de 1973 de la OIT sobre la edad mínima para trabajar. En Colombia es aprobada a través de la Ley 515 de 1999 y ratificada en 2001 (Ley 515 de 1999, 1999). Esta ley ajusta el límite inferior de la población en edad de trabajar: 15 años y más.

De esta manera, es evidente el desafío metodológico que el documento técnico propone resolver. Esto a través de un *método simple* que se estructura alrededor de dos frentes de armonización (unificación o estandarización) que están orientados a garantizar la comparabilidad intertemporal de la GEIH.

En primer lugar, se propone *armonizar o estandarizar* el factor de expansión de la GEIH a lo largo del horizonte temporal considerado en el documento. Lo anterior se logra al articular tres grandes tramos: (i) la base de datos que incorpora el factor de expansión del marco metodológico 2005 (CNPV 2005) para el periodo 2007-2021, (ii) el factor de expansión asociado a la base de datos con las series empalmadas para 2010-2020, y (iii) el factor de expansión asociado a la base de datos bajo el marco metodológico 2018 (CNPV 2018) para 2022-2024.

En segundo lugar, se plantea armonizar o estandarizar las definiciones operativas o conceptos de clasificación de las categorías que conforman el mercado laboral entre el marco 2005 y el marco 2018. De esta forma, se pretende que las categorías sostengan definiciones operativas que sean consistentes en el periodo estudiado. La estandarización aborda las definiciones de población ocupada, población desocupada, población económicamente activa o fuerza de trabajo, población en edad de trabajar y población total. También se incluye las poblaciones ocupadas informal y formal con un tratamiento particular, el cual conecta directamente con el tratamiento para los ingresos laborales promedio de los informales y los formales. A continuación, se presentan los cambios más importantes que tuvo la GEIH.

Figura 1. Línea de tiempo: cambios en la encuesta GEIH.



Fuente: (DIMPE & DANE, 2023).

III. Trilogía de datos maestros: GEIH-M05, GEIH-M18 y GEIH-EMP

Para el método de empalme propuesto, es necesario aclarar con rigor el cimiento principal del proceso. La trilogía de datos maestros consiste en tres bases de datos, cada una con las variables del mercado laboral que integran los conceptos metodológicos y conceptuales apropiados a su contexto. La primera de ellas es la *GEIH-M05*. Esta se compone de siete variables poblacionales y tres de ingresos. Todas estas bajo el amparo del factor de expansión derivado del *censo nacional de población y vivienda 2005 (CNPV 2005)* y las definiciones conceptuales de clasificación de población del marco 2005. La segunda es la *GEIH-M18*. Integra las mismas variables. La diferencia fundamental está en su construcción: a partir del factor de expansión derivado del *censo nacional de población y vivienda 2018 (CNPV 2018)* y las definiciones conceptuales de clasificación de población del marco 2018. Por último, y que es fundamental en este proceso, la *GEIH-EMP*. Integra las series de tiempo de las mismas variables ajustadas con el factor de expansión empalmado que propone el DANE (2022a). Esta base de datos permite conectividad y comparabilidad muestral entre las dos primeras. Aunando a lo anterior, dado su periodo de disponibilidad, entre 2010 y 2020, solo es posible construirla conceptualmente a partir del marco 2005. A continuación, se realiza una breve descripción de cada una de las bases de datos.

1. Gran Encuesta Integrada de Hogares – Marco 2005 (GEIH-M05)

Las series trimestrales de esta base de datos están disponibles desde el primer trimestre del 2007 hasta el último trimestre del 2025. Cabe aclarar que los trimestres se toman de forma natural, es decir, desde enero a marzo y así sucesivamente.

1.1. Diseño temático y metodológico

Para este periodo, la GEIH responde a la necesidad de información para la medición de la estructura de la fuerza de trabajo (DIMPE & DANE, 2016). De esta información se derivan análisis de la situación del empleo, el desempleo y las características sociodemográficas de la población colombiana. En concreto, dados ciertos conceptos y definiciones por la *Conferencia Internacional de Estadísticos del Trabajo (CIET)* de 1982, 1993 y 1998 (DANE, 2022b), se pretende clasificar la población de cada uno de los dominios; y con lo anterior, calcular los principales indicadores del mercado laboral: tasa global de participación (TGP), tasa de ocupación (TO), tasa de desempleo (TD), tasa de subempleo (TS) y proporción de informalidad (PI) y su variación a lo largo del tiempo. Además de lo anterior, la GEIH permite medir características de la población, la vivienda, el acceso a los servicios públicos, variables demográficas, características del empleo, del desempleo y de la inactividad, información básica sobre microempresas e ingresos de los hogares tanto en dinero como especie (vital para las mediciones de la pobreza). Asimismo, la información que se recolecta en este periodo yace en trece ciudades y sus áreas metropolitanas, once ciudades intermedias, ocho capitales de nuevos departamentos; cabeceras y centros poblados y rural disperso en más de cuatrocientos cuarenta y tres municipios del país (DIMPE & DANE, 2016).

1.2. Diseño del cuestionario

La GEIH abarca una cobertura nacional y como se menciona anteriormente, tiene una desagregación geográfica y temporal. Esta última es de profundo interés. El cuestionario se aplica mensualmente para las trece grandes ciudades capitales y sus áreas metropolitanas, las once ciudades capitales intermedias y para el total nacional. Trimestralmente por ciudad capital, grandes e intermedias, con su correspondiente área metropolitana y para el total del país por zona, esto es: cabecera, centros poblados y rural disperso. Y de forma semestral por ciudad capital, grandes e intermedias, con su correspondiente área metropolitana; por grandes regiones, Atlántica, Oriental, Central, Pacífica y Bogotá; para cabeceras, centros poblados y rural disperso; y, por último, para el total nacional por zona (**DIMPE & DANE, 2016**).

El cuestionario de la GEIH se compone por los siguientes capítulos:

- **Datos de la vivienda.** Corresponde a la información acerca de características físicas de la vivienda. Esta permite determinar su calidad y su acceso a servicios públicos domiciliarios.
- **Datos del hogar.** Permite determinar el grado de hacinamiento de los hogares, sus condiciones en términos de servicios públicos y salubridad. Se tiene en cuenta para ello, el abastecimiento de agua, el tipo de servicio sanitario y el manejo de basuras.
- **Registro de personas.** Recolecta información sobre las personas que conforman el hogar.
- **Características generales.** Esta información permite caracterizar las variables demográficas de la población: edad, sexo, estado civil y parentesco con el jefe de hogar.
- **Seguridad social en salud.** Permite identificar si los individuos tienen acceso al sistema de salud: se indaga de esta forma, por la afiliación y el tipo de régimen de seguridad social.
- **Educación.** Con esta información, se identifica el alfabetismo, la asistencia escolar formal, el nivel educativo alcanzado y el grado de estudio de los encuestados.
- **Fuerza de trabajo.** Permite clasificar la fuerza de trabajo entre ocupados, desocupados e inactivos.
- **Ocupados.** Permite conocer las características de la población ocupada y de sus empleos.
- **Desocupados.** Permite conocer las características de la población desocupada y su búsqueda de empleo.

- **Inactivos.** Posibilita conocer las características de la población inactiva.
- **Otras actividades y ayudas en la semana pasada:** Capta actividades realizadas por los miembros de los hogares y por las cuales no obtienen remuneración. Además, permite captar las actividades de trabajo comunitario voluntario.
- **Otros ingresos:** Determina cuáles personas recibieron otros ingresos distintos a los ingresos laborales.

1.3. Estructura de los microdatos

Para el periodo comprendido entre el 2007 y el 2021, el DANE ofrece la posibilidad de descargar los microdatos en frecuencia mensual para los siguientes módulos:

- Vivienda y hogares.
- Características generales (personas).
- Fuerza de trabajo.
- Ocupados.
- Desocupados.
- Inactivos.
- Otras actividades y ayudas en la semana.
- Otros ingresos.

En pro de los cambios metodológicos y conceptuales descritos anteriormente, a partir del 2022, y por ende hasta 2024, los módulos sufren un pequeño cambio:

- Datos del hogar y la vivienda.
- Características generales, seguridad social en salud y educación.
- Fuerza de trabajo.
- Ocupados.
- No ocupados.
- Otras formas de trabajo.
- Otros ingresos e impuestos.
- Migración.

Si bien el cambio es evidente, y dada la agrupación de ciertos módulos para incorporar otros, las preguntas fundamentales para el cálculo de las poblaciones persisten. Por lo tanto, no amerita un esfuerzo explicarlo a profundidad.

Los microdatos se estructuran a partir de una estructura jerárquica compuesta por vivienda, hogar y personas. En una vivienda pueden coexistir uno o varios hogares, y a su vez, dentro de cada hogar hay una o varias personas. En consecuencia, la integración de los distintos

módulos de la encuesta exige construir identificadores acordes a dicha jerarquía que permitan enlazar correctamente la información. Para este caso, la variable DIRECTORIO identifica la vivienda, SECUENCIA_P distingue el hogar al interior de la vivienda y ORDEN identifica a cada persona dentro del hogar.

Con base a estas variables se generan los identificadores que sirven como llaves de empalme entre los módulos. En primer lugar, se genera *id_vivienda* a partir de DIRECTORIO. Luego, se genera *id_hogar* mediante la concatenación de *id_vivienda* con SECUENCIA_P. Esto permite identificar de forma única cada hogar. Finalmente, en los módulos cuyo nivel de observación es la persona, se construye *id_persona* concatenando *id_hogar* con ORDEN. De este modo, cada individuo queda plenamente identificado dentro de la estructura de la encuesta.

Por lo tanto, cada módulo es tratado de forma independiente. Es decir, se carga cada archivo original, se construyen los identificadores correspondientes según el nivel de observación del módulo y se organiza la base a partir de los identificadores. Este procedimiento se aplica al módulo de *Vivienda y Hogares* y a los diferentes módulos de personas, entre ellos: *Características generales*, *Fuerza de trabajo*, *Inactivos*, *Desocupados*, *Ocupados*, *Otras actividades y ayudas en la semana* y *Otros ingresos*.

Como se mencionó anteriormente, la diferencia entre los módulos radica principalmente en su unidad de observación. Por ejemplo, para *Vivienda y Hogares*, la información está asociada al nivel del hogar. De forma que basta con construir *id_vivienda* e *id_hogar*. En contraste, en los módulos de personas resulta necesario construir también *id_persona*, dado que la información se registra de manera individual.

Ahora, una vez preparados todos los módulos, procedemos con la consolidación de la base mensual. El archivo “base” es el módulo de *Características generales*. Este funciona como estructura principal sobre la cual se integran los demás módulos individuales. A esta base se anexan de forma sucesiva los módulos de *Fuerza de trabajo*, *Inactivos*, *Desocupados*, *Ocupados*, *Otras actividades y ayudas en la semana* y *Otros ingresos*. Lo anterior se logra a través de un empalme de tipo uno a uno con base en *id_persona*. De forma interpretativa, este proceso significa que, para cada persona identificada en la base principal, se agrega la información correspondiente proveniente de los demás módulos individuales. Esto reúne en un solo registro sus características sociodemográficas, laborales y de ingresos.

Luego, se incorpora el módulo de *Vivienda y Hogares* mediante un empalme tipo muchos a uno al utilizar *id_hogar*. La lógica del procedimiento propuesto es la siguiente: en la base principal existen múltiples personas pertenecientes a un mismo hogar, mientras que en el módulo de vivienda y hogares existe una sola observación por hogar. De esta manera, la información del hogar se asigna a todas las personas que lo conforman. Esto es, integrar una sola base las características personales como las condiciones del entorno doméstico.

Completado este procedimiento para cada uno de los meses que conforman el trimestre, el siguiente paso consiste en realizar una integración vertical de las bases mensuales. Esta etapa se apoya en un proceso de apilamiento, en el que las bases de cada uno de los meses del trimestre se concatenan por filas, siempre que compartan la más estructura de variables. Por último, se divide el factor de expansión por el número de meses con el fin de evitar sobreestimaciones. Así, el proceso total de construcción de la base trimestral comprende dos momentos: primero, la consolidación de cada mes a partir del empalme de módulos; y segundo, la unión secuencial de los meses para obtener una base trimestral.

2. Gran Encuesta Integrada de Hogares – Marco 2018 (GEIH-M18)

Las series trimestrales pertenecientes a esta base de datos están disponibles en frecuencia trimestral, desde el primero del 2022 al último de 2025. Es decir, se ofrece la información actualizada hasta este momento. Esta base de datos es importante porque recoge dos cambios fundamentales: el metodológico y conceptual. Además, integra las series que se utilizan oficialmente por el DANE para el cálculo de indicadores del mercado laboral colombiano. Las mejoras en la medición, tanto metodológica como conceptualmente, permiten un acercamiento a la realidad óptimo que antes, y por supuesto, con estándares internacionales actualizados. A continuación, se establecen los elementos más importantes de esta base de datos.

2.1. Diseño temático y metodológico

Esta base de datos se caracteriza principalmente por la actualización metodológica y conceptual a partir de diferentes bases. Entre las más importantes, que son de carácter internacional, se encuentran las recomendaciones de la Organización Internacional del Trabajo (OIT), que, a su vez, están plasmadas en los lineamientos dados por las Conferencias Internacionales de Estadísticos del Trabajo (CIET). En este orden de ideas, en el rediseño de la operación estadística incorporar las recomendaciones de la CIET 17^a (2003), 19^a (2013) y 20^a (2018). De manera complementaria, algunos requerimientos nacionales en concordancia con las necesidades de política pública, también se incluyeron. Inicialmente, el Plan Nacional de Desarrollo fue uno de los cimientos del rediseño. Esto permitió que el DANE trabajara junto con el Departamento Nacional de Planeación (DNP) para identificar las necesidades de política pública con base a la GEIH y establecer criterios que permitieran un rediseño transparente (DIMPE & DANE, 2023).

Como se ha mencionado a lo largo del documento, igualmente el marco 2018 tiene el objetivo de brindar información estadística significativa para generar indicadores del mercado laboral como: porcentaje de población en edad de trabajar (% PET), tasa global de participación (TGP), tasa bruta de participación (TBP), tasa de desempleo (TD), tasa de ocupación (TO), proporción de informalidad (PI), entre otros; al igual que generar información relacionada con el ingreso de los hogares, los indicadores de pobreza monetaria y de pobreza monetaria extrema y por último, medir las características demográficas de los colombianos.

Esto último resulta muy interesante porque, en aras de cumplir con los lineamientos del PND 2018-2022, se integran preguntas que permiten dar visibilidad estadística a grupos poblacionales como campesinado, personas con autorreconocimiento étnico, LGBT, etc. El alcance, por lo tanto, permite caracterizar a los colombianos en diferentes variables sociodemográficas. Los resultados que surgen de esto permiten una desagregación geográfica de la siguiente manera: total nacional, total de cabeceras municipales, total de los centros poblados y rural disperso, cada una de las 23 ciudades capitales y áreas metropolitanas, San Andrés, Buenaventura, Barrancabermeja, Soacha, Tumaco, Rionegro y las 8 ciudades capitales de los departamentos de la Amazonía y Orinoquía.

2.2. Diseño del cuestionario

Desde el 2021, luego del rediseño de la encuesta, se aplican los siguientes capítulos:

- **Datos de la vivienda:** permite conocer las características físicas de la vivienda que permitan determinar la calidad y el acceso a los servicios públicos domiciliarios.
- **Datos del hogar:** determina el grado de hacinamiento de los hogares, sus condiciones y calidad del suministro de los servicios públicos domiciliarios, las condiciones de salubridad y los productos financieros con los que cuenta el hogar.
- **Registro de personas:** se registran todas las personas que forman parte del hogar.
- **Características generales:** caracteriza la población conforme a su edad, sexo, estado civil y parentesco con el jefe de hogar. Identifica, además, la población campesina y los distintos grupos étnicos. De igual forma, en este marco se incluyen preguntas asociadas a las consecuencias que dejó la pandemia del COVID – 19.
- **Seguridad social en salud:** identifica si las personas tienen acceso al sistema de salud e indaga el tipo de afiliación y el tipo de régimen de seguridad social. En cuanto a personas en condición de discapacidad, se incluyen preguntas que permiten caracterizar las dificultades que esto provoca.
- **Educación:** identifica el alfabetismo, la asistencia escolar al sistema educativo formal, el mayor nivel educativo alcanzado y el grado de estudio de los encuestados.
- **Fuerza de trabajo:** permite clasificar la población en edad de trabajar (PET) entre la población en la fuerza de trabajo (FT) o fuera de la fuerza de trabajo (FFT).
- **Ocupados:** permite conocer las características de la población ocupada y sus empleos. Se incluyen, de forma novedosa, preguntas que logran identificar elementos importantes de asalariados, independientes con o sin negocio (con o sin registro mercantil) que finalmente, se asocian a una mejor clasificación de informalidad de los ocupados.

- **No ocupados:** permite conocer las características de la población no ocupada y de su búsqueda de empleo.
- **Otras formas de trabajo:** identifica las actividades realizadas por miembros del hogar y por las cuales no consiguen remuneración. Además, permite captar las actividades de trabajo comunitario voluntario.
- **Otros ingresos e impuestos:** determina las personas que recibieron otros ingresos distintos a los laborales y capta los impuestos pagados durante los últimos doce meses.
- **Migración:** identifica las dinámicas migratorias de la población dentro y fuera del país y la migración laboral internacional. Este capítulo es nuevo. Surge a partir de las recomendaciones de la CIET 20ª (2018).

2.3. Estructura de los microdatos

La construcción de la base de datos parte de un proceso similar al anterior descrito para el marco 2005. En términos generales, el proceso tiene tres etapas complementarias: i) el empalme de módulos dentro de cada mes, ii) la consolidación mensual de la información a nivel de personas y hogares, y iii) la agregación de los meses para conformar trimestres.

En este caso, la GEIH se organiza en ocho módulos que capturan dimensiones distintas de la población. Entre ellos se encuentran: *datos del hogar y la vivienda, características generales, seguridad social en salud y educación, fuerza de trabajo, migración, no ocupados, ocupados, otras formas de trabajo y otros ingresos e impuestos*. Cada uno de estos archivos tiene información concreta y complementaria. Para ejemplificar, el módulo de hogar y vivienda reúne las características estructurales de la unidad habitacional y del hogar, el de características generales incorpora atributos sociodemográficos y de educación, y los módulos de fuerza de trabajo, ocupados, no ocupados y otras formas de trabajo, aportan información necesaria para clasificar a la población según su relación con el mercado laboral. Igualmente, el módulo de otros ingresos e impuestos amplía la caracterización económica de los individuos y hogares, mientras que el módulo de migración introduce elementos relevantes para el análisis poblacional y territorial.

La información se encuentra fragmentada en ocho módulos con microdatos. El primer paso consiste en construir un sistema de identificadores que permita el enlace correcto de las observaciones. De esta forma, la integración se apoya en una estructura jerárquica similar a la descrita anteriormente: vivienda, hogar y persona. Esta jerarquía organiza el empalme y define el nivel de incorporación de cada módulo. Así, la integración no se realiza de forma indiferenciada. Por el contrario, respeta la naturaleza de cada archivo y unidad estadística que representa.

Desde una mirada metodológica, se parte de la variable DIRECTORIO para definir el identificador de vivienda. Con base a esta, se construye el identificador del hogar por medio

de la concatenación con la variable del hogar, que en unos casos es SECUENCIA_P y en otros HOGAR. Luego, el identificador de la persona se obtiene al añadir la variable ORDEN. Además, en este proceso aparece la variable FEX_C18, que corresponde al factor de expansión del marco 2018 y que posteriormente es ajustado con el fin de mantener la sintonía entre tres meses consecutivos.

Una vez se determina el empalme de los módulos, se obtiene una base de datos mensual consolidada. Esta representa una unidad intermedia en este proceso. El siguiente paso corresponde al empalme de meses para formar trimestres. Es decir, una vez construidas las bases mensuales, estas se agregan de forma vertical en grupos de tres meses consecutivos. Esto da pie a una base trimestral única para cada periodo del año. La agregación vertical permite ampliar el tamaño de la base y adaptar la información a una periodicidad trimestral. Sin embargo, la formación de trimestres implica un tratamiento específico al factor de expansión. Se divide la variable FEX_C18 entre el número de meses del trimestre. Se realiza este ajuste con el fin de evitar que la suma de tres meses produzca una sobreestimación de la población representada.

3. Gran Encuesta Integrada de Hogares – Series de Empalme (GEIH-EMP)

Esta base de datos surge con la necesidad de empalmar las series del marco 2005 con el marco 2018. En el 2021 se realizó una encuesta paralela para recolectar información con la aplicación del formulario rediseñado de forma simultánea con la encuesta del marco 2005. Luego, se contrastaron los resultados de la encuesta tradicional y la encuesta paralela. Se identificaron diferencias considerables entre la población ocupada y la población fuera de la fuerza laboral; para la población desocupada, no se encontraron cambios significativos.

La situación mostró un punto de inflexión en las series de tiempo que afectaban la comparabilidad desde el año 2021 respecto a los periodos antecesores. Como respuesta, el DANE desarrolló un método de ajuste de dichas series. Este consiste en calcular un factor de empalme que se aplica a las series de ocupados y fuera de la fuerza laboral, desde el 2010 hasta el 2020.

El método de empalme se divide en tres pasos. Primero, se divide el país en cuarenta bloques geográficos acorde al diseño muestral actual. Estos bloques se definen en 32 cabeceras capitales de departamento, 4 regiones para el resto de las cabeceras y otras 4 regiones para los centros poblados y rural disperso. Las regiones son: Atlántica, Pacífica, Oriental y Central. Segundo, luego de realizar la división geográfica, se calcula un ponderador para cada bloque mencionado anteriormente. Este, al multiplicarlo con la población ocupada del marco 2005 promedio del año 2021, da exactamente el promedio de ocupados del marco 2018 para el año 2021.

$$\text{Ponderador} = \frac{\text{población ocupada marco 2018}}{\text{población ocupada marco 2005}} = \frac{PO_{18}}{PO_{05}}$$

Tercero, se construye un factor por cada mes a partir de enero de 2010 hasta diciembre de 2020 que ajusta de forma secuencial una parte del nivel de la serie de ocupados. Por lo tanto, se define este paso de la siguiente forma:

$$\theta = 1 - \text{Ponderador}$$

De manera adicional, θ se divide en 132 partes, los meses entre enero de 2010 y diciembre de 2020. Entonces:

$$\text{Ponderador}_i = 1 - \left(\frac{i\theta}{132} \right)$$

Por último, como cuarto paso, luego de calcular la población ocupada ajustada con el ponderador se calcula la población fuera de la fuerza laboral y la población desempleada. Para la población desempleada, el DANE consideró que el cambio de marco no tuvo un impacto significativo (DIMPE & DANE, 2022).

IV. Marco conceptual: definiciones operativas

Es importante tener claro el marco conceptual y normativo que abordan las definiciones operativas para la clasificación poblacional del mercado laboral colombiano. Esta sección se enfocará en dos frentes amplios: las definiciones bajo el marco 2005, y las definiciones bajo el marco 2018. Cada uno de estos cuenta con siete categorías poblacionales. Además, se incluye un mapa esquemático para cada categoría con el fin de contrastar la evolución de las definiciones. Existen algunas categorías que no tienen cambios de un marco conceptual a otro, estas serán mencionadas a continuación.

1. Población total (PT)

La población total se define como la suma de las observaciones del trimestre procesado, cada una ponderada por su factor de expansión, de forma que el agregado trimestral corresponde al total poblacional estimado para ese periodo. Esta definición, de elaboración propia, se aplica tanto al *marco 2005*, al *marco 2018* y las *series de empalme*.

2. Población en edad de trabajar (PET)

La población en edad de trabajar se define desde dos enfoques. El primero, bajo el marco 2005, define que la PET la integran personas a partir de los doce años y más en la parte urbana, y de diez años y más en la parte rural (DANE, 2022d). En la actualización conceptual para el marco 2018, el intervalo de edad se modifica. Esto es, el conjunto de personas de 15 años o más que a su vez, se divide en dos grandes grupos: la fuerza de trabajo y la población fuera de la fuerza de trabajo (DANE, 2026b). Este cambio se adoptó en concordancia con el convenio 138 de la OIT sobre la edad mínima de admisión al empleo, aprobado en Colombia

a través de la Ley 515 de 1999. Es importante mencionar que, en la actualización de esta definición, esta abarca tanto el contexto urbano como el rural.

$$\text{Población en edad de trabajar} = \text{Fuerza de trabajo} + P. \text{fuera de la fuerza de trabajo}$$

3. Población económicamente activa o fuerza de trabajo (FT)

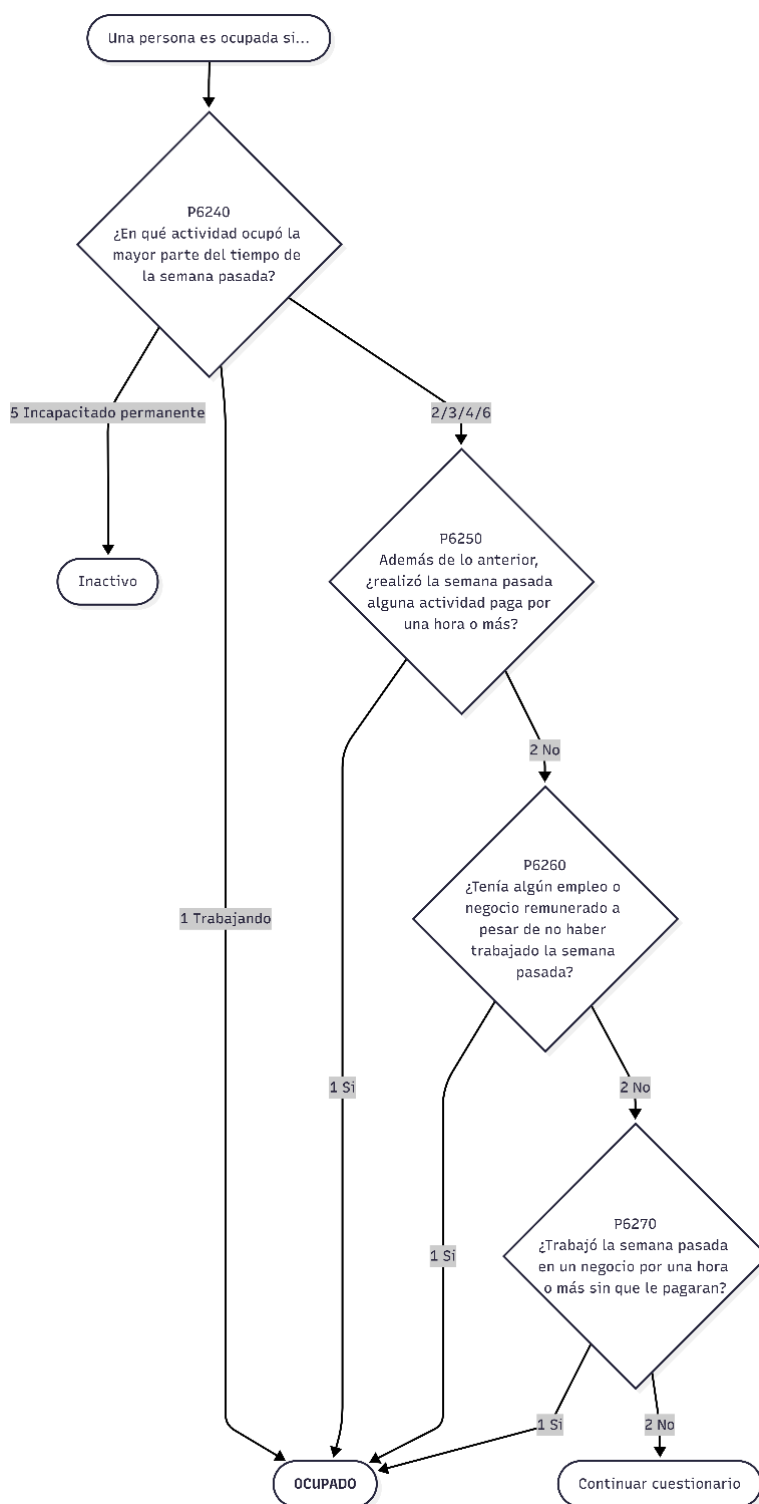
Para el marco 2005, la población económicamente activa, denominada también como fuerza laboral, se conforma por las personas en edad de trabajar que participan en el mercado laboral, ya sea porque se encuentran ocupadas o porque buscan activamente empleo. Operativamente hablando, se desagrega en ocupados y desocupados (DANE, 2022d). La anterior definición se basa completamente en la CIET 13 (International Labour Office, 1982). En la actualización conceptual, se cambia la terminología, esto al abordar los conceptos de la CIET 19 (International Labour Office, 2013). En este orden de ideas, la PEA se denomina ahora como fuerza de trabajo. La definición por su lado es similar: comprende a las personas en edad de trabajar que se encuentran ocupadas o que buscan activamente trabajo y están disponibles para empezar a laborar. En consecuencia, se desagrega en población ocupada y población desocupada (DANE, 2026b). A partir de ahora, se tomará solo el término de fuerza de trabajo (FT).

$$\text{Fuerza de trabajo} = \text{Población ocupada} + \text{Población desocupada}$$

4. Población ocupada (PO)

Según DANE (2021), para el marco 2005, una persona se encuentra empleada cuando durante en el periodo de referencia se encontraba en alguna de las siguientes situaciones: (i) trabajó al menos una hora remunerada; (ii) no trabajó en el periodo de referencia, pero tenía un trabajo; y (iii) aquellos trabajadores sin remuneración que trabajaron al menos una hora en la semana de referencia. A continuación, el siguiente diagrama de flujo establece el camino a seguir hasta clasificar la población empleada:

Figura 2. Diagrama de flujo para la clasificación de la población con empleo para el marco 2005.



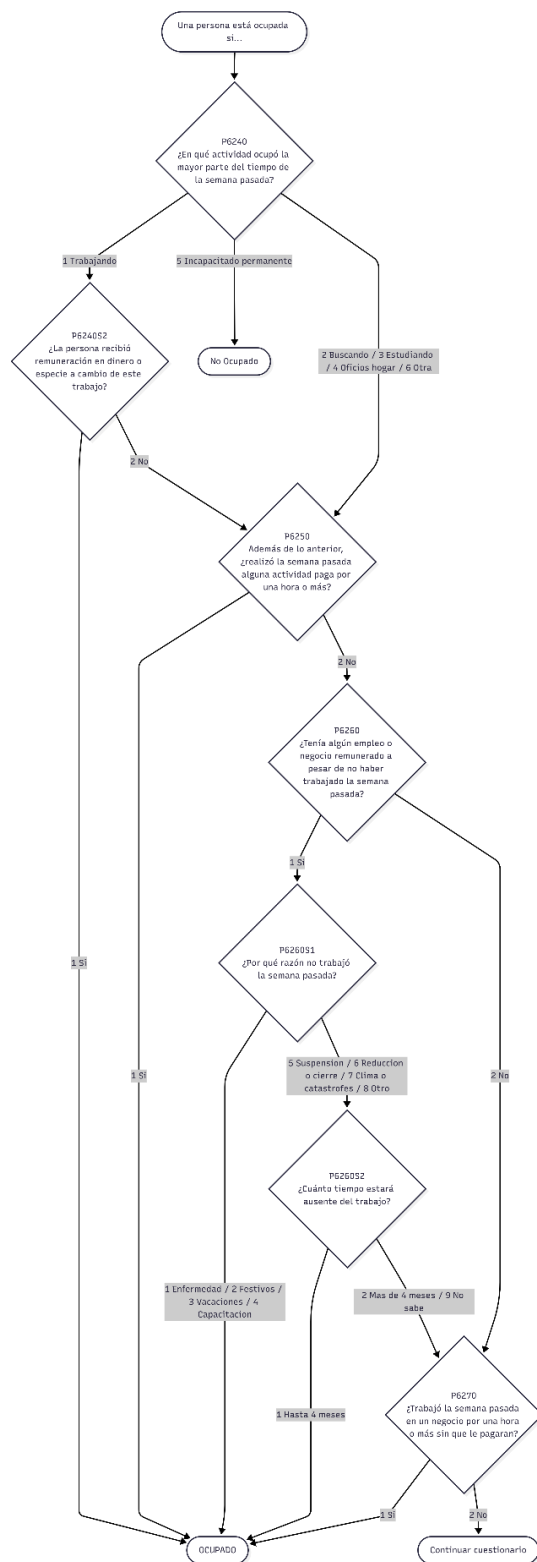
Fuente: Elaboración propia a partir de DANE (2019).

De igual forma, en la actualización conceptual para el marco 2018, a través de la inclusión de las recomendaciones de la CIET 19 (International Labour Office, 2013), se cambian las terminologías de empleo a ocupación, y de población con empleo a población ocupada. Este cambio se materializa al incluir tres preguntas clave en el módulo de fuerza de trabajo. El objetivo de ellas es, alcanzar una mayor precisión en la medición del fenómeno laboral.

En este nuevo enfoque, la precisión del concepto recae en identificar la ocupación como aquel trabajo realizado estrictamente a cambio de una remuneración. Esto permite un seguimiento más detallado y técnico de la actividad económica. Y de forma complementaria, este marco robustece y permite una identificación de los *ocupados ausentes* al enfatizar en la clasificación de las ausencias por razones no tradicionales y en el análisis de la duración de la ausencia. Lo anterior garantiza que la estadística refleje con mayor fidelidad el vínculo real de las personas con el mercado laboral (DANE, 2022b).

Ahora, el DANE (2026b) define la población ocupada como el conjunto de individuos de quince años o más que mantienen una participación en el mercado laboral. En este criterio se contabiliza a aquellos que hayan trabajado de forma remunerada al menos una hora durante el periodo de referencia. De igual forma, se tiene en cuenta a aquellos que, a pesar de estar ausentes temporalmente, conservan un vínculo firme con su empleo. Por último, se integra a las personas que realizaron labores sin remuneración por lo menos una hora. En el siguiente diagrama de flujo, se presenta el criterio de elección para integrar la población ocupada bajo el marco 2018:

Figura 3. Diagrama de flujo para la clasificación de la población ocupada para el marco 2018.



Fuente: Elaboración propia a partir de DANE (2024).

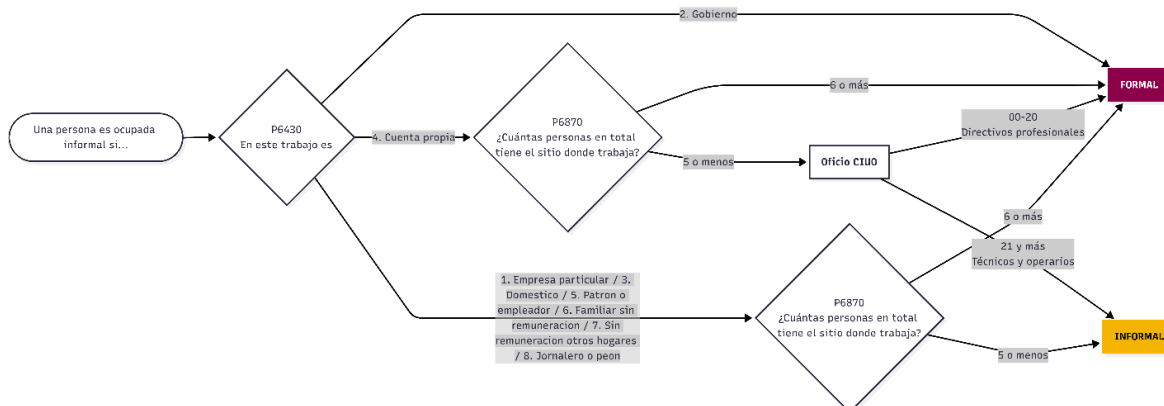
5. Población ocupada informal (POI)

La conceptualización de la informalidad sigue un recorrido bastante interesante. A partir del artículo *Informal Income Opportunities and Urban Employment in Ghana*, de Hart (1973) se esboza una definición inicial de la informalidad: aquellas actividades económicas autoorganizadas y no reguladas, que funcionan como una estrategia de supervivencia ante el subempleo. A partir de ello, la OIT desarrolla una definición estadística para la medición de la ocupación en el sector informal (International Labour Office, 1993). Esta misma, es implementada en la **GEIH-M05**. De esta forma, el DANE (2014) define la ocupación informal como aquellas personas que en el periodo de referencia se encontraban en alguna de las siguiente ocho situaciones:

- Empleados particulares y obreros que trabajan en establecimientos, negocios o empresas que ocupen hasta cinco personas en todas sus agencias y sucursales, incluido el patrono y/o socio;
- Trabajadores familiares sin remuneración en empresas de cinco trabajadores o menos;
- Trabajadores sin remuneración en empresas o negocios de otros hogares;
- Empleados domésticos en empresas de cinco trabajadores o menos;
- Jornaleros o peones en empresas de cinco trabajadores o menos;
- Trabajadores por cuenta propia que laboran en establecimientos de hasta cinco personas;
- Trabajadores por cuenta propia que laboran en establecimientos hasta cinco personas, excepto los independientes profesionales;
- Patronos o empleadores en empresas de cinco trabajadores o menos;
- Se excluyen los obreros o empleados del gobierno.

En el siguiente diagrama, se esboza la estructura algorítmica para establecer si una persona empleada es formal o informal:

Figura 4. Diagrama de flujo para determinar la población informal en el marco 2005.



Fuente: Elaboración propia a partir de DANE (2014) y DIMPE & DANE (2016).

En cuanto a la actualización conceptual para el marco 2018, se introducen grandes cambios. Al utilizar las *Directrices sobre una definición estadística de la ocupación informal* de la CIET 17 (International Labour Office, 2003) se amplía el marco conceptual del empleo informal. Es decir, se incluye el trabajo informal tanto dentro como fuera del sector informal, esto al basarse en la falta de protección social. En la definición anterior, de la CIET 15 (1993), la definición tenía un único enfoque: el sector informal basado en características de la empresa, como el registro y el tamaño.

El DANE (2022b) introduce nuevas preguntas para identificar asalariados e independientes con y sin negocio. Lo anterior se realiza con el fin de garantizar variables para la aplicación de la definición del empleo informal en cuanto a empleo en el sector informal, empleo informal y empleo informal fuera del sector informal. De esta forma, la medición de la informalidad en Colombia se realiza a través de dos enfoques: (i) la unidad de producción, referente al sector informal; y (ii) el puesto de trabajo, referente al empleo informal.

El primer enfoque se define en términos de la naturaleza jurídica y otras características relacionadas a la empresa. Lo anterior se mide por medio del registro mercantil, el RUT, la personería jurídica, la contabilidad y la separación de gastos. Ahora, el segundo enfoque se define en términos de situación en el empleo y de otras características relacionadas con el empleo. Se cuantifica a través de la afiliación a salud, pensión, ARL y caja de compensación, de derechos prestacionales y de la formalidad de la empresa en la que trabaja.

Para evidenciar lo anterior, se utiliza un dispositivo analítico útil: la matriz de Hussmanns adaptada a Colombia. Esta matriz integra los dos enfoques complementarios para estudiar la informalidad laboral: unidad productiva o sector informal y puesto de trabajo o empleo informal. En su estructura la matriz cruza, por un lado, el tipo de unidad en la que se genera el empleo (empresas del sector formal, empresas del sector informal y hogares como

empleadores), y por otro, la categoría ocupacional (trabajadores por cuenta propia, empleadores familiares no remunerados, asalariados y servicio doméstico). Esto permite distinguir entre puestos formales e informales, y, además, se evita una equivalencia que simplifique la relación entre empleo informal y empresa informal. Permite también observar que el fenómeno puede originarse tanto en la estructura productiva como en las condiciones institucionales y contractuales de los puestos de trabajo.

Desde esta mirada, la matriz vislumbra la heterogeneidad del fenómeno de la informalidad. En tanto que es susceptible a un diagnóstico más elaborado que permita una tasa agregada más precisa. Esto se logra en dos pasos. En primer lugar, se identifica el núcleo clásico de la informalidad, es decir, los puestos informales dentro del sector informal que estén asociados a unidades productivas de baja escala y productividad, restricciones de financiamiento, capacidades administrativas limitadas y altos costos al incurrir en el cumplimiento de las normas. En segundo lugar, separa la informalidad dentro del sector formal. Esto es, puestos informales que existen en empresas que fueron clasificadas como formales. Lo anterior revela que la informalidad no solo depende exclusivamente de la condición empresarial sino también de prácticas de contratación, tercerización, incumplimiento de normas laborales y limitaciones de fiscalización. Además, esta matriz permite combinaciones intermedias (formalidad dentro del sector informal) que reflejan procesos de formalización “parcial”. Por último, se distingue la informalidad vinculada a hogares como empleadores, relevante sobre todo para el servicio doméstico, en el cual prevalecen arreglos privados con barreras específicas para el aseguramiento y el control institucional. La siguiente figura permite una visualización de lo explicado.

Figura 5. Matriz de Hussmanns adaptada a Colombia.

Unidades por tipo de producción	Tipo de trabajadores según categoría en la ocupación								
	Trabajadores por Cuenta Propia		Empleadores		Familiares no Remunerados del hogar	Asalariados		Trabajadores de Servicio Doméstico	
	Informal	Formal	Informal	Formal	Informal	Informal	Formal	Informal	Formal
Empresas del sector formal		a		b	1	2	c		
Empresas del sector informal	3		4		5	6	7		
Hogares como empleadores	d					8	e	9	f

■ Empleos que por definición no existen

■ Empleos formales

Celdas en blanco: Empleos informales

1. Empleo informal: Suma de las casillas 1 a 6 y de 8 y 9

2. Empleo en el sector informal: Suma de las casillas 3 a 7

3. Empleo informal fuera del sector informal: Suma de casillas 1,2,8 y 9

4. Empleo formal: suma de las casillas a,b,c,e y f

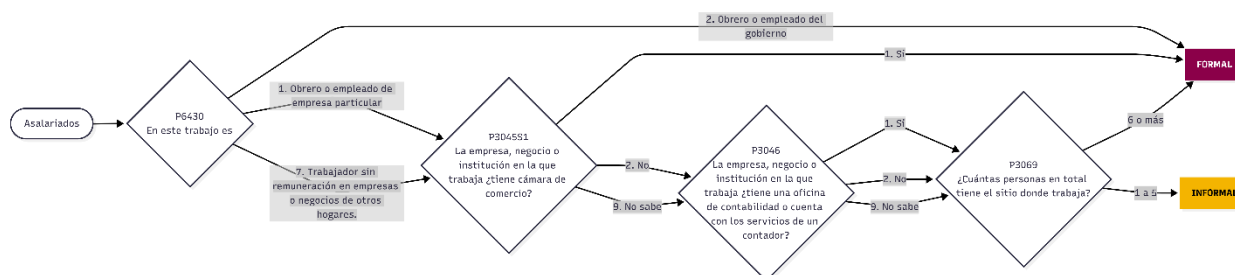
Fuente: Tomada de (DANE, 2022b).

Con esto en mente, el siguiente paso es definir la población ocupada informal para el marco 2018. El sector informal comprende todas las empresas no registradas ante las autoridades competentes que, al carecer de una contabilidad completa o simplificada, no pueden separar efectivamente sus gastos de los del hogar ni clasificarse como cuasisociedades. Estas unidades operan generalmente a pequeña escala con una organización rudimentaria, en la que la distinción entre trabajo y capital como factores de producción es escasa o inexistente; quedan excluidas las entidades de gobierno y las organizaciones sin fines de lucro.

A partir de esta definición, la población ocupada informal incluye: los asalariados y empleados domésticos sin cotizaciones al sistema de salud ni a pensión derivadas de su vínculo laboral; los trabajadores por cuenta propia y los patrones o empleadores clasificados en el sector informal; los trabajadores familiares sin remuneración; y las personas sobre quienes no se dispone de información suficiente para ubicarlas en alguna de las categorías anteriores. Se excluyen los obreros y empleados del gobierno (DANE, 2026a).

Ahora, se aborda cada uno de los elementos de las definiciones presentadas. Esta misión se realizará a través de diagramas de flujo de decisiones algorítmicas. Primero, se abordarán los diagramas por sectores: asalariados, independientes sin negociación, independientes con negocio y con registro mercantil e independientes con negocio y sin registro mercantil. luego, se abordarán los diagramas por ocupación: ocupación informal a través de salud y pensión.

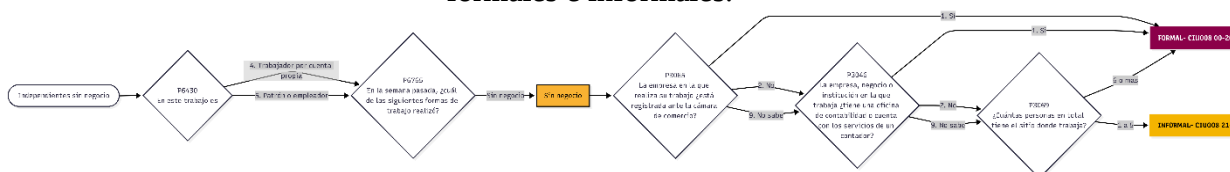
Figura 6. Diagrama de flujo para determinar los asalariados formales e informales.



Fuente: Elaboración propia a partir de DANE (2022c).

El árbol de decisión de los asalariados sugiere una operacionalización de la informalidad para este rubro basándose en señales institucionales del empleador como la presencia de la cámara de comercio y servicios contables. En ausencia de estas, se dirige el enfoque al tamaño del sitio de trabajo.

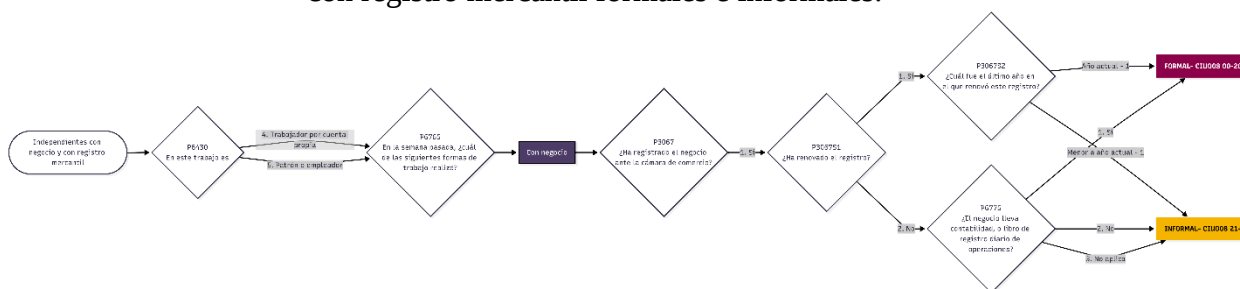
Figura 7. Diagrama de flujo para determinar los trabajadores independientes sin negocio formales e informales.



Fuente: Elaboración propia a partir de DANE (2022c).

En cuanto a los independientes sin negocio, la formalidad se define de manera jerárquica. Primero, se alude al registro ante la cámara de comercio sobre la unidad donde se realiza la actividad. Si esta no se cumple o no se conoce, se realiza por la disponibilidad de servicios contables. En última instancia, se utiliza el tamaño del sitio de trabajo. A través de esta lógica, el sector formal se asocia con registro, soporte contable o con establecimientos de seis personas o más. En cambio, el sector informal se concentra en unidades pequeñas de una y cinco personas cuando no se observan los marcadores institucionales previos.

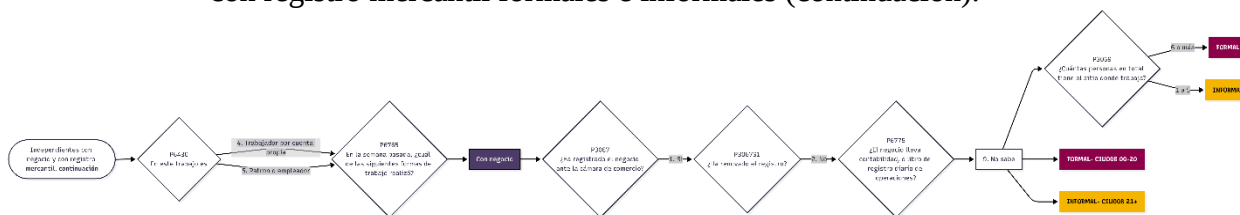
Figura 8. Diagrama de flujo para determinar los trabajadores independientes con negocio y con registro mercantil formales e informales.



Fuente: Elaboración propia a partir de DANE (2022c).

Ahora, para los independientes con negocio y con registro mercantil, el criterio operativo del sector formal se apoya en señales de institucionalización y trazabilidad administrativa. Es decir, la clasificación a formal se activa cuando el registro mercantil aparece renovado de forma reciente según el umbral (año anterior al actual). Si no hay renovación, se observa si el negocio lleva contabilidad o un libro de registro diario de operaciones. En contraste, el sector informal se asocia a escenarios en los cuales la última renovación es anterior al umbral establecido y donde no se llevan dichos registros contables.

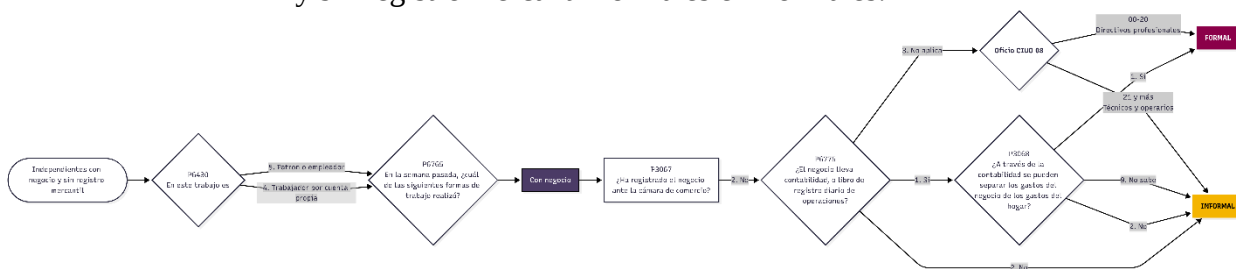
Figura 9. Diagrama de flujo para determinar los trabajadores independientes con negocio y con registro mercantil formales e informales (continuación).



Fuente: Elaboración propia a partir de DANE (2022c).

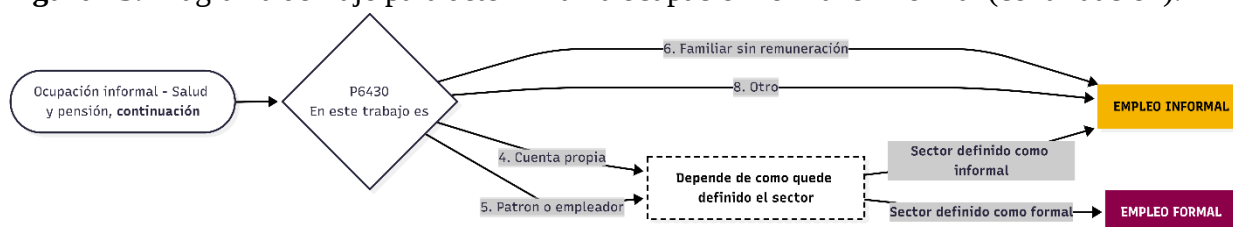
De forma análoga, como continuación del diagrama anterior, esta figura muestra que, dentro del grupo de independientes con negocio y registro mercantil, la clasificación del sector formal se condiciona a la vigencia del registro. Esto es, ante ausencia de renovación, se recurre a la existencia de soportes administrativos como la contabilidad o un libro diario de operaciones. Cuando esta información no está disponible o no se tiene conocimiento de ella, se incorpora una clasificación asociada a los rangos de la actividad económica (CIUO 08) y un criterio de tamaño del sitio de trabajo. La primera, se integra para diferenciar a los directivos profesionales (sector formal) de los técnicos y operarios (sector informal). El segundo, aborda empresas de seis o más personas como sector formal, y de una a cinco personas como informal.

Figura 10. Diagrama de flujo para determinar los trabajadores independientes con negocio y sin registro mercantil formales e informales.



Fuente: Elaboración propia a partir de DANE (2022c).

Como continuación, esta figura operacionaliza el sector formal a partir de una secuencia de filtros administrativos. Primero, se tiene en cuenta que hay una ausencia del registro mercantil que conduce a la evaluación de la existencia de registros contables o un libro diario de operaciones. Cuando estos están presentes, se introduce un criterio adicional de separación contable entre los gastos del negocio y los del hogar. Si tal separación se confirma, la clasificación se apoya luego en un criterio de actividad, a través de la clasificación internacional uniforme de oficios (CIUO 08), que distingue entre directivos profesionales y técnicos y operarios como umbrales de asignación. Este diagrama sugiere algo interesante: aun sin registro mercantil, la clasificación al sector formal se puede aproximar mediante la capacidad de registro y contabilidad. Además, en esta ruta el componente ocupacional o de actividad se usa como un mecanismo de refinamiento de la clasificación.

Figura 13. Diagrama de flujo para determinar la ocupación formal e informal (continuación).

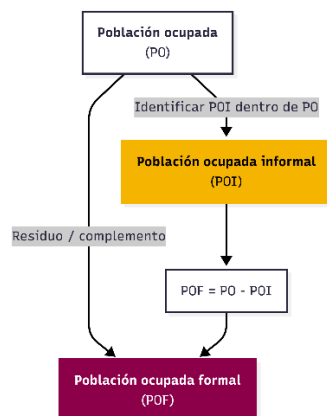
Fuente: Elaboración propia a partir de DANE (2022c).

Finalmente, para trabajadores por cuenta propia, patrones o empleadores, trabajadores familiares sin remuneración y otros, se utiliza un criterio diferente. Para familiares sin remuneración y para otros, se clasifican como empleados informales directamente. En cambio, para los trabajadores por cuenta propia y patrones o empleadores, se utiliza un criterio intermedio que se basa en la definición del sector formal o informal. De este modo, el resultado final depende de cómo haya quedado clasificado en ese sector.

De esta manera, la actualización metodológica y conceptual en cuanto a la definición de la población ocupada informal, se ha visualizado como un cambio importante que ayuda a encontrar más personas que hacen parte de este fenómeno.

6. Población ocupada formal (POF)

En las anteriores secciones se ha definido la población ocupada y la población ocupada informal. Estos dos conceptos son fundamentales para definir la población ocupada formal. Tanto para el marco 2005 como para el marco 2018, en este documento técnico se define a la población ocupada formal como aquella que resulta luego de clasificar la población ocupada informal dentro del total de ocupados. De esta forma, es válido afirmar que ésta resulta de forma residual. De forma operativa, el DANE lo realiza de esa forma a través del código en STATA/SAS que comparte en su página web.

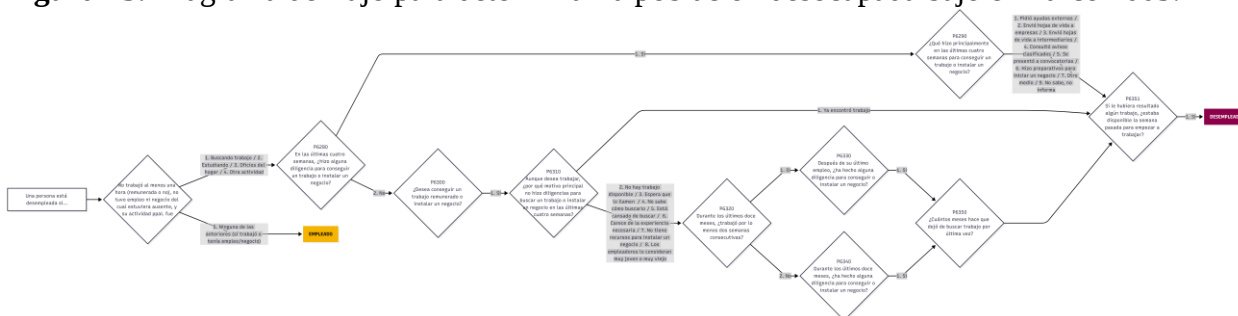
Figura 14. Diagrama de flujo para determinar la población ocupada formal.

Fuente: Elaboración propia a partir de DANE (2022c).

7. Población desempleada o desocupada (PDSO)

Para el marco 2005, la población desempleada se identifica a partir de dos conceptos fundamentales: el desempleo abierto y el desempleo oculto. En ambos casos se usa un periodo de referencia en el cual se cumpla alguna de las condiciones establecidas. En primer lugar, se considera desempleo abierto cuando una persona: (i) no tuvo empleo en la semana de referencia, (ii) realizó diligencias para conseguir empleo en el último mes y (iii) estaba disponible para trabajar. En segundo lugar, se define el desempleo oculto de forma similar. La diferencia es que la segunda condición cambia: la persona no realizó diligencias en el último mes, pero si en los últimos doce meses, y además presenta una razón válida de desaliento (DANE, 2022d). En el siguiente diagrama se observan las preguntas que llevan a la clasificación de este grupo poblacional.

Figura 15. Diagrama de flujo para determinar la población desocupada bajo el marco 2005.



Fuente: Elaboración propia a partir de (DANE, 2022d).

Este diagrama resume las primeras condiciones. Son las mismas condiciones para clasificar a una persona como empleada, con la diferencia de que las respuestas necesarias para este caso son todas las contrarias.

En cuanto al marco 2018, se mantiene el mismo enfoque y se utilizan las mismas preguntas. Solo que, en este caso, la definición operativa se modifica y se suprimen los términos de desempleo abierto y desempleo oculto. Por lo tanto, la población desocupada está conformada por las personas de quince años o más que en el periodo de referencia estuvieron simultáneamente: (i) sin empleo, es decir, que no tenían un empleo asalariado o un trabajo independiente ni se desempeñaron como trabajador familiar sin remuneración; (ii) buscando empleo, es decir que habían tomado medidas concretas para buscar un empleo asalariado o independiente en las últimas cuatro semanas; y (iii) actualmente disponibles para trabajar, esto es, disponibles para trabajar en empleo asalariado o empleo independiente en el periodo de referencia (DANE, 2026b). Además, conforme a las recomendaciones propuestas en la *Resolución I sobre las estadísticas del trabajo, la ocupación y la subutilización de la fuerza de trabajo*, de la 19 CIET (International Labour Office, 2013), se ajustan los términos desocupados por población desocupada. Para este marco, la anterior figura sirve también como referencia.

V. Construcción del diccionario maestro

Se organizan las variables en dos grandes categorías: (i) *poblacional* y (ii) *ingresos*. La nomenclatura de las variables se construye a través de un *prefijo*, referente a la familia conceptual, y un *sufijo*, que evidencia la versión de la serie. De esta forma, la nomenclatura de cada variable es la combinación de:

$$\text{variable} = \text{familia conceptual} + \text{versión de la serie}$$

1. Convención de nombres (prefijos y sufijos)

Los prefijos hacen referencia a las familias conceptuales. Estas últimas se dividen en dos:

- **Categoría poblacional:** PO, POI, POF, PDSO, FT, PET, PT.
- **Categoría de ingresos:** INGLABO_PO, INGLABO_POI, INGLABO_POF.

En cuanto a los sufijos, estos aluden a la versión de la serie. En la siguiente tabla se resume el significado:

Sufijo	Descripción operativa
_05	Serie construida bajo GEIH-M05 .
_18	Serie construida bajo GEIH-M18 .
_emp	Serie construida bajo GEIH-EMP .
_final	Serie final, construida bajo el método de empalme definido en este documento.
_p	Serie preliminar que necesita ser ajustada.
_final_sa	Serie final ajustada estacionalmente bajo el método X-13ARIMA-SEATS .

Tabla 1: Leyenda de sufijos (versión de las series). Elaboración propia.

2. Categoría poblacional

En esta categoría se agrupan las familias conceptuales asociadas a agregados poblacionales. Para cada familia se reportan las versiones disponibles, esto con el fin de estandarizar la lectura del diccionario maestro.

2.1. Familia PO

Nombre largo: Población ocupada (PO).

Unidad: Personas.

Variable	Versión
PO_05	<i>GEIH-M05.</i>
PO_18	<i>GEIH-M18.</i>
PO_emp	<i>GEIH-EMP.</i>
PO_final	Final, empalme propuesto.
PO_final_sa	Final y ajuste estacional.

Tabla 2: Estados disponibles para PO. Elaboración propia.

2.2. Familia POI

Nombre largo: Población ocupada informal (POI).

Unidad: Personas.

Variable	Versión
POI_05	<i>GEIH-M05.</i>
POI_18	<i>GEIH-M18.</i>
POI_emp	<i>GEIH-EMP.</i>
POI_p	Preliminar, sin ajuste de imputación de valores faltantes con modelos SARIMA.
Pinf	Proporción de informalidad. Variable necesaria para el ajuste de imputación de valores faltantes con modelos SARIMA:
POI_final	Final, empalme propuesto.
POI_final_sa	Final y ajuste estacional.

Tabla 3: Estados disponibles para POI. Elaboración propia.

2.3. Familia POF

Nombre largo: Población ocupada formal (POF).

Unidad: Personas.

Variable	Versión
POF_05	<i>GEIH-M05.</i>
POF_18	<i>GEIH-M18.</i>
POF_emp	<i>GEIH-EMP.</i>
Pfor	Residual: $Pfor = 1 - Pinf^{ajustada}$
POF_final	Final, empalme y ajuste propuesto.
POF_final_sa	Final y ajuste estacional.

Tabla 4: Estados disponibles para POF. Elaboración propia.

2.4. Familia PDSO

Nombre largo: Población desocupada (PDSO).

Unidad: Personas.

Variable	Versión
PDSO_05	<i>GEIH-M05.</i>
PDSO_18	<i>GEIH-M18.</i>
PDSO_emp	<i>GEIH-EMP.</i>
PDSO_final	Final, empalme propuesto.
PDSO_final_sa	Final y ajuste estacional.

Tabla 5: Estados disponibles para PDSO. Elaboración propia.

2.5. Familia FT

Nombre largo: Fuerza de trabajo (FT).

Unidad: Personas.

Variable	Versión
FT_05	<i>GEIH-M05.</i>
FT_18	<i>GEIH-M18.</i>
FT_emp	<i>GEIH-EMP.</i>
FT_final	Final, empalme propuesto.
FT_final_sa	Final y ajuste estacional.

Tabla 6: Estados disponibles para FT. Elaboración propia.

2.6. Familia PET

Nombre largo: Población en edad de trabajar (PET).

Unidad: Personas.

Variable	Versión
PET_05	<i>GEIH-M05.</i>
PET_18	<i>GEIH-M18.</i>
PET_emp	<i>GEIH-EMP.</i>
PET_final	Final, empalme propuesto.
PET_final_sa	Final y ajuste estacional.

Tabla 7: Estados disponibles para PET. Elaboración propia.

2.7. Familia PT

Nombre largo: Población total (PT).

Unidad: Personas.

Variable	Versión
PT_05	<i>GEIH-M05.</i>
PT_18	<i>GEIH-M18.</i>
PT_emp	<i>GEIH-EMP.</i>
PT_final	Final, empalme propuesto.
PT_final_sa	Final y ajuste estacional.

Tabla 8: Estados disponibles para PT. Elaboración propia.

3. Categoría de ingresos

En esta categoría se agrupan las variables asociadas a los ingresos laborales definidas por grupo poblacional. Para cada familia, se listan las versiones disponibles.

3.1. Familia INGLABO_PO

Nombre largo: Ingresos laborales de la población ocupada (INGLABO_PO).

Unidad: Pesos colombianos (COP).

Variable	Versión
INGLABO_PO_05	<i>GEIH-M05.</i>
INGLABO_PO_18	<i>GEIH-M18.</i>
INGLABO_PO_emp	<i>GEIH-EMP.</i>
INGLABO_PO_p	Preliminar, sin ajuste de imputación de valores faltantes con modelos SARIMA.
INGLABO_PO_final	Final, empalme propuesto.
INGLABO_PO_final_sa	Final y ajuste estacional.

Tabla 9: Estados disponibles para INGLABO_PO. Elaboración propia.

3.2. Familia INGLABO_POI

Nombre largo: Ingresos laborales de la población ocupada informal (INGLABO_POI).

Unidad: Pesos colombianos (COP).

Variable	Versión
INGLABO_POI_05	<i>GEIH-M05.</i>
INGLABO_POI_18	<i>GEIH-M18.</i>
INGLABO_POI_emp	<i>GEIH-EMP.</i>
INGLABO_POI_p	Preliminar, sin ajuste de imputación de valores faltantes con modelos SARIMA.
INGLABO_POI_final	Final, empalme propuesto.
INGLABO_POI_final_sa	Final y ajuste estacional.

Tabla 10: Estados disponibles para INGLABO_POI. Elaboración propia.

3.2. Familia INGLABO_POF

Nombre largo: Ingresos laborales de la población ocupada formal (INGLABO_POF).

Unidad: Pesos colombianos (COP).

Variable	Versión
INGLABO_POF_05	<i>GEIH-M05.</i>
INGLABO_POF_18	<i>GEIH-M18.</i>
INGLABO_POF_emp	<i>GEIH-EMP.</i>

INGLABO_POF_p	Preliminar, sin ajuste de imputación de valores faltantes con modelos SARIMA.
INGLABO_POF_final	Final, empalme propuesto.
INGLABO_POF_final_sa	Final y ajuste estacional.

Tabla 11: Estados disponibles para INGLABO_POF.

VI. Procedimiento operativo para la construcción de la Serie Trimestral del Mercado Laboral Colombiano: El camino a la unificación total, una serie única y consistente

En esta sección se presenta una propuesta metodológica que, a partir del sustento teórico expuesto anteriormente, sea consistente para la construcción de series de tiempo únicas, consistentes y que arrojen los cambios metodológicos y conceptuales que los años han presentado. Este proceso se divide en dos vertientes. Inicialmente, este se propone para los grupos poblacionales. Luego, otro muy similar para los ingresos laborales, que, como nos daremos cuenta a continuación, necesitan de un proceso mucho más riguroso en la búsqueda de la máxima transparencia.

Para los grupos poblacionales, se proponen dos pasos. El primero surge de una pregunta interesante: *¿cómo se juntan las series de la GEIH-M05 con las series de la GEIH-EMP?* La respuesta lo es aún más. Se dejan los valores de la GEIH-EMP desde el primer trimestre de 2010 al último trimestre de la muestra, en este caso, hasta el último trimestre del 2025. Luego, para las observaciones que faltan, es decir, desde el primer trimestre de 2007 hasta el último trimestre del 2009, se propone una fórmula de empalme basada en tasas de crecimiento. A continuación, se muestra el procedimiento paso a paso.

Sea $y = year$ que representa el año de las observaciones y $q = quarter$ que representa los trimestres, se definen como:

$$y = 2007, \dots, 2025$$

$$q = 1, 2, 3, 4$$

Esto es, datos trimestrales. Sea además $\alpha = grupo\ poblacional\ de\ aplicación$, la cual puede ser cualquiera de las siete planteadas, y los sufijos expuestos en el diccionario maestro, la fórmula de este paso es:

$$\alpha_{05_emp_{y,q}} = \frac{\alpha_{emp_{y+1,q}}}{\left(\frac{\alpha_{05_{y+1,q}}}{\alpha_{05_{y,q}}} \right)}$$

para $y < 2010$

$$\alpha_{05_emp_{y,q}} = \alpha_{emp_{y,q}}$$

para $y \geq 2010$

De este proceso resultan las variables $\alpha_{05_emp_{y,q}}$. Estas deben tener observaciones desde el primer trimestre de 2007 al último trimestre del 2025. Con lo anterior, se procede con el segundo paso para llegar a la variable final. Esto es, se deben juntar las series resultantes en el paso uno con las que residen en la GEIH-M18. En otras palabras, se necesita una fórmula para unificar $\alpha_{05_emp_{y,q}}$ con $\alpha_{18_{y,q}}$. De esta forma, de una forma muy similar, y con la notación propuesta, la fórmula está definida por:

$$\alpha_{final_{y,q}} = \frac{\alpha_{18_{y+1,q}}}{\left(\frac{\alpha_{05_emp_{y+1,q}}}{\alpha_{05_emp_{y,q}}} \right)}$$

para $y < 2022$

$$\alpha_{final_{y,q}} = \alpha_{18_{y,q}}$$

para $y \geq 2022$

Con este proceso, se consolida la base de datos final con siete grupos poblacionales. Sin embargo, variables como la población ocupada informal y la población ocupada formal necesitan de un tratamiento especial. En la siguiente sección se aborda de forma específica. Ahora, para los ingresos laborales de la población ocupada, ocupada formal y ocupada informal el procedimiento es similar. La única diferencia radica en el horizonte temporal de las series. En lugar de comenzar a partir del primer trimestre del 2007, estas comienzan en el primer trimestre del 2009. Sea $\beta = \text{variable de ingresos de aplicación}$. En concordancia con lo dicho, las fórmulas son exactamente las mismas:

$$\beta_{05_emp_{y,q}} = \frac{\alpha_{emp_{y+1,q}}}{\left(\frac{\alpha_{05_{y+1,q}}}{\alpha_{05_{y,q}}} \right)}$$

para $y < 2010$.

$$\beta_{05_emp_{y,q}} = \alpha_{emp_{y,q}}$$

para $y \geq 2010$.

Como resultado se obtiene $\beta_{05_emp_{y,q}}$. Estas variables tienen observaciones desde el primer trimestre del 2009 hasta el último trimestre del 2025. Ahora, para llegar a la variable final y capturar tanto los cambios metodológicos como los conceptuales, resulta necesario

utilizar las variables residentes en GEIH-M18. Es decir, unificar $\beta_{05_emp_{y,q}}$ con $\beta_{18_{y,q}}$. La fórmula propuesta es:

$$\beta_{final_{y,q}} = \frac{\beta_{18_{y+1,q}}}{\left(\frac{\beta_{05_emp_{y+1,q}}}{\beta_{05_emp_{y,q}}}\right)}$$

para $y < 2022$.

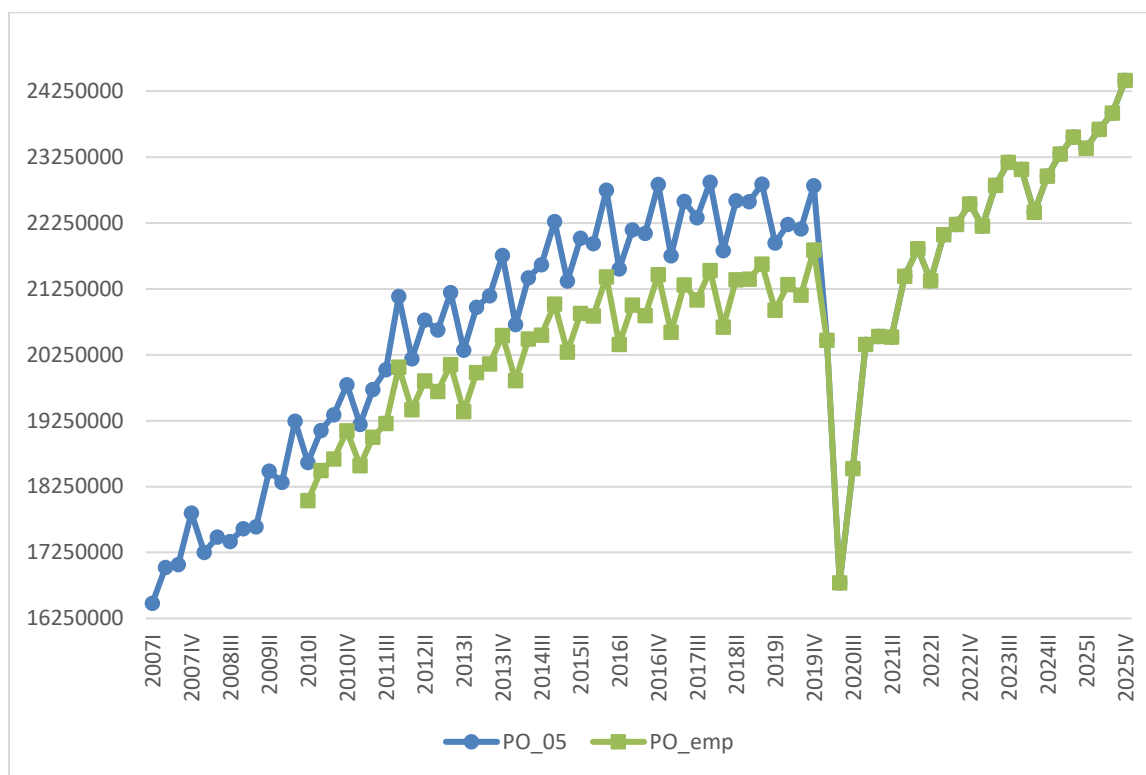
$$\beta_{final_{y,q}} = \beta_{18_{y,q}}$$

para $y \geq 2022$.

De esta forma se obtienen series unificadas y consistentes para los ingresos laborales de tres grupos poblacionales. Sin embargo, para los ingresos de formales e informales es necesario realizar un procedimiento extra que se expondrá en las siguientes subsecciones. En estas, que vienen a continuación, se presentan individualmente cada una de las variables. El objetivo de esto es abordar de forma concreta cada caso con sus características.

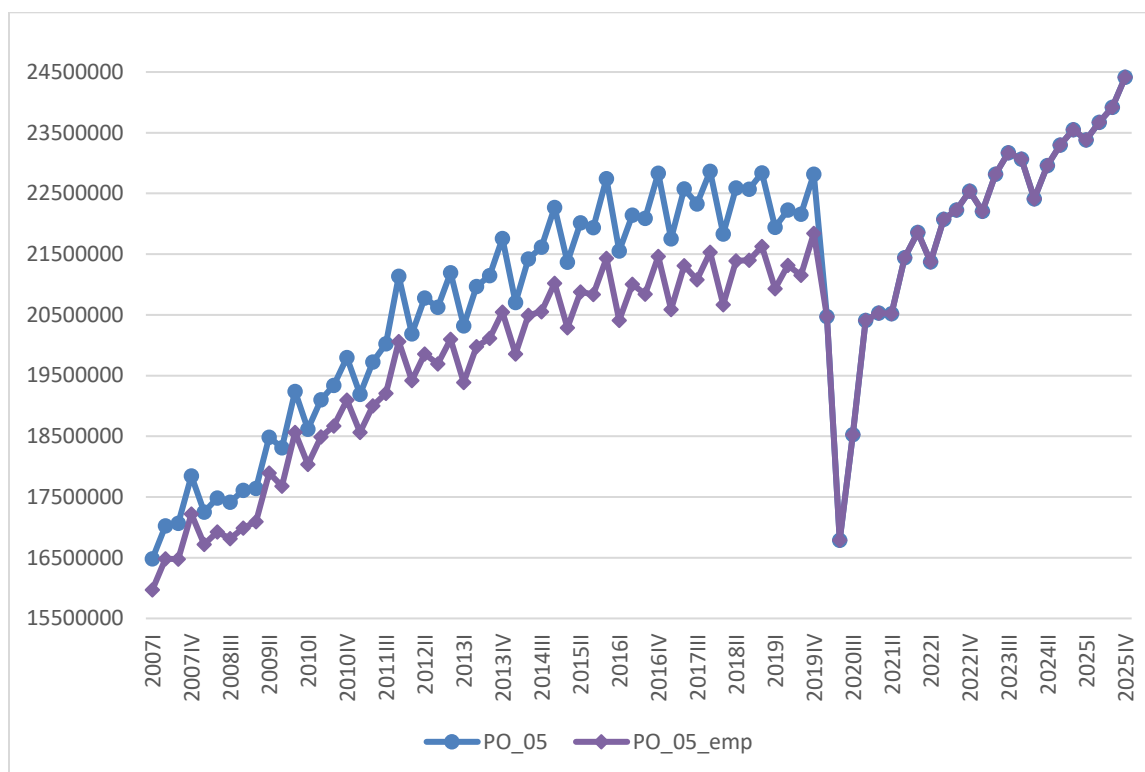
1. Población Ocupada

Se inicia con la comparación entre las variables PO_05 y PO_EMP, cuyas características se describen anteriormente. En la figura 16 se observa una brecha importante en los años que anteceden a la pandemia del COVID-19. Esta diferencia se explica, principalmente, por la reestimación del factor de expansión utilizada en las series construidas bajo el marco 2005. Es decir, aun cuando ambas variables responden a una misma definición operativa sobre población ocupada, las series iniciales del marco 2005 probablemente sobreestimaban sus niveles. No obstante, dada la disponibilidad de información, la serie de empalme únicamente puede construirse desde 2010.

Figura 16. Población Ocupada: M05 y EMP.

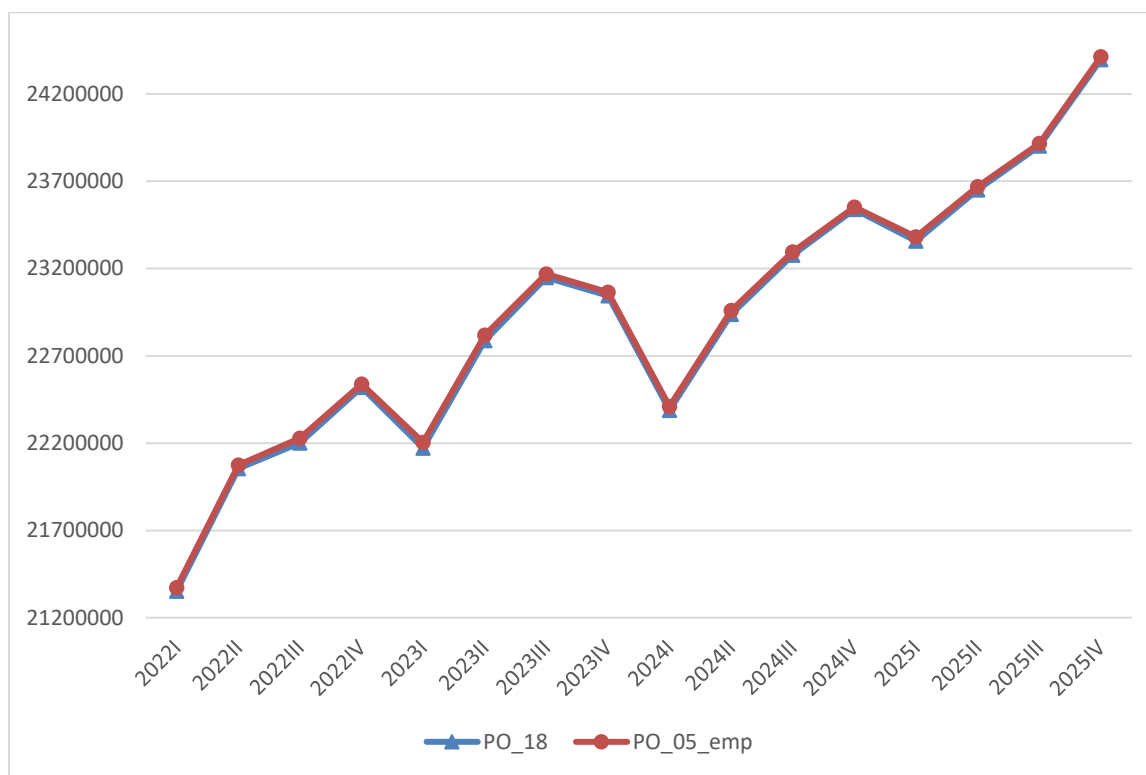
Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE.

A partir de este diagnóstico, la metodología propuesta integra las variables PO_05 y PO_EMP, dando lugar a la serie PO_05_EMP. Este primer empalme permite corregir la discontinuidad asociada al componente muestral, de modo que la nueva serie mantenga la trayectoria histórica de la población ocupada, pero con ajuste más consistente frente a las diferencias derivadas del cambio de marco. Sin embargo, este resultado aún no considera el cambio conceptual que se introdujo en el rediseño de la GEIH, el cual se encuentra reflejado en la variable PO_18.

Figura 17. Población Ocupada: M05 y M05 + EMP.

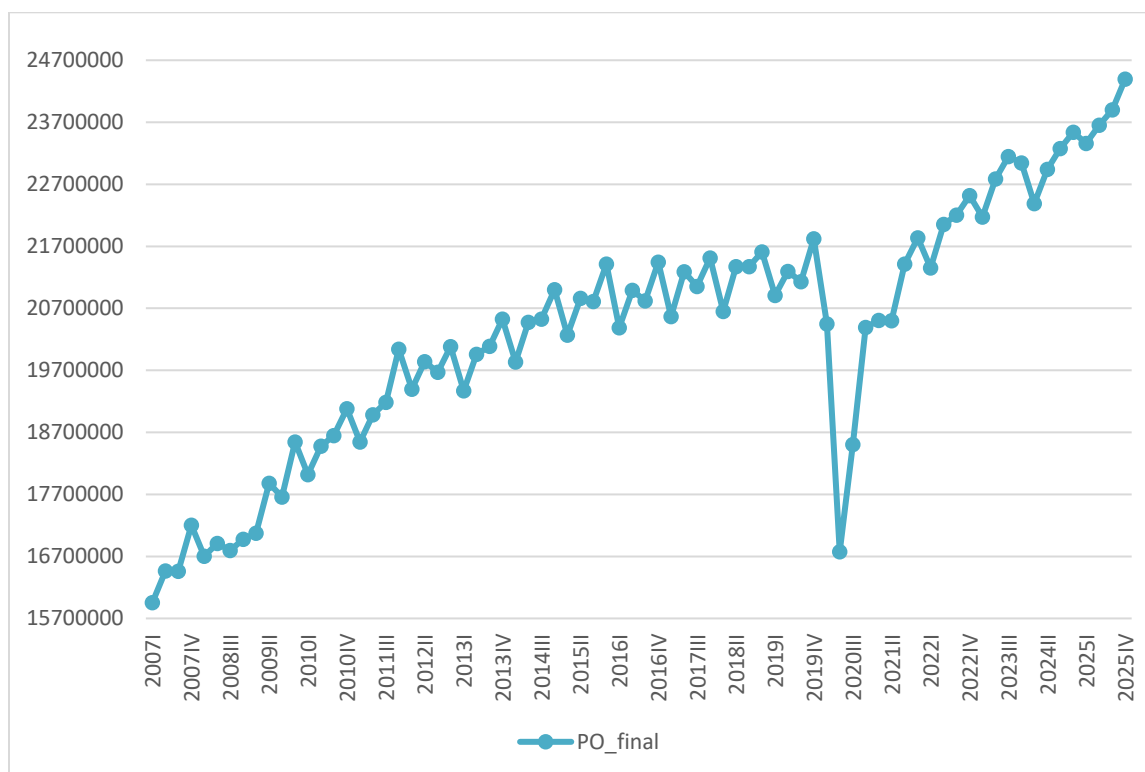
Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE

En consecuencia, el siguiente paso consiste en comparar la serie PO_05_EMP con PO_18. Esta segunda comparación es crucial porque permite identificar la ruptura que ya no se asocia al diseño muestral, sino a la nueva definición de clasificación de la población ocupada adoptada para el marco 2018. Un desafío es que PO_18 solo está disponible desde el primer trimestre de 2022, pero se resuelve a través de la fórmula expuesta anteriormente.

Figura 18. Población ocupada: M05 + EMP y M18.

Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE

De esta forma, y con base en la metodología de empalme propuesta, se integra PO_05_EMP con PO_18 para construir una serie final homogénea. Con ello, la variable resultante no solo corrige las diferencias originadas por el cambio de marco muestral, sino que también incorpora el nuevo criterio conceptual de clasificación de la población ocupada. Esta serie podría constituir una mejor aproximación disponible para estudiar la evolución de la población ocupada en el tiempo bajo un criterio estadísticamente comparable.

Figura 19. Población Ocupada: Variable final.

Fuente: Elaboración propia a partir de la propuesta metodológica.

Este procedimiento de empalme se desarrolla en dos etapas complementarias. La primera corrige el componente muestral, mediante la integración de PO_05 y PO_EMP. La segunda ajusta el componente conceptual, mediante la integración de PO_05_EMP y PO_18. Así, la variable final recoge simultáneamente los efectos del rediseño metodológico de la GEIH y ofrece una serie más consistente para el análisis de largo plazo.

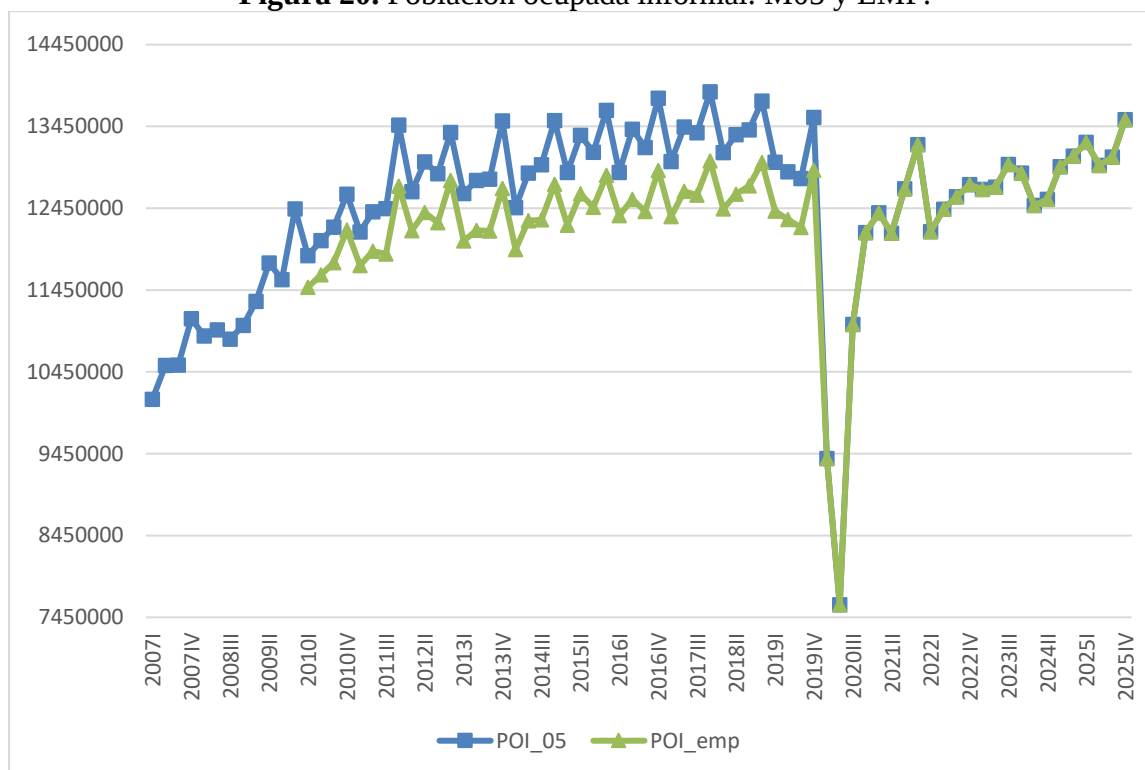
2. Población Ocupada Informal

Para construir la serie final de la población ocupada informal se toman en cuenta varias etapas. El propósito es corregir las diferencias asociadas al cambio muestral, conceptual y la ruptura observada en los primeros trimestres de la pandemia. Este procedimiento, por lo tanto, permitió obtener una serie homogénea y comparable en el tiempo adecuada para el análisis de la informalidad bajo distintos marcos metodológicos de la GEIH.

En primer lugar, se compararon las series POI_05 y POI_EMP, cuyas características se describieron anteriormente. Como se observa en la figura 20, existe una brecha sistemática entre ambas series en el periodo que antecede la pandemia. Esto se observa especialmente desde el 2010, que es el punto a partir del cual se dispone información para el empalme. Ahora, dado que ambas series conservan una misma lógica conceptual de medición de la informalidad, esta diferencia se asocia al componente muestral. En concreto, se refiere a la

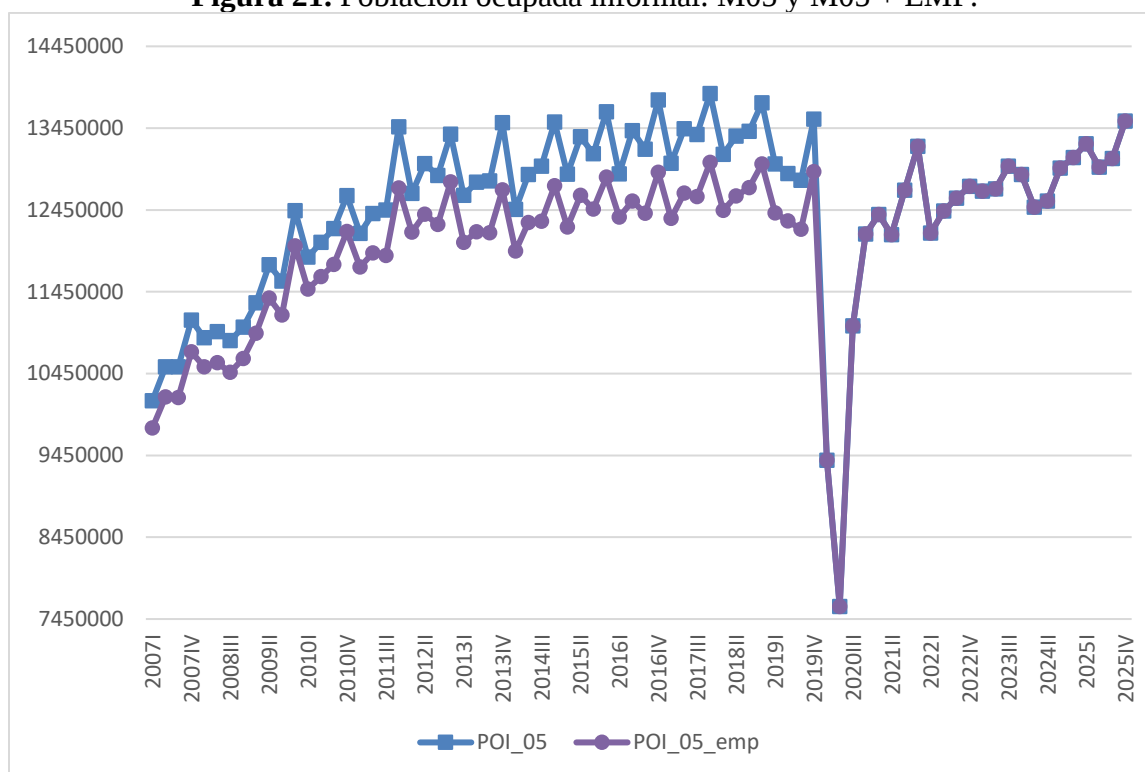
estimación del factor de expansión empleado en las series del marco 2005. La evidencia gráfica sugiere que POI_05 se ubica por encima de POI_EMP, esto podría indicar una posible sobreestimación de los niveles de informalidad en la anterior metodología.

Figura 20. Población ocupada informal: M05 y EMP.



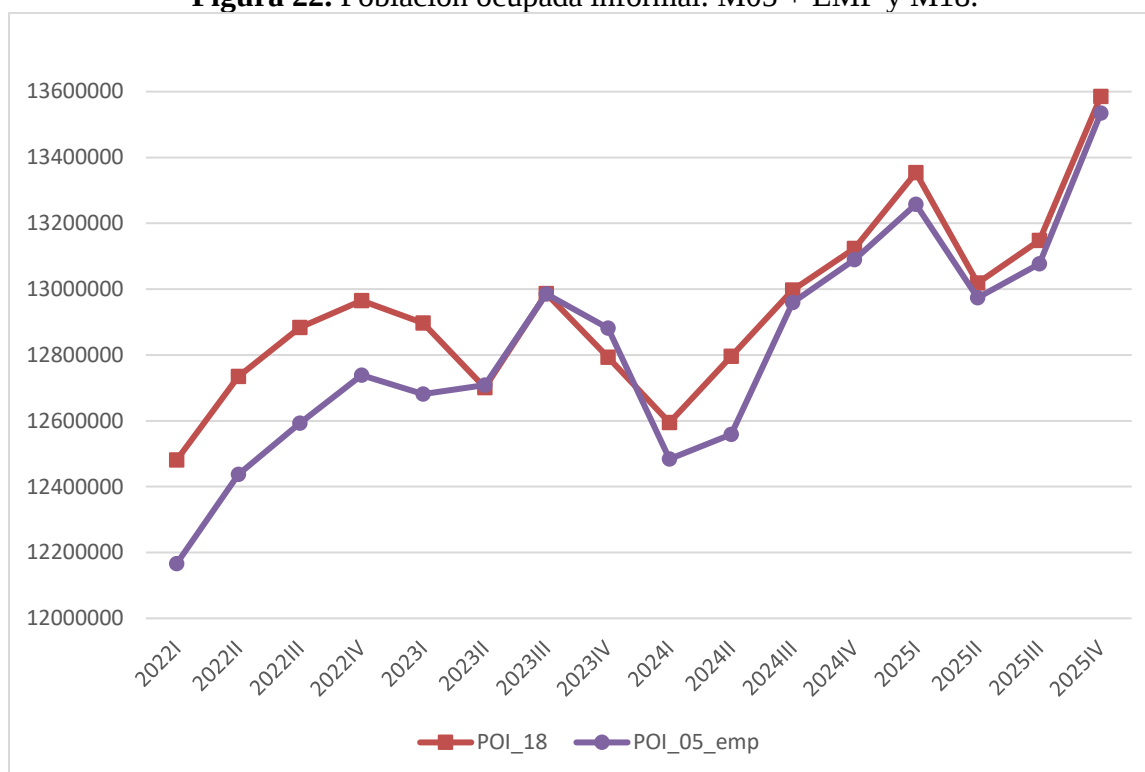
Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE.

Teniendo en cuenta la primera comparación, se integran las series POI_05 y POI_EMP para obtener POI_05_EMP. Este empalme permite corregir la discontinuidad asociada al cambio en el marco muestral y, de igual forma, conservar la trayectoria completa de la serie. Tal como se observa en la figura 21, la serie empalmada se sitúa por debajo de POI_05. Esto es consistente con la corrección de la brecha que se identificó anteriormente. Como consecuencia, esta primera armonización incorpora parte fundamental del ajuste metodológico. Sin embargo, aún no recoge el cambio conceptual introducido por el rediseño de la GEIH.

Figura 21. Población ocupada informal: M05 y M05 + EMP.

Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE.

Como siguiente paso, se incorpora el componente conceptual contenido en la serie POI_18. Para lograr este fin, se compara la serie POI_05_EMP con POI_18 en el periodo de traslape disponible. La figura 22 muestra que las trayectorias son relativamente cercanas, aunque persisten diferencias que no se pueden atribuir al marco muestral. Por el contrario, se atribuyen al cambio en la definición para clasificar la población ocupada informal. Dado esto, el traslape entre estas series constituye la justificación para el segundo empalme a través del cual la serie ajustada inicialmente, se traslada al nuevo marco conceptual. De esta manera, la nueva serie no solo resulta consistente en término muestrales, sino también comparable con la medición conceptual más reciente.

Figura 22. Población ocupada informal: M05 + EMP y M18.

Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE.

Ahora, una vez integrados los componentes muestral y conceptual, fue necesario tratar la ruptura observada en los tres primeros trimestres de 2020 asociada a la pandemia. Con el fin de corregir esta inconsistencia, el procedimiento a seguir no se aplica de forma directa sobre el número de ocupados informales, sino sobre la proporción de informalidad. En primer lugar, esta proporción se construye como el cociente entre la población ocupada informal preliminar, esto es, la serie que aún no se ajusta por imputación, y la población ocupada final que se obtuvo con anterioridad a través de la propuesta de empalme. La lógica de este procedimiento consiste en intervenir primero la relación entre ambas variables, dado que esto permite capturar la dinámica de la informalidad dentro del total de ocupados y, luego, trasladar ese ajuste a la variable en niveles. De esta manera, la imputación no se realiza sobre la variable en niveles, sino sobre una medida proporcional adecuada para modelar la alteración observada durante la pandemia.

Con base en esta proporción, se identifica una inconsistencia en los trimestres 2020I, 2020II y 2020III. Lo anterior motiva proponer una imputación de valores faltantes mediante un modelo SARIMA⁴. Esta estrategia constituye una propuesta de este documento técnico y se fundamenta en un enfoque desarrollado por (Castaño, 2010) para la identificación de modelos ARIMA en presencia de observaciones faltantes. Bajo este criterio, el ajuste busca

⁴ Se utiliza el componente estacional con el fin de ofrecer una consistencia metodológica pura. Esto es, realizar un ajuste sobre las variables en bruto con el fin de capturar de la mejor forma los cambios.

reemplazar la trayectoria atípica de la proporción por una evolución más consistente con la estructura temporal de la serie. Así, este procedimiento podría evitar que el quiebre en pandemia distorsione de manera desproporcionada la medición de la informalidad.

Figura 23. Proporción de informalidad: imputación de valores faltantes con un modelo SARIMA (2020I, 2020II y 2020III).

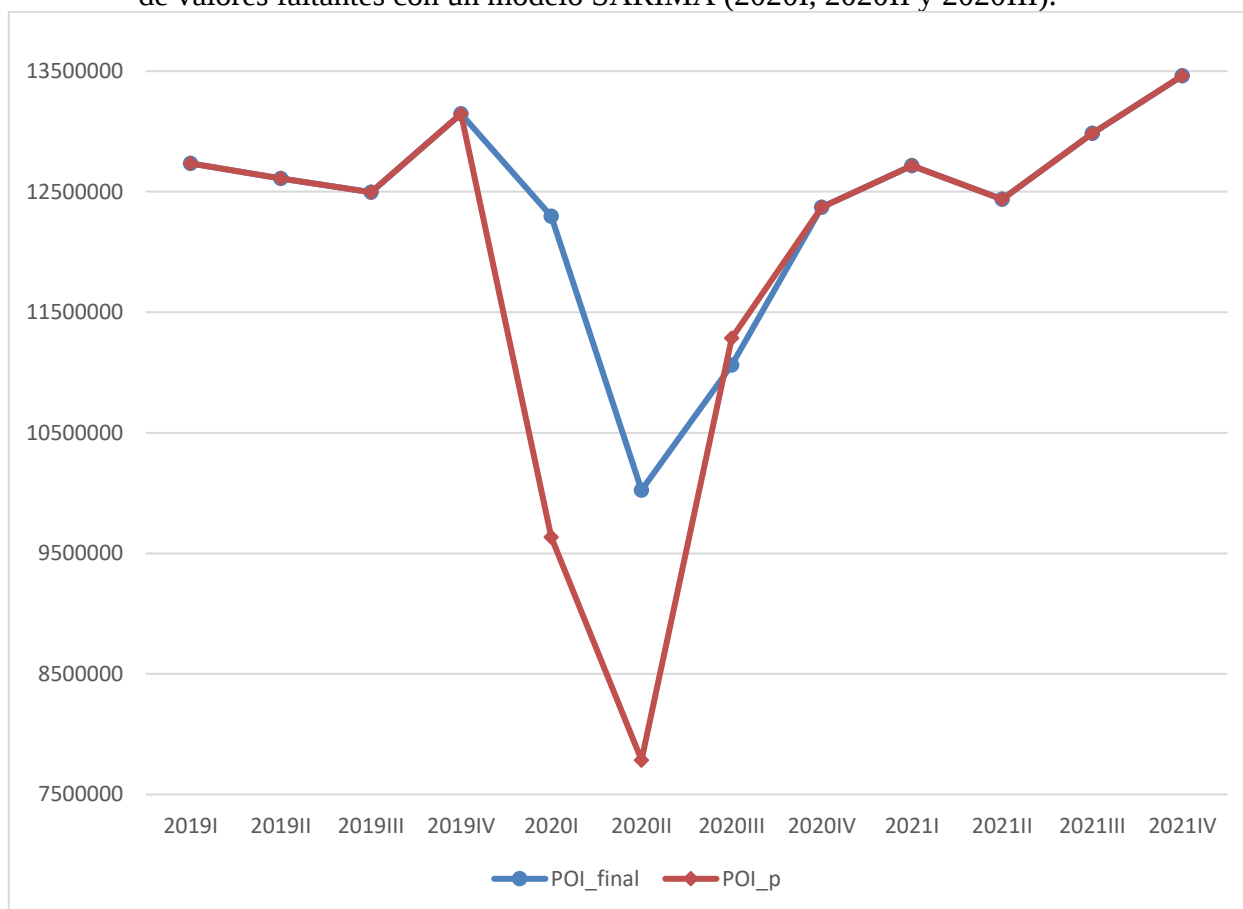


Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE.

La figura 23, muestra el resultado de la primera etapa de ajuste. Allí se compara la proporción de informalidad originalmente observada (P_{inf_p}) con la proporción ajustada mediante la imputación (P_{inf_final}). Se puede observar que la serie inicial muestra una caída abrupta e inconsistente, mientras que la trayectoria imputada restablece un comportamiento más suave dentro del contexto de la serie. Es decir, el modelo no elimina el efecto de la pandemia, pero sí corrige la magnitud extrema de este choque, lo que produce una trayectoria plausible.

Una vez obtenida la proporción de informalidad ajustada, el siguiente paso consiste en transformarla en niveles. Para este propósito, dicha proporción se multiplica por la población ocupada final, de forma que el resultado expresa el número de personas ocupadas informales bajo el ajuste propuesto. De esta manera, la imputación sobre la proporción se incorpora directamente en la construcción de la nueva serie de ocupación informal.

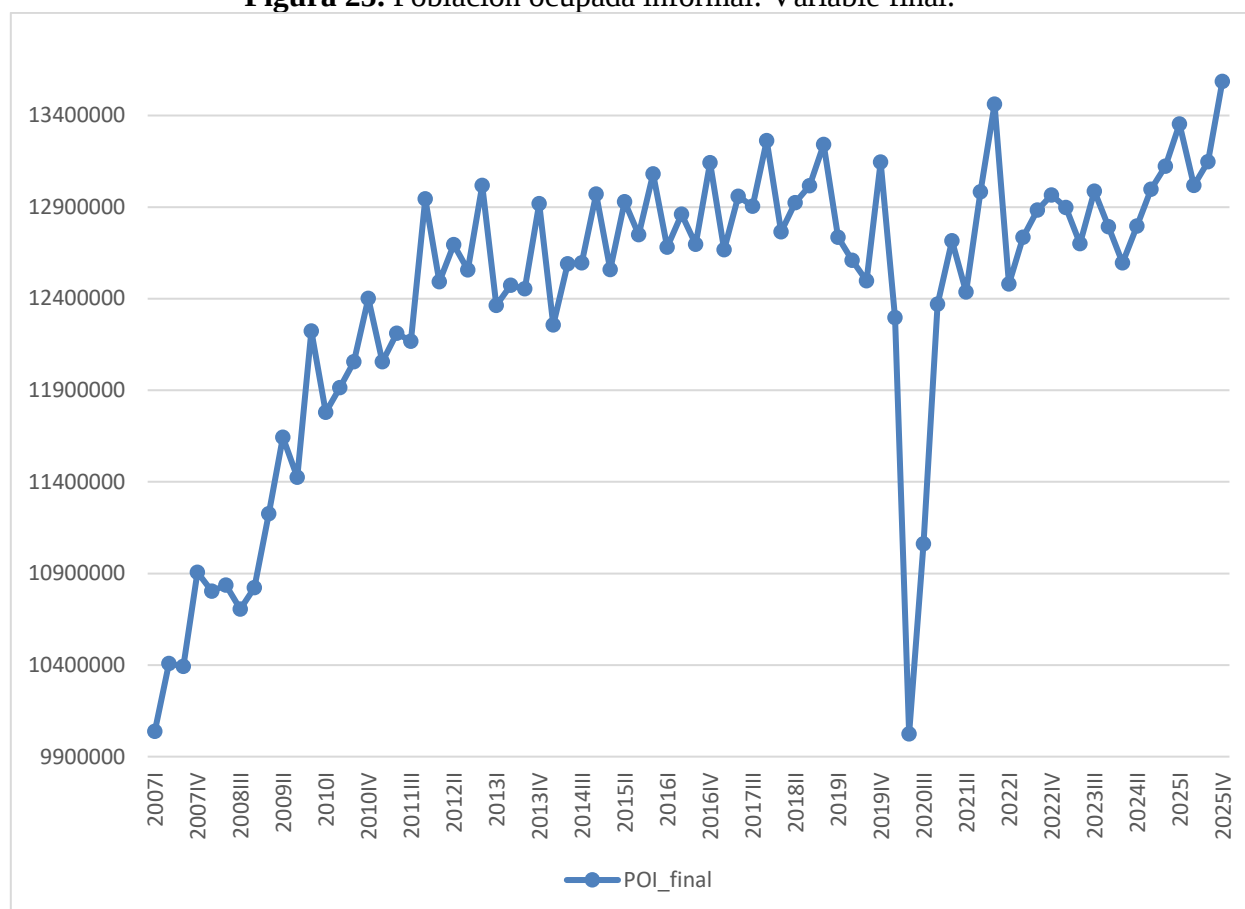
Figura 24. Población ocupada informal: Proporción de informalidad a partir de imputación de valores faltantes con un modelo SARIMA (2020I, 2020II y 2020III).



Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE.

La figura 24, presenta el efecto de esta operación sobre la variable en niveles. Al multiplicar la proporción ajustada por la población ocupada final, se observa que la caída de la informalidad en 2020 se suaviza de forma considerable respecto a la trayectoria preliminar (POI_p). Este resultado es interesante, y por supuesto, relevante, porque produce una evolución más realista de la informalidad durante el periodo crítico de la pandemia. Es decir, en lugar de mantener una contracción excesiva y difícilmente creíble empíricamente, la serie ajustada refleja una reducción significativa, pero al final compatible con la dinámica que se puede esperar del mercado laboral en un contexto de choques.

Finalmente, la serie que resulta se adopta como la variable final de la población ocupada informal. En este punto, el procedimiento para esta variable ha integrado tres elementos de manera conjunta: *i)* la armonización de la población ocupada total y ocupada informal a través del empalme, *ii)* la estimación de la proporción de informalidad sobre las series armonizadas y *iii)* la imputación de los valores faltantes de dicha proporción mediante un modelo SARIMA.

Figura 25. Población ocupada informal: Variable final.

Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE.

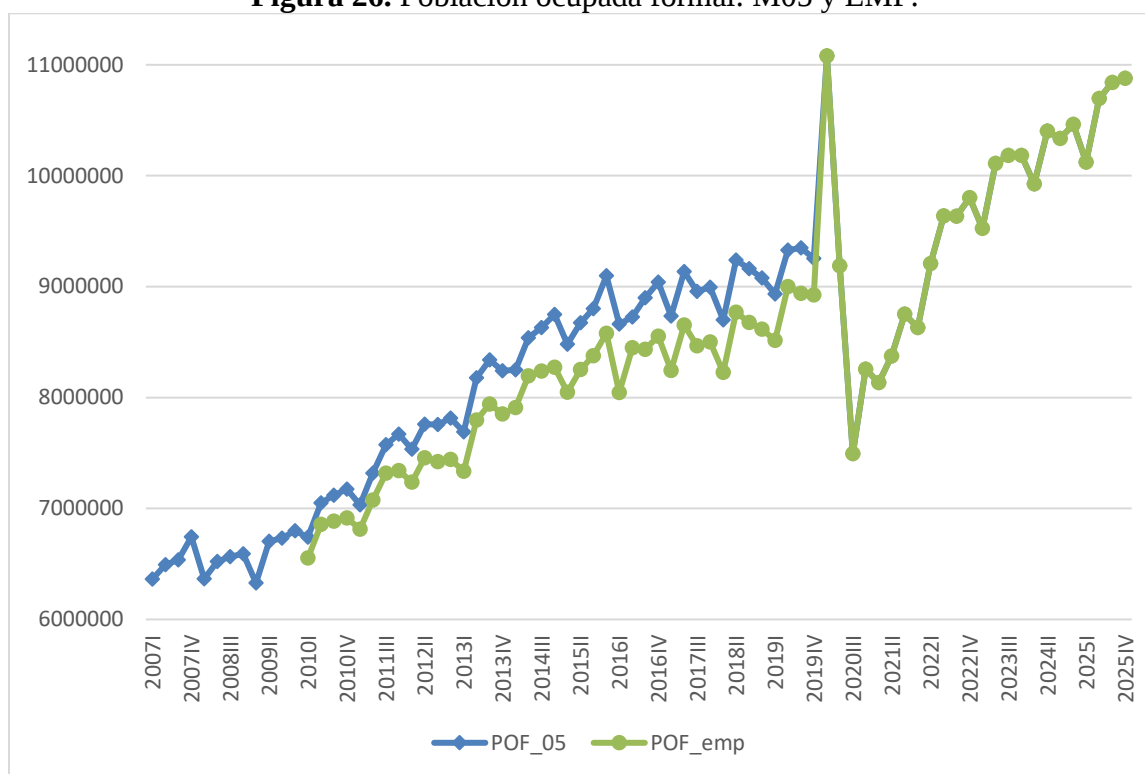
2.1. Población Ocupada Formal.

En cuanto a la serie de la población ocupada formal, se sigue el mismo proceso metodológico aplicado anteriormente. En concreto, este procedimiento consta de tres pasos: *i)* la corrección de la discontinuidad asociada al cambio muestral, *ii)* la incorporación del cambio conceptual producto del rediseño de la GEIH y *iii)* la definición de una variable final más consistente. Sin embargo, este caso toma otro rumbo, pues se opta por definir la población ocupada formal como un residuo calculado como la diferencia entre la población ocupada total final (producto de la metodología propuesta) y la población ocupada informal final.

Primero, se compararon las series POF_05 y POF_EMP. Como se observa en la figura ##, entre ambas series existe una brecha persistente en el periodo previo a la pandemia, en especial desde 2010, que es el tramo para el cual se dispone información de las series de empalme. Dado que ambas variables conservan una lógica conceptual comparable vía definición, la discrepancia se interpreta principalmente como el resultado del componente muestral. Así, la serie PO_05 tiende a situarse por encima de PO_EMP en buena parte del

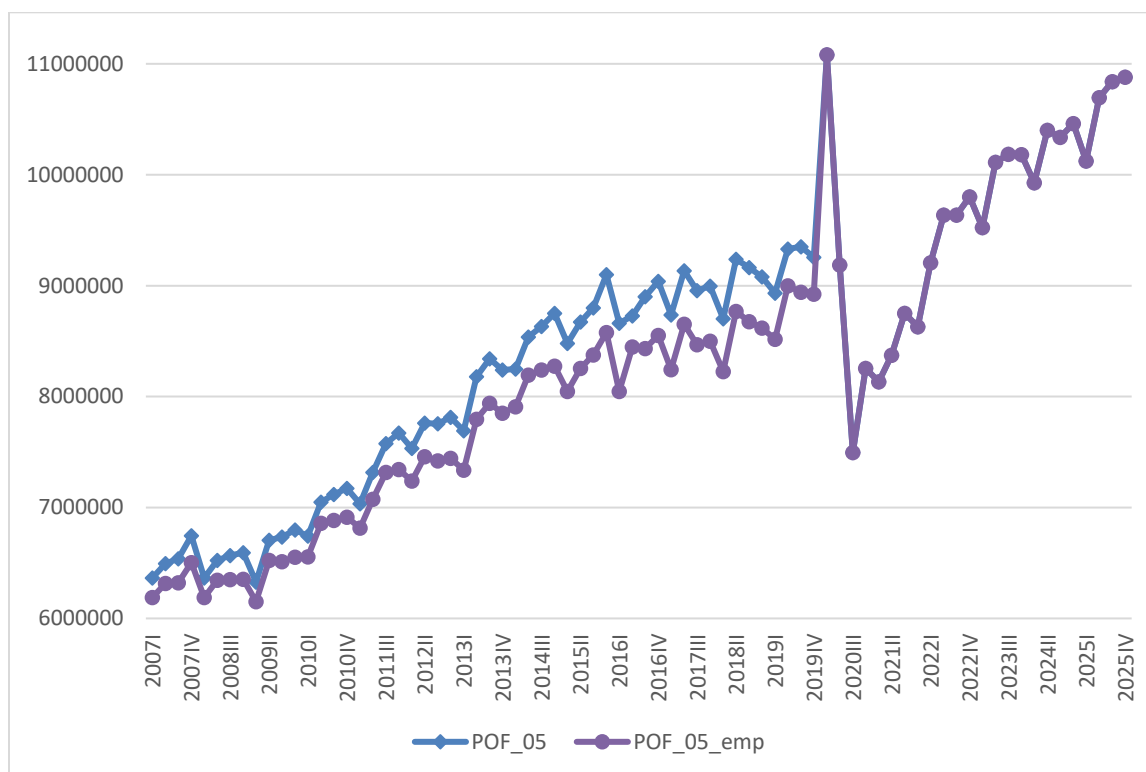
periodo a 2020, lo que puede sugerir una posible sobreestimación de la ocupación formal bajo la anterior metodología.

Figura 26. Población ocupada formal: M05 y EMP.



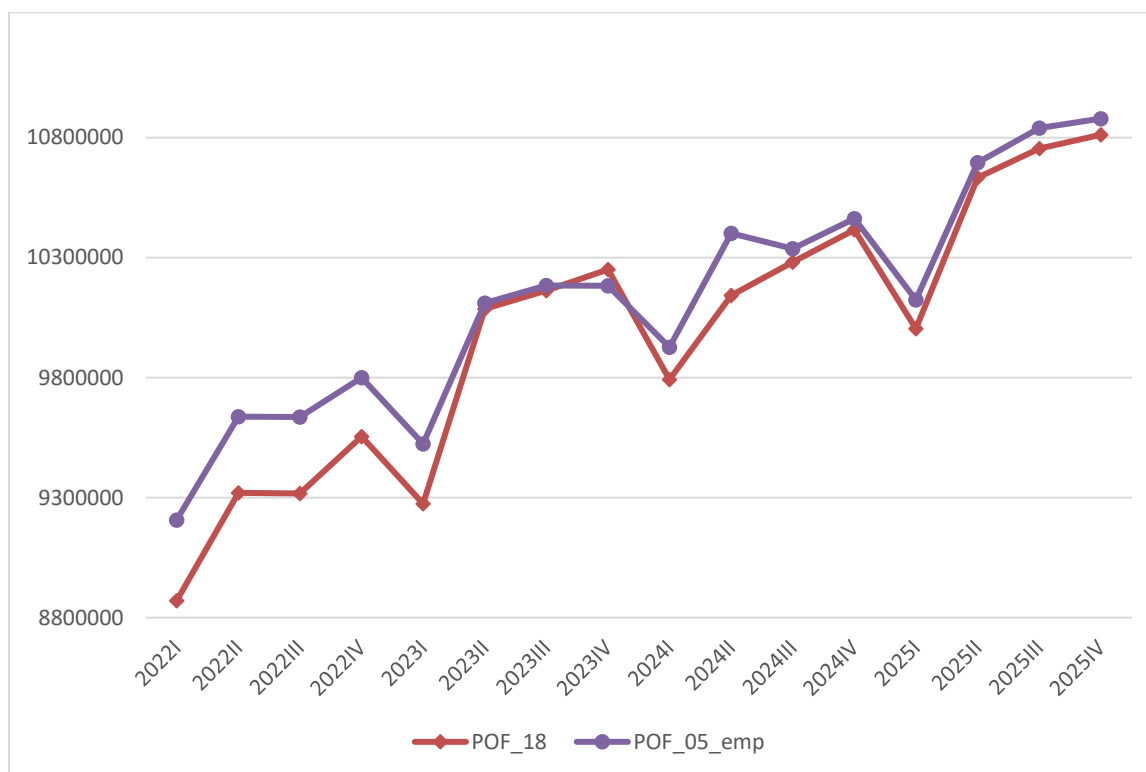
Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE.

Teniendo en cuenta esta evidencia, se integran las variables POF_05 y POF_EMP, con lo que se obtiene la serie POF_05_EMP. Este primer empalme corrige la discontinuidad asociada al cambio en el marco muestral y permite conservar la trayectoria histórica de la serie de forma más consistente. En la figura 27, la serie resultante permite establecer qué sucedió antes de 2010. Sin embargo, aunque este ajuste mejora la comparabilidad, todavía no se incorpora el cambio conceptual introducido en el rediseño de la GEIH.

Figura 27. Población ocupada formal: M05 y M05 + EMP.

Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE.

Luego, se incorpora el componente conceptual a través de la integración de POF_05_EMP y POF_18 en el periodo disponible. En la figura 28, ambas series presentan trayectorias próximas. Se observa ciertas diferencias en nivel que ya no responden al componente muestral, sino al cambio en la definición y clasificación de la población ocupada formal bajo el marco 2018. Como consecuencia, el segundo empalme permite trasladar la serie al nuevo marco conceptual. El resultado se puede entender como una serie final preliminar.

Figura 28. Población ocupada formal: M05 + EMP y M18.

Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE.

No obstante, para esta variable se sigue un procedimiento análogo al utilizado en la construcción de la población ocupada informal, con la diferencia de que, en este caso, el ajuste parte de la proporción de formalidad que se obtiene como el complemento de la proporción de informalidad ajustada. Es decir, una vez estimada la proporción de informalidad ajustada corregida para los tres primeros trimestres del 2020, se calcula la proporción ajustada de la formalidad como:

$$P_{for} = 1 - P_{inf}^{ajustada}$$

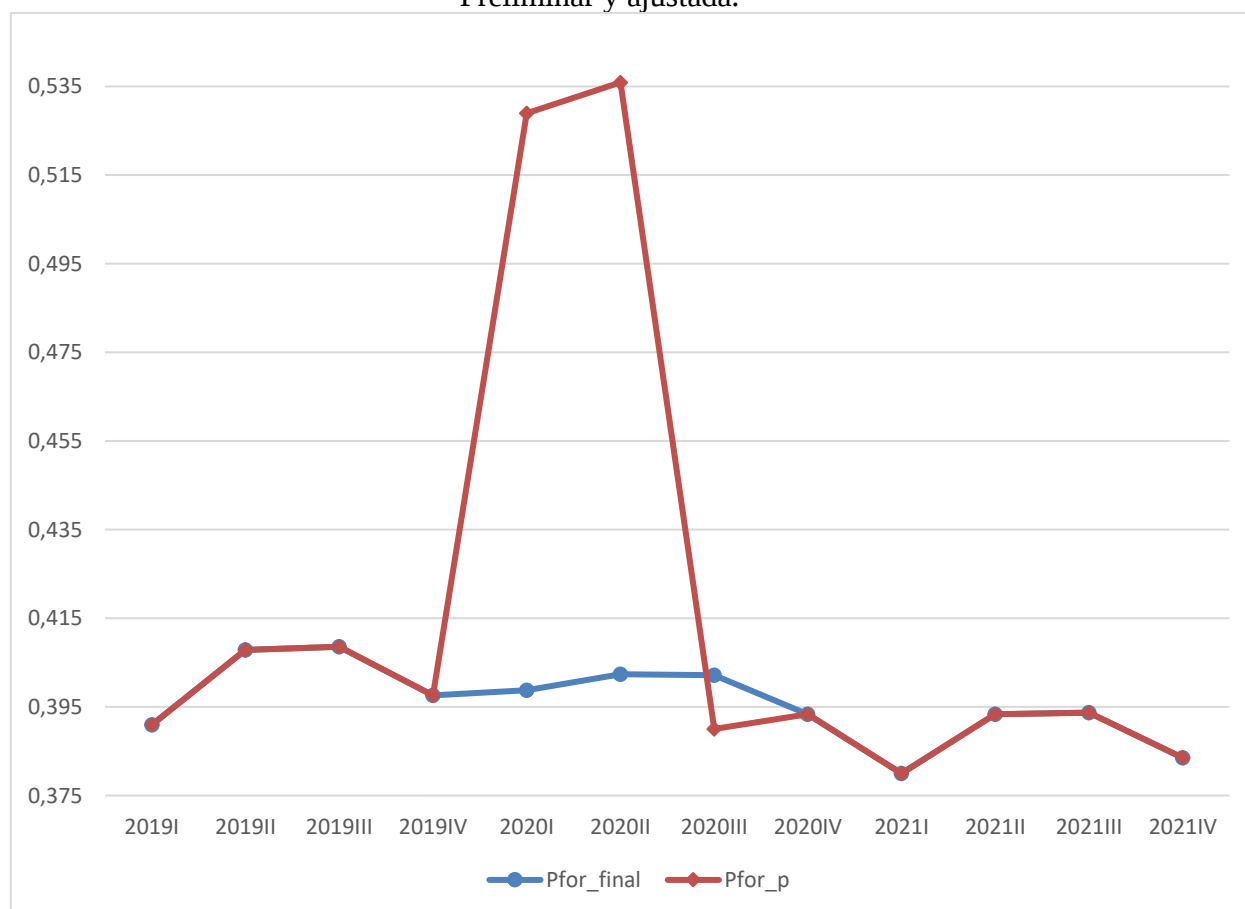
De esta forma, el ajuste aplicado a la informalidad se traslada de forma residual a la formalidad⁵, lo que garantiza una coherencia interna entre ambas proporciones y preserva la identidad básica según la cual la suma entre proporción de formalidad e informalidad debe ser igual a uno, el total de ocupados.

Este paso es metodológicamente importante porque evita tratar la formalidad como una serie aislada. Contrariamente, al derivarla como un complemento de la proporción de la informalidad ajustada, se asegura que ambos componentes del empleo permanezcan

⁵ Consistente con la definición de formalidad, que surge de forma residual a partir del cálculo de la informalidad.

consistentes entre sí. De esta manera, el ajuste se realiza sobre la participación relativa de los ocupados formales dentro del total de ocupados.

Figura 29. Proporción de formalidad: Residual ($1 - \text{proporción de formalidad}$). Preliminar y ajustada.

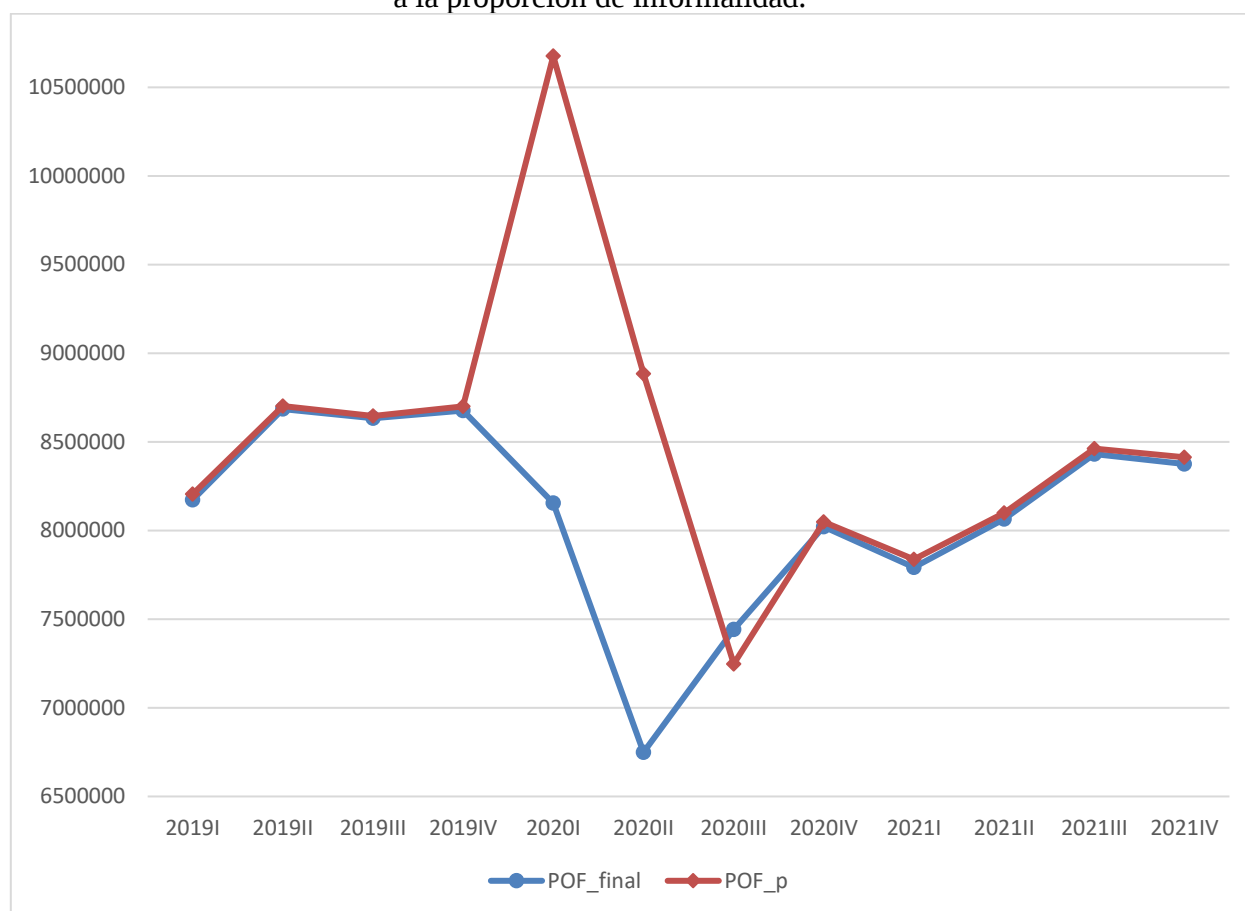


Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE.

La figura 29 muestra el resultado de este procedimiento sobre la proporción de formalidad. Se puede observar que, en los primeros tres trimestres de 2020, la trayectoria ajustada se separa de la preliminar. Esto refleja que el efecto residual del ajuste efectuado anteriormente sobre la proporción de informalidad. La diferencia es especialmente visible en los dos primeros trimestres de 2020, cuando la corrección de la serie de informalidad modifica de forma inmediata la participación relativa de los formales dentro del total de ocupados.

Una vez obtenida la proporción de formalidad ajustada, esta se multiplica por la población ocupada final. Así, se obtiene una nueva estimación del número de formales bajo el ajuste propuesto. Este procedimiento permite trasladar la corrección realizada en términos proporcionales a una serie en niveles coherente tanto con la armonización inicial como con la corrección específica.

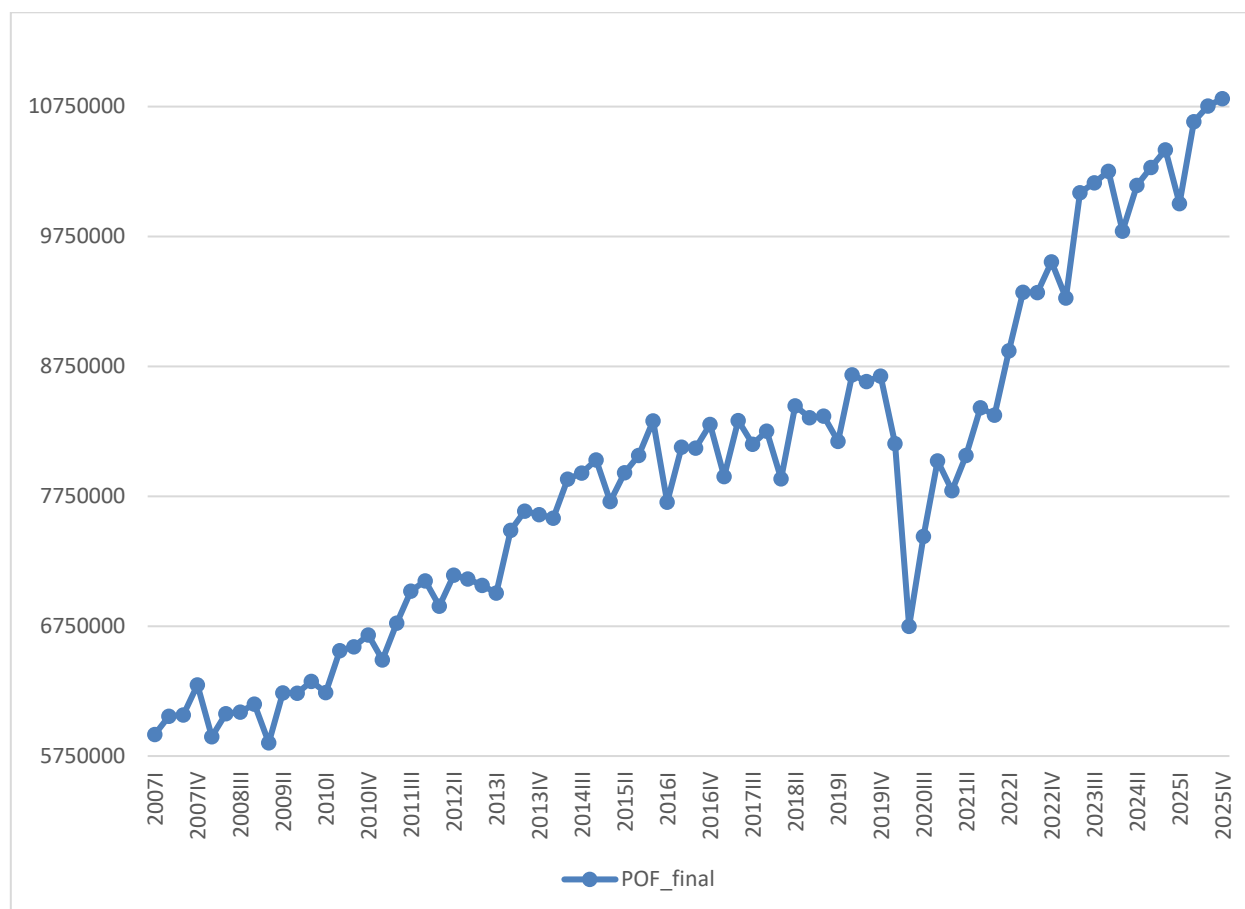
Figura 30. Población ocupada formal: Ajuste a partir de la imputación de valores faltantes a la proporción de informalidad.



Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE.

La figura 30 presenta el efecto del ajuste en la serie en niveles. Se observa que el ajuste altera significativamente la trayectoria de la formalidad en los trimestres más críticos de la pandemia. En consecuencia, la serie corregida evita replicar automáticamente la distorsión original y genera una trayectoria consistente con la nueva distribución proporcional entre formales e informales.

Finalmente, la serie resultante se adopta como la variable final. En este punto, la construcción de esta integra dos elementos metodológicos complementarios: la armonización previa de la población ocupada y ocupada formal, y por otra, la corrección de la composición interna entre formalidad e informalidad a partir del ajuste de la proporción de informalidad.

Figura 31. Población ocupada formal: Variable final.

Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE.

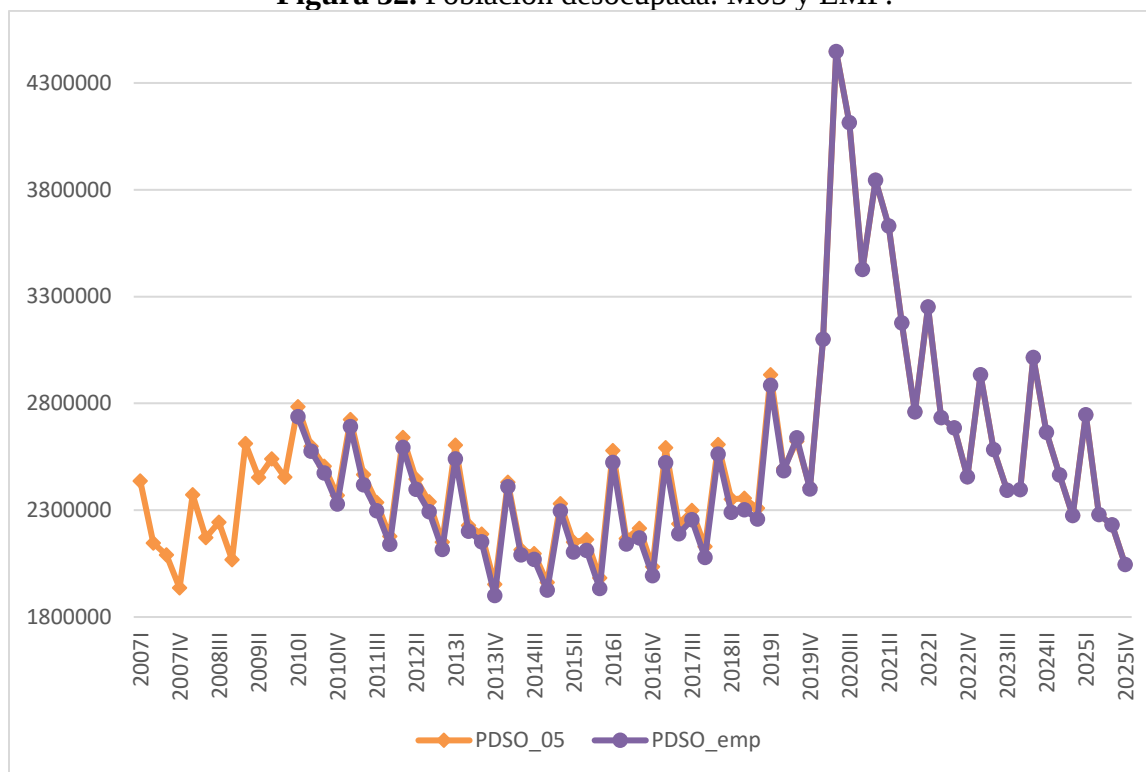
3. Población Desocupada

En cuanto a la serie de la población desocupada se siguió, en principio, el mismo esquema metodológico aplicado a las demás variables del mercado laboral: *i)* corrección de la discontinuidad asociada al cambio muestral, *ii)* evaluación de la necesidad de incorporar un ajuste conceptual y *iii)* la definición de la variable final. Este caso resulta interesante porque presenta una particularidad importante: aunque se desarrolla el procedimiento completo de comparación y empalme, la evidencia empírica mostrará que el primer empalme es suficiente para obtener la serie definitiva. En estos términos, la variable PDSO_05_EMP resulta análoga a PDSO_final.

En primer lugar, se comparan las variables PDSO_05 y PDSO_EMP. Como se aprecia en la figura 32, ambas series presentan una trayectoria muy similar en el periodo de traslape, aunque con ciertas diferencias atribuibles, posiblemente, al componente muestral. Estas discrepancias responden, principalmente, al cambio en el factor de expansión. Además, no

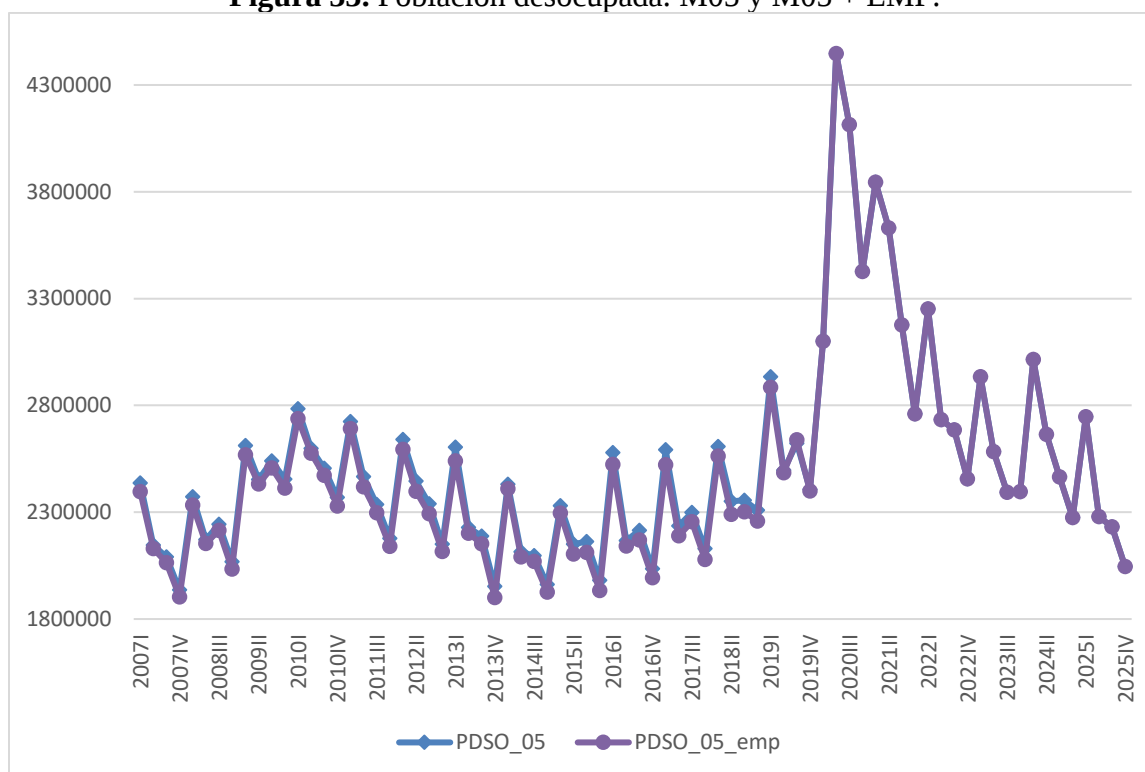
se observa una ruptura estructural persistente entre ambas series antes de la pandemia. Esto es, las diferencias de nivel están relativamente acotadas, lo que sugiere que el problema de comparabilidad para la población desocupada se concentra, ante todo, en el componente muestral.

Figura 32. Población desocupada: M05 y EMP.



Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE.

Con base a esta evidencia, se integran las variables PDSO_05 y PDSO_EMP, con lo que se obtiene la serie PDSO_05_EMP. Como muestra la figura 33, el resultado del empalme permite corregir las diferencias observadas entre ambas series y, sobre todo, establecer un horizonte temporal completo desde 2007 hasta 2025. Este ajuste cumple, entonces, la función de armonizar la serie frente al cambio en términos muestrales y constituye una versión consistente y comparable a lo largo del tiempo.

Figura 33. Población desocupada: M05 y M05 + EMP.

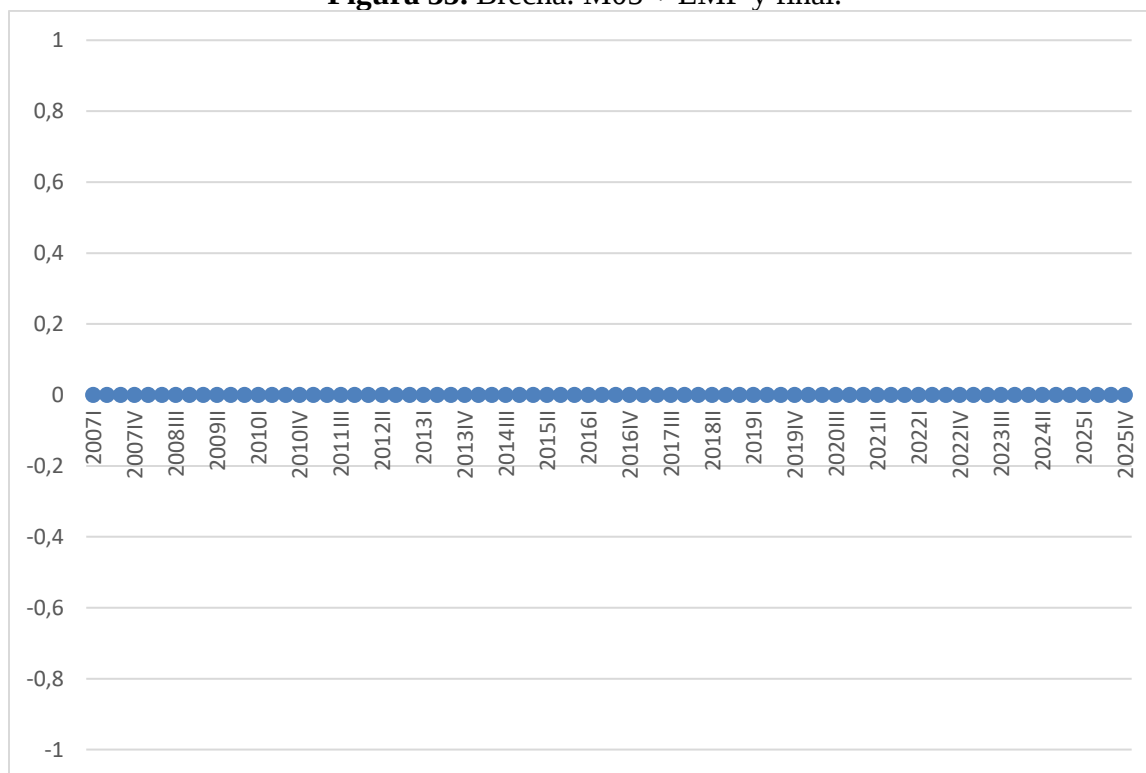
Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE.

Luego, se compara la serie PDSO_05_EMP con PDSO_18. Como se aprecia en la figura 34, ambas series son iguales. Este resultado es importante en términos metodológicos. Esto puede indicar que, aunque en el rediseño de la GEIH se introdujeron modificaciones más amplias en la encuesta, la construcción de la población desocupada no experimentó una ruptura conceptual importante. En particular, la pregunta de clasificación, **P6351**, se mantiene esencialmente igual. En consecuencia, no se requiere un segundo ajuste de empalme para trasladar la serie al marco 2018 en términos conceptuales.

Figura 34. Población desocupada: M05 + EMP y M18.

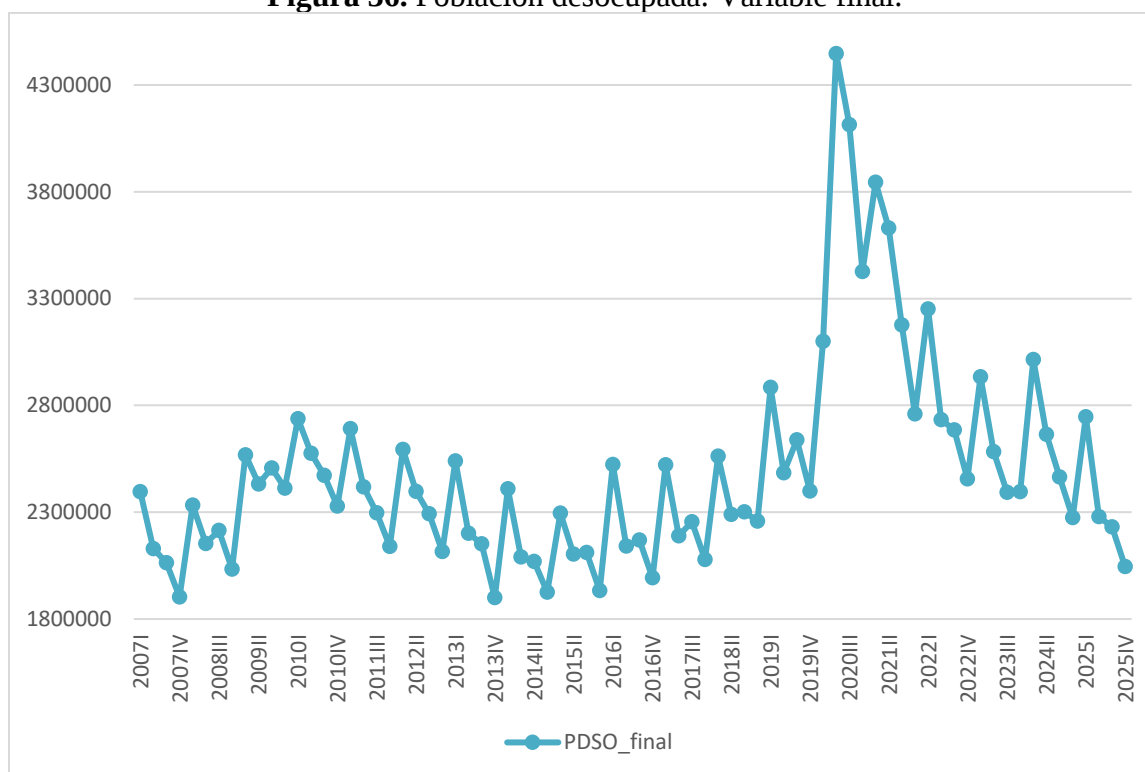
Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE.

La equivalencia descrita, puede apreciarse con mayor claridad en la siguiente figura. En esta se presenta la brecha entre PDSO_05_EMP y la variable final. La brecha es igual a cero a lo largo del horizonte temporal, lo que confirma que no existe diferencia alguna entre la serie empalmada tras el ajuste muestral y la serie que finalmente se adopta. Metodológicamente, esto implica que el procedimiento completo ha sido verificado, pero que la evidencia empírica muestra que el primer empalme es más que suficiente.

Figura 35. Brecha: M05 + EMP y final.

Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE.

Finalmente, en la figura 36 se presenta la serie definitiva para la población desocupada. Dado que esta serie coincide exactamente con PDSO_05_EMP, la variable final recoge únicamente la corrección asociada al cambio muestral sin necesidad de incorporar un ajuste conceptual adicional. En este sentido, la serie final es consistente y comparable en el tiempo. Su principal fortaleza radica en que preserva la continuidad estadística sin introducir transformaciones innecesarias.

Figura 36. Población desocupada: Variable final.

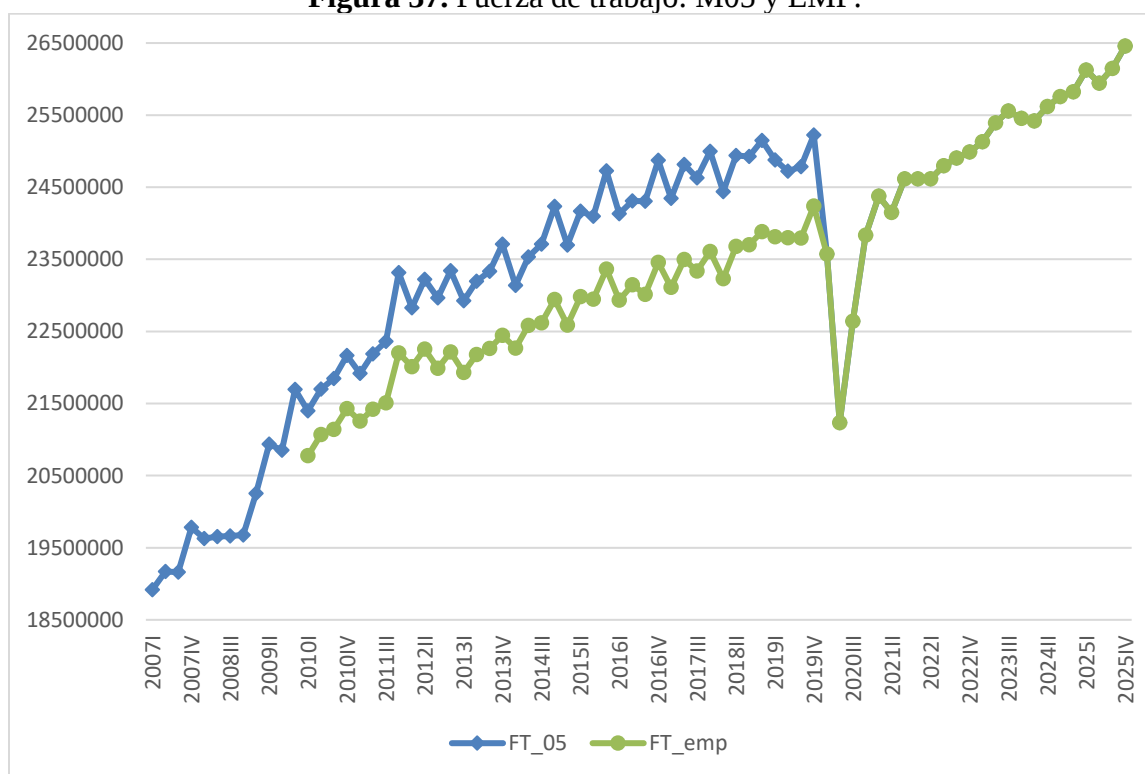
Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE.

4. Población Económicamente Activa o Fuerza de Trabajo

Para la construcción de la serie de la fuerza de trabajo, se toma el mismo esquema metodológico aplicado anteriormente. Primero, se realiza la corrección de la discontinuidad asociada al cambio muestral. Segundo, se incorpora el componente conceptual introducido en el marco 2018⁶ y por último, se define la variable final más consistente desde el punto de vista contable y conceptual. Esto último resulta interesante porque la decisión final no consistió en adoptar la serie empalmada, sino en recalcular la fuerza de trabajo como la suma de la población ocupada final y la población desocupada final calculadas anteriormente. Este criterio conserva de mejor manera la coherencia conceptual.

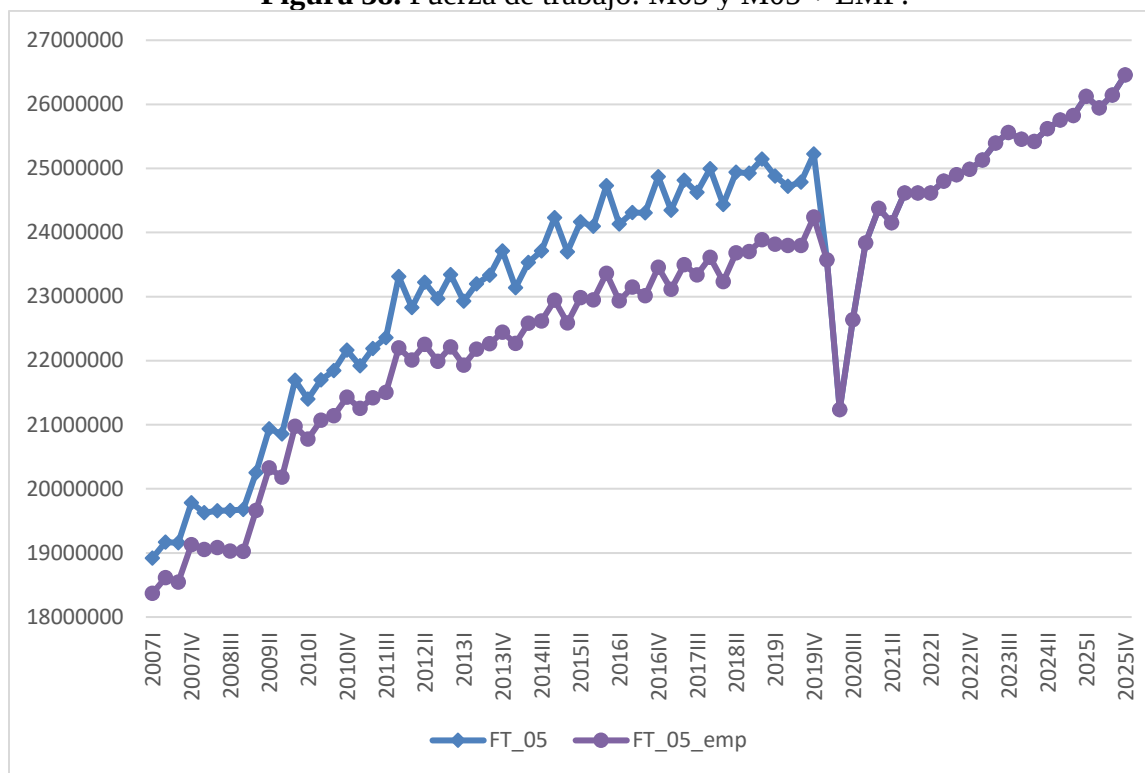
En la primera etapa, se comparan las series FT_05 y FT_EMP. Como se puede apreciar en la figura 37, entre ambas existe una brecha considerable antes de la pandemia. Al igual que antes, la diferencia se puede asociar principalmente al componente muestral. Esto es, los cambios en el factor de expansión. Empíricamente, la serie del marco 2005 tienen ubicarse por encima de la serie de empalme en buena parte del tiempo antes de 2020. Podría sugerir entonces, que bajo el marco 2005 se sobreestima el nivel de la fuerza de trabajo.

⁶ La definición de la fuerza de trabajo es la misma. Sin embargo, varía porque su construcción depende de la población ocupada y desocupada. Por lo tanto, ante el cambio en las definiciones de estas dos variables entre marcos conceptuales, tiene impacto directo en la clasificación de la población como fuerza de trabajo.

Figura 37. Fuerza de trabajo: M05 y EMP.

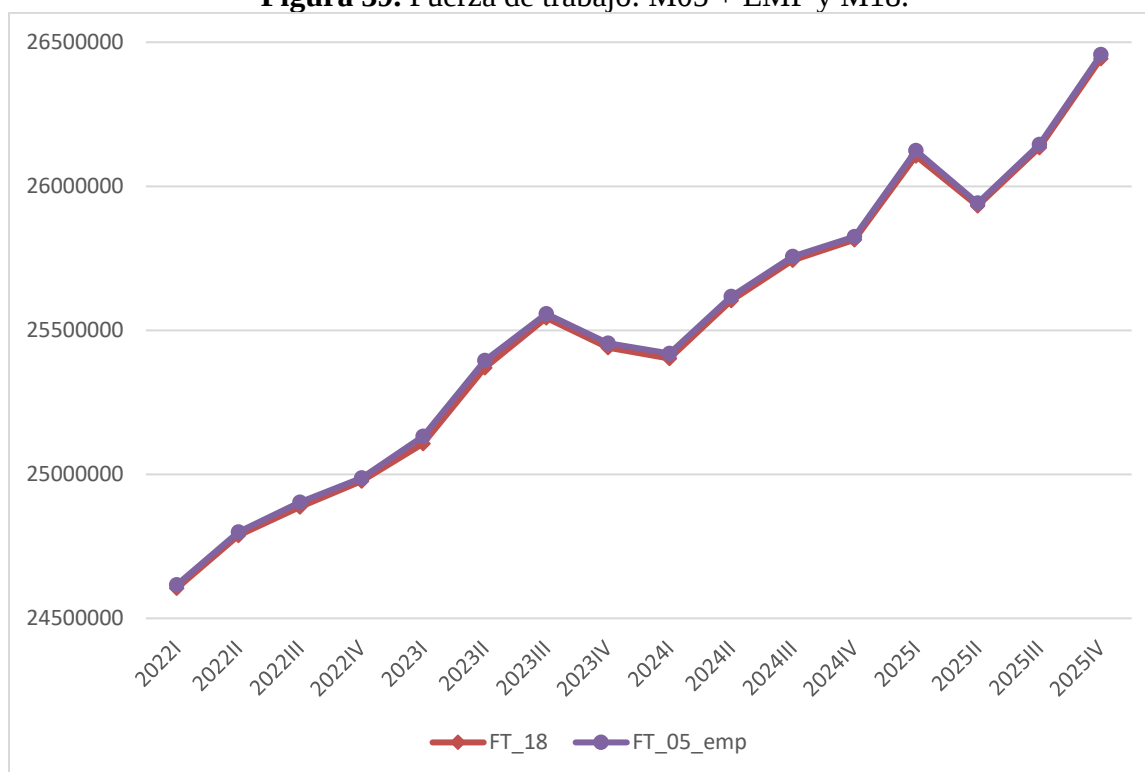
Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE.

Al seguir esta evidencia, resulta necesario integrar las variables FT_05 y FT_EMP para obtener la serie FT_05_EMP. El empalme permite ampliar el horizonte temporal desde 2007 hasta el final de la serie. En la siguiente figura se observa el resultado.

Figura 38. Fuerza de trabajo: M05 y M05 + EMP.

Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE.

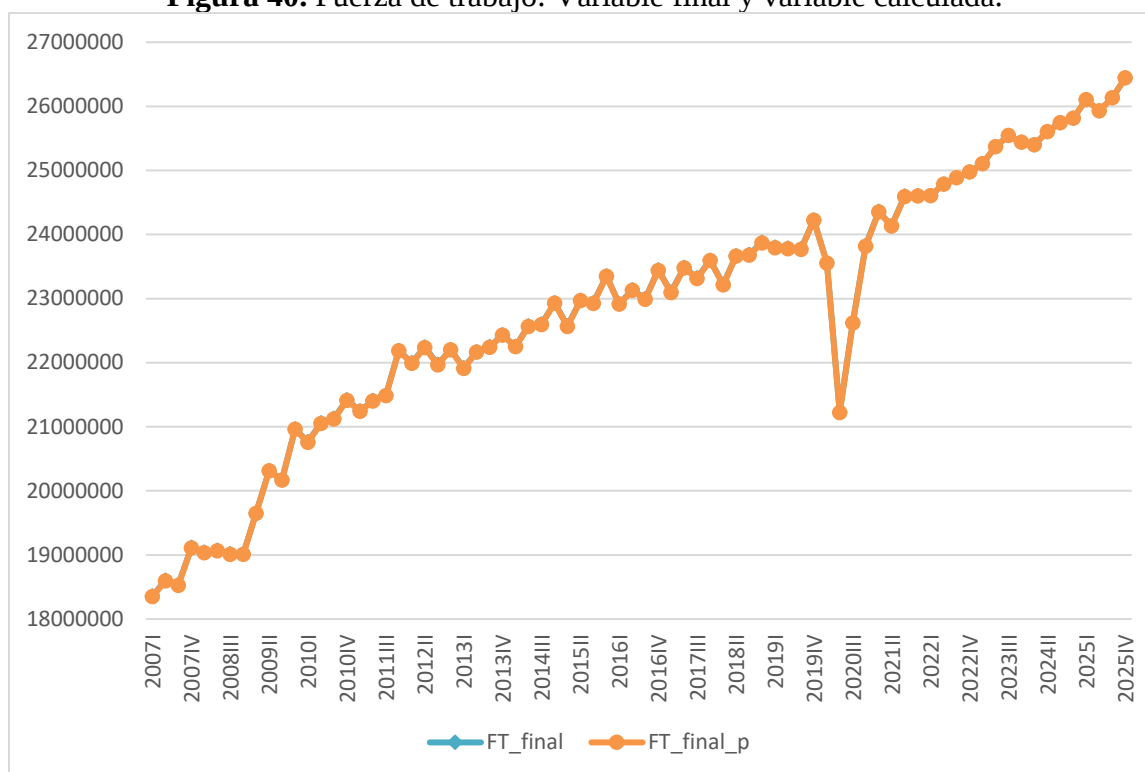
Luego, se realiza un segundo empalme integrando las variables FT_05_EMP y FT_18. Como se observa en la figura 39, ambas series presentan trayectorias similares con diferencias en niveles muy pequeñas. Esto puede indicar que, una vez corregido el componente muestral, la brecha frente al marco 2018 es relativamente pequeña. En consecuencia, este segundo empalme permite traer la serie al nuevo marco conceptual y obtener una primera versión final bajo la metodología propuesta.

Figura 39. Fuerza de trabajo: M05 + EMP y M18.

Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE.

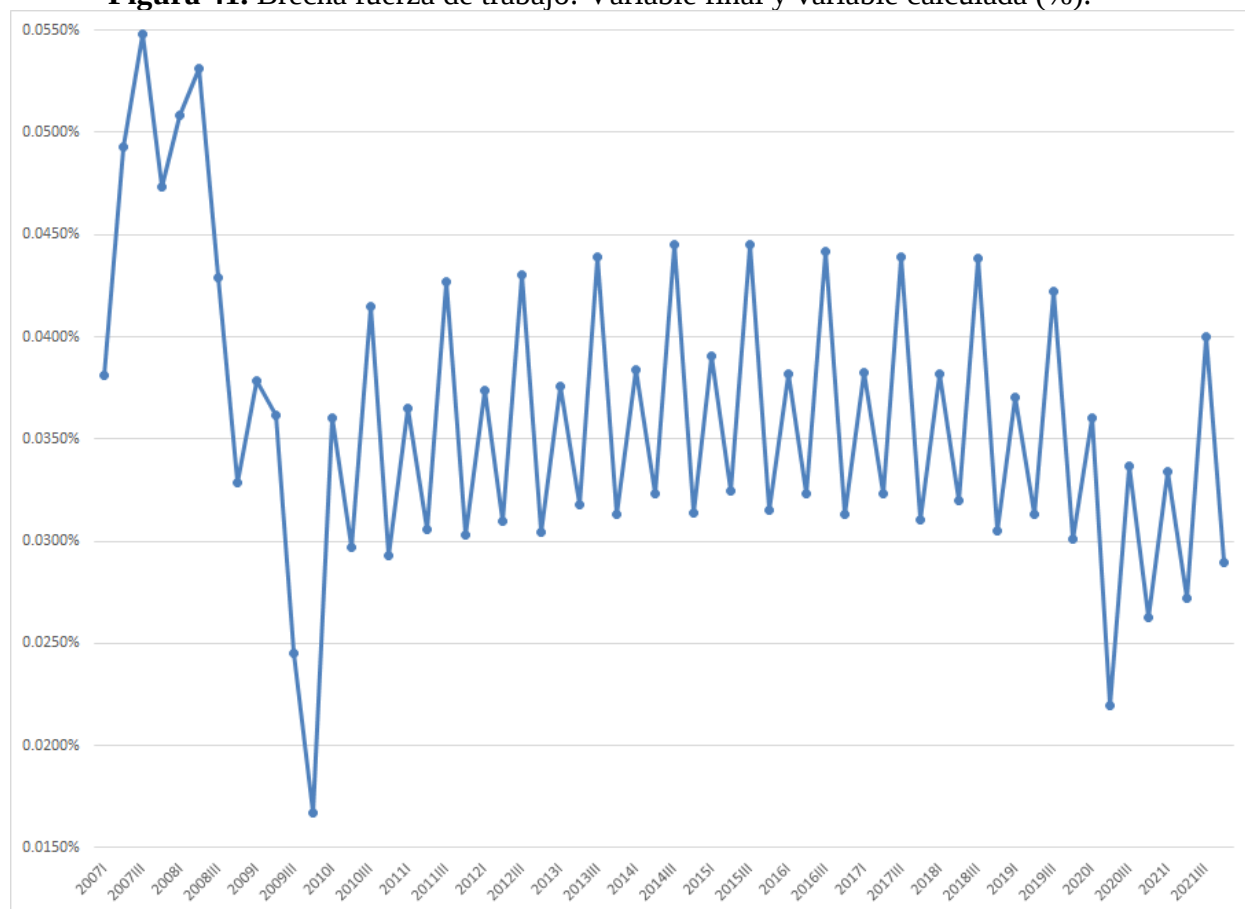
Sin embargo, para la definición de la serie final se adopta un criterio adicional consistente con la conceptualización de la fuerza de trabajo. En lugar de preservar únicamente la serie final producto del empalme directo, se recalcula la fuerza de trabajo como la suma de la población ocupada final y la población desocupada final. La principal razón es que, desde un punto de vista conceptual, la fuerza de trabajo debe corresponder a la agregación de estos dos componentes. Es por eso por lo que, aun cuando la serie obtenida a través de la metodología de empalme y la serie recalculada son muy similares, se considera óptimo adoptar esta última como serie definitiva.

La siguiente figura permite vislumbrar con claridad la comparación entre la variable final producto de la metodología de empalme y la variable calculada a partir de la suma de los dos agregados. Como se puede apreciar, ambas trayectorias se superponen. Esto confirma que nuestra metodología produce resultados consistentes. Sin embargo, la coincidencia no es perfecta, de manera que aún hay una pequeña diferencia entre ambas series.

Figura 40. Fuerza de trabajo: Variable final y variable calculada.

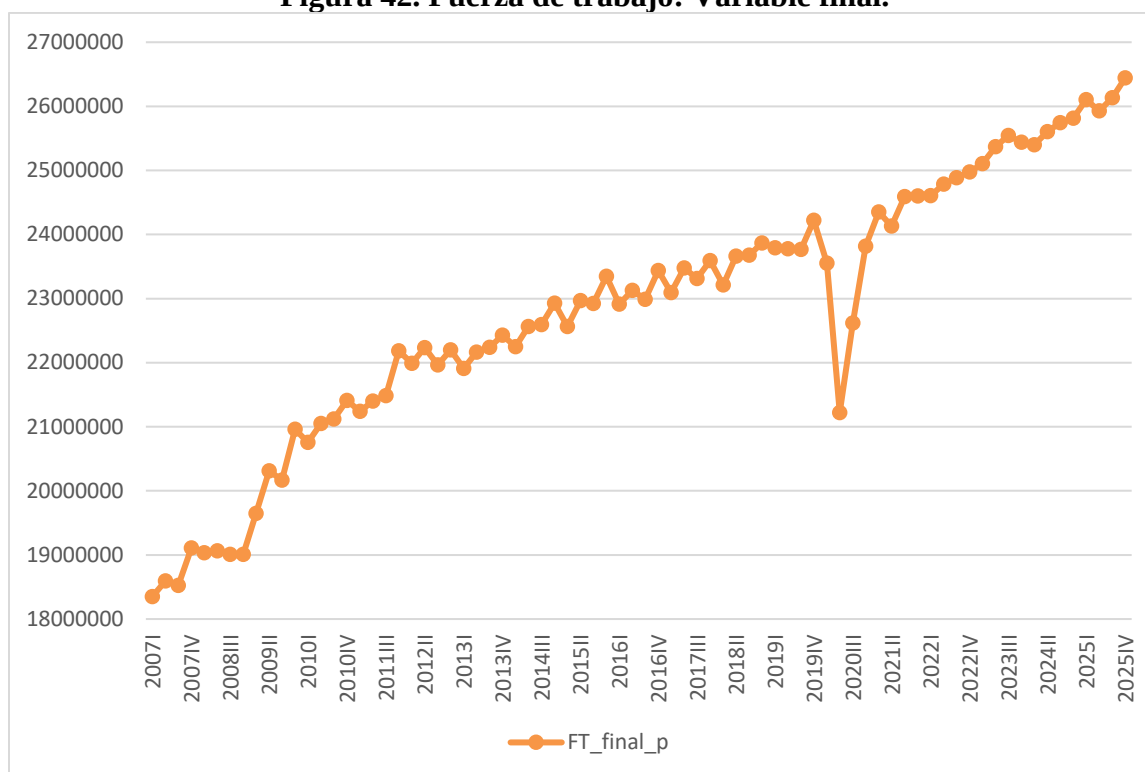
Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE.

La diferencia se examina a mayor detalle en la figura 41, en la cual se presenta la brecha entre la variable final de la metodología y la variable recalculada. En términos absolutos la brecha es reducida y no altera la tendencia de la serie. La existencia de la brecha indica que ambas construcciones no son del todo equivalentes. Ante esto, y dado que conceptualmente la fuerza de trabajo debe satisfacer una identidad, se opta por privilegiar la serie recalculada.

Figura 41. Brecha fuerza de trabajo: Variable final y variable calculada (%).

Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE.

Finalmente, la figura 42 presenta la serie definitiva de fuerza de trabajo, correspondiente a la variable recalculada. La trayectoria recoge, de forma indirecta, los ajustes ya incorporados en las series finales de ocupación y desocupación.

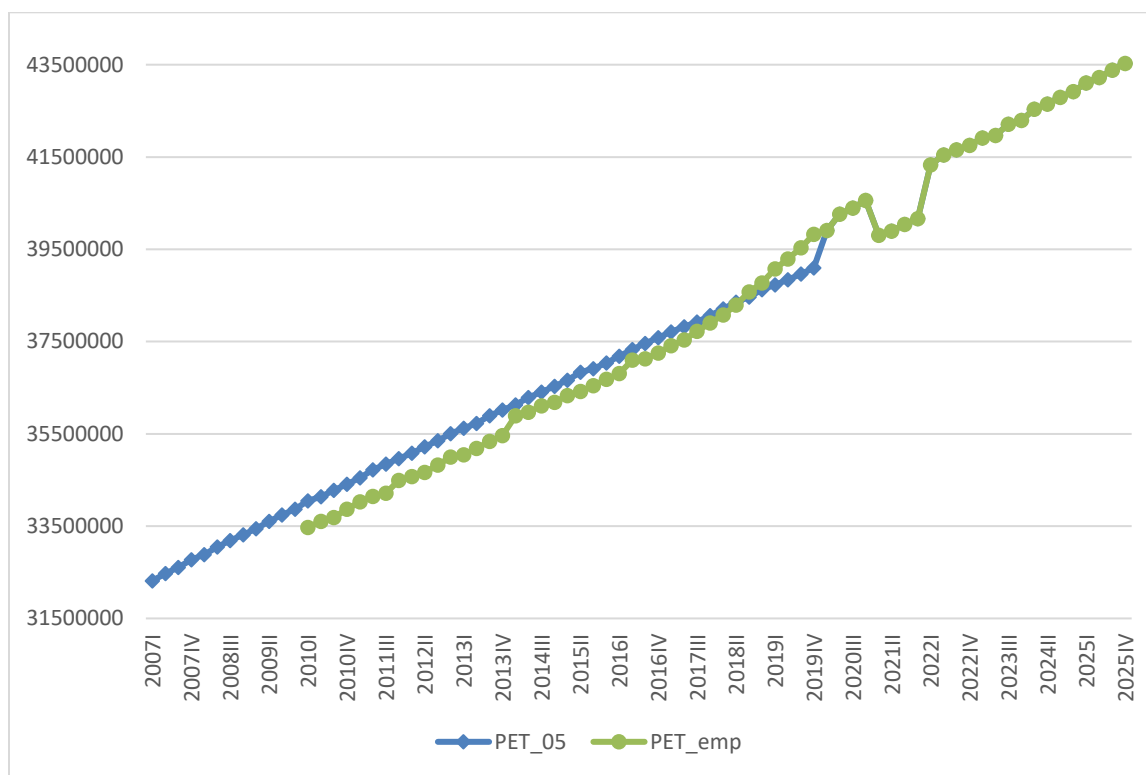
Figura 42. Fuerza de trabajo: Variable final.

Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE.

5. Población en Edad de Trabajar

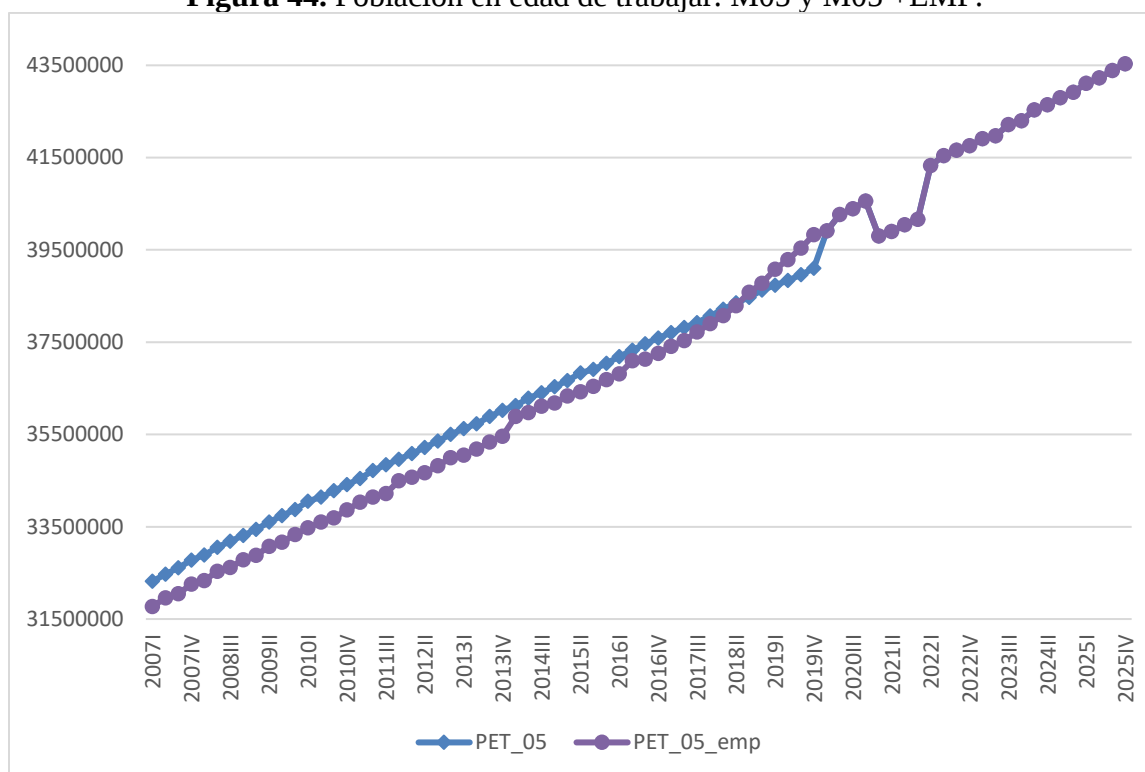
Para construir la población en edad de trabajar, se sigue el mismo esquema general aplicado anteriormente. No obstante, a diferencia de las otras variables, en este caso el componente conceptual adquiere una relevancia central. Esto dado que entre marcos se modifica de forma abrupta uno de los elementos básicos de la definición: el criterio de edad utilizado para delimitar la población en edad de trabajar. De esta manera, el segundo empalme constituye una transformación fundamental para garantizar una propuesta interesante en términos de comparabilidad de la serie.

Primero, se comparan las series PET_05 y PET_EMP. Como se observa en la siguiente figura, la tendencia de ambas series es similar. Sin embargo, con los datos disponibles, se observa cierta diferencia, sobre todo en el periodo que antecede la pandemia. Al igual que antes, esta diferencia se asocia al componente muestral: los cambios en el factor de expansión.

Figura 43. Población en edad de trabajar: M05 y EMP.

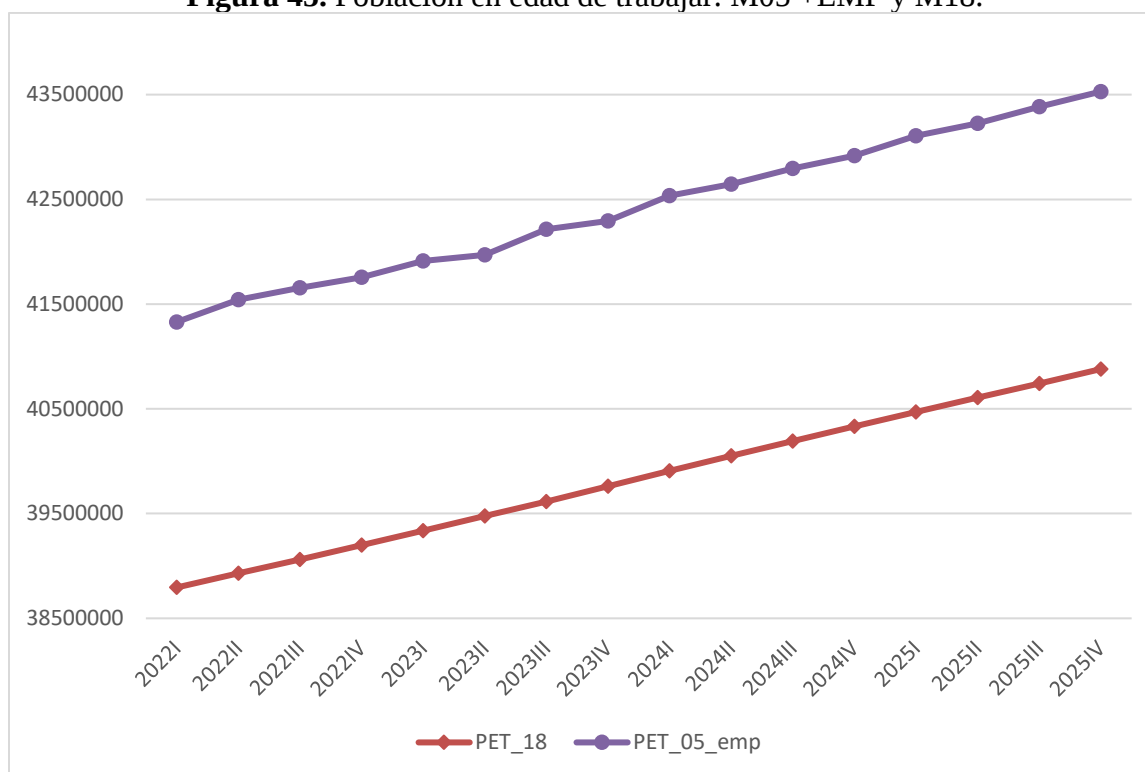
Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE.

Con esto claro, se integran ambas variables para conseguir PET_05_EMP. Es importante aclarar en este punto que, las dos variables comparten el mismo componente conceptual. Esto es, la misma definición para clasificar la población: doce años o más. En la siguiente figura se observa que con esta integración se logra una serie completa en términos temporales, desde 2007 hasta 2025. Este resultado es útil, pero es aún insuficiente para garantizar la comparabilidad entre periodos, pues no se incorpora el cambio conceptual en la definición etaria de la variable.

Figura 44. Población en edad de trabajar: M05 y M05 +EMP.

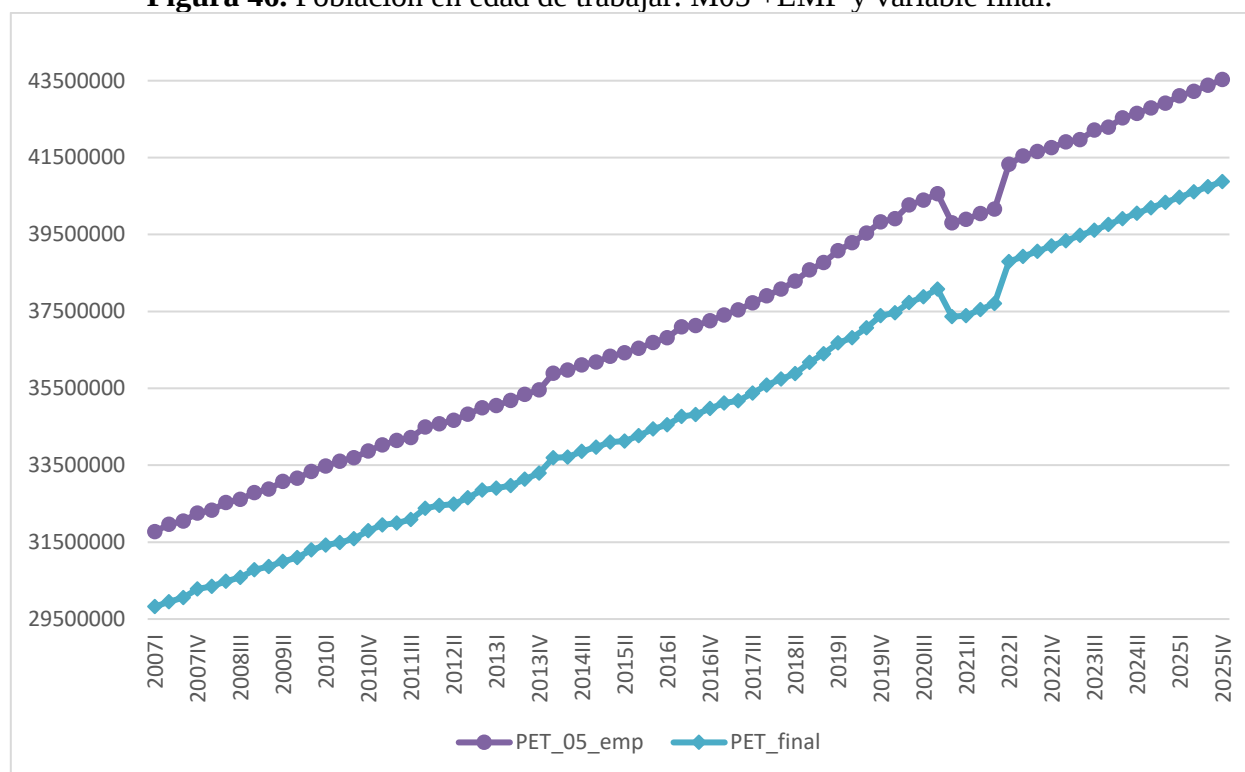
Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE.

Ahora, la variable resultante se compara con la del marco 2018 desde el periodo en el que se dispone información. La figura 45 muestra que, a diferencia de lo observado en variables como la población desocupada o la fuerza de trabajo, aquí la discrepancia entre las dos series es amplia, persistente y sistemática. La brecha es una manifestación directa del cambio conceptual introducido en el marco 2018. En particular, la modificación en el umbral de edad, abordado en secciones anteriores, redefine el universo de referencia de la población en edad de trabajar. De este modo, ambas series no serían directamente comparables. Sin embargo, con la metodología propuesta, un segundo empalme podría resolver esta situación.

Figura 45. Población en edad de trabajar: M05 +EMP y M18.

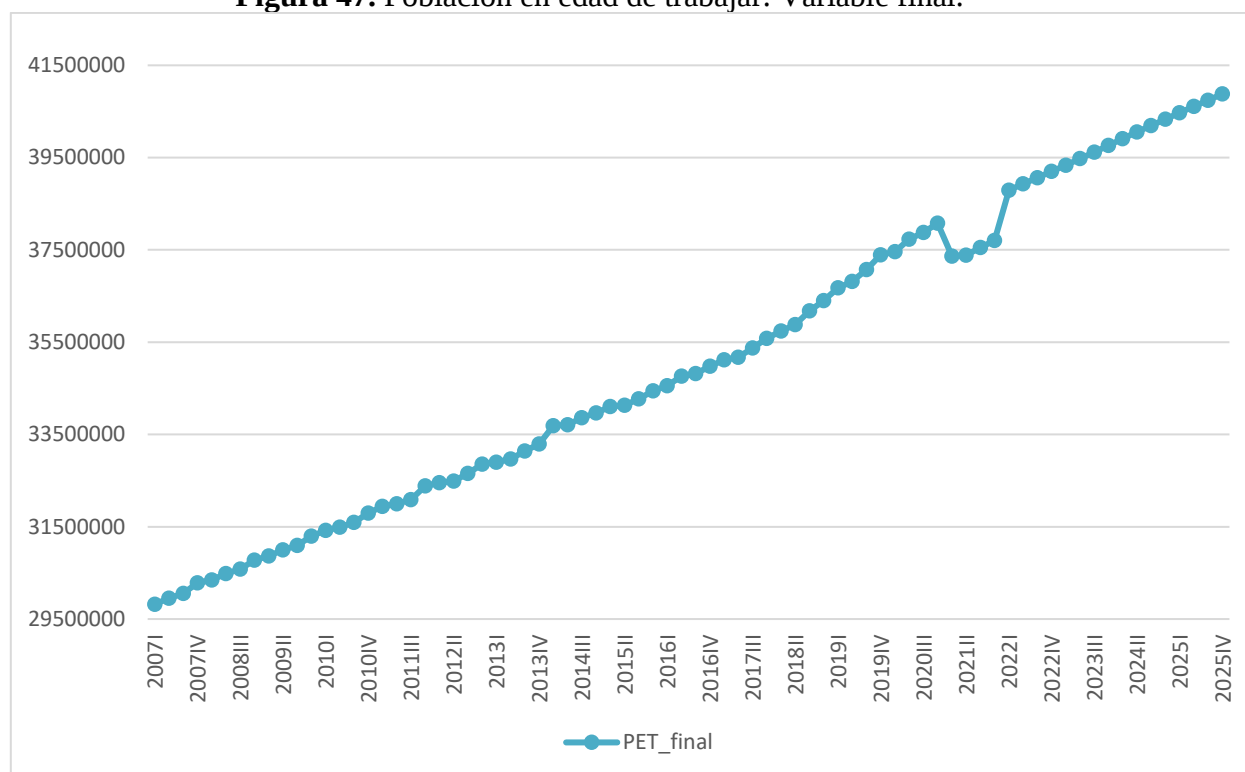
Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE.

Con este contraste claro, se realiza el segundo empalme al integrar PET_05_EMP y PET_18. La magnitud de este ajuste se puede observar en la figura 46, en la cual se compara la serie resultado del primer empalme, con la serie final. La diferencia entre ambas es considerable a lo largo del horizonte temporal. Esto puede sugerir que el cambio conceptual no es un simple detalle técnico menor, por el contrario, una transformación sustantiva en la definición de la PET. Todo esto implica que la serie final no solo corrige una diferencia de escala derivada del cambio muestral, sino que permite expresar nuevamente toda la trayectoria histórica de la variable bajo el nuevo criterio etario adoptado en el marco 2018.

Figura 46. Población en edad de trabajar: M05 +EMP y variable final.

Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE.

Finalmente, la figura 47 presenta la serie final de la población en edad de trabajar. Esta variable recoge, de forma simultánea, los dos ajustes requeridos para garantizar comparabilidad histórica: la corrección muestral entre la serie del marco 2005 y la serie de empalme; y la incorporación del componente conceptual derivado del cambio en la definición por edad. Por lo tanto, esta serie constituye una mejor aproximación disponible para el análisis temporal de la población en edad de trabajar, dado que expresa una trayectoria homogénea y consistente bajo un único marco conceptual.

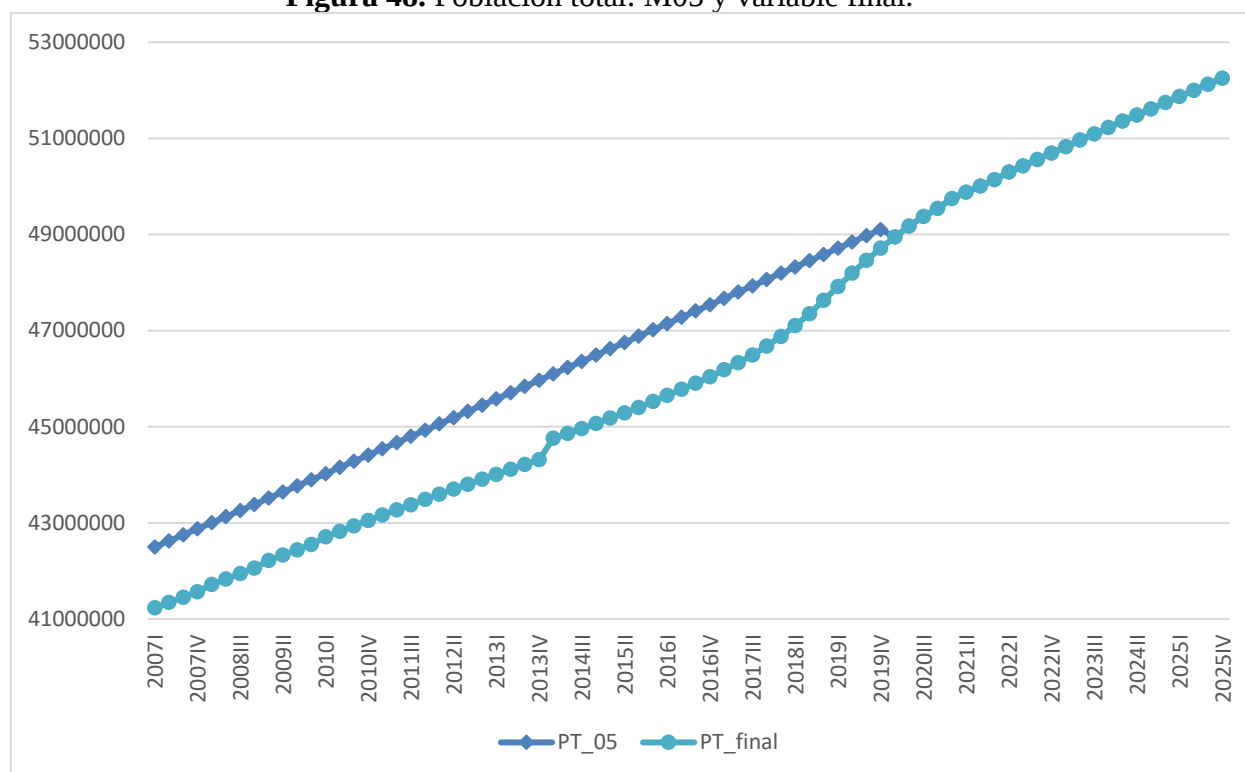
Figura 47. Población en edad de trabajar: Variable final.

Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE.

6. Población Total

Para construir la serie de la población total, se sigue el mismo esquema descrito a lo largo de este documento. En este caso, solo se menciona un elemento central del ajuste: este no radica en un cambio conceptual fuerte en la definición de la variable, sino en la necesidad de corregir el componente muestral. De esta forma, la trayectoria histórica de la población total puede resultar estadísticamente comparable a lo largo del tiempo.

Como se observa en la siguiente figura, la serie del marco 2005 se ubica por encima de la variable final durante buena parte del horizonte temporal. Esta diferencia puede reflejar los efectos del diseño muestral y del factor de expansión asociado al marco 2005. En este sentido, nuestra metodología de empalme permite ajustar la serie histórica y trasladarla a una trayectoria consistente con la nueva base muestral. La evidencia gráfica puede sugerir que, luego de aplicar el procedimiento, la variable final presenta una evolución continua y coherente, pues corrige la sobreestimación de la serie original y, además, puede garantizar una mejor comparabilidad intertemporal.

Figura 48. Población total: M05 y variable final.

Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE.

Si se piensa en términos analíticos, esta serie puede entenderse como el resultado de la metodología orientada fundamentalmente a depurar las diferencias derivadas del cambio en el marco muestral. Es por eso por lo que, aunque el procedimiento general es análogo a las demás variables, en este caso no resulta necesario enfatizar en un ajuste conceptual adicional. Por el contrario, se destaca que la metodología permite tener una serie de la población total homogénea, consistente y adecuada para el análisis de largo plazo.

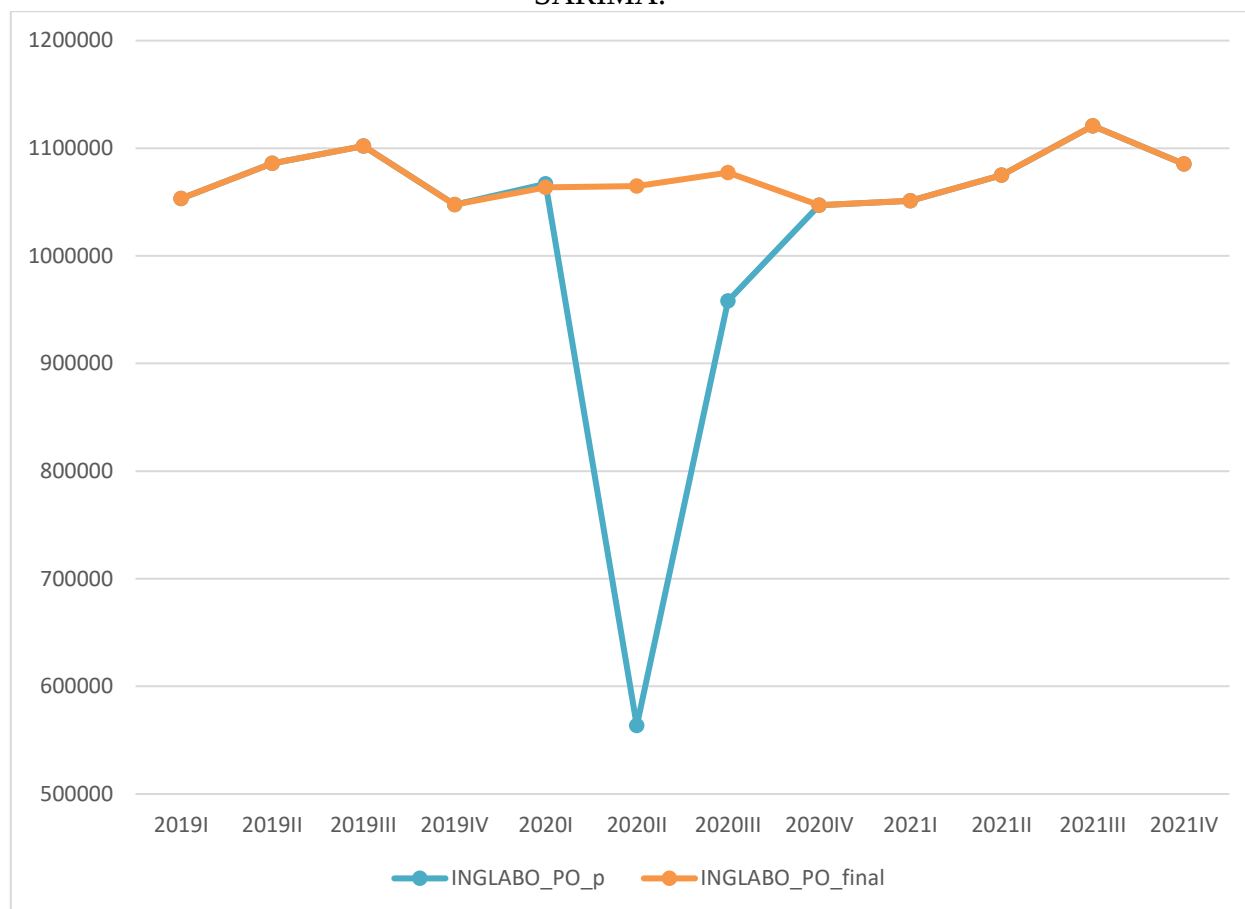
7. Ingresos laborales

7.1. Ingresos laborales de la población ocupada

En el caso de los ingresos laborales promedio de la población ocupada, la serie también se somete al mismo proceso de empalme y armonización descrito para las demás variables del documento. Sin embargo, en esta sección no se detallará nuevamente ese procedimiento. Por el contrario, se enfatiza principalmente en el ajuste por imputación de valores faltantes mediante modelos SARIMA (descrito anteriormente), dado que este constituye un tratamiento concreto, y con resultados interesantes, que es necesario para corregir la inconsistencia observada en el periodo de pandemia. De esta manera, el objetivo de esta etapa es garantizar una variable final de los ingresos que preserve coherencia temporal y resulte con la lógica general de armonización adoptada en este documento.

Se identifica una alteración atípica en la serie en los tres primeros trimestres de 2020 que se podrían asociar al momento de mayor disrupción de la pandemia. Frente a esto, se aplica el mismo procedimiento utilizado, por ejemplo, en la población ocupada informal: reemplazar los valores inconsistentes por estimaciones coherentes con la estructura temporal de la serie, se considera tanto su tendencia como su componente estacional.

Figura 49. Ingresos laborales de ocupados: Imputación de valores faltantes con un modelo SARIMA.

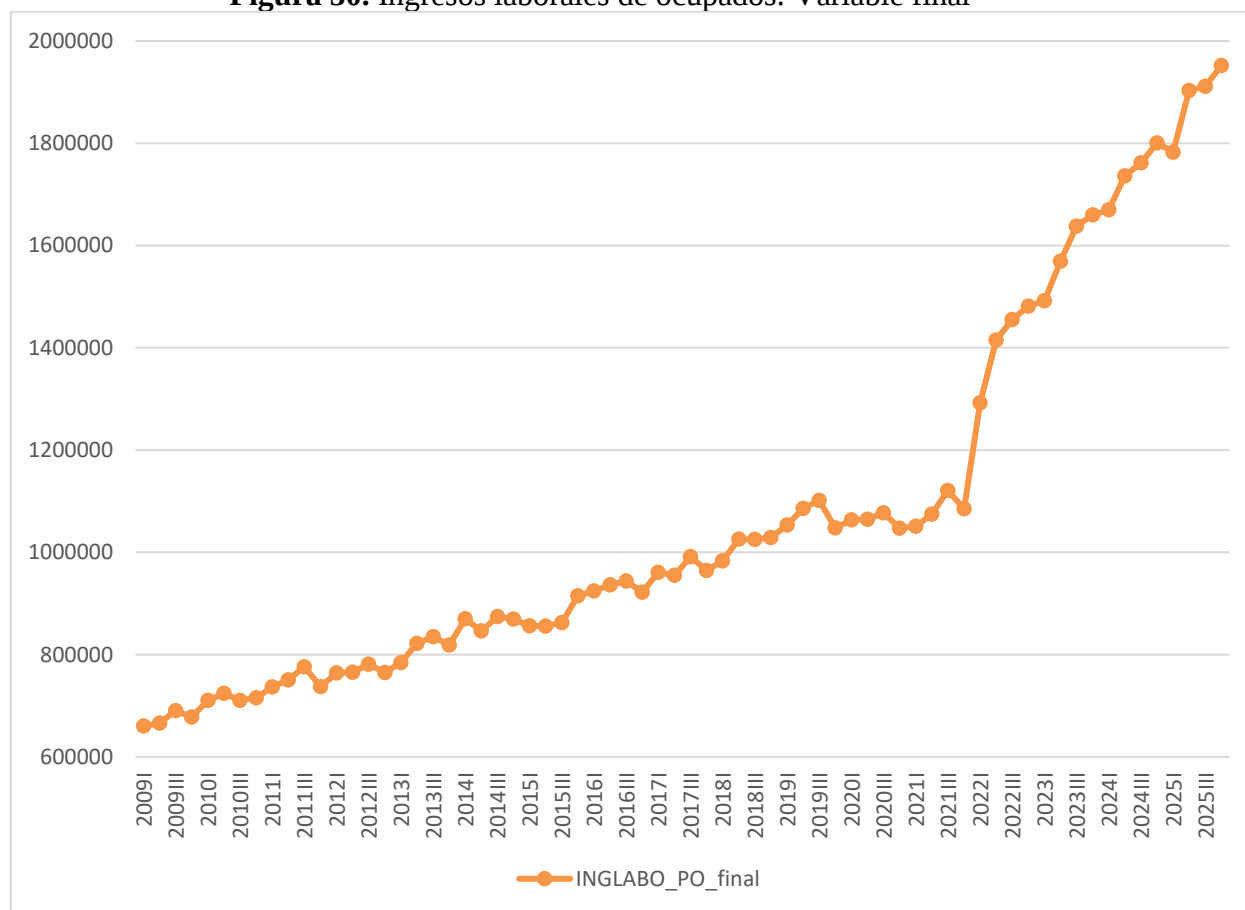


Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE.

La figura 49 permite observar con claridad el procedimiento. Se compara la serie empalmada con la serie empalmada (INGLABO_PO_p) y ajustada con la imputación (INGLABO_PO_final). La diferencia se concentra en los tres primeros trimestres afectados por la inconsistencia, en la cual, la serie preliminar presenta una caída abrupta que rompe con la evolución general de los ingresos laborales promedio de los ocupados. En contraste, la trayectoria ajustada restablece una dinámica mucho más estable y plausible al suavizar ese quiebre.

Luego, la serie resultante se toma como variable final de ingresos laborales promedio de la población ocupada. En la siguiente figura, se observa una evolución consistente a lo largo del tiempo, en la cual el ajuste aplicado evitaba una ruptura artificial en el periodo de pandemia y permite conservar una senda consistente. Para un mayor análisis, es necesario quitar el componente estacional y por supuesto, deflactar la variable.

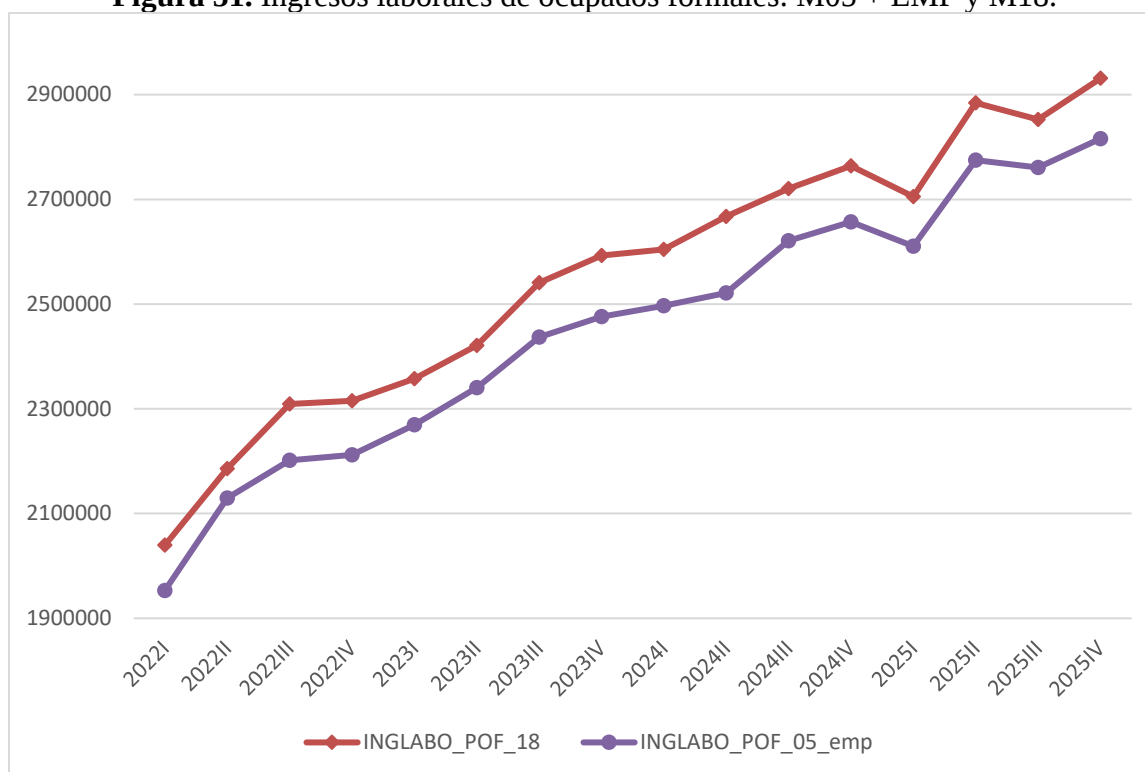
Figura 50. Ingresos laborales de ocupados: Variable final



Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE.

7.2. Ingresos laborales de la población ocupada formal

Para los ingresos laborales promedio de los ocupados formales, la construcción de la serie sigue el mismo esquema de empalme y armonización aplicado anteriormente. Sin embargo, en este caso resulta interesante destacar el comportamiento observado en el segundo empalme. Esto es, la comparación entre la serie construida bajo el primer empalme (M05+EMP) y la serie correspondiente al marco 2018. Esto es relevante porque se identifica que, aun después de corregir el componente muestral, persiste una diferencia apreciable en el nivel de los ingresos laborales promedio de los formales. Lo anterior puede sugerir una presencia adicional asociado al nuevo marco conceptual de medición.

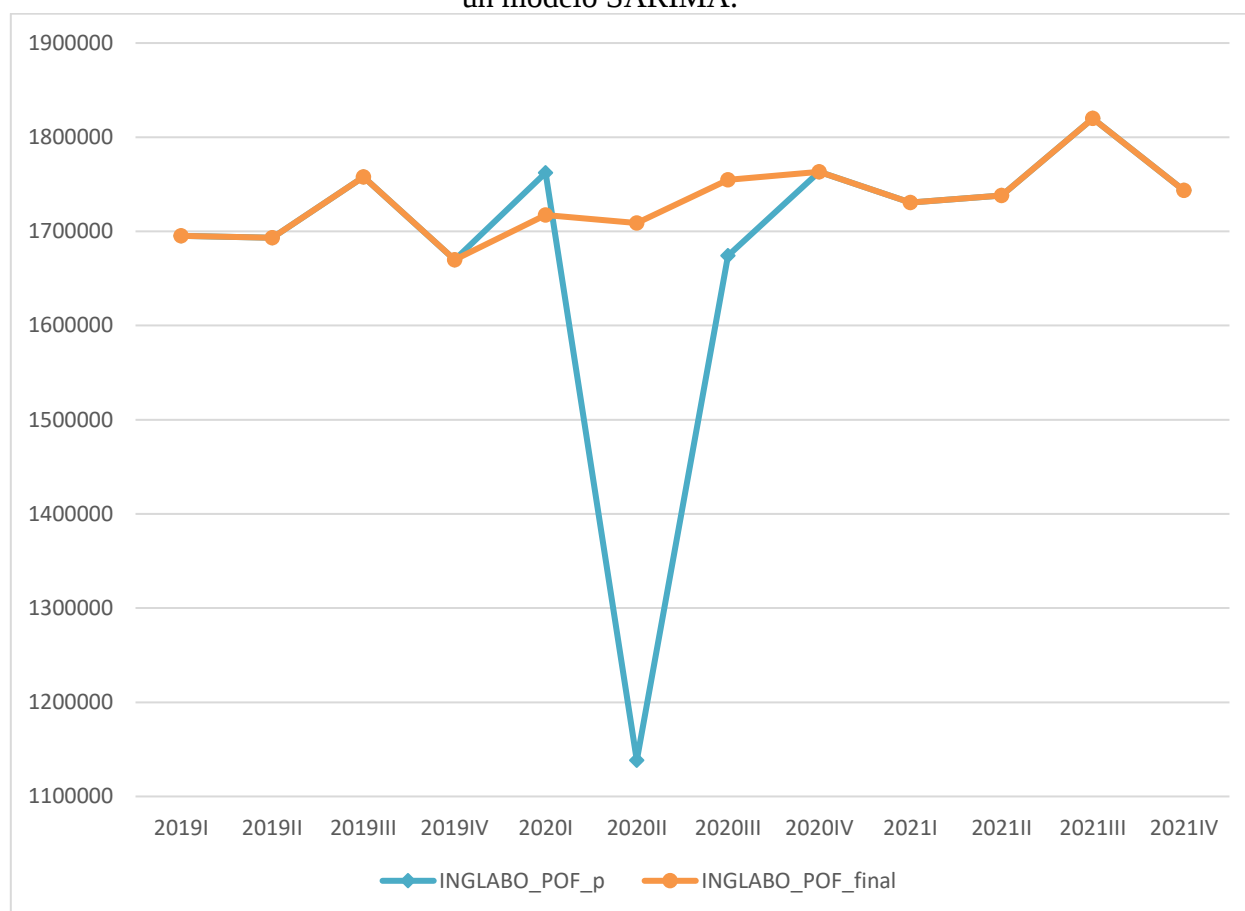
Figura 51. Ingresos laborales de ocupados formales: M05 + EMP y M18.

Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE.

La figura 51, permite observar que la serie del marco 2018 (INGLABO_POF_18) se ubica por encima de la serie resultante del primer empalme (INGLABO_POF_05_EMP). Por lo tanto, se evidencia una brecha importante entre ambas trayectorias. Desde el punto de vista metodológico, este resultado indica que el segundo empalme es una armonización necesaria para hacer ambos tramos comparables. Es decir, la diferencia que se observa confirma que el cambio de marco introduce una modificación relevante en la medición de los ingresos laborales de los formales, por lo cual la serie final debe incorporar este ajuste para garantizar una continuidad analítica.

Una vez se realiza el segundo empalme, se obtiene una serie preliminar para los ingresos laborales de los formales. Sin embargo, al igual que la variable de ingresos laborales para los ocupados, esta trayectoria presenta una inconsistencia importante en el periodo inicial de la pandemia. Por lo tanto, es necesario aplicar el mismo ajuste por imputación de valores faltantes con modelos SARIMA. El propósito no es eliminar el choque asociado a la pandemia, sino evitar que se observe sobredimensionado dentro de la serie y altere la interpretación de la dinámica empírica de los ingresos laborales promedio.

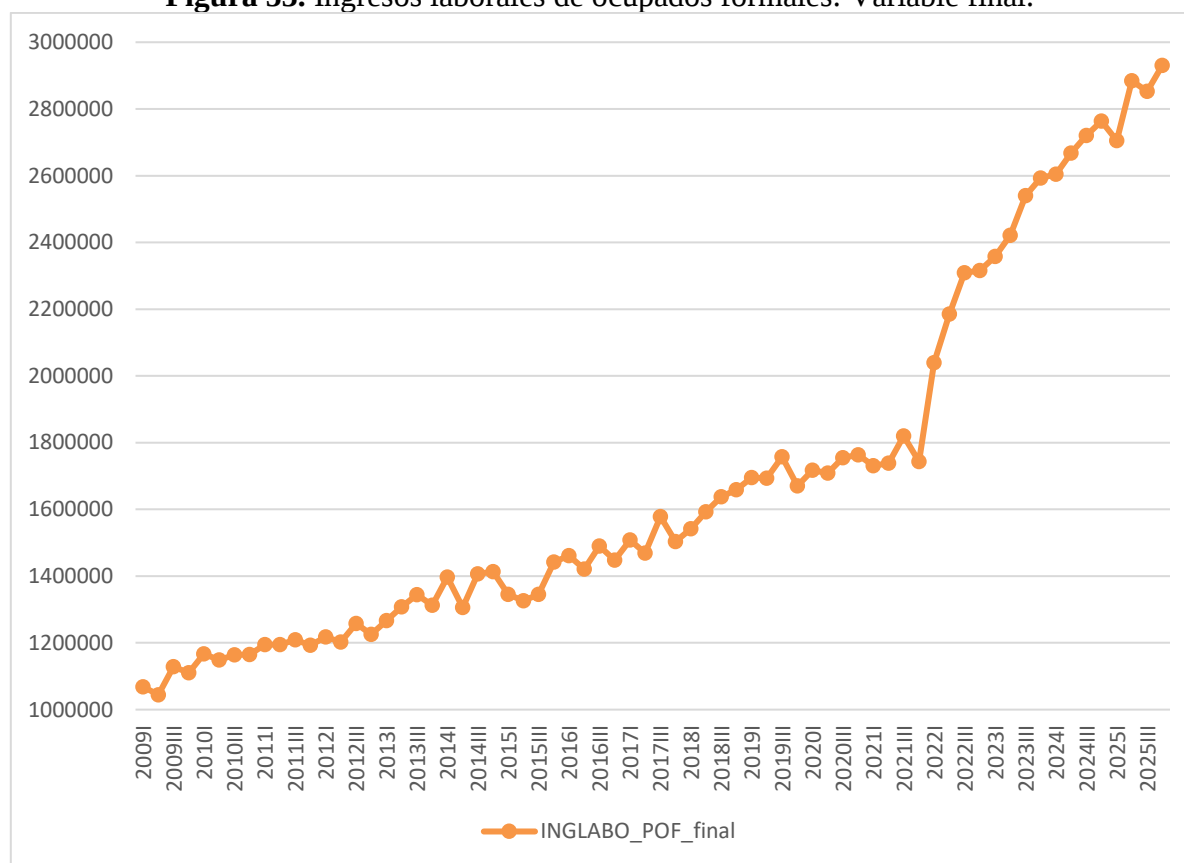
Figura 52. Ingresos laborales de ocupados formales: Imputación de valores faltantes con un modelo SARIMA.



Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE.

La figura 52 permite apreciar el efecto del ajuste. Se puede observar que la serie preliminar exhibe una caída abrupta en los trimestres más críticos de la pandemia, mientras que la serie ajustada restablece una trayectoria mucho más estable y plausible. Este procedimiento suaviza la magnitud del quiebre y respeta al mismo tiempo la evolución general de la serie antes y después del choque.

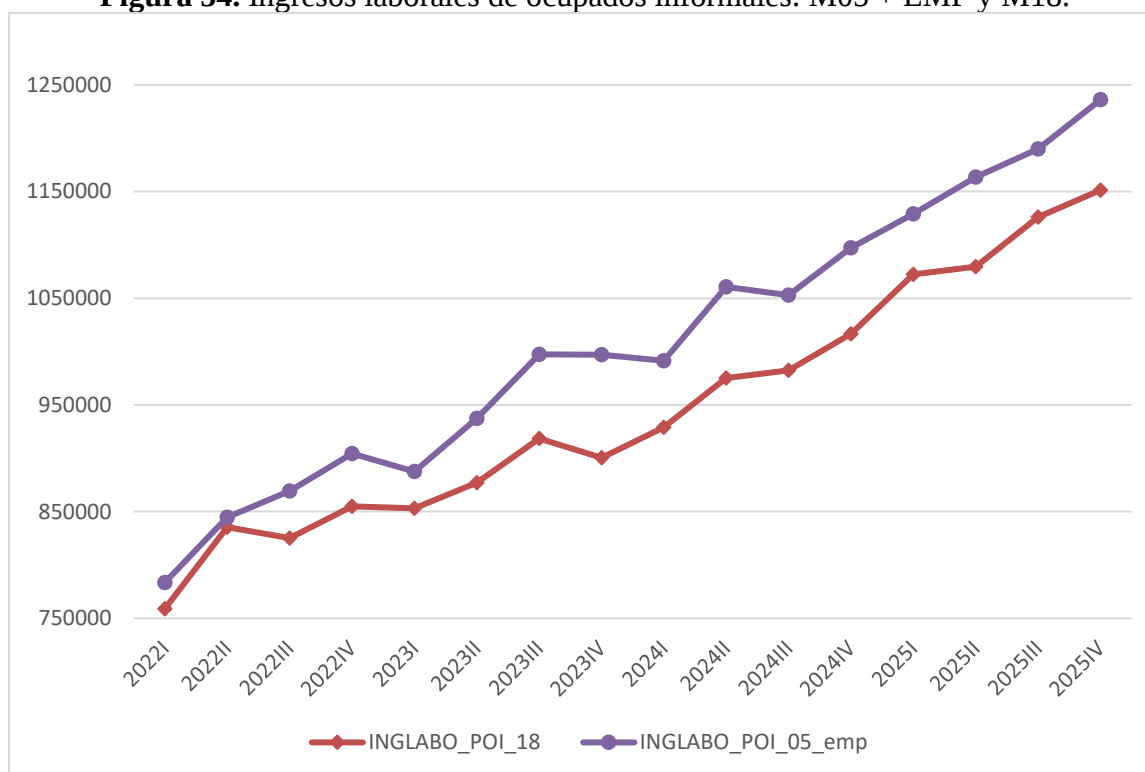
En consecuencia, la variable final de ingresos laborales promedio de los ocupados formales surge de una combinación de dos etapas complementarias: primero, el empalme muestral y conceptual entre las series disponibles bajo distintos marcos operativos y, segundo, el ajuste puntual de la inconsistencia observada en la pandemia a partir de la imputación con modelos SARIMA. La doble corrección permite obtener una serie continua y coherente, que incorpora la información contenida en los distintos marcos y, además, evita un valor atípico que condicione de forma desproporcional la lectura de la serie.

Figura 53. Ingresos laborales de ocupados formales: Variable final.

Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE.

7.3. Ingresos laborales de la población ocupada informal

En cuanto a los ingresos laborales promedio de los ocupados informales, se sigue un procedimiento análogo al anterior. Se parte del mismo esquema de empalme y armonización entre las series disponibles bajo distintos marcos estadísticos. En este caso, resulta interesante observar el comportamiento del segundo empalme, ya que esta superposición permite identificar el cambio de nivel que introduce la nueva medición.

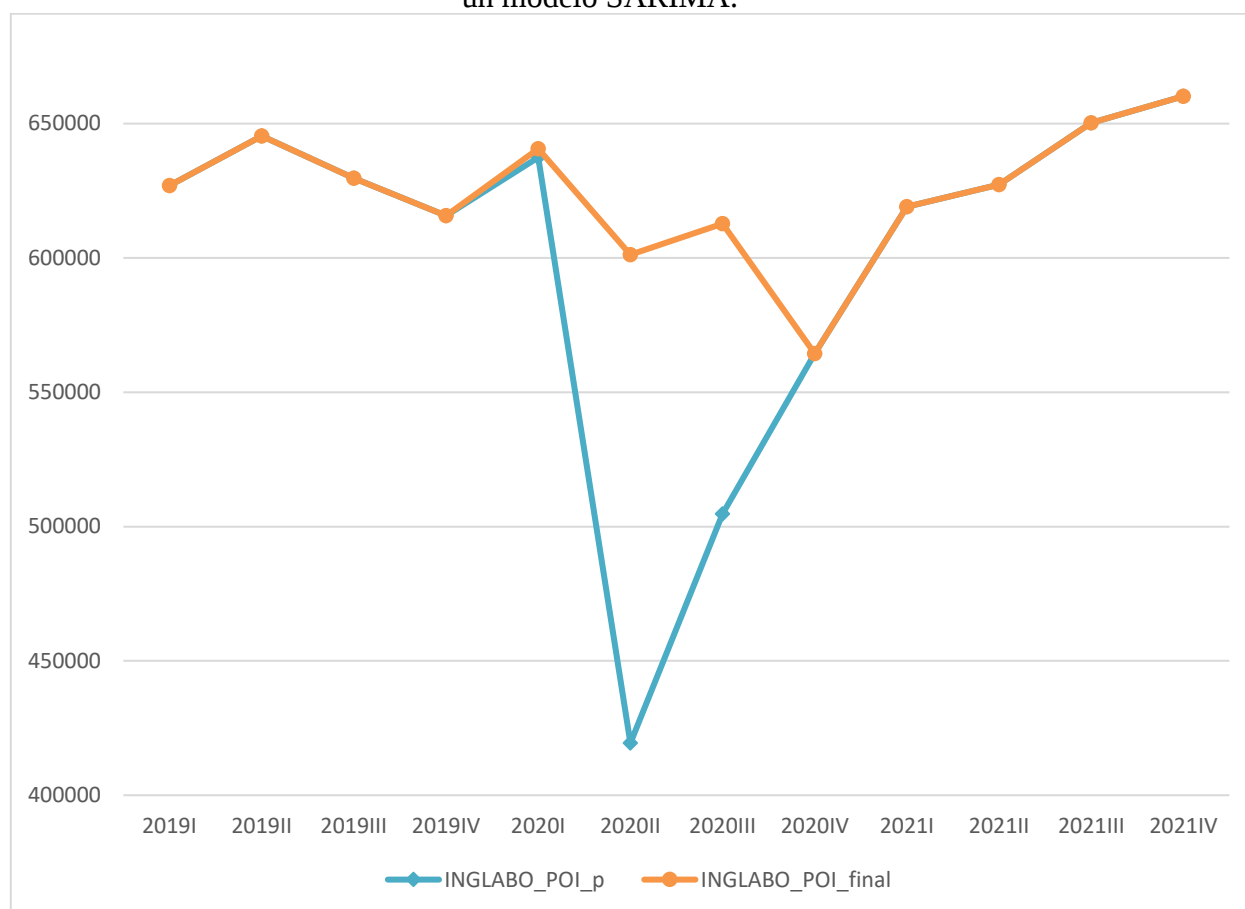
Figura 54. Ingresos laborales de ocupados informales: M05 + EMP y M18.

Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE.

La figura 54 muestra que, durante el periodo de traslape, la serie del primer empalme (INGLABO_POI_05_EMP) se ubica por encima de la serie del marco 2018 (INGLABO_POI_18), lo que evidencia una brecha sistemática entre ambas trayectorias. Esto puede sugerir que persiste una diferencia asociada al cambio en el marco conceptual. Por lo tanto, el segundo empalme es necesario para integrar ambos tramos en una sola trayectoria comparable.

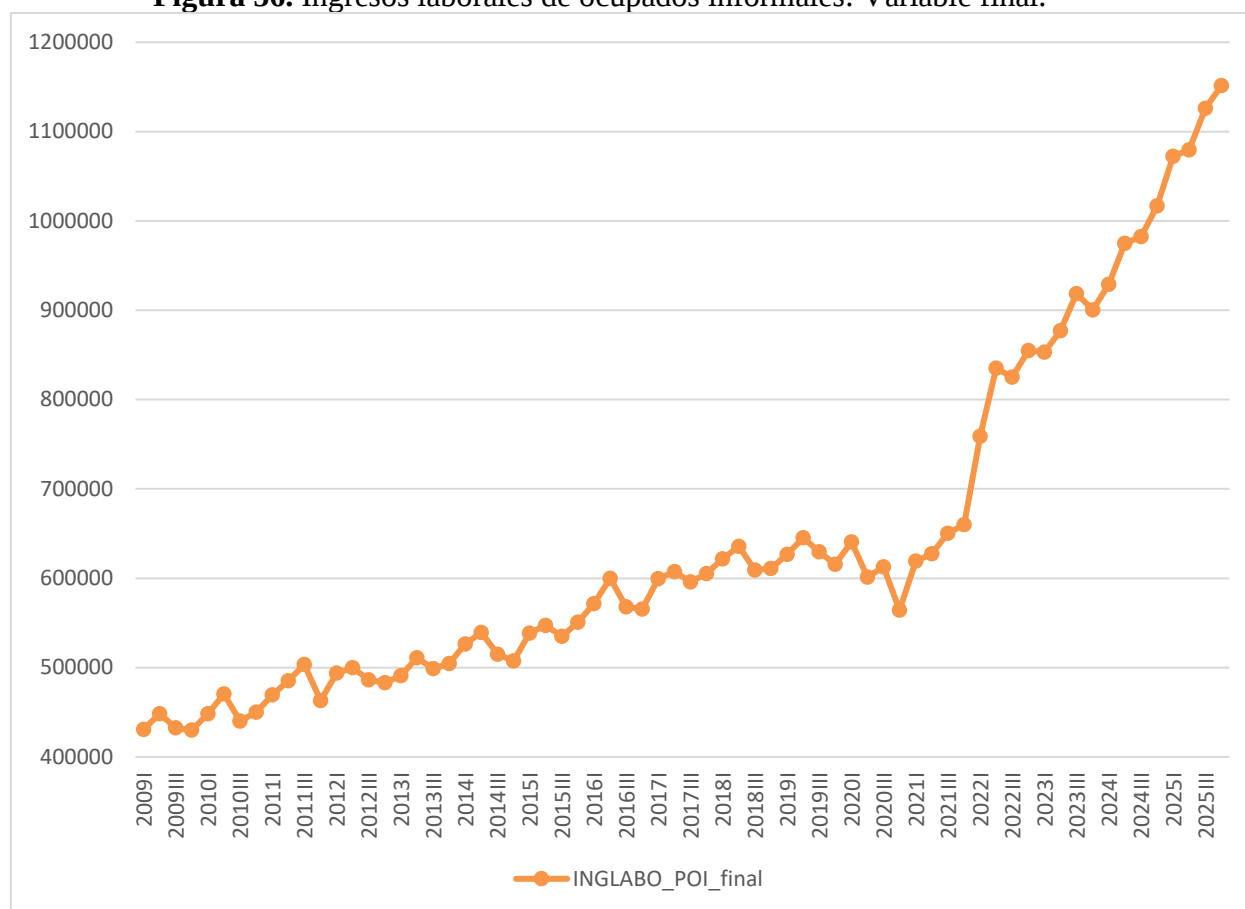
Luego del empalme, se aplica el ajuste adicional por imputación de valores faltantes con la identificación de modelos SARIMA, con el fin de corregir la inconsistencia que se ha observado en el periodo inicial de pandemia.

Figura 55. Ingresos laborales de ocupados informales: Imputación de valores faltantes con un modelo SARIMA.



Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE.

En la figura 55 se puede observar la caída pronunciada en los trimestres más críticos de la serie preliminar, en contraparte, la serie ajustada suaviza ese descenso y restablece una evolución más coherente con el comportamiento previo de la variable. En consecuencia, la imputación reduce este efecto y mejora la comparabilidad sin desconocer el choque externo. De esta manera, se toma este resultado como la variable final de ingresos laborales promedio de los ocupados formales. Esta variable, como la anterior, también resulta de la combinación de las dos grandes etapas ya mencionadas. Así, la serie final integra la información de diferentes marcos, en términos muestrales y conceptuales y, además, está corregida y apta para el análisis. En la siguiente figura, se puede observar esta trayectoria final.

Figura 56. Ingresos laborales de ocupados informales: Variable final.

Fuente: Elaboración propia a partir de la propuesta metodológica con base en GEIH-DANE.

VII. Conclusiones

En primer lugar, este trabajo se propuso caracterizar completamente la Gran Encuesta Integrada de Hogares. Desde sus componentes muestrales, asociados a los factores de expansión derivados del Censo Nacional de Población y Vivienda; conceptuales, asociados a las definiciones internacionales sobre el mercado laboral y adaptadas a Colombia; y metodológicos, asociados a la recolección, procesamiento y cálculo de datos para las variables presentadas. Esto es valioso puesto que, desde un inicio, se plantea una guía técnica completa que permita entender desde sus cimientos hasta la aplicación metodológica propuesta, la estructura de la GEIH.

Luego, la metodología de empalme propuesta y en aplicación, permitió construir una base de datos con series armonizadas, comparables a lo largo del tiempo y por supuesto, metodológica y conceptualmente consistentes para el análisis del mercado laboral colombiano. Esto se logra a partir de la integración de diversa información que yace en los distintos marcos de la GEIH. Esta propuesta parte del reconocimiento de que las rupturas

observadas en las series a lo largo del tiempo no responden a un solo cambio, como las fluctuaciones de los ciclos económicos, sino también a las siguientes transformaciones: el factor de expansión, el diseño muestral y, en algunos casos, en los criterios conceptuales de medición y clasificación de la población del mercado laboral colombiano. El procedimiento se puede resumir, fundamentalmente, en dos etapas. Primero, se realizó la corrección del componente muestral al integrar las series del marco 2005 con las series de empalme. Esto permitió depurar, consistentemente, las diferencias asociadas al factor de expansión y el rediseño muestral. Este paso resulta en una primera versión armonizada de las series. Segundo, se incorpora en la serie inicialmente corregida el componente conceptual. Esto se logra a través de la integración entre la integración de las series resultantes del primer paso y las series del marco 2018. En este paso surge una conclusión interesante: la magnitud del ajuste conceptual no es homogéneo entre variables. Esto es, mientras en algunas series es reducida o casi nula, en otras resulta decisiva para garantizar la comparabilidad histórica.

Como resultado de la metodología propuesta, se consigue una serie comparable y consistente para cada rubro del mercado laboral. Para la población ocupada total, la metodología corrigió tanto el componente muestral como el conceptual. Pero resulta mucho más interesante el caso de la población ocupada informal, que además de ser corregida en ambos componentes, fue necesario introducir un ajuste adicional sobre las observaciones de los tres primeros trimestres del 2020, donde se identificó una inconsistencia considerable asociada al choque del COVID-19. En este caso, se calculó inicialmente la proporción de informalidad respecto de la población ocupada —ambas variables con la aplicación de la metodología de empalme propuesta—, para luego imputar sus valores faltantes a través de modelos SARIMA, y con este resultado, multiplicar la proporción ajustada por la población ocupada total: de esta forma se obtuvo una serie de informalidad más realista y consistente. Con esta corrección, se reconstruyó la serie final de la población ocupada formal: se calculó la proporción de formalidad como el complemento de la proporción ajustada de informalidad para luego calcular el valor de la serie en niveles. Se observa sobre todo un efecto de re-distribución de las cifras de ocupación informal y formal, sin cambiar las cifras de ocupación total resultantes de la aplicación metodológica propuesta. La necesidad y justificación de este ajuste de imputación es la siguiente: las cifras originales de la GEIH para esos trimestres muestran dos hechos contraintuitivos: un aumento exagerado del empleo formal y una caída igualmente exagerada del empleo informal. Es inconsistente que, durante tres trimestres de encierro y parálisis económica, el empleo formal haya aumentado, habida cuenta del proceso de destrucción de empresas que esto generó. De igual manera, tampoco es plausible la gran caída que se observa en el empleo informal, que con certeza también disminuyó. Como resultado, la proporción de formalidad aumentó abruptamente, con la consecuente caída también abrupta de la tasa de informalidad.

En cuanto a los desocupados, la aplicación metodológica propuesta mostró un resultado distinto y particularmente importante para el trabajo. Esto es, aunque se desarrollara el procedimiento de empalme completo, el primer resultado resultó más que suficiente —es decir, el resultado del primer empalme es análogo, si se realiza, con el resultado del segundo empalme—. Esto se podría explicar porque, pese a los cambios muestrales, no hay cambios conceptuales de medición considerables. En la fuerza de trabajo también se siguió el

procedimiento completo de armonización. Sin embargo, se adoptó un criterio contable coherente con el componente conceptual: la suma de las series finales de la población ocupada y desocupada. Es importante mencionar también que, si bien la brecha entre la serie con los empalmes y la serie recalculada era pequeña, esta última construcción garantizó mayor consistencia interna. Para la población en edad de trabajar, el componente conceptual fue determinante dado el gran cambio de rango de edad entre marcos. Por lo tanto, el segundo empalme fue fundamental para la comparabilidad histórica. Por último, en cuanto a la población total, el rasgo central de su ajuste fue sobre el componente muestral. Esto permitió depurar la sobreestimación relativa de la serie original y obtener una trayectoria final consistente.

Para los ingresos laborales promedio, se aplicó de igual forma el mismo esquema general de armonización complementado, cuando fue necesario, con ajustes por imputación de valores faltantes. Para los ingresos laborales promedio de la población ocupada, se aplicó un ajuste mediante modelos SARIMA sobre los tres primeros trimestres —asociados también a la pandemia— para obtener una serie final más consistente; esto fue necesario porque al ajustar por medio de valores imputados las cifras de empleo formal e informal, los promedios salariales calculados directamente desde la información de la GEIH, para los primeros tres trimestres de 2020, también son inconsistentes. Para el caso de los formales e informales, luego de los empalmes propuestos, fue necesario también un ajuste con modelos SARIMA.

Entonces, como se evidencia a lo largo de este trabajo, la propuesta metodológica no solo se limita a unir, de forma mecánica, las series provenientes de distintos marcos estadísticos, sino que ofrece una ruta de armonización capaz de distinguir entre cambios los componentes muestrales y conceptuales, y además valores atípicos como los observados en pandemia. La aplicación de esta metodología mostró que cada variable exige un análisis concreto, esto es, para algunas el ajuste muestral es suficiente, en otras, el componente conceptual redefine completamente la comparabilidad, y en algunas más, se requirieron criterios metodológicos adicionales de consistencia contable o ajustes por imputación. El resultado final es, por tanto, un conjunto de series de población e ingresos laborales promedio mucho más sólido, coherente, comparable y disponible para los análisis descriptivos o econométricos del mercado laboral colombiano.

Referencias bibliográficas

Castaño, E. (2010). Identificación de un modelo ARIMA cuando existen observaciones faltantes. *Lecturas de Economía*, 47, 25–45. <https://doi.org/10.17533/udea.le.n47a4935>

DANE. (2014). *Medición del empleo informal y seguridad social*. https://www.dane.gov.co/files/investigaciones/boletines/ech/ech_informalidad/bol_ech_informalidad_ene_mar2014.pdf

DANE. (2019). *Formulario Gran Encuesta Integrada de Hogares*. <https://microdatos.dane.gov.co/index.php/catalog/599/related-materials>

DANE. (2022a). *Empalme de series: GEIH 2005 - GEIH 2018*. https://www.dane.gov.co/files/investigaciones/boletines/ech/ech/Empalme_%20GEIH_%202005-2018.pdf

DANE. (2022b). *Enfoque conceptual y metodológico de la Gran Encuesta Integrada de Hogares - GEIH*. https://www.dane.gov.co/files/investigaciones/fichas/empleo/Presentacion_nuevo_enfoque_GEIH.pdf

DANE. (2022c). *Nueva medición de informalidad laboral*. https://www.dane.gov.co/files/investigaciones/boletines/ech/ech/Nueva_medicion_informalidad.pdf

DANE. (2022d). *Principales indicadores del mercado laboral - Diciembre de 2021*. https://www.dane.gov.co/files/investigaciones/boletines/ech/ech/bol_empleo_dic_21.pdf

DANE. (2024). *Formulario Gran Encuesta Integrada de Hogares*. <https://microdatos.dane.gov.co/index.php/catalog/819/related-materials>

DANE. (2026a). *Ocupación informal. Trimestre octubre - diciembre 2025*. <https://www.dane.gov.co/files/operaciones/GEIH/bol-GEIHEISS-oct-dic2025.pdf>

DANE. (2026b). *Principales indicadores del mercado laboral. Diciembre 2025*. <https://www.dane.gov.co/files/operaciones/GEIH/bol-GEIH-dic2025.pdf>

DIMPE, & DANE. (2016). *Ficha metodológica Gran Encuesta Integrada de Hogares*. https://www.dane.gov.co/files/investigaciones/fichas/empleo/ficha_metodologica_GEIH-01_V10.pdf

DIMPE, & DANE. (2022). *Nota técnica: Empalme series GEIH marco 2005 - GEIH marco 2018*. <https://www.dane.gov.co/files/investigaciones/boletines/ech/ech/Nota-tecnica-empalme-series-GEIH.pdf>

DIMPE, & DANE. (2023). *Metodología general Gran Encuesta Integrada de Hogares GEIH*. <https://www.dane.gov.co/files/operaciones/GEIH/DSO-GEIH-MET-001-V11.pdf>

Hart, K. (1973). Informal Income Opportunities and Urban Employment in Ghana. *The Journal of Modern African Studies*, 11(I), 61–89. <https://www.jstor.org/stable/159873>

Hermida, D., & Pulido-Mahecha, K. (2022). *Recuadro 2: Actualización metodológica de la Gran encuesta integrada de hogares (GEIH) y sus principales cambios*. <https://www.banrep.gov.co/es/publicaciones-investigaciones/informe-politica-monetaria/recuadro2-abril-2022>

International Labour Office. (1982). *Statistics of Labour Force, Employment, Unemployment and Underemployment*. https://webapps.ilo.org/public/libdoc/ilo/1982/82B09_438_engl.pdf

International Labour Office. (1993). *Statistics of employment in the informal sector*. https://webapps.ilo.org/public/libdoc/ilo/1992/92B09_385_engl.pdf

International Labour Office. (1998). *The measurement of underemployment*. https://webapps.ilo.org/public/libdoc/ilo/1998/98B09_207_engl.pdf

International Labour Office. (2003). *Guidelines concerning a statistical definition of informal employment*. https://www.ilo.org/sites/default/files/wcmsp5/groups/public/@dgreports/@stat/documents/normativeinstrument/wcms_087622.pdf

International Labour Office. (2013). *Resolución I: Resolución relativa a las estadísticas de trabajo, la ocupación y la subutilización de la fuerza de trabajo*.

International Labour Office. (2018). *Resolución sobre las estadísticas de las relaciones de trabajo*. https://www.ilo.org/sites/default/files/wcmsp5/groups/public/@dgreports/@stat/documents/meetingdocument/wcms_647367.pdf

Ley 515 de 1999 (1999). <https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=6300>

Borradores del CIE

No.	Título	Autor(es)	Fecha
01	Organismos reguladores del sistema de salud colombiano: conformación, funcionamiento y responsabilidades.	Durfari Velandia Naranjo Jairo Restrepo Zea Sandra Rodríguez Acosta	Agosto de 2002
02	Economía y relaciones sexuales: un modelo económico, su verificación empírica y posibles recomendaciones para disminuir los casos de sida.	Marcela Montoya Múnera Danny García Callejas	Noviembre de 2002
03	Un modelo RSDAIDS para las importaciones de madera de Estados Unidos y sus implicaciones para Colombia	Mauricio Alviar Ramírez Medardo Restrepo Patiño Santiago Gallón Gómez	Noviembre de 2002
04	Determinantes de la deserción estudiantil en la Universidad de Antioquia	Johanna Vásquez Velásquez Elkin Castaño Vélez Santiago Gallón Gómez Karoll Gómez Portilla	Julio de 2003
05	Producción académica en Economía de la Salud en Colombia, 1980-2002	Karem Espinosa Echavarría Jairo Humberto Restrepo Zea Sandra Rodríguez Acosta	Agosto de 2003
06	Las relaciones del desarrollo económico con la geografía y el territorio: una revisión.	Jorge Lotero Contreras	Septiembre de 2003
07	La ética de los estudiantes frente a los exámenes académicos: un problema relacionado con beneficios económicos y probabilidades	Danny García Callejas	Noviembre de 2003
08	Impactos monetarios e institucionales de la deuda pública en Colombia 1840-1890	Angela Milena Rojas R.	Febrero de 2004
09	Institucionalidad e incentivos en la educación básica y media en Colombia	David Fernando Tobón Germán Darío Valencia Danny García Guillermo Pérez Gustavo Adolfo Castillo	Febrero de 2004
10	Selección adversa en el régimen contributivo de salud: el caso de la EPS de Susalud	Johanna Vásquez Velásquez Karoll Gómez Portilla	Marzo de 2004
11	Diseño y experiencia de la regulación en salud en Colombia	Jairo Humberto Restrepo Zea Sandra Rodríguez Acosta	Marzo de 2004
12	Economic Growth, Consumption and Oil Scarcity in Colombia: A Ramsey model, time series and panel data approach	Danny García Callejas	Marzo de 2005
13	La competitividad: aproximación conceptual desde la teoría del crecimiento y la geografía económica	Jorge Lotero Contreras Ana Isabel Moreno Monroy Mauricio Giovanni Valencia Amaya	Mayo de 2005
14	La curva Ambiental de Kuznets para la calidad del agua: un análisis de su validez mediante raíces unitarias y cointegración	Mauricio Alviar Ramírez Catalina Granda Carvajal Luis Guillermo Pérez Puerta Juan Carlos Muñoz Mora Diana Constanza Restrepo Ochoa	Mayo de 2006
15	Integración vertical en el sistema de salud colombiano: Aproximaciones empíricas y análisis de doble marginalización	Jairo Humberto Restrepo Zea John Fernando Lopera Sierra Sandra Rodríguez Acosta	Mayo de 2006
16	Cliometrics: a market account of a scientific community (1957-2005)	Angela Milena Rojas	Septiembre de 2006
17	Regulación ambiental sobre la contaminación vehicular en Colombia: ¿hacia dónde vamos?	David Tobón Orozco Andrés Felipe Sánchez Gandur Maria Victoria Cárdenas Londoño	Septiembre de 2006

18	Biology and Economics: Metaphors that Economists usually take from Biology	Danny García Callejas	Septiembre de 2006
19	Perspectiva Económica sobre la demanda de combustibles en Antioquia	Elizeth Ramos Oyola Maria Victoria Cárdenas Londoño David Tobón Orozco	Septiembre de 2006
20	Caracterización económica del deporte en Antioquia y Colombia: 1998-2001	Ramón Javier Mesa Callejas Rodrigo Arboleda Sierra Ana Milena Olarte Cadavid Carlos Mario Londoño Toro Juan David Gómez Gonzalo Valderrama	Octubre de 2006
21	Impacto Económico de los Juegos Deportivos Departamentales 2004; el caso de Santa Fe De Antioquia	Ramón Javier Mesa Callejas Ana Milena Olarte Cadavid Nini Johana Marín Rodríguez Mauricio A. Hernández Monsalve Rodrigo Arboleda Sierra	Octubre de 2006
22	Diagnóstico del sector deporte, la recreación y la educación física en Antioquia	Ramón Javier Mesa Callejas Rodrigo Arboleda Sierra Juan Francisco Gutiérrez Betancur Mauricio López González Nini Johana Marín Rodríguez Nelson Alveiro Gaviria García	Octubre de 2006
23	Formulación de una política pública para el sector del deporte, la recreación y la educación física en Antioquia	Ramón Javier Mesa Callejas Rodrigo Arboleda Sierra Juan Francisco Gutiérrez Betancur Mauricio López González Nini Johana Marín Rodríguez Nelson Alveiro Gaviria García	Octubre de 2006
24	El efecto de las intervenciones cambiarias: la experiencia colombiana 2004-2006	Mauricio A. Hernández Monsalve Ramón Javier Mesa Callejas	Octubre de 2006
25	Economic policy and institutional change: a context-specific model for explaining the economic reforms failure in 1970's Colombia	Angela Milena Rojas	Noviembre de 2006
26	Definición teórica y medición del Comercio Intraindustrial	Ana Isabel Moreno M. Héctor Mauricio Posada D	Noviembre de 2006
Borradores Departamento de Economía			
27	Aportes teóricos al debate de la agricultura desde la economía	Marleny Cardona Acevedo Yady Marcela Barrero Amortegui Carlos Felipe Gaviria Garcés Ever Humberto Álvarez Sánchez Juan Carlos Muñoz Mora	Septiembre de 2007
28	Competitiveness of Colombian Departments observed from an Economic geography Perspective	Jorge Lotero Contreras Héctor Mauricio Posada Duque Daniel Valderrama	Abril de 2009
29	La Curva de Engel de los Servicios de Salud En Colombia. Una Aproximación Semiparamétrica	Jorge Barrientos Marín Juan Miguel Gallego Juan Pablo Saldarriaga	Julio de 2009
30	La función reguladora del Estado: ¿qué regular y por qué?: Conceptualización y el caso de Colombia	Jorge Hernán Flórez Acosta	Julio de 2009
31	Evolución y determinantes de las exportaciones industriales regionales: evidencia empírica para Colombia, 1977-2002	Jorge Barrientos Marín Jorge Lotero Contreras	Septiembre de 2009
32	La política ambiental en Colombia: Tasas retributivas y Equilibrios de Nash	Medardo Restrepo Patiño	Octubre de 2009
33	Restricción vehicular y regulación ambiental: el programa “Pico y Placa” en Medellín	David Tobón Orozco Carlos Vasco Correa Blanca Gómez Olivo	Mayo de 2010

34	Corruption, Economic Freedom and Political Freedom in South America: In Pursuit of the missing Link	Danny García Callejas	Agosto de 2010
35	Karl Marx: dinero, capital y crisis	Ghislain Deleplace	Octubre de 2010
36	Democracy and Environmental Quality in Latin America: A Panel System of Equations Approach, 1995-2008	Danny García Callejas	Noviembre de 2010
37	Political competition in dual economies: clientelism in Latin America	Angela M.Rojas Rivera	Febrero de 2011
38	Implicaciones de Forward y Futuros para el Sector Eléctrico Colombiano	Duvan Fernando Torres Gómez Astrid Carolina Arroyave Tangarife	Marzo de 2011
39	Per Capita GDP Convergence in South America, 1960-2007	Danny García Callejas	Mayo de 2011
40	Efectos del salario mínimo sobre el estatus laboral de los jóvenes en Colombia	Yenny Catalina Aguirre Botero	Agosto de 2011
41	Determinantes del margen de intermediación en el sector bancario colombiano para el periodo 2000 – 2010	Perla Escobar Julián Gómez	Septiembre de 2011
42	Tamaño óptimo del gasto público colombiano: una aproximación desde la teoría del crecimiento endógeno	Camilo Alvis Cristian Castrillón	Septiembre de 2011
43	Estimación del stock de capital humano bajo la metodología Jorgenson-Fraumeni para Colombia 2001-2009	Juan David Correa Ramírez Jaime Alberto Montoya Arbeláez	Septiembre de 2011
44	Estructura de ingresos para trabajadores asalariados y por cuenta propia en la ciudad de Ibagué	José Daniel Salinas Rincón Daniel Aragón Urrego	Noviembre de 2011
45	Identificación y priorización de barreras a la eficiencia energética: un estudio en microempresas de Medellín	Juan Gabriel Vanegas Sergio Botero Botero	Marzo de 2012
46	Medición del riesgo sistémico financiero en estudios de historia económica. Propuesta metodológica y aplicación para la banca libre en Antioquia, 1888	Javier Mejía Cubillos	Mayo de 2012
47	El tiempo, el éter que lo cubre todo: Un análisis de la temporalidad en la economía política de Karl Marx	Germán Darío Valencia Agudelo	Septiembre de 2012
48	Características de la Población Ocupada en Colombia: Un análisis del perfil de los formales e informales	José Daniel Salinas Rincón Sara Isabel González Arismendy Leidy Johana Marín	Octubre de 2012
49	Desarrollo económico Territorial: El caso del Cluster TIC, Medellín y Valle de Aburrá. Propuesta de fomento y consolidación de la industria de Contenidos Digitales	Felipe Molina Otálvaro Pablo Barrera Bolaños Tulio Montemiranda Aguirre	Noviembre de 2012
50	Análisis de la interacción entre las autoridades monetaria y fiscal en Colombia (1991-2011). Una aplicación desde la teoría de juegos	Sebastián Giraldo González Edwin Esteban Torres Gómez Ana Cristina Muñoz Toro	Enero de 2013
51	Tangible Temptation in the Social Dilema: Cash, Cooperation, and Self Control	Kristian Ove R. Myrseth Gerhard Riener Conny Wollbrant	Mayo de 2013
52	Análisis de las disparidades regionales en Colombia: una aproximación desde la estadística espacial, 1985 – 2010	Jhonny Moncada Osmar Leandro Loaiza Quintero	Octubre de 2013
53	Modelo VECM para estimar relaciones de largo plazo de un indicador de liquidez y sus determinantes	Wilman A. Gómez John F. Lopera	Noviembre de 2013
54	Informality and Macroeconomic Volatility: Do Credit Constraints Matter?	Catalina Granda Carvajal	Enero de 2015
55	¿Debería la Historia del Pensamiento Económico ser incluida en los Planes de Estudio de Economía en Pregrado?	Alessandro Roncaglia	Junio de 2015
56	A Comparative Analysis of Political Competition and Local Provision of Public Goods: Brazil, Colombia and Mexico (1991-2010)	Ángela M. Rojas Rivera Carlos A. Molina Guerra	Octubre de 2015

57	Economía, gestión y fútbol: de la pasión a la sostenibilidad financiera	Ramón Javier Mesa Callejas Jair Albeiro Osorio Agudelo Carlos Eduardo Castaño Ríos	Julio de 2016
58	Desarrollo económico y espacial desigual: panorama teórico y aproximaciones al caso colombiano	Angela Milena Rojas Rivera Juan Camilo Rengifo López	Noviembre de 2016
59	Extent of Expected Pigouvian Taxes and Permits for Environmental Services in a General Equilibrium Model with a natural capital constraint	David Tobón Orozco Carlos Molina Guerra John Harvey Vargas Cano	Noviembre de 2016
60	Riesgo idiosincrático y retornos en el mercado accionario de Colombia	Carlos Andrés Barrera Montoya	Enero de 2017
61	Incidencia de los flujos de capital en la política monetaria de Colombia, 1996-2011	Deivis Agudelo Hincapié Alexis Arias Saavedra Julián Jiménez Mejía	Enero de 2017
62	Sobre los fundamentales del precio de la energía eléctrica: evidencia empírica para Colombia	Jorge Barrientos Marín Monica Toro Martínez	Marzo de 2017
63	Desarrollo económico local y género en ámbitos territoriales rurales: el caso de la zona Liborina-Sabanalarga, Antioquia, Colombia	Harold Cardona Trujillo Jorge Lotero Contreras Paula Andrea Galeano Morales Alix Bibiana Gómez Robinson Garcés Marín	Mayo de 2017
64	Recursos y capacidades para el desarrollo económico local en Buriticá Antioquia	Tatiana María Colorado Marín Juan David Franco Henao Yesica Rangel Villada	Junio de 2017
65	Panel de VAR: Una aplicación en la movilidad de factores de producción en la integración económica Alianza del Pacífico	Carlos Andrés Villarreal Restrepo	Junio de 2017
66	Cálculo de un WACC diferenciado por región para proyectos de generación de electricidad con fuentes renovables en Colombia	Jorge Barrientos Marín Fernando Villada Duque	Agosto de 2017
67	La determinación de los precios en la teoría económica de Sir James Steuart	Alexander Tobon Arias	Agosto de 2017
68	La teoría macroeconómica de John Maynard Keynes	Ghislain Deleplace	Octubre de 2017
69	Revisión general de la producción académica en historia empresarial colombiana publicada en revistas académicas 1984-2016	Tatiana González Lopera	Noviembre de 2017
70	Una regla empírica de tasa de interés de política monetaria para una economía emergente, pequeña y abierta	Jaime Montoya Ramirez	Noviembre de 2017
71	Los salarios y la fatiga acumulada: una revisión de la teoría de la oferta de trabajo	Carlos Andrés Vasco Correa	Diciembre de 2017
72	Modelo cualitativo para estudiar la internacionalización de las multilatinas colombianas	Ramón Javier Mesa Callejas Mauricio Lopera Castaño Paola Melisa Valencia Guzmán Mónica Andrea Álvarez Marín Paula Andrea Uribe Polo	Febrero de 2018
73	Mediciones del crecimiento económico regional y local en Colombia, 1950-2017: una revisión	Jaime Vallecilla G.	Febrero de 2018
74	Planteamiento de la cuestión agraria en la historiografía agraria colombiana: 1936 – 2016	Juan Carlos Velásquez Torres	Marzo de 2018
75	Los estudios en historia fiscal de Colombia sobre el siglo xx	Angela Milena Rojas R.	Noviembre de 2018
76	Can environmental taxes and payments for ecosystem services regulate pollution when the resilience of water bodies is surpassed?	David Tobón-Orozco Carlos Molina Harvey Vargas	Noviembre de 2018
77	Sobre la estructura de gasto y la curva de Engel de los hogares urbanos: evidencia empírica para Medellín	Jorge Barrientos Marín Efraín Arango Sánchez	Noviembre de 2018

78	Determinantes de la productividad multifactorial: los casos de las principales economías latinoamericanas y emergentes de Asia (1960 - 2015)	Wilman Arturo Gómez Carlos Esteban Posada Remberto Rhenals	Diciembre de 2018
79	Implementación de una evaluación por competencias académicas en el pregrado de Economía de la Universidad del Magdalena, Colombia	Rafael García José González Porto Luz Helena Díaz Álvaro Acevedo Alexander Tobón	Mayo de 2019
80	Determinantes del ahorro interno en Colombia: un acercamiento desde las Cuentas Nacionales Trimestrales para el período 1994-2017	Jaime Montoya Ramirez	Junio de 2019
81	Álgebra de un modelo simple IS-MR-AD-AS: Notas de clase	Jaime Alberto Montoya Remberto Rhenals	Agosto de 2019
82	¿Las diferencias importan? Heterogeneidad y dilemas sociales en recursos naturales, aportes desde la Economía experimental y del comportamiento	Yady Marcela Barrero	Septiembre de 2019
83	Concentración de tierras, paz territorial e impuesto predial rural en Antioquia	Cristian Sánchez Salazar	Septiembre de 2019
84	Una breve aplicación a la predicción de la fragilidad de empresas colombianas, mediante el uso de modelos estadísticos	Jorge Iván Pérez García Mauricio Lopera Castaño Fredy Alonso Vásquez Bedoya	Septiembre de 2019
85	Diseño e implementación de resultados de aprendizaje para pregrados en Economía	Alexander Tobón	Octubre de 2019
86	Corrupción, incentivos y contrabando técnico en Colombia, 1998 – 2013	Edwin Esteban Torres Gómez Luis Ricardo Argüello Cuervo	Noviembre de 2019
87	Efecto de los programas educativos en pruebas estandarizadas. Un análisis por cuartiles de la política educativa "Antioquia la más educada".	Diana Lucia López López Edwin Esteban Torres Gómez Cristian Sánchez Salazar	Diciembre de 2019
88	Identificación de las principales restricciones operativas al crecimiento en Antioquia, Caldas, Risaralda y Quindío	Harold Cardona-Trujillo Estefany Peña Rojas	Diciembre de 2020
89	La teoría económica: ¿un monumento en peligro?	Jean Cartelier	Febrero de 2021
90	Caída y convergencia mundial de las tasas de inflación	Wilman Gómez Carlos Esteban Posada Remberto Rhenals	Marzo de 2021
91	¿Es posible explicar la crisis colombiana de 1998-2003 a partir de la teoría austríaca del ciclo económico?	Andrés Mauricio Rosero Sánchez	Mayo de 2021
92	La estructura de la propiedad de la tierra y su relación con la inversión social local en Colombia, 2000 – 2010	Mariana Rincón Orozco	Junio de 2021
93	Conferencia de Carlo Benetti con motivo de su investidura como Doctor Honoris Causa de la Universidad Metropolitana de México en 2015	Carlo Benetti	Septiembre de 2021
94	Crédito, producción y consumo en la teoría monetaria de Hawtrey (1919)	Carlos Andrés Villarreal Restrepo	Octubre de 2021
95	Entre el poder y la resistencia. Asesinato de líderes sociales y restitución de tierras en Colombia	Juan Fernando Zabala Hincapié	Diciembre de 2021
96	La estructura lógica de la teoría del equilibrio general dinámico estocástico	Alexander Tobón	Agosto de 2022
97	Transición demográfica, acumulación y uso del capital humano: ¿por qué muchos países siguen siendo pobres?	Ana Isabel Suárez García	Octubre de 2023
98	Metodología para la construcción y evaluación de resultados de aprendizaje para programas de pregrado y posgrado	Alexander Tobón Arias	Abril de 2024
99	La interacción gráfica en Python como herramienta para mejorar el aprendizaje en microeconomía y macroeconomía	Carlos Andrés Palacio Gómez	Octubre de 2024

100	Sector informal como amortiguador del ciclo económico: el papel de la aversión a la pérdida	Wilman Gómez M. Cristian Castrillón G. Jaime Montoya A.	Octubre de 2024
101	Evaluación de reglas fiscales a través de un modelo de equilibrio general dinámico estocástico DSGE estimado para la economía colombiana	Geraldine Grimaldo Álvarez Wilman Arturo Gómez Muñoz Juan Pablo Saldarriaga Muñoz	Octubre de 2024
102	Notas sobre el marxismo de Maurice Dobb	Ramon Boixadera Bosch	Noviembre de 2024
103	Ciclos reales de negocios: una nota didáctica	Wilman Arturo Gómez Muñoz	Noviembre de 2024
104	El ciclo del ahorro de los hogares en Colombia: una aproximación trimestral usando variables macroeconómicas, laborales y demográficas	Jaime Horacio Montoya Ramírez	Diciembre de 2024
105	Transformación digital en la cadena de valor cafetera: análisis comparativo de plataformas tecnológicas para el cumplimiento del Reglamento Europeo sobre Productos Libres de Deforestación (EUDR)	Harold Cardona-Trujillo	Enero de 2025
106	Entre buenas intenciones e incentivos perversos: ¿El Régimen Simple de Tributación impulsa la formalización o solo reduce el recaudo en Colombia?	Alejandro Montoya Castaño Luis Fernando Gamboa Mónica Mogollón Luis Carlos Reyes	Junio de 2025
107	Asociatividad Territorial y Provincias Administrativas y de Planificación en Antioquia: Una estrategia para el desarrollo regional	Maria Juliana Muñoz Cardona	Septiembre de 2025
108	Un modelo de inflación para el caso colombiano (2001-2025)	Wilman Arturo Gómez Carlos Esteban Posada	Marzo de 2026
109	Una guía técnica para la GEIH	José Manuel Bastidas Burbano Wilman Arturo Gómez	Mayo de 2026

**LECTURAS
DE
ECONOMÍA**

Universidad de Antioquia
Facultad de Ciencias Económicas
Departamento de Economía