Universidad de Buenos Aires
Facultad de Ciencias Económicas
Escuela de Negocios y Administración Pública

MAESTRÍA EN ECONOMÍA APLICADA

TRABAJO FINAL DE MAESTRÍA

Subcaptación de salarios en Argentina. La EPH y sus divergencias con SIPA en múltiples dimensiones (1996-2021).

AUTOR: IGNACIO LAUTARO PAOLA

DIRECTOR: DAMIÁN KENNEDY

CODIRECTOR: MARTÍN TROMBETTA

MAYO DE 2025

Agradecimientos

A mi familia y amigos, por ser quienes me hacen ser quién soy. Desde mi vieja Gabriela, mis hermanos Santiago y Juan, Adrián y mis abuelos Jorge y Cristina; hasta todos mis tíos y primos. Desde mis amigos del barrio, la primaria y la secundaria; hasta mis amigos de la universidad, de trabajos pasados y del club. Sin su compañía y apoyo, no podría haber realizado ningún tipo de avance en este camino.

A Lourdes, por acompañarme día a día. Y en particular, por escuchar siempre con mucho amor -con sumo interés y ninguna referencia- acerca de los hallazgos expuestos en esta tesis.

A mis compañeros del Centro de Estudios sobre Población, Empleo y Desarrollo (CEPED), por estar siempre dispuestos a escuchar, debatir y aconsejar acerca de las temáticas que aborda este trabajo.

A mis directores Damián y Martín, por acompañarme con paciencia y dedicación en la elaboración de esta tesis (y más allá también).

A Daniel Pandolfi, responsable del temprano término del año futbolístico para el Club Ferro Carril Oeste, razón por la cual dispuse del tiempo necesario para elaborar la parte final de este trabajo.

A la Universidad de Buenos Aires en particular, por haberme permitido realizar la Maestría en el marco de su programa de becas de investigación; y a la educación pública de la República Argentina en general, por resultar parte íntegra de mi formación como profesional y -sobre todo- como persona.

Resumen

Los problemas de captación de ingresos en las encuestas de hogares conllevan sesgos en indicadores derivados, generalmente causando sobreestimaciones de pobreza y subestimaciones de desigualdad. En Argentina, es de reciente reconocimiento un nuevo período de creciente brecha en el ingreso de asalariados registrados observado por encuestas de hogares en su comparación con fuentes de registro. Sin embargo, el emparejamiento con fines de análisis estadístico entre estos tipos de fuentes adolece de distintas dificultades para su realización. En este trabajo se plantea un emparejamiento entre la Encuesta Permanente de Hogares (EPH) y la Muestra Longitudinal de Empleo Registrado (MLER) en todas sus dimensiones posibles: género, edad, región, tamaño del establecimiento y sector productivo. Los resultados indican que para todas las dimensiones se replica una mayor distancia entre fuentes a mayor nivel de ingresos. A su vez, se re caracteriza esta diferencia entre fuentes para la década de 1990, a partir de la consideración de sobrecaptación -subdeclaración a las fuentes de registro- en bajos niveles de ingresos. En consecuencia, se plantea una nueva lectura sobre el crecimiento inicial de esta brecha a mediados de la década del 2000, guiado por la reducción del salario "en gris" que trajo aparejada la reducción de la informalidad laboral.

Palabras clave: Salarios, Encuestas, Fuentes de información, Subcaptación.

Códigos JEL: J31, C81, C51.

Abstract

Income capture issues in household surveys lead to biases in derived indicators, typically resulting in overestimation of poverty and underestimation of inequality. In Argentina, a new period of a widening gap in the registered workers income, as observed in household surveys compared to administrative records, has been recently recognized. However, matching these types of data sources for statistical analysis purposes faces various challenges. This study proposes a matching approach between the Encuesta Permanente de Hogares (EPH) and the Muestra Longitudinal de Empleo Registrado (MLER) across all possible dimensions: gender, age, region, establishment size, and productive sector. The results indicate that for all dimensions, the gap between data sources increases with higher income levels. Additionally, the difference between sources during the 1990s is recharacterized by considering overreporting (underreporting to administrative records) for low-income levels wages. Consequently, a new interpretation of the initial growth of this gap in the early 2000s is proposed, driven by the reduction of "off-the-books" wages that accompanied the decline in labor informality.

Key words: Wages, Surveys, Information Sources, Underreporting.

JEL codes: J31, C81, C51.

Índice

1. INTRODUCCIÓN: De nuestra encuesta y sus ingresos 1
2. ESTADO DE LA CUESTIÓN DE LA SUBCAPTACIÓN: Antecedentes bibliográficos y evolución reciente para asalariados privados registrados 4
2.1. Las encuestas, nuestra encuesta y sus problemas 4
2.1.1. Una historia de cambios en las costumbres de corrección 7
2.1.2. Se puede corregir para ver que hay 13
2.2. La evolución de la brecha de ingresos entre EPH y SIPA 15
3. CONSIDERACIONES METODOLÓGICAS: Emparejar todo lo que sea emparejable. 24
3.1. Sobre la compatibilización de dimensiones entre EPH y MLER 24
3.2. Otras consideraciones sobre el emparejamiento propuesto 28
4. EL ANÁLISIS DE A UNA DIMENSIÓN ENTRE EPH Y MLER 31
4.1. La comparación a nivel general 31
4.2. La comparación según atributos personales 34
4.3. La comparación según atributos del puesto de trabajo 38
4.4. La brecha depende del nivel de ingresos, pero existen matices 46
4.5. Hallazgos desprendidos del análisis individual de las dimensiones 49
5. EL ANÁLISIS DE LAS DIFERENCIAS ENTRE EPH Y MLER A PARTIR DE UN MODELO PARAMÉTRICO 53
5.1. La unidad en grupos socioocupacionales 53
5.2. Un vistazo dentro de los grupos socioocupacionales 54
5.3. Una aproximación paramétrica para contemplar todas las dimensiones 57
6. CONCLUSIONES, DISCUSIÓN Y TRAZOS A FUTURO 65
Bibliografía 70
Anexo I: Criterios operativos para la compatibilización entre EPH y MLER 74

1. INTRODUCCIÓN: De nuestra encuesta y sus ingresos

La información sobre ingresos monetarios de la población para el análisis empírico en las ciencias sociales se recolecta a través de distintas fuentes. Entre ellas, las encuestas de hogares, que se caracterizan por su mayor amplitud a la hora de captar distintos perfiles sociolaborales y sus diferencias en términos de ingresos. Por lo tanto, las estimaciones de indicadores socioeconómicos relevantes -como de la distribución del ingreso y de la proporción de individuos u hogares con ingresos inferiores a los umbrales de pobreza- se basan en la información recolectada a través de estas encuestas. En nuestro país, la encuesta que funciona con estos objetivos es la Encuesta Permanente de Hogares (EPH) que realiza el Instituto Nacional de Estadística y Censos (INDEC).

Acerca de las otras fuentes que recolectan información sobre ingresos, existen las cuales lo hacen a través de información derivada de registros administrativos o tributarios. Estas se diferencian de las anteriores en tanto la información disponible en las encuestas reviste de carácter declarativo. Particularmente para asalariados registrados, en nuestro país las declaraciones tributarias laborales son centralizadas en el Sistema Integrado Previsional Argentino (SIPA)[1] hacia donde se exige a las empresas declarar mensualmente la totalidad del personal en relación de dependencia y sus remuneraciones para el pago de los aportes personales y las contribuciones patronales.

En esta clave, los registros administrativos o tributarios se presentan siempre como un buen insumo para contrastar con las encuestas de hogares en miras de detectar sesgos en las últimas. La información de ingresos contenida en los registros combina varias ventajas: mayor fidelidad de la información recolectada, mayor masividad en su alcance, mayor periodicidad y regularidad de recolección, entre otras. No obstante, las comparaciones estadísticas de la información de ingresos entre encuestas y registros administrativos tienden a ser difíciles de realizar. Principalmente, a causa de los distintos universos capturados, la disímil disponibilidad de información en términos temporales y de tamaño de muestra, o la imposibilidad de emparejar unidades observadas entre fuentes, entre otras dificultades. El carácter anónimo y declarativo de la encuesta, con fines

[1] Hasta 2008, este sistema era denominado "Sistema Integral de Jubilaciones y Pensiones" (SIJyP). Por simplicidad, en este trabajo se toma -salvo aclaración explícita- SIPA como denominación común para todo el período analizado.

estadísticos, es cualitativamente diferente al de registros, identificados y derivados de obligaciones tributarias o administrativas.

Actualmente en Argentina es reconocida la existencia de una creciente brecha entre el ingreso salarial promedio recolectado por EPH y el capturado en SIPA (Müller, 2023; Kennedy *et al*, 2024; Albina *et al*, 2024). Anteriormente, la subcaptación de ingresos en EPH fue reconocida y abordada en la década de 1980, profundizada y diversificada en sus técnicas de corrección en la década de 1990, rebatida su necesidad de corrección a principios de los 2000 y vuelta a la relevancia, las discusiones sobre su mala captación y a las necesidades de correcciones en los últimos años (Altimir 1986, Gasparini 1999, Roca y Pena, 2001; Sánchez, Kennedy y Pacífico, 2016).

Este trabajo propone aportar al conocimiento acerca de la de la subcaptación de salarios en la EPH, a partir de un análisis de la información sobre el ingreso de los asalariados registrados a través de características del individuo (género, edad) y del puesto laboral (tamaño del establecimiento, región y sector productivo). Para esto, se realiza un emparejamiento entre la EPH (en sus versiones puntual y continua) y la Muestra Longitudinal de Empleo Registrado (MLER), fuente de información derivada del SIPA. Se contempla el período abarcado entre los años 1996 y 2021, máximo alcanzable según la información disponible en MLER.

Los resultados alcanzados son relevantes en dos sentidos. Por un lado, refuerzan la tendencia general documentada en la literatura que vincula mayor nivel de ingresos con peor captación en la encuesta; comportamiento replicado para cada particular dimensión considerada, con distinto nivel de dispersión según el caso. Por el otro, sugieren una nueva interpretación acerca de la captación del ingreso en la encuesta en la década de 1990. Este subperíodo, generalmente clasificado como de ausencia de subcaptación en asalariados, puede alternativamente caracterizarse como uno de mala captación de salarios de alto nivel de ingresos, pero compensada por una relevante subdeclaración al SIPA (sobrecaptación en EPH) en salarios de bajo nivel de ingresos. En línea con este hallazgo, se propone una nueva lectura acerca del ensanchamiento de la brecha entre fuentes, parcialmente causado -paradójicamente- por la mejora en la captación de los bajos salarios (o reversión de su sobrecaptación) a partir del descenso de la tasa de no registro dada a principios de la década del 2000.

El trabajo se estructura de la siguiente forma. El Capítulo 2 se dedica al estado del arte del fenómeno de la subcaptación de ingresos, y se divide en dos secciones. En una primera sección se encuentra una revisión de la literatura dedicada a documentar y -sobre todo- corregir los sesgos en la subcaptación de ingresos en encuestas de hogares, mientras que en la segunda sección se ofrece una introducción a la evolución cuantitativa de la brecha agregada de salarios en EPH respecto a SIPA. En el Capítulo 3 se detallan los aspectos metodológicos vinculados al emparejamiento entre los datos de EPH y MLER, lo que constituye la base para el trabajo presentado en los dos capítulos siguientes. En el Capítulo 4 se comienza analizando de forma agregada la evolución del salario en niveles para cada fuente y su correspondiente brecha; para luego replicar en detalle el mismo procedimiento a partir de desagregar la información de a una característica a la vez (ya sea del individuo o del puesto de trabajo). En el Capítulo 5 se describe la creación de grupos ocupacionalmente similares en múltiples dimensiones, para con ese insumo desarrollar un modelo de comparación paramétrica entre fuentes. Finalmente, en el Capítulo 6 se detallan los resultados, conclusiones y líneas de trabajo futuras desprendidas.

2. ESTADO DE LA CUESTIÓN DE LA SUBCAPTACIÓN: Antecedentes bibliográficos y evolución reciente para asalariados privados registrados

Para abordar integralmente el estado de la cuestión del fenómeno de la subcaptación de salarios en la EPH, se opta por dividir este capítulo en dos secciones. En una primera sección, se revisa la literatura existente vinculada con el fenómeno en Argentina, remarcando hechos estilizados y clasificando las correcciones empleadas en los ingresos de la encuesta según su finalidad. En una segunda sección se realiza un análisis empírico de la brecha de ingresos entre SIPA y EPH. Para este análisis, se contemplan también características metodológicas de generación y publicación de la información en estas fuentes, relevantes para dar cuenta algunas de las diferencias explicitadas tanto en nivel como en movimiento de la brecha entre fuentes. En concreto, esta revisión en dos facetas insume el análisis específico a realizar por este trabajo, en donde -como fuera explicitado- se busca profundizar en las características de la subcaptación de salarios en las distintas dimensiones socio ocupacionales disponibles.

2.1. Las encuestas, nuestra encuesta y sus problemas

A nivel global, los primeros desarrollos vinculados a las encuestas y censos con fines de obtención de información socioeconómica datan de principios del siglo XX. De ahí en adelante, el tipo de información objetivo a captar por las encuestas ha variado. Entre los años 30 y 40 buscaban principalmente dimensiones macroeconómicas. Desde entonces, creció progresivamente la relevancia de la información a menor escala a partir de su alcance granular en términos de universidades observadas. Así, sobre todo desde la década de 1990, se centran en recolectar y construir información relevante a nivel de hogares como unidad de información (Cattaneo, 2001).

En la literatura especializada existen consensos acerca de las dificultades para captar correctamente los ingresos de la población objetivo que se presentan en las encuestas de hogares, dificultades documentadas desde su origen como fuente de información socioeconómica (Kusnetz, 1953; Okner, 1975; Atkinson, 2007). Por un lado,

dado el limitado alcance muestral que caracteriza a esta fuente de información[2], la representatividad de las unidades captadas puede no ser la adecuada, en detrimento de las unidades que no son encuestadas y se dan por representadas en base al caso testigo recolectado. Este problema suele estar asociado a la "calidad" muestral en términos de representatividad, al proceso de diseño de la ejecución de la encuesta previo a su realización. Por otro lado, se presentan problemas particulares sobre la captación de ingresos en hogares que efectivamente son alcanzados por la encuesta. Estos son, a saber: la no respuesta al total de la encuesta por parte del hogar, la falta de respuesta sólo a las preguntas de ingresos, o la declaración de ingresos por debajo de los valores reales percibidos. Dicho de otra forma: subcaptación de perceptores, no respuesta de ingresos y subdeclaración de ingresos. A su vez, en torno al segundo problema pueden definirse dos problemas distintos: la falta de respuesta total a la encuesta o la falta de respuesta sólo en torno a preguntas de ingresos. El primer caso resulta más asimilable a una situación de no captación. En el segundo, las características disponibles permiten contar con -al menos algo de- información de dicho relevamiento. De este modo, podemos clasificar a los hogares según su situación en la encuesta, de menor a mayor en términos de información recolectada: no captado, no respuesta total, no respuesta de ingresos, subdeclarador de ingresos. En conjunto, todos estos problemas -tanto de la muestra como de las respuestas de los encuestados- comprenden el fenómeno agregado de la subcaptación de ingresos en las encuestas de hogares.

Cada una de estas dificultades introduce sesgos en la información recolectada y, por ende, en los indicadores socioeconómicos que se derivan de ella. En busca de morigerar su influencia, el procedimiento de corrección posible de realizarse varía según la dificultad diagnosticada. En primer lugar, los problemas de no captación de unidades representativas en tanto hogares no encuestados como característica de la muestra resultan imposibles de tratar, implicando necesariamente lidiar con los sesgos de discrepancias entre encuestas y otras fuentes que esto pueda causar. En segundo lugar, la no respuesta de ingresos es plausible de imputar a través de la información que reportan las unidades que sí declaran, en tanto se dispone de información socioeconómica vinculada con los niveles de ingresos. Finalmente, la subdeclaración puede ser corregida

[2] Particularmente en nuestro país, podemos ubicar a las encuestas de hogares en tercer lugar en términos de alcance y representatividad, por detrás de censos y registros administrativos.

en tanto se disponga de otras fuentes de información para estimar el grado de la corrección necesaria.

Es de largo análisis, a su vez, que no resulta trivial el vínculo entre nivel y fuente de ingresos con la aparición de estos problemas. La tendencia general, y los estudios en particular que se centran en nuestro país -generalmente en torno a la EPH, pero también respecto a la Encuesta Nacional de Gasto de los Hogares (ENGHo)-, marcan que individuos y hogares de mayores ingresos tienden a no ser captados, a no responder o a subdeclarar su nivel de ingresos (Gasparini, Marchionni y Sosa Escudero, 2000; INDEC, 2006; Jiménez y Rossignolo 2019; entre otros).

En estrecha vinculación a este fenómeno distributivo, tampoco resulta trivial las fuentes del origen del ingreso en relación con su captación. Las retribuciones al capital y las rentas tienden a ser peor captadas que los ingresos laborales. En buena parte, a causa de ser fuentes de ingreso con mayor participación relativa en altos niveles de ingreso, donde los individuos tienden a ser no respondientes o sub declarantes (Altimir, 1986; Salvia y Donza, 1999; Sosa Escudero y Alaimo, 2000; entre otros); aunque también se le atribuye a esta población mayores sesgos cognitivos vinculados al desconocimiento sobre su nivel de ingresos, tanto personales como del hogar, dada la falta de estabilidad de estos en términos temporales (Beccaria, 1998). El ingreso salarial, en cambio, aparece -dada su estabilidad- con mayor facilidad de ser informado con precisión; excepto en casos de asalariados que no poseen un ingreso regular a lo largo del año. Tanto la vinculación del nivel de ingresos como de la fuente de éstos con su captación en las encuestas aparecen como hechos estilizados de gran relevancia para un primer acercamiento al fenómeno de la subcaptación de ingresos.

En lo particular de nuestro país, con el correr de los años se han llevado a cabo distintos estudios abordando esta problemática. Estos trabajos pueden clasificarse en dos grupos para su revisión. En un primer grupo, mayoritario, se encuentran los estudios que realizan correcciones en pos de aminorar los sesgos introducidos por la subcaptación de ingresos en EPH, en busca de recalcular otros indicadores socio económicos (generalmente, distribución del ingreso o pobreza). Para estos trabajos, la subcaptación es un mero obstáculo que saltear, y las correcciones realizadas funcionan como un instrumento para un fin derivado. En un segundo grupo, minoritario, se ubican los trabajos que indagan en las características en distintas dimensiones de la subcaptación, no necesariamente con otro fin más que de construir nociones sobre la subcaptación como

fin en sí mismo. Resulta conveniente repasar particularmente los trabajos de este segundo grupo, en tanto -en distinta medida- se asimilan al objetivo planteado en este trabajo.

2.1.1. Una historia de cambios en las costumbres de corrección

El origen de las metodologías de corrección de subcaptación de ingresos de encuestas en Argentina puede situarse en trabajos de Altimir (1979, 1986)[3]. En el primer trabajo, el autor se propone construir líneas de pobreza para distintos países de Latinoamérica hacia el año 1970 en base a los patrones de consumo de cada país, pero con criterios normativos equiparables, de tal modo que cobre sentido la comparación entre líneas. A la hora evaluar la cantidad de hogares situades por debajo de ese umbral según encuestas, realiza un ajuste de ingresos en base a las cuentas nacionales, "intentando neutralizar los sesgos por subdeclaración y por omisión" (Altimir 1979, p. 68). Obtiene que, en general, la población debajo del umbral de pobreza para cada país es menor a la calculada por institutos oficiales[4]. En el segundo trabajo analiza la distribución personal del ingreso en la región, para lo cual documenta el estado de las encuestas de hogares, severamente afectadas por la subdeclaración en base a la comparación agregada con las Cuentas Nacionales (CCNN). Para Argentina, analiza resultados de distintas encuestas realizadas antes de 1980 en pos de su corrección. Así lo hace con EPH, pero también con EED (Encuesta de Empleo y Desempleo, 1970) y EPF (Encuesta de Presupuestos Familiares, 1969/70). Los resultados indican un crecimiento de la desigualdad por ingresos a lo largo de la década de 1970, fenómeno que no es documentado por los datos oficiales, que muestran una relativa estabilidad en la distribución del ingreso para ese período. En base a esta discrepancia, el autor realza la relevancia de aplicar correcciones a los ingresos obtenidos en las encuestas de hogares (Altimir, 1986).

Esta metodología de corrección vía ajustar el nivel agregado de ingresos según las CCNN ha sido replicada en múltiples trabajos durante la década de 1990. La misma

[3] Cabe destacar, de todas formas, que el mismo autor ya había documentado los problemas de la captación de ingresos en encuestas de hogares y sus consecuentes sesgos en la distribución del ingreso en Altimir (1975). Allí aborda cuestiones como las discrepancias con los agregados macroeconómicos, problemas de cobertura (subcaptación de perceptores) y de contenido (subdeclaración de ingresos), o las dificultades para la comparación entre fuentes acarreadas por distintas mediciones de tipos de salario (neto/bruto), entre otras.

[4] Como consideración adicional, resulta relevante que el autor toma consumo per cápita del hogar en lugar de ingresos, al considerarlo una aproximación más certera al nivel de vida de la población de menor nivel de ingresos.

permite correcciones diferenciales -aunque lineales- por fuente de ingreso, en estrecha vinculación con el diagnóstico sobre los diferenciales de afección de estos problemas según fuente de ingresos. En este sentido, Gasparini (1998) busca estimar la incidencia de la estructura tributaria Argentina en términos de distribución del ingreso, para lo cual corrige EPH y ENGHo vía CCNN, construyendo distintos coeficientes por fuentes de ingreso, aunque asumiendo niveles de subcaptación constantes entre deciles, al diferenciar sólo por fuente de ingreso. En un trabajo posterior, el mismo autor analiza la evolución del bienestar y desigualdad para Argentina en comparación con el resto del mundo entre 1988 y 1998, para lo cual repite metodológicamente la decisión de ajustar ingresos de la EPH vía coeficientes lineales por fuente de ingreso obtenidos para empardar las CCNN. Concluye así que hacia fines de la década de 1990 aún no se había detenido la fase ascendente de la desigualdad iniciada a mediados de la década de 1970 (Gasparini, 1999).

Por su parte, Llach y Montoya (1999) reestiman pobreza y desigualdad para Argentina entre 1974 y 1998. Para esto realizan dos grandes correcciones: la reponderación de los hogares que brindan información de ingresos, luego del descarte de los que lo no hacen; y el ajuste de ingresos también vía cocientes derivados de las CCNN (concretamente, la Cuenta Generación del Ingreso -CGI- de 1984 y 1993) para equiparar el total agregado de ingresos entre esta fuente y los arrojados por las encuestas. Tras esta primera corrección, adicionan el efecto neto de la política fiscal para evaluar su impacto distributivo y en la pobreza. Los resultados obtenidos indican menores índices de pobreza, pero mayores de desigualdad respecto a las estadísticas oficiales. Este resultado se profundiza -en ambos sentidos- al incluir el impacto fiscal neto. En la misma línea, Santiere, Gómez Sabaíni, y Rossignolo, (2000) evalúan la carga tributaria y su vínculo con la distribución general de los ingresos hacia fines de la década de 1990. Para esto se valen de la ENGHo a la hora de obtener datos de ingresos, los cuales ajusten para equiparar los totales derivados de las CCNN[5].

En la misma época, Salvia y Donza (1999) parten de postular la no declaración como más dañina para la fidelidad de los datos de ingresos respecto a la subdeclaración, argumentando que nada asegura que las CCNN capten efectivamente un nivel de ingresos

[5] Una comparación entre los trabajos revisados resulta llamativa -dado que estos emplean metodologías razonablemente parecidas- en el sentido que los coeficientes de corrección obtenidos resultan divergentes entre sí. Para un desarrollo de mayor detalle en torno a las diferencias cuantitativas resultantes entre los distintos ajustes, ver Gómez Sabaíni y Rossignolo (2014) o Sánchez, Pacífico y Kennedy (2016).

más certero. Así, se concentran en corregir el sesgo de captación introducido al no considerar de los datos sin respuesta de ingresos. Lo hacen imputando un nivel de ingresos para las unidades no respondientes mediante una regresión multivariada de mínimos cuadrados ordinarios considerando dimensiones socioeconómicas de respondientes relevantes para determinar su nivel de ingresos. Tras esta corrección, reajustan estimaciones de pobreza.

De forma no enfrentada sino complementaria al método de ajuste vía cocientes calculados desde CCNN -realizan ambos procedimientos-, Gasparini y Sosa Escudero (1999) ajustan el cálculo de la distribución del ingreso de las décadas de 1980 y 1990, corrigiendo el sesgo introducido por la no declaración a través de imputar ingresos estimándolos con una regresión de mínimos cuadrados ordinarios. Así, buscan estimar el bienestar como una combinación de nivel de ingreso y distribución, y optan por elegir dos fuentes distintas: CCNN para nivel de ingresos, y EPH para distribución de estos. En un procedimiento similar, un trabajo posterior de los mismos autores ajusta el cálculo de la distribución del ingreso, pero limitándose exclusivamente a la provincia de Buenos Aires, lo cual permite una mayor extensión temporal para la corrección. Concretamente, a los aglomerados de esta que relevaba la EPH en aquel entonces -como lo hace hoy-: Buenos Aires, La Plata, Mar del Plata y Bahía Blanca. En esta ocasión, aplican corrección de Heckman[6] para considerar la probabilidad de no declaración de forma intrínseca al modelo. Su significatividad refuerza la noción de que no es trivial la decisión de no declara en términos de nivel de ingresos (Gasparini *et al*., 2000).

Cabe resaltar, de todas formas, que en todos los trabajos con esta metodología empleada se da lugar a la consideración que, dada la vinculación entre no respuesta y nivel de ingreso, esta imputación tiende a sesgarse -en tanto se realiza con los datos de quienes sí responden-. Es por esto por lo que, si bien el ingreso en promedio de las unidades imputadas es mayor al de los respondientes, resulta razonable esperar que esto sea una aproximación que aún subestima la información real.

En contraposición a los trabajos previamente comentados, otras miradas desestimaban la necesidad de correcciones para los ingresos de los asalariados en EPH para la década de 1990. Roca y Pena (2001) comparan la EPH con CCNN y también con registros administrativos del SIJyP, y concluyen no encontrar indicios de subdeclaración

[6] Para la especificación del método, ver Heckman (1979).

en asalariados ni en perceptores de ingresos fijos -jubilados- durante la década de 1990. Aborda a este resultado a partir de limitar la información de registro proveniente del SIJyP de forma equiparable a la de la muestra de los aglomerados que captura EPH -considerar el 70% de población urbana, excluyendo aglomerados pequeños y rural- y comparando los ingresos agregados en ambas fuentes, que resultan razonablemente similares. En base a este procedimiento remarca que, dado que las CCNN captan el total de la población y no es posible seccionar su información geográficamente sobre lo que cubre EPH, resulta estéril la comparación entre estas fuentes. De este modo, reafirma que la subdeclaración es un problema exclusivo de perceptores de ingresos por cuenta propia, capital y rentas; y resalta la importancia de la distinción entre subcaptación y subdeclaración para no atribuirle toda diferencia de ingresos medios a la subdeclaración, sino considerar también lo no captado. Así, los autores postulan que la disparidad en los resultados de correcciones marca las arbitrariedades que se pueden cometer al intentar corregir los ingresos provenientes de ella EPH basándose en CCNN o según diferentes metodología y criterios.

A inicios de los años 2000, la crítica de Roca y Pena (2001) se combinó con las implementaciones metodológicas para EPH en su versión continua, que ayudaron a aminorar la necesidad de las correcciones (INDEC, 2003), llevando a un período de "des recomendación" de correcciones para los ingresos para EPH. Es así explicitado en trabajos como el de Gasparini, Cicowiecz y Sosa Escudero (2013) en donde -luego de realizar una detallada revisión de correcciones sobre subcaptación de ingresos en encuestas y sus impactos en la reestimación de pobreza y desigualdad- plantea no realizar corrección alguna por la disparidad de resultados de posible obtención dada las distintas metodologías existentes, y recomienda así usar los datos en bruto.

Para la década de 2010, es habitual leer como un consenso la utilización de datos en bruto a lo largo del continente al encontrar más desventajas que ventajas a la corrección, principalmente por la disparidad de impacto a la hora de corregir ingresos con distintos métodos. Esta des recomendación se da incluso en pleno entendimiento con los procesos de corrección vía CCNN realizados en el pasado, a partir de la idea de que la mayor confiabilidad en esta fuente respecto a las encuestas resultaba un consenso hacia fines de la década de 1990, cuando los países de América Latina apenas comenzaban a recolectar información de manera sistemática mediante encuestas (Amarante, 2013). En esta etapa, las recomendaciones de organismos internacionales se centran en mejorar la captación de las encuestas a sabiendas de los problemas de subcaptación que

inherentemente traen las mismas, pero también dada la dificultad de homogeneizar metodologías y resultados de correcciones en estas (Villatoro, 2015; CEPAL, 2018).

Merecen una mención aparte los trabajos centrados exclusivamente en la correcta captación de los altos ingresos (*top incomes*), que por su objetivo diferencial continúan durante esta época realizando procesos de corrección en pos de dimensionar la participación de los percentiles superiores en el ingreso agregado. En esta familia de estudios se recurre a tabulaciones o bases de microdatos de registros tributarios para corregir la información de los mayores percentiles de ingresos de las encuestas a través del método de Interpolación de Pareto[7]. Estos estudios resaltan como las encuestas de hogares están imposibilitadas de captar el 1% superior de la distribución del ingreso, ya sea por su ausencia en la muestra relevada (no captación) pero también por su no respuesta, ya sea general o particular sobre la información de ingresos.

En esta línea, Alvaredo (2010) arma series de la participación de los deciles superiores en el total de los ingresos de nuestro país entre 1932 y 2004. Usa estadísticamente información de impuestos a los ingresos personales, aunque advierte que la relevancia e impacto de estos tributos sufrió distintas transformaciones a lo largo de los años. Compara esta información con la arrojada por la EPH para el año 1997 y marca lo deficiente que resulta la encuesta para la captación de los *top incomes*: encuentra que los ingresos del 0,1% más rico directamente no es captado en la encuesta, pese a que entre 25% y 30% de los ingresos de este percentil provienen de fuentes salariales. Por su parte, Jiménez y Rossignolo (2019) toman datos de declaraciones juradas del Impuesto a las Ganancias, y corrigen el índice de Gini oficial resultante de EPH vía Interpolación de Pareto. Concretamente, realizan tres correcciones: amplificando el ingreso del 1%, 0,5% y 0,1% superior, lo que resultan en una progresiva mayor desigualdad por ingresos. Adicionalmente, cabe destacar el trabajo de Gómez Sabaíni y Rossignolo (2014), en el cual los autores comparan resultados de correcciones por subcaptación de ingresos con este método contra las realizadas a través de coeficientes derivados de las CCNN, tendiendo puentes entre metodologías.

Retomando el análisis cronológico, tras el período de "des recomendación" de correcciones de ingresos en la década del 2000, hacia mediados del 2010 aparecen trabajos que alertan sobre una creciente subcaptación de ingresos particularmente en

[7] Para revisar en profundidad detalles de esta metodología, ver Jiménez y Rossignolo (2019).

asalariados. En Sánchez, Pacífico y Kennedy (2016) se propone actualizar temporalmente una serie unificada de distribución funcional del ingreso, y para eso debe revisa metodologías y datos de fuentes de salarios y sus correcciones. Así, advierten cómo incluso siguiendo la metodología propuesta por Roca y Pena (2001), para la cual en la década de 1990 y a inicios de los 2000 no hay señales de subcaptación de salarios vía ingresos medios, hacia 2014 se expande una brecha entre fuentes que ubica al ingreso agregado de EPH un 30% debajo del derivado de SIPA. Del mismo grupo de autores, en Kennedy *et al.* (2024) se presenta un análisis íntegro de las series oficiales de salario vigentes en Argentina. Allí se aprecia, con actualización hasta el año 2022, la creciente brecha existente en particular para ingresos de asalariados registrados entre EPH y registros administrativos.

Esta evidente documentación de una brecha de ingresos en asalariados cada vez más relevante motivó una reciente nueva ola de propuestas de corrección en los ingresos de la encuesta para refinar estimaciones derivadas. Müller (2023) reestima la línea de pobreza recuperando los métodos habituales de la década de 1990: calcular coeficientes para que el ingreso agregado de las encuestas iguale al agregado de las CCNN. Se basa particularmente en el cambio metodológico suscitado para la línea de pobreza en 2016[8], como piedra basal desde donde existe evidente sobreestimación del fenómeno. Propone dos correcciones con la misma metodología, una primera ajustando los ingresos de toda la muestra de forma uniforme en base a los coeficientes obtenidos de CCNN (se asume subdeclaración pura, con nula subcaptación), y una segunda igualando los agregados de CCNN disminuyendo la representatividad de las unidades de bajos ingresos para dar lugar a las omitidas de altos ingresos (se asume subcaptación pura, con nula subdeclaración)[9]. Además de las correcciones monetarias, destaca las transferencias en especie como motivo de subestimación de la pobreza, afectando particularmente la captación del ingreso de los deciles inferiores.

En la misma línea, el trabajo de Albina *et al.* (2024) realiza una minuciosa corrección de ingresos de EPH por distintas fuentes, en pos de reestimar íntegramente la distribución del ingreso. Para esto, los autores hacen uso de información de distintas

[8] Acerca de los detalles del cambio de metodología para estimar la pobreza por ingresos, ver Arakaki (2016).
[9] En este caso, resulta interesante mencionar que con la primera metodología de corrección se reduce la población bajo la línea de pobreza para el período 2004-2022 de un promedio anual de 35% a 13%. Con la segunda, de 35% a 27%.

fuentes administrativas disponibles[10] en pos de corregir el ingreso de toda la muestra de la encuesta, y no restringir el ajuste sólo a una parte de la distribución. Adicionalmente, ajustan los ponderadores de la EPH para replicar la población nacional (para el año 2022). Como resultado del ajuste, obtienen un ingreso equivalente (ajustando por adultos por hogar) 62% superior al de EPH que, a partir de coeficientes de ajustes moderadores en ingresos bajos, tendiente a nulos en ingresos medios y elevados en ingresos altos, resultan en un crecimiento sustancial de la desigualdad.

La nueva adaptación de datos corregidos, que deja atrás las recomendaciones de no corrección, goza actualmente de creciente aceptación, al punto de que incluso se publican tableros de datos para fines de investigación y divulgación con explicaciones sobre estas correcciones y el detalle de los datos obtenidos de las mismas (Tornarolli, Gasparini, y Schteingart, 2024; Gasparini, 2024)[11].

2.1.2. Se puede corregir para ver que hay

Durante la primera etapa de correcciones revisada, Camelo (1998) busca dimensionar la subcaptación en EPH a partir de una corrección de los ingresos de la encuesta a través de coeficientes para igualar los agregados de las CCNN, del mismo modo que ya fuera revisado. Diferencia por fuente de ingresos y también por quintil. Al mismo tiempo, compara ingresos salientes de la ENGH contra tabulaciones provenientes del SIJyP. Encuentra subdeclaración para todas las fuentes, aunque de distinta intensidad: baja en salarios, elevada en ingresos provenientes de rentas. Al mismo tiempo, encuentra diferente grado de subdeclaración por quintil de ingreso: a mayor quintil de ingresos mayor subcaptación, para todas las distintas fuentes de ingreso consideradas. Por el contrario, no encuentra que las distintas ramas de actividad sean relevantes en esta clave, en tanto prima el efecto de los quintiles de nivel de ingresos.

En la misma etapa, Sosa Escudero y Alaimo (2000) buscan dimensionar la subdeclaración en la Argentina como porción conformante de la "economía oculta". Lo

[10] Se emplea MLER derivado de SIPA para salarios registrados (se asume misma proporción para no registrados), registros administrativos del Ministerio de Trabajo para ingresos por jubilaciones y de cuentapropistas -monotributistas y autónomos-, CCNN para ingresos de capital, y adicionalmente reajuste general con Interpolación de Pareto para los *top incomes* del 0,5% superior en base a información tributaria.

[11] Pueden consultarse en https://argendata.fund.ar/topico/salarios-e-ingresos/ y https://argendata.fund.ar/topico/desigualdad/.

hacen insumiéndose de encuestas de gastos de hogares (puntualmente, la ENGH) vía método de discrepancias en el consumo: emparejando familias con similar estructura e ingresos, pero distintos consumos. Las diferencias en términos de consumo sirven como aproximación de la subdeclaración de un grupo de familias respecto a otro. Estrictamente, si ambas subdeclaran, sólo es estimada la diferencia entre ambas subdeclaraciones.

Posteriormente, Felcman, Kidyba y Ruffo (2004) busca ajustar los ingresos de EPH, pero sin pasar por las CCNN, sino que usando ENGH y SIJyP. Mezcla las estrategias empleadas en dos de los trabajos revisados: aplican el mismo procedimiento que Gasparini (1999) para imputar un ingreso a quienes no responden, y el mismo procedimiento que Sosa Escudero y Alaimo (2000) para corregir subdeclaración en no asalariados vía discrepancias de gastos en la ENGH. Particularmente, corrige asalariados contra SIJyP en los cincuentiles que la diferencia es más de 5%, y aplica esos coeficientes contra todos los asalariados, no sólo los registrados. Este trabajo aparece como antecedente del de INDEC (2006), en donde se detalla el armado de la CGI para 1993-2005 estimando ingresos por categoría ocupacional. En base a EPH y SIJyP, parte de postular a la subdeclaración como un problema generalizado, pero con mayor relevancia en patrones y cuentapropistas, donde la alta variabilidad en sus ingresos aparece como un factor relevante para la subdeclaración. Nuevamente, como en el trabajo anterior, el supuesto es la ausencia de subdeclaración en asalariados. Corrigen también no respuesta comparando asalariados registrados EPH con SIJyP tomando remuneración neta de puesto principal del CUIL sumando asignaciones familiares y restando pagos, IIGG y otras deducciones. Consideran también correcciones en el sentido inverso: de sectores que, por sus características, subdeclaran al SIJyP, de modo que registran mayores ingresos en EPH. Aplican también corrección de Heckman, siguiendo a Gasparini, Marchionni y Sosa Escudero (2000). En el análisis por sector productivo, no encuentra subdeclaración relevante en Industria (D) ni Electricidad, Gas y Agua (E); pero sí subdeclaración al SIJyP -sobrecaptación en EPH- en Construcción (F), Comercio (G) y Actividades Empresariales (K).

En esta línea, Groppa (2009) busca simultáneamente dimensionar subdeclaración y corregir desigualdad. Inicialmente realiza correcciones al consumo siguiendo patrones obtenidos al comparar los ingresos de la ENGH con información de registro por percentiles. En un segundo momento, asigna un vector de consumo a tramos de ingresos elevados no captados por las encuestas, pero sí mediante la declaración jurada del

Impuesto a las Ganancias. Realiza inicialmente correcciones a la ENGH desde SIJyP, para luego incorporar información proveniente de DDJJ de IIGG. Entre los resultados obtenidos, destaca una notable subcaptación de ingresos para los percentiles de mayores ingresos particularmente para la región del NEA.

Finalmente, Minoldo y Born (2019) desarrollan una minuciosa contrastación para la EPH del período 2007-2015, período suscitado de desconfianza de sus datos dada la intervención del INDEC. Tras contrastar las distintas acusaciones, concluyen que no hay señales sólidas para pensar que EPH fuera manipulada, e incluso que los ingresos se encuentran particularmente más subestimados en este subperíodo respecto a otros.

Como resumen de esta revisión, la subcaptación de ingresos se ha tratado de forma diferencial en distintas etapas temporales. Generalmente, como un obstáculo a evadir, en pos de instrumentalizar su corrección para recalcular la desigualdad por ingreso o la pobreza, con distintos y disímiles resultados. De distinto impacto, pero generalmente las correcciones llevan a mayor desigualdad y menor pobreza respecto a las cifras oficiales que publica el INDEC. No obstante, también existen trabajos buscan dimensionar el fenómeno como un fin en sí mismo. Incluso dentro de este grupo también existe divergencia en términos de relevancia de las dimensiones como región o sector productivo como relevantes en sí mismas a la subcaptación.

2.2. La evolución de la brecha de ingresos entre EPH y SIPA

Una noción cuantitativa agregada sobre la subcaptación de ingresos en la EPH puede obtenerse al delimitar[12] sus microdatos sobre los asalariados registrados privados capturados por SIPA y comparar los movimientos agregados en el largo plazo de una medida de ingreso comparable con la que capta el universo SIPA. A partir de este punto, todo lo referido a información devenida de la encuesta debe entenderse también como limitado al mundo de los asalariados registrados privados.

Como fuera comentado en la Introducción, la EPH tiene por objeto relevar las características socioeconómicas y demográficas de la población de los principales aglomerados urbanos del Argentina. En su modalidad puntual se aplicó entre 1974 y 2003 de forma bianual, con un relevamiento en mayo y otro en octubre. En su modalidad

[12] Le metodología sobre este procedimiento se detalla el siguiente capítulo.

continua, vigente desde 2003, es realizada ininterrumpidamente, pero publica datos con frecuencia trimestral[13].

Si bien los datos recolectados en el SIPA no son de dominio público, los mismos son publicados en forma agregada por el Observatorio de Empleo y Dinámica Empresarial (OEDE)[14] de forma mensual y continua desde enero de 1995 hasta la actualidad, con un trimestre de rezago hacia el dato más actualizado. La serie de remuneraciones por puesto de trabajo publicada por OEDE es de cobertura nacional e incluye todas las relaciones laborales (combinaciones CUIT-CUIL) reportadas por las empresas al SIPA para el pago de aportes y contribuciones. Se contemplan para esta serie todas las ramas de actividad del sector privado, excluyendo el servicio doméstico. Aunque en SIPA también se pueden presentar problemas de captación de ingresos a través de la falsabilidad de las declaraciones juradas realizadas, es razonable suponer que sus datos son más fidedignos que los de la EPH y que representan mejor los movimientos del salario, tanto en variación como en niveles. Incluso no sólo respecto a EPH, sino también frente al resto de las series oficiales disponibles (Kennedy *et al.*, 2024).

En adición a las dificultades generales de compatibilización de información entre registros tributarios y encuestas de hogares en términos de unidad de análisis y de ingresos, ya revisadas, cabe considerar algunas otras cuestiones concretas de dificultad de compatibilización entre EPH y SIPA. En primer lugar, en términos de unidad de información. En el caso de los registros tributarios, la unidad es la relación laboral. Concretamente, la relación que empareja el documento tributario del empleador (CUIT) con el del empleado (CUIL). De manera que un mismo CUIT puede tener emparejados para el mismo período de tiempo muchos CUIL, como habría de esperarse para establecimientos que cuentan con más de un obrero u empleado; pero al mismo tiempo un CUIL puede estar emparejado con varios CUIT. Es el caso del pluriempleo. En el caso de las encuestas de hogares, en cambio, la unidad de análisis es el individuo encuestado. De modo que, como definición en general y en el caso concreto de la EPH en particular, el foco está puesto en la ocupación principal del individuo por sobre ocupaciones secundarias.

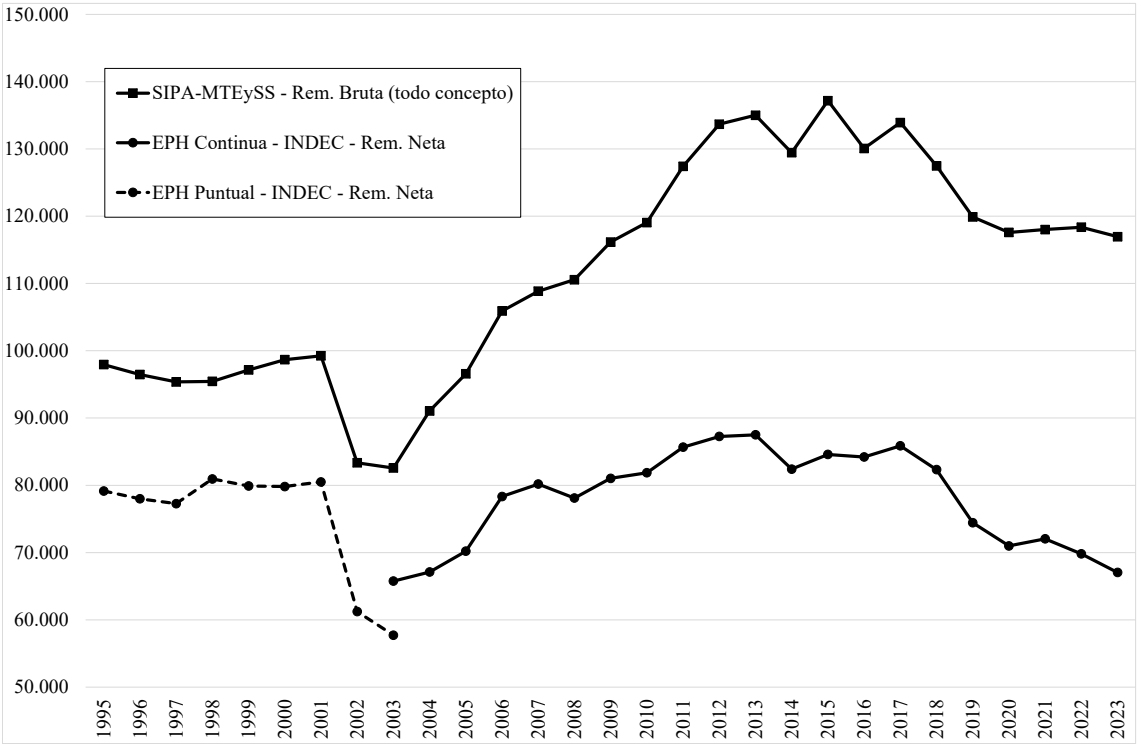
[13] Para un detalle integral de la concepción de la encuesta en su modalidad continua, ver INDEC (2003).
[14] Antes dependiente del Ministerio de Trabajo y Seguridad Social (MTEySS), hoy de la secretaria del mismo nombre, dependencia del Ministerio de Capital Humano (MCH).

Otra diferencia entre fuentes de información radica en la periodización de los datos. A tono con los formularios que demandan las obligaciones tributarias, la información de SIPA está disponible mes a mes. Si bien las encuestas pueden tener periodización mensual, en el caso de la EPH en su modalidad continua -si bien los relevamientos son continuos, ininterrumpidos en el tiempo- se publica a partir de agrupar los datos trimestralmente. De modo que las respuestas de un individuo corresponden al ingreso percibido en uno sólo de los tres meses contemplados en la publicación de los datos de la encuesta. Esto puede resultar particularmente problemático en contextos de elevada inflación.

A este punto, a la hora de comprar ingresos entre fuentes no debe perderse de vista la diferencia de carácter entre el ingreso obtenido por la encuesta y el presente en los registros administrativos. En la EPH el ingreso es auto reportado por el individuo encuestado. Se trata de una declaración del salario "neto" o de bolsillo[15]. Es decir, sin incluir -dado que nos circunscribimos al universo de los asalariados registrados- los aportes personales. Los registros tributarios referidos, en cambio, contienen información de salario bruto, de modo que incluyen los aportes personales. Particularmente debe considerarse, además, que entre salario bruto y neto medían otros tributos, de cambiante incidencia temporal, que pueden afectar esa distancia (al menos en niveles). Principalmente las percepciones vinculadas al pago del Impuesto a las Ganancias (4ta categoría) y las asignaciones familiares. De modo que, movimientos de expansión o contracción en la brecha observada por las medidas de tendencia central de ambas fuentes podrían explicarse por variaciones en estos componentes.

[15] Estrictamente, entre el salario neto y el de bolsillo medían las asignaciones familiares y tributos como el Impuesto a las Ganancias. Por simplicidad, en este trabajo “neto” y “de bolsillo” son considerados sinónimos.

Gráfico 1. Promedio anual de ingresos de asalariados registrados sin servicio doméstico, por fuente. En $ constantes de nov-21. 1995-2023.



Fuente: elaboración propia en base a MTEySS e INDEC.

La distancia que existe en los promedios en ambas fuentes muestra un ensanchamiento en la diferencia entre las mismos. A tono con lo revisado (Roca y Pena, 2001; Sánchez, Pacífico y Kennedy, 2016), la brecha entre fuentes para la etapa de la EPH puntual hasta el año 2000 estaría acorde a los niveles "teóricos", que se deducen a partir de considerar la estabilidad en la alícuota de aportes a cargo del trabajador en torno a 17% -11% previsional, 3% al instituto de la seguridad social y 3% por obras sociales- durante la década del 90 (Bertranou *et al.*, 2011). Sin embargo, a partir de nuevo milenio -y con mayor profundidad con el cambio de modalidad de EPH- la distancia en el ingreso promedio captado entre fuentes no hace más que agrandarse. El último dato disponible muestra un nivel salarial muy distinto entre fuentes, lo que dimensiona la relevancia del problema. En el caso de SIPA, se encuentra en un nivel similar al que la misma fuente muestra para 2009. En cambio, en EPH el ingreso asalariado privado promedio de 2023 se ubica un 17% por debajo de su nivel en 2009. El nivel de 2023 para EPH resulta asimilable, en cambio, al de la misma fuente de información para el año 2004. En SIPA, 2004 se ubicaba en un incremento de 22% por debajo del nivel de esta fuente en 2023.

Cabe considerar, de aquí en adelante se referirá durante todo el trabajo a "brecha" como el cociente de la división del salario promedio en la EPH con el salario promedio en la MLER, para un año determinado.

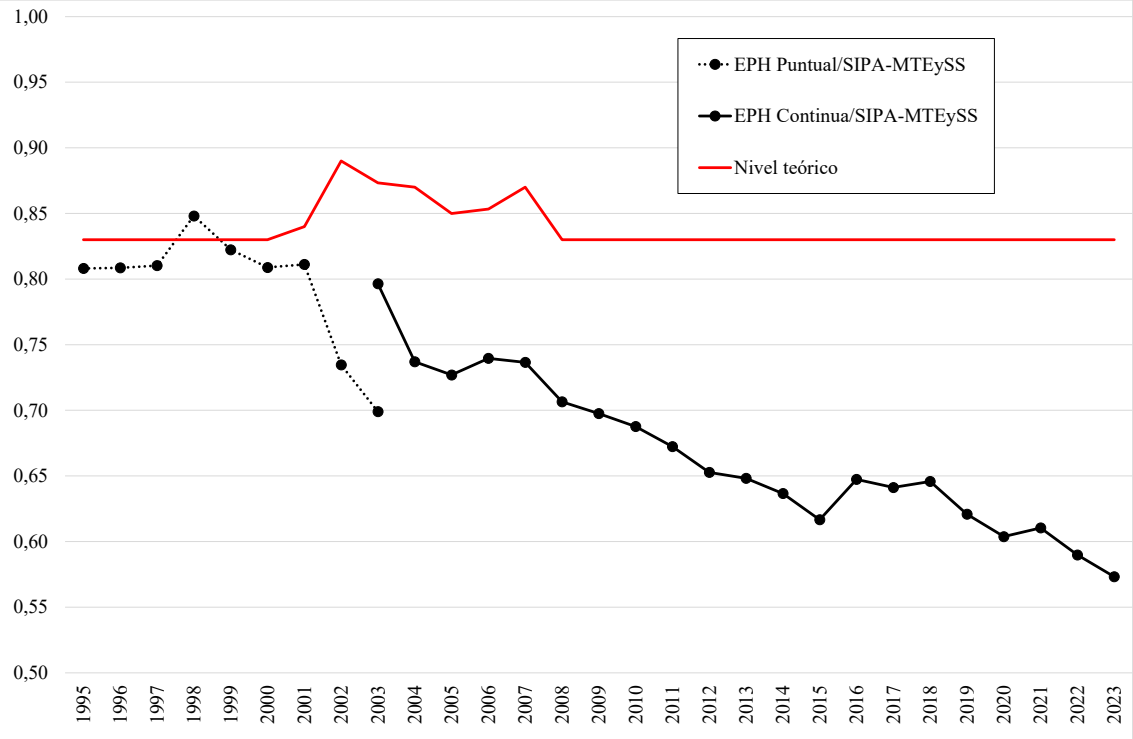
$$Br_t = \frac{WEPH_t}{WMLER_t} \quad \textbf{(1)}$$

Donde se llama $WEPH$ al salario promedio en EPH, $WMLER$ al salario promedio en MLER y el subíndice t refiere a cada período -año- determinado. De este modo, un descenso en el ratio Br_t implicará un incremento de la brecha entre los ingresos promedio de ambas fuentes, y viceversa. Así, $\boldsymbol{Br_t}$ **es un indicador del nivel de captación de ingresos en la encuesta**: más alto, mejor captación, y viceversa.

En este sentido, en el Gráfico 2 se visualiza este ratio entre salario promedio de ambas fuentes a lo largo del período considerado. Se grafica también una serie que refleja el valor teórico que debería tomar la diferencia entre fuentes de no existir subcaptación ni otra mediación entre bruto y neto que no sean los aportes personales. De modo que, para la mayoría de los años, donde la alícuota de aportes personales se sitúa en 17%, el nivel teórico de Br_t debería ubicarse en 0,83; fluctuando ante variaciones de esta alícuota. Así, serie añade al análisis una referencia de los cambios suscitados en este sentido entre 2001 y 2007. Esta serie es construida en base a información de Bertranou *et al*. (2011) y CIFRA (2009).

Cabe considerar aquí, de todas formas, dos cuestiones. En primer lugar, que existen topes sobre el haber sujeto a aportes, lo que implica que los salarios más altos -que excedan el tope- tributan un porcentaje de aportes personales menor al del nivel teórico considerado. Del mismo modo sucede con salarios con gran participación de componentes no remunerativos. En estos casos, el salario bruto de SIPA debería estar más cerca del neto de EPH-la distancia entre fuentes debería ser menor- que lo que la referencia construida para el nivel teórico indica.

Gráfico 2. Ratio entre el salario promedio de EPH y SIPA para asalariados registrados sin servicio doméstico, por etapa. Comparación respecto a nivel teórico. 1995-2023.



Fuente: elaboración propia en base a MTEySS, INDEC, Bertranou *et al*. (2011) y CIFRA (2009).

Tal como fue anticipado, los datos disponibles para la década del 1990 muestran una brecha entre encuesta y registro muy cercana a la que marca el nivel teórico dada la alícuota de aportes a cargo del asalariado. El ratio Br_t se ubica en 0,81 entre 1995 y 1997, para luego subir sorpresivamente por encima del nivel teórico hasta 0,85 en 1998, y retornar gradualmente hasta 0,81 en 2001. A este año, como marca la serie de color rojo, la alícuota de los aportes personales comienza un temporal camino de reducción. El año 2002, signado por la salida del de la convertibilidad cambiaria y un fuerte salto del tipo de cambio acompañado parcialmente por una aceleración inflacionaria, marca un fuerte descenso de la captación de ingresos en la EPH puntual respecto a los datos de SIPA, ubicando Br_t en 0,73. Como agravante, la alícuota de aportes personales que grava el ingreso bruto para los asalariados se reduce aquí hasta su mínimo (11%), lo que dicta un valor teórico para Br_t no de 0,83 como en la década anterior, sino de 0,89. Hacia el final de este período de modalidad de la encuesta en el año 2003 la brecha vuelve a ensancharse hasta alcanzar EPH un 0,7 del ingreso promedio capturado en SIPA[16]. Este nuevo

[16] Cabe considerar aquí que 2003 contiene una sola onda para EPH puntual en lugar de las dos; y dos trimestres, los terceros y cuartos, para la continua.

descenso es sólo parcialmente acompañado por un retroceso en la alícuota de los aportes personales, lo que resulta en promedio para este año de 12,7%, lo que sitúa el valor teórico en 0,87.

En el año en que conviven ambas modalidades de la encuesta, la EPH continua recupera la distancia perdida en los últimos años de la puntual, en torno a una captación del 0,8 de SIPA, apenas abajo del nivel teórico. Sin embargo, y pese al gradual descenso en la alícuota de aportes personales, tras el 2003 la brecha vuelve a ensancharse fuertemente. Entre 2004 y 2007 Br_t se encuentra estable cerca de 0,75, aunque con cambios en la alícuota de aportes personales que varía entre 13% y 15%. En 2008 la alícuota de aportes personales se estabiliza en 17%, lo que marca un valor teórico fijo hasta el final de la serie en 0,83. Sin embargo, desde este punto aparece un proceso continuo de ensanchamiento de la brecha. Hacia 2015, el salario promedio capturado en EPH representa sólo el 0,62 de lo que captura SIPA en el mismo período. Este proceso creciente es sólo parcialmente interrumpido con el cambio metodológico de la EPH en 2016, donde Br_t se sitúa en 0,65. Sin embargo, vuelve a deteriorarse y en 2019 ya iguala el 0,62 de 2015. Finalmente, para el año 2023 -último año disponible con datos anuales completos- la encuesta captura un ingreso promedio que representa el 0,57 de lo que captura SIPA, mínimo histórico de la serie.

En este punto, cabe presentar una serie de consideraciones acerca de factores que podrían estar agudizado o relativizando la brecha entre fuentes como manifestación de la subcaptación de ingresos en la encuesta.

En primer lugar, debe tenerse en cuenta la cambiante incidencia en el período considerado del del Impuesto a las Ganancias, cuya cuarta categoría grava las rentas del trabajo, impactando a los asalariados. De este modo, un aumento en su alícuota o en la cantidad de asalariados que lo tributan podría estar explicando cierto ensanchamiento de la brecha, al influir en la distancia entre salario bruto de SIPA y el reportado en EPH[17]. En este sentido, resulta relevante repasar la información respecto a cambios en ambas dimensiones tomando en consideración MECON (2023). En términos de puestos laborales registrados alcanzados por el impuesto, hacia fines de la década de 1990 y principios de los 2000 se encontraban de manera estable en 5%. Para 2005 se da un

[17] Si bien según las CCNN no se trata de un tributo que medie entre salario bruto y neto, si lo hace entre el neto y el percibido "de bolsillo".

abrupto aumento, llegando al 13% de los, pero revirtiéndose esta suba hacia 2007. Entre 2008 y 2012, las tensiones inflacionarias y el rezago en actualizaciones incrementaron significativamente la cantidad de asalariados contribuyentes, alcanzado a casi el 20% de los puestos. De 2013 a 2015, se buscó reducir la cantidad de retenidos sin resignar ingresos, lo que aumentó la carga tributaria particularmente para los salarios de alto nivel de ingresos. Entre 2015 y 2019, se recompuso gradualmente la cantidad de asalariados perceptores, particularmente desde 2018 vía inflación y necesidades fiscales. Tras la pandemia, se volvió a reducir la cantidad de contribuyentes manteniendo la recaudación, incrementando la carga sobre los retenidos. En términos de tasa efectiva, entre 1999 y 2007 se mantenía entre el 20% y el 25%, y entre 2008 y 2011 se estabilizó en torno al 25%. A partir de allí crece sostenidamente, superando el 30% para 2015. En 2017 se introducen modificaciones la reduce temporalmente, que son revertidas con el repunte inflacionario dado desde 2019, ubicando nuevamente la tasa efectiva en torno al 30% para 2021 (MECON, 2023). En este sentido, si bien la incidencia del tributo podría explicar parte del crecimiento de la brecha particularmente desde 2011, no parece sustentar la totalidad del constante y creciente rezago.

Sobre este tributo cabe considerar, de todas formas, que una porción no menor de los asalariados alcanzados por el mismo no son captados por la encuesta (Alvaredo, 2010), matizando su impacto en la brecha entre fuentes. Ahora bien, dado que el Impuesto a las Ganancias se tributa por individuo y no por puesto laboral, la combinación entre pluriempleo y este tributo puede ampliar la brecha de ingresos entre fuentes para individuos cuyo ingreso de ocupación principal no es necesariamente elevado -mayor al del umbral determinado para el impuesto-. De todas maneras, como será descrito posteriormente, excepto en ciertos sectores productivos la incidencia del pluriempleo en la muestra construida es relativamente baja.

En segundo lugar, debe considerarse el cambio en la modalidad de pago de las Asignaciones Familiares. Esto es, mientras inicialmente eran pagadas por los empleadores junto con los salarios pero de forma separada -resultando percibido en el salario “de bolsillo” pero no dentro del salario en bruto-, posteriormente comenzaron a ser pagadas en forma directa por el Estado.

Finalmente, deben tenerse presente las diferentes capacidades de captación del denominado “salario gris” (esto es, la porción del salario percibido por un trabajador registrado que se paga ilegalmente sin ser incluido en su liquidación). Mientras éste no es

captado por MLER al provenir la información de declaraciones juradas de las empresas, en la EPH sí podría encontrarse declarado por el propio trabajador.

3. CONSIDERACIONES METODOLÓGICAS: Emparejar todo lo que sea emparejable.

Este capítulo se destina a las cuestiones metodológicas a considerar antes de avanzar en una exploración cuantitativa. En una primera sección se revisa la compatibilidad de dimensiones disponibles entre fuentes empleadas y sus implicancias en términos del emparejamiento efectuado, mientras que en una segunda sección se especifican las consideraciones desprendidas de este emparejamiento vinculadas principalmente a los ingresos.

3.1. Sobre la compatibilización de dimensiones entre EPH y MLER

Los datos públicos provenientes del SIPA no permiten aprovechar un desgrane mayor en la clasificación de los ingresos que sí permiten las múltiples dimensiones socio ocupacionales que brinda EPH. Si bien el registro tributario del SIPA posee, en crudo, características de individuo, el empleador y la relación laboral que determinan distintos niveles de ingresos, por lógicas cuestiones de secreto estadístico estas bases no son de dominio público[18].

La disponibilidad de información a partir de la Muestra Longitudinal de Empleo Registrado (MLER), derivada de la información del SIPA, aparece como una interesante alternativa a abordar en el problema de la ausencia de multidimensionalidad en los datos de SIPA. La MLER contiene información a frecuencia mensual para el período abarcado entre enero de 1996 y diciembre de 2021 de una porción (3%) del total de relaciones laborales registradas contenidas en el SIPA luego de pasar por un proceso de anonimización para poder poner en dominio público la base de datos sin violar cuestiones de secreto estadístico (OEDE, 2022)[19]. De esta manera, mediante MLER se ofrece acceso a microdatos sobre salarios registrados junto con otras variables adicionales relacionadas tanto con características socioeconómicas del individuo tales como su género y edad, así

[18] Si bien el organismo que provee la información de SIPA no lo hace únicamente de manera agregada sino también a partir de algunas de las dimensiones disponibles en los datos registrales, la presentación tabulada limita, tanto en periodización como en desagregación, un potencial emparejamiento con EPH.

[19] Cabe considerar adicionalmente que se realizan distintos procedimientos para anonimizar parte de los ingresos. Concretamente, se distorsionan todas las remuneraciones de la muestra en un 3% hacia arriba o debajo de forma aleatoria; y se micro agregan a través de promedios de 3 salarios continuos todas las remuneraciones de los dos percentiles superiores de cada actividad a dos dígitos de la CIIU.

como también detalles del empleador como el tamaño de la empresa y la rama de actividad a la que corresponde. De este modo, se permite en términos generales un análisis más detallado sobre los movimientos de los salarios en SIPA en distintos subgrupos; y en términos particulares una mayor granularidad en el estudio particular de la subcaptación de ingresos.

Si bien la EPH cubre muchas más dimensiones -la base de microdatos actualmente cuenta con 177 columnas por individuo encuestado[20]-, las disponibles en la MLER pueden compatibilizarse de tal manera de encontrar poblaciones de características similares entre bases de datos. Un detalle de cada dimensión disponible entre bases y su posible compatibilización entre ellas se detalla en el Cuadro 1.

Cuadro 1. Dimensiones de MLER emparejables con EPH.

	MLER	EPH	
		Puntual	Continua
Temporal	Mensual	Onda	Trimestre
Remuneración	Bruta, del puesto de trabajo (par CUIL-CUIT en SIPA)	Neta, ingreso de bolsillo de la ocupación principal del encuestado	
Tamaño empresa empleadora	Disponible en 4 categorías: 9 o menos, 10-49, 50-200, más de 200	Disponible en 12 categorías: 1-5 exacto, 6-10, 11-25, 26-40, 41-100, 101-200, 201-500, 500 o +	
Ubicación	Provincia de ubicación legal del empleador	Región o aglomerado del hogar del individuo	
Rama de actividad	CLANAE 1997 (INDEC) a 4 dígitos	CIIU rev3	CAES-Mercosur (2003-2011) o CAES-Mercosur 1.0 (2011-2021)
Género	Disponible	Disponible	
Edad	Disponible	Disponible	

Fuente: elaboración propia en base a MTEySS (2022) e INDEC (1995, 2003, 2024).

En términos temporales, para la etapa puntual de la EPH las encuestas se realizaban sólo en dos momentos del año (ondas): mayo u octubre. En la versión continua, las encuestas se realizan ininterrumpidamente, reunidas en frecuencia trimestral. En este sentido, resulta relevante considerar un mes de rezago de ingresos para las encuestas de hogares. Es decir, lo declarado en las ondas de mayo y octubre corresponde a salarios que en MLER figuran por abril o por septiembre, mientras que los trimestrales de la EPHC

[20] Para un mayor detalle de la disponibilidad de información en EPH al momento de la elaboración de este documento, ver INDEC (2024).

corresponden a salarios de MLER de diciembre-febrero (1t), marzo-mayo (2t), junio-agosto (3t) y septiembre-noviembre (4t). Esto se debe a que los individuos en las respuestas responden por su último ingreso percibido en su ocupación principal, el cual corresponde al trabajo realizado a mes vencido. En los registros administrativos, en cambio, los meses corresponden al mes trabajado, no al mes de percepción del ingreso. A su vez, esta diferencia metodológica en términos de períodos de captura de información configura dos series cualitativamente diferentes, por lo que de aquí en adelante el análisis se realizará de forma separada entre ambas etapas.

Sobre el género de los asalariados, debe considerarse que en MLER no existe información sobre esta dimensión en el 0,1% de las observaciones del total de la base, registros que son desechados. Para EPH, en cambio, la disponibilidad de información es completa.

En términos de edad, al tratarse de una variable discreta pero tendiente a continua, se opta por reducir sus dimensiones posibles a 3 grupos: individuos menores de 25 años, de entre 25 y 45 años, y mayores de 45 años. Adicionalmente, se debe considerar que para esta desagregación se desestima una pequeña porción de MLER en donde no es posible reconstruir la edad de los individuos (entre 1% y 2%).

La categorización del tamaño de la empresa empleadora aparece diferenciada. En MLER se disponen sólo 4 grandes categorías, asimilables a como coloquialmente se seccionan empresas por tamaño en micro, pequeñas, medianas y grandes. En EPH en cambio la disponibilidad de información en esta dimensión es de mayor granularidad, e igual en sus dos versiones. Los establecimientos pequeños de menos de 5 personas se identifican en su cantidad exacta, y en las grandes empresas la categoría mayor incluye a los establecimientos de más de 500 personas. Además, no sólo en términos de cantidad de categorías sino también por la ausencia de un solapamiento perfecto, de modo que el agrupamiento de las categorías de EPH en determinado sentido no deviene en obtener exactamente las categorías disponibles en MLER. Dicho de otro modo, las categorías no son estrictamente comparables en tanto se superponen en algunos segmentos. Por ejemplo, para la categoría 41-100 de EPH algunas observaciones corresponden a la categoría 10-49 de MLER, mientras que otras a la categoría 50-200, y es imposible una delimitación exacta entre estas.

Para emparejarlas, se optó metodológicamente por trasladar las categorías EPH a categorías MLER a través de asignaciones aleatorias suponiendo la misma probabilidad de pertenencia por cantidad de trabajadores para cada categoría. De este modo, a la categoría de EPH de 6 a 15 se asignó 50% a la primera categoría de MLER (menos de 10 personas) y otro 50% a la segunda categoría MLER de (entre 11 y 50 personas), repitiendo este proceso con todo el resto de los casos de solapamiento imperfecto. Cabe considerar también que la muestra de EPH se reduce dada la imposibilidad de reconstruir esta dimensión en algunos casos. Esto se da entre el 5% y el 10% de los casos a lo largo del período considerado, observaciones que resultaron desechadas.

Cabe resaltar aquí, de todas formas, la diferencia cualitativa entre el tamaño del establecimiento vinculado EPH y el de cantidad de asalariados asociados a una firma en MLER. En este último caso, su distribución puede darse en más de un establecimiento asociado al mismo empleador en términos legales.

En torno a la dimensión de la ubicación geográfica se opta por trabajar a un nivel de alta agregación de modo no resulte problemático dado tamaño muestral disponible. Se consideran las 6 regiones dispuestas por el INDEC: GBA, Pampeana, Patagónica, NEA, NOA y Cuyo. Desde MLER se agregan las provincias según su pertenencia[21] a estas regiones determinadas. Cabe considerar, a su vez, que la región construida por MLER puede verse sesgada al tratarse de la provincia de pertenencia legal de la empleadora en la relación laboral, mientras que en EPH en cambio, el aglomerado o región se corresponde al del hogar en donde se encuesta al individuo. Particularmente la región de GBA, dada la existencia de múltiples firmas que operan en otras regiones, pero cuentan con su domicilio legal en la Ciudad de Buenos Aires, puede contar con casos considerados en ella que realmente pertenecen a otra región. A su vez, la cobertura de MLER, derivada de SIPA, capta puestos laborales de cualquier lugar del país; mientras que EPH lo hace sólo en los aglomerados donde se efectúa la encuesta.

La rama de actividad se presenta en distintos clasificadores. En MLER aparece para todo el período uniformemente presentada a nivel de clases (4 dígitos) del CLANAE 1997 de INDEC. Este clasificador resulta empalmable a distintos niveles al Clasificador

[21] Partidos de GBA y Ciudad Autónoma de Buenos Aires en MLER para GBA en EPH; Buenos Aires, Córdoba, Entre Ríos, La Pampa y Santa Fe para Pampeana; Chubut, Neuquén, Río Negro, Santa Cruz y Tierra del Fuego para Patagónica; Corrientes, Chaco, Formosa y Misiones para NEA; Catamarca, Jujuy, La Rioja, Salta, Santiago del Estero y Tucumán para NOA; y Mendoza, San Juan y San Luis para Cuyo.

Industrial Internacional Uniforme (CIIU) en su tercera revisión -disponible en EPH puntual- como también se deriva de este último el CAES-Mercosur -disponible en la primera etapa de EPH continua-. Tras la actualización del CIIU a revisión 4 (ONU, 2009), y por ende la del clasificador del bloque regional, EPH emplea desde el año 2011 el CAES-Mercosur 1.0.

En torno a esta dimensión, cabe resaltar varias cuestiones en términos de compatibilización y elección de los clasificadores de actividad. En primer lugar, por cuestiones de cantidad de datos y agregación, el análisis se realiza a nivel de sección (a 1 dígito o letra). A su vez, a esta agregación es posible clasificar todas las observaciones en ambas fuentes (a toda vez que la información se encuentre disponible, claro está). En segundo lugar y en consecuencia de lo anterior, mientras que a este nivel de agregación los clasificadores del CLANAE y CAES Mercosur son compatibles con el CIIU rev. 3, para la información posterior a 2011 contenida en EPH en su etapa continua con el clasificador CAES Mercosur 1.0, es necesario retroceder a sus equivalencias en CAES Mercosur. En tercer lugar, considerar las observaciones de las dos fuentes que deben ser desechadas por no encontrarse contenidas en la otra. En concreto, Administración Pública y Organismos Extraterritoriales están ausentes -por definición- en MLER, por lo que se descartan de EPH. En espejo, dada la cobertura urbana de la encuesta, Agro y Pesca es desechado de ambas fuentes (mayor presencia en MLER, residual en EPH). A su vez, se desestiman en ambas fuentes se desestiman los individuos cuya clasificación de rama productiva recae en la categoría residual Otros. Finalmente, en EPH se quitan trabajadoras del servicio doméstico (alcanzadas por el SIPA pero en un régimen distinto de información, con lo cual están ausentes en MLER). De esta manera, se consideran 12 distintas ramas para el análisis: Minería, Industria, Energía, Construcción, Comercio, Hoteles/restaurantes, Transporte, Finanzas, Servicios Empresariales, Enseñanza, Salud y Servicios Comunitarios.

3.2. Otras consideraciones sobre el emparejamiento propuesto

Más allá de las particularidades en las dimensiones socioocupacionales consideradas, existen otros detalles metodológicos del emparejamiento que deben ser tenidos en cuenta. En primer lugar, a la hora de efectuar la encuesta existen individuos que optan por no declarar sus ingresos. En segundo lugar, en ambas fuentes de

información existen registros suficientemente bajos que pueden tratarse de errores y/o falsaciones, lo que sesga la información. Por ambas razones, tras el corte grueso del universo EPH la compatibilización de bases demanda un refinamiento mayor en la dimensión de ingreso. En cuanto al primer problema, para no incurrir en correcciones, aquellos individuos que no declaren ingresos en EPH no son tenidos en cuenta. De todas formas, se reconoce esta no respuesta como parte del mismo problema a indagar en tanto la no respuesta está asociada a mayores niveles de ingreso, como ya fuera revisado anteriormente (Salvia y Donza,1999; INDEC, 2006). En cuanto al segundo, se delimita un umbral arbitrario de ingresos de ocupación principal para eliminar observaciones con valores exageradamente bajos. De $10.- netos para ingreso de bolsillo en EPH o $12.- brutos de los registros de MLER de abr-03 (a donde se deflacta inicialmente en la etapa de EPH puntual) y su equivalente deflactado en nov-21: $750.- netos o $877.- brutos. Por ser el último mes disponible en EPH en su etapa comparable con MLER (salarios del 4to trimestre de 2021, que a lo sumo se declararon en diciembre, correspondiendo al devengado en el mes de noviembre), todo el análisis de este trabajo se realiza en precios constantes de noviembre de 2021.

A diferencia de lo revisado con SIPA, el período a contrastar se delimita entre 1996 y 2021, en tanto este intervalo es el que determina la disponibilidad de MLER. De este modo, se pierde la referencia para los años 1995, 2022 y 2023. A su vez, deben tenerse que los valores promedio anual para EPH en algunos casos no corresponden al año completo según ondas (puntal) o trimestres (continua). Esto es, el año 2003 en la etapa puntual sólo contempla la onda de mayo, mientras que el año 2003 en la etapa continua sólo los trimestres tercero y cuarto. Tampoco se encuentra disponible el tercer trimestre de 2007, el tercero y cuarto de 2015 ni el primero de 2016, por lo que los promedios anuales se realizan sin esta información (en ambas fuentes).

Como fuera mencionado, dado que en MLER (SIPA) se registran relaciones laborales CUIL-CUIT y en EPH se encuesta a individuos perceptores de ingresos, es necesario compatibilizar también el ingreso de ocupación principal que se reporta en EPH con el registro de salario bruto de MLER. Ambas fuentes captan el pluriempleo, aunque MLER sólo en caso de que las ocupaciones secundarias sucedieran en condición registrada. En términos de relevancia e incidencia, se identifica en ambas fuentes como al pluriempleo principalmente relevante en el sector de enseñanza (entre 15% y 30% en MLER; 30 y 50% en EPH). Metodológicamente se opta por realizar ninguna corrección

en este sentido. Se conserva el ingreso de ocupación principal en EPH, y el puesto con mayor nivel de ingresos por mes por individuo en MLER.

Particularmente debe considerarse que la existencia de ingresos no registrados pero provenientes de una ocupación sí registrada tributariamente (coloquialmente conocido como "salario en gris") podría actuar como factor de contrapeso de la subcaptación. De modo que, parte del ingreso de la ocupación esté siendo efectivamente declarado por el trabajador en la EPH, pero no figura en los registros previsionales del SIPA ni en MLER[22].

[22] Por un detalle en términos de operaciones realizadas sobre las bases de datos de las fuentes de información empleadas, ver Anexo I.

4. EL ANÁLISIS DE A UNA DIMENSIÓN ENTRE EPH Y MLER

Al delimitar la muestra de la EPH al universo que capta MLER de asalariados registrados, para cada período temporal y para cada combinación de características asimilables entre fuentes, es posible construir series de promedio de ingresos desagregados por distintas dimensiones socioocupacionales. De mostrar comportamientos diferenciables, los movimientos evidenciados por las series en esta apertura son un primer paso exploratorio para afinar una caracterización de las características de la subcaptación de ingresos en EPH.

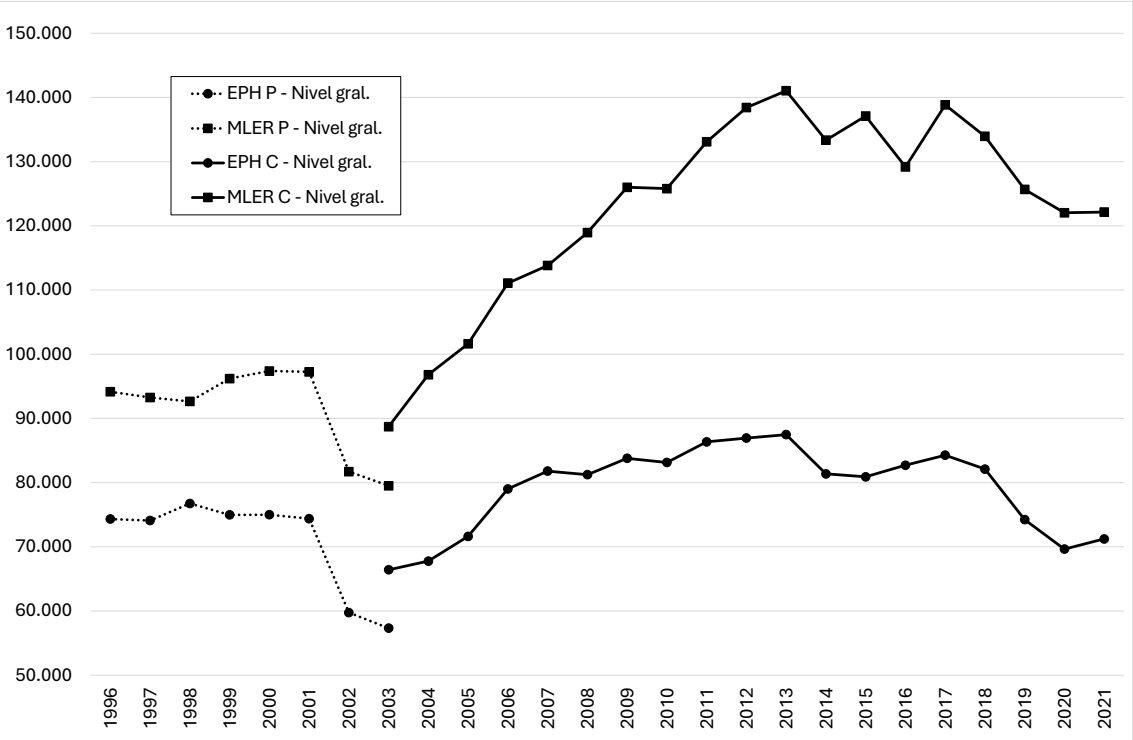
Para la contrastación entre los datos se considerarán los ingresos promedio anuales reflejados en cada fuente, tal como se ha realizado con SIPA en la sección anterior. A su vez, las comparaciones con cada nueva dimensión socio ocupacional disponible se realizarán una a una, en tanto se consideran para cada caso las múltiples categorías que componen cada variable, de a una variable a la vez.

De este modo, en una primera sección del capítulo se revisan las diferencias agregadas entre el recorte de ambas fuentes. En una segunda, se replica el análisis considerando de forma separada características vinculadas al individuo: género y edad. En la tercera sección, se consideran las características vinculadas al puesto laboral: tamaño del establecimiento, región y sector productivo. En la cuarta sección se presenta un resumen cuantitativo y se recuperan coincidencias con la literatura revisada. En una última sección se ofrecen los hallazgos desprendidos del análisis realizado.

4.1. La comparación a nivel general

En el Gráfico 3 se visualizan los movimientos que exhiben los promedios salariales anuales de MLER y EPH para el período seleccionado de forma análoga a como fuera realizado con SIPA anteriormente, en busca de contrastar con los datos obtenidos desde MLER.

Gráfico 3. Nivel general de ingresos de asalariados registrados sin SSD de EPH y MLER. Promedio anual en $ constantes de nov-21. 1996-2021.



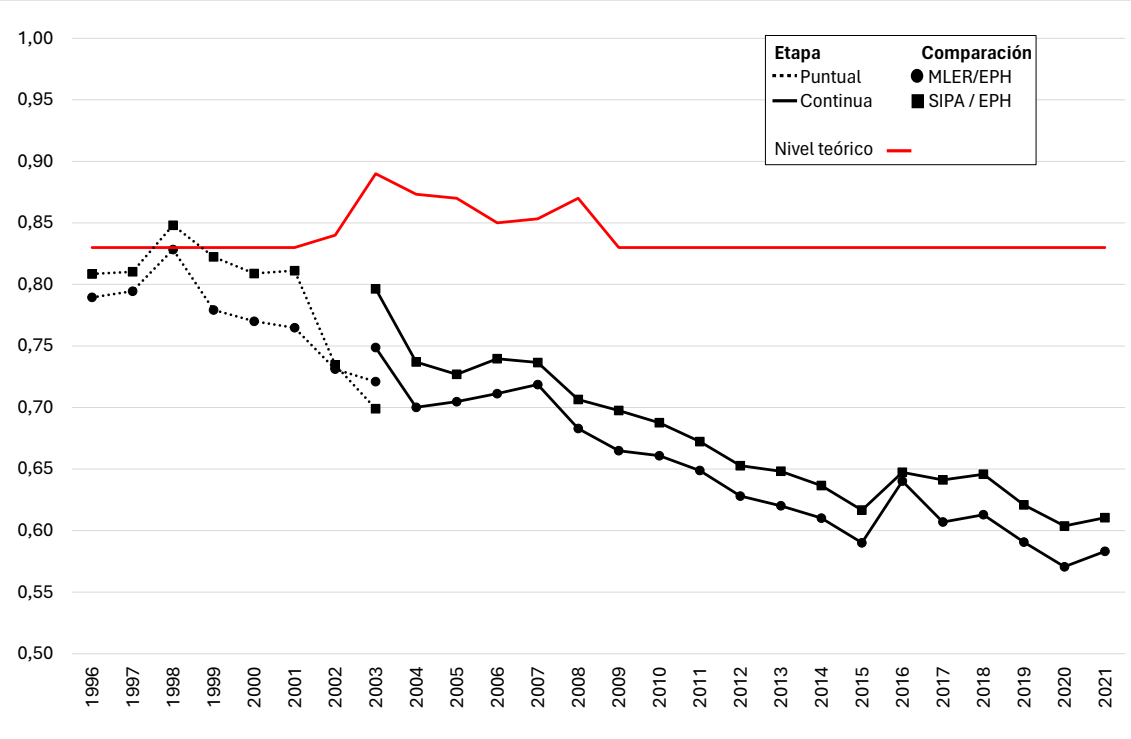
Fuente: elaboración propia en base a MTEySS e INDEC.

MLER -en tanto muestra de SIPA, y a su vez en la submuestra determinada para emparejar con EPH- replica con mucha fidelidad el nivel y los movimientos agregados de la serie de SIPA (Gráfico 1). Adicionalmente, el salto entre versiones puntual y continua se ve ahora reflejado en ambas series, a causa de separar la información de MLER para ese año en etapa puntual (sólo el mes de abril, emparejable con la onda de mayo) y etapa continua (los meses de junio a noviembre, emparejables con el tercer y cuarto trimestre).

En términos de diferencias en niveles salariales, se pone en relieve nuevamente la disparidad entre fuentes. En concreto, para la etapa continua el último dato disponible de la encuesta ubica un nivel salarial de similar al reflejado para el año 2005. En cambio, los datos de MLER para 2021 ubican un salario registrado promedio un 20% mayor al de la misma fuente para el año 2005, asimilable en cambio a niveles de entre 2008 y 2009. A su vez, la irregularidad del subperíodo 2013-2017 intercalando años de crecimiento con años de caída en los salarios se visualiza de forma marcada en MLER, mientras que en EPH tiende a suavizarse. De este modo, en este subperíodo se denota también un comportamiento diferente en términos de variaciones.

Para visualizar con claridad los cambios en la distancia relativa entre ambas series, en el Gráfico 4 se ilustra una comparación del ratio Br_t dado por el cociente resultante entre los promedios de EPH y MLER con la que fuera anteriormente construida para SIPA. Al igual que en el Gráfico 2, se considera también una serie de distancia teórica entre salario bruto y neto.

Gráfico 4. Ratio entre el salario promedio de EPH y SIPA para asalariados registrados sin servicio doméstico según fuente, por etapa. Comparación respecto a nivel teórico. 1996-2021.



Fuente: elaboración propia en base a MTEySS, INDEC, Bertranou *et al.* (2011) y CIFRA (2009).

La brecha agregada entre ambas fuentes se replica de forma similar, de modo que puede tomarse por válida la aproximación hacia SIPA a través de MLER. Caben destacar dos cuestiones. En primer lugar, en la brecha respecto a MLER resulta más relevante el achicamiento de la brecha reflejado en el incremento del ratio Br_t entre 2015 y 2016 (5 puntos porcentuales -de 0,64 a 0,59- en MLER vs EPH; y 3pp. -de 0,62 a 0,65- en SIPA vs EPH)[23]. En segundo lugar, que durante todo el período considerado -con excepción del año 2003 de la etapa puntual- la distancia entre MLER y EPH es mayor que a la que existe entre SIPA y EPH. En promedio, 2 pp. mayor en la etapa puntual y 3 pp. en la etapa

[23] Este comportamiento notable en ambas fuentes puede atribuirse a la revisión metodológica de la encuesta realizada en 2016, principalmente a partir de retrotraer cambios en las ponderaciones introducidos en 2013 con origen el Censo Nacional de Población de 2010 (Arakaki, 2016).

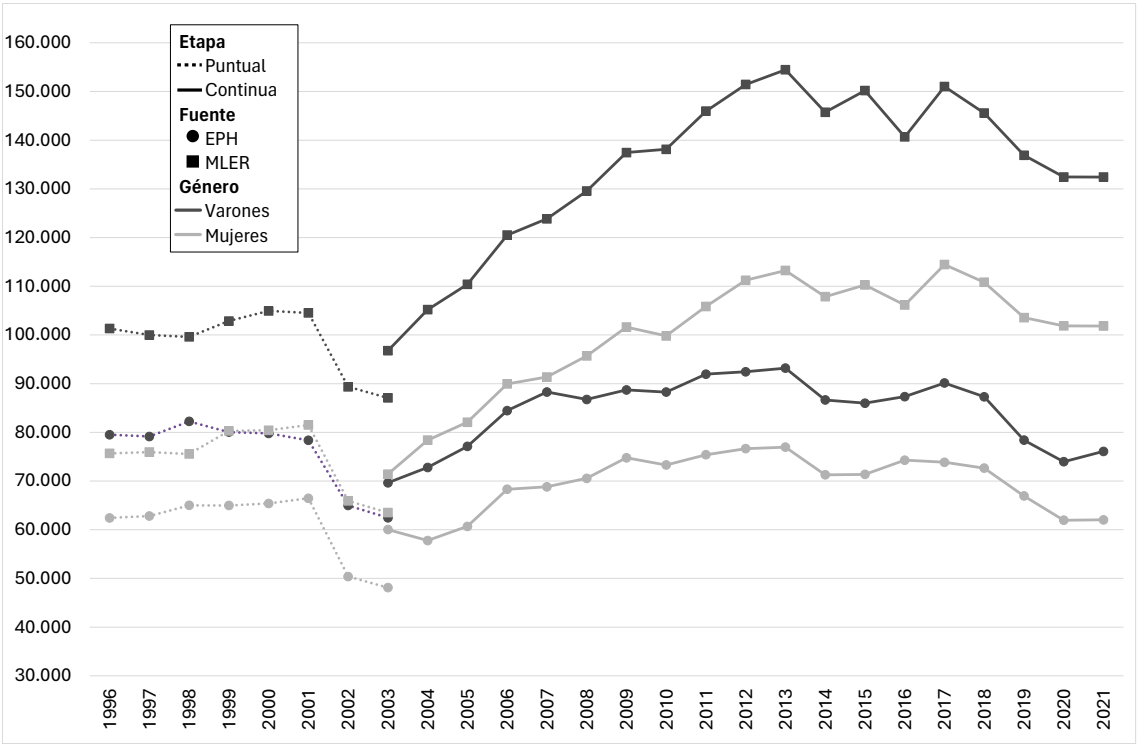
continua. Esto puede deberse al recorte pormenorizado realizado para emparejar ambas fuentes, tal como fuera descrito en la sección anterior, tras desechar por completo algunos sectores productivos que no se disponen en ambas fuentes.

Más allá de estas diferencias entre brechas, su similar comportamiento valida la comparación entre fuentes a realizar a continuidad, donde se consideran otras de las dimensiones en términos de atributos personales y del puesto de trabajo.

4.2. La comparación según atributos personales

Inicialmente, se realizan comparaciones entre fuentes análogas a las ya realizadas a nivel agregado, pero a partir del género de los perceptores de salario. En el Gráfico 5 se observan los ingresos diferenciados de varones y mujeres, para cada etapa y en cada fuente.

Gráfico 5. Ingresos de asalariados registrados sin servicio doméstico de EPH y MLER, por género. Promedios anuales en $ constantes de nov-21.



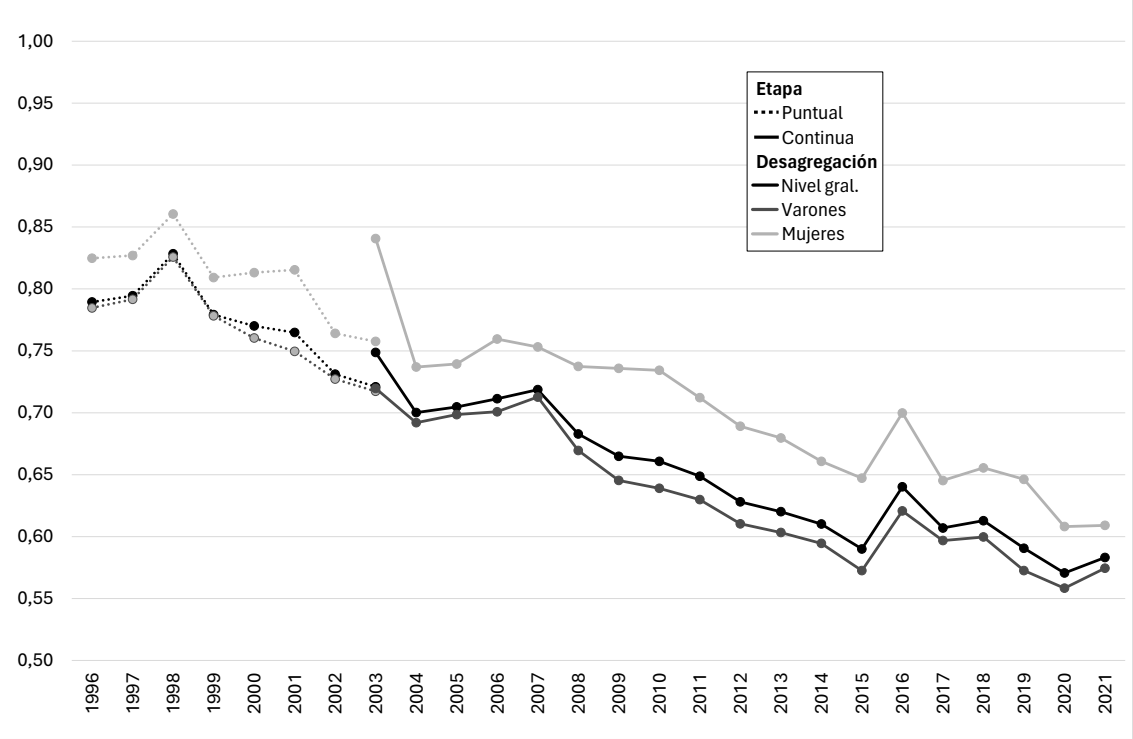
Fuente: elaboración propia en base a MTEySS e INDEC.

Como es ampliamente documentado en la literatura, en la ocupación principal el ingreso de los varones es mayor al de las mujeres, cuestión que es parcialmente reducida a la hora de considerar otros factores para la comparación, principalmente la cantidad de

horas trabajadas (Trombetta y Cabezón Cruz, 2020). Sus evoluciones, a grandes rasgos, siguen tendencias similares, pero a niveles diferentes.

De forma análoga a como fuera realizado anteriormente para los niveles generales, en el Gráfico 6 se visualiza la evolución del ratio Br para cada género. Se mantiene, a su vez, la brecha observada para el nivel agregado, en miras de tomarla como referencia para la comparación en nivel y en movimiento.

Gráfico 6. Ratio entre el salario promedio de EPH y SIPA para asalariados registrados sin servicio doméstico por género, según fuente y etapa. Comparación respecto a nivel agregado.

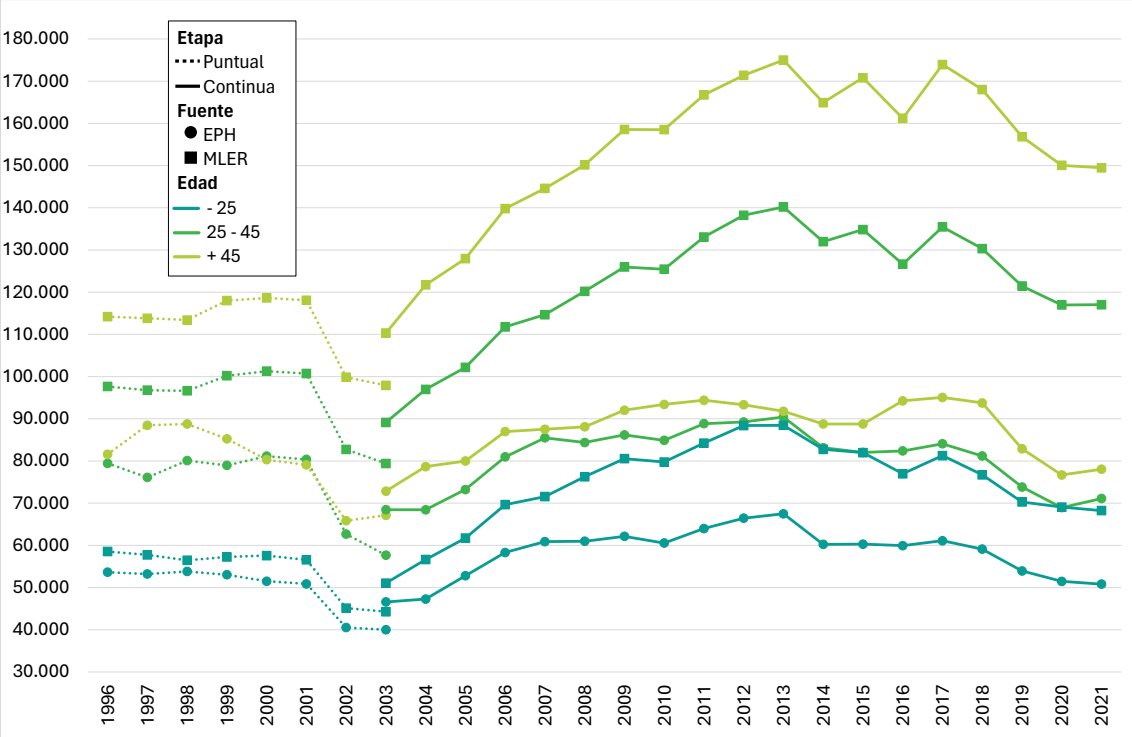


Fuente: elaboración propia en base a MTEySS e INDEC.

La serie de varones además de mostrar un mayor nivel de ingresos en promedio durante toda la extensión de la serie muestra también una brecha mayor -por ende, un ratio Br_t menor- a la del nivel general. En promedio, 0,5 pp. menor para la etapa puntual y 1,5 pp. menor para la etapa continua. La brecha diferenciada de mujeres se ubica, en oposición, en niveles de EPH más cercanos a los de MLER: 3,5 pp. mayor en promedio para la etapa puntual respecto al nivel general, y 5 pp. mayor en la etapa continua. De esta manera, resaltan dos cuestiones. Por un lado, **el mayor nivel de brecha (menor Br_t) dado en la categoría de mayores ingresos en niveles** -en este caso, varones-. Por el otro, el ensanchamiento de la diferenciación en la brecha en términos de genero para la etapa continua respecto de la puntual.

El otro atributo personal disponible entre fuentes de información es el de la edad del asalariado. En el Gráfico 7 se visualizan tanto los ingresos en niveles como los ratios Br entre fuentes para todo el período considerado por fuente, etapa y estas tres categorías de edad.

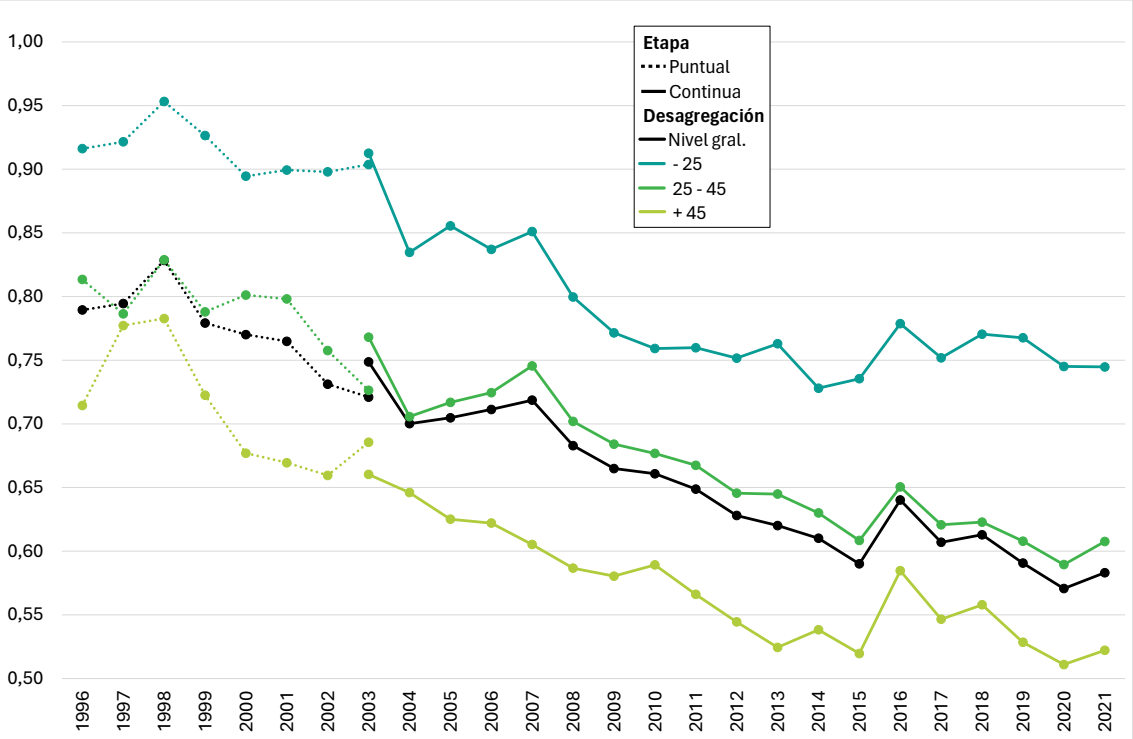
Gráfico 7. Ingresos de asalariados registrados sin servicio doméstico de EPH y MLER, por edad. Promedios anuales en $ constantes de nov-21.



Fuente: elaboración propia en base a MTEySS e INDEC.

Las categorías de edad se reparten en tres distintas franjas de ingresos para todo el período, donde los salarios más altos se corresponden con una mayor edad, y los más bajos con los comienzos de la vida laboral.

Gráfico 8. Ratio entre el salario promedio de EPH y MLER para asalariados registrados sin servicio doméstico por edad, según fuente y etapa. Comparación respecto a nivel agregado.



Fuente: elaboración propia en base a MTEySS e INDEC.

Resalta el hecho de que, en la etapa puntual, etapa de baja diferencia entre la brecha entre fuentes y el nivel teórico, como fuera revisado; la categoría más joven muestra un ingreso muy parecido entre fuentes -Br_t en torno a 0,9-. En contraste, en la misma etapa para la categoría de 45 años se revela una distancia entre fuentes en promedio mayor en un -Br_t promedio alrededor de 0,7- aunque también de mayor volatilidad. Se repite el patrón observado en términos de género: la correspondencia dada en términos de menor Br_t para categorías de mayor nivel de ingresos.

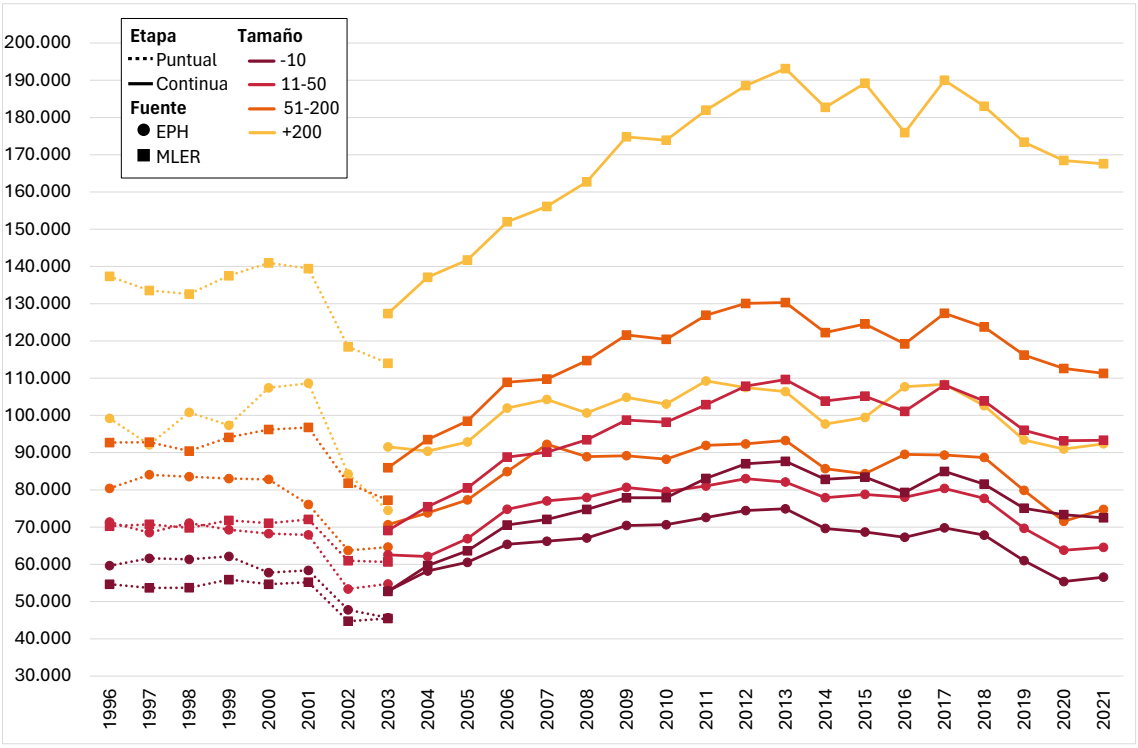
Considerando la evolución de las brechas por categoría de edad respecto al nivel general, mientras que la categoría intermedia de entre 25 y 45 años mantiene un coeficiente Br_t mayor -por ende, una captación mejor en la encuesta- que el del nivel general (1,5 pp. mayor estable entre puntual y continua), la categoría de menores de 25 años aparece siempre con un ratio Br_t marcadamente superior. Esto es, 15 pp. mayor en promedio en la etapa puntual y 13 pp. mayor en promedio en la etapa continua. Análogamente, la de mayores de 45 resulta inferior: 6 pp. menor promedio en la etapa puntual, 7,5 pp. menor promedio en la etapa continua. La correspondencia principal vuelve a aparecer en torno al nivel salarial. **A mayor edad se corresponde un mayor**

nivel de ingresos promedio -en ambas fuentes-, y a mayor nivel de ingresos una mayor brecha -menor ratio Br_t- en el salario captado entre fuentes.

4.3. La comparación según atributos del puesto de trabajo

Se ubica en atributos del puesto de trabajo a aquellos que corresponden a la dimensión productiva del puesto de trabajo (tamaño del establecimiento empleador, sector productivo), como también al lugar en donde se lleva a cabo el trabajo (región). Continuando la dinámica de revisar desde dimensiones de mayor agregación hacia dimensiones de mayor desagregación, inicialmente se trata lo referido a tamaño del establecimiento (4 categorías disponible), luego región (6 categorías) y finalmente sector productivo (12 categorías).

Gráfico 9. Ingresos de asalariados registrados sin servicio doméstico de EPH y MLER, por tamaño de establecimiento. Promedios anuales en $ constantes de nov-21.

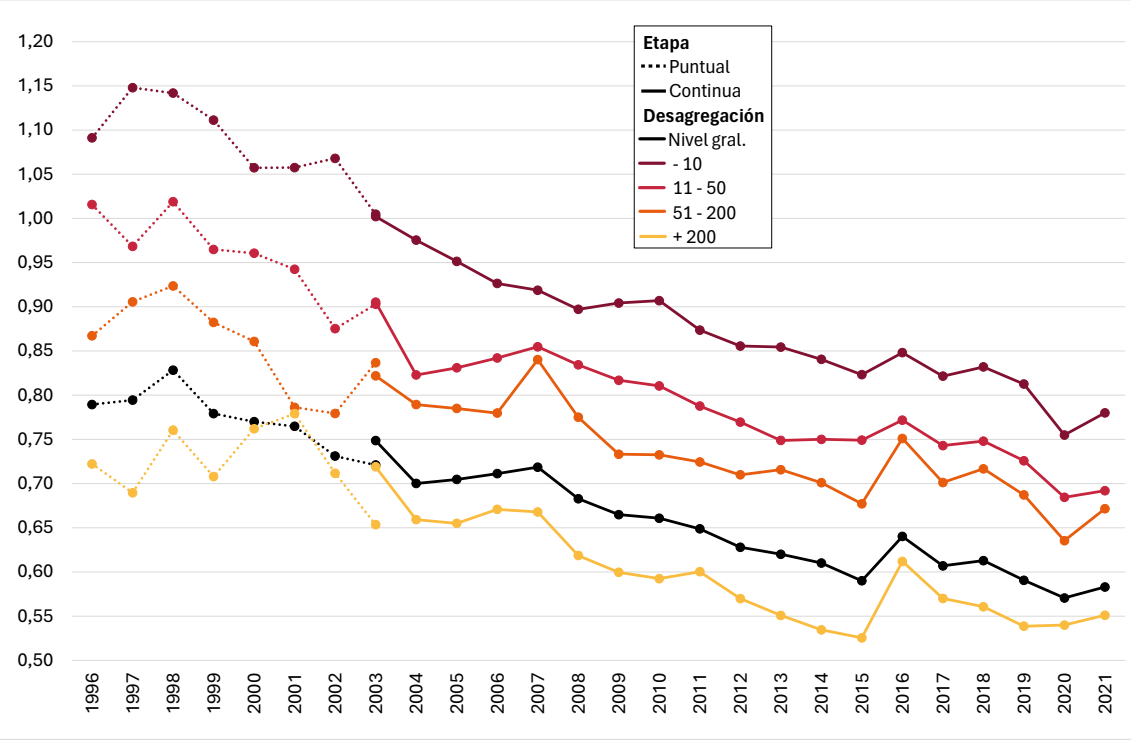


Fuente: elaboración propia en base a MTEySS e INDEC.

En el Gráfico 9 se visualiza el nivel de ingresos promedio de cada grupo por fuente, etapa y tamaño del establecimiento empleador. Establecimientos más grandes se corresponden con niveles de ingresos (en ambas fuentes) mayores. Resalta particularmente la categoría menor, la de pequeños establecimientos -de menos de 10

empleados-, que muestra que en la etapa puntual los ingresos de la encuesta se encontraban por encima de los de MLER.

Gráfico 10. Ratio entre el salario promedio de EPH y MLER para asalariados registrados sin servicio doméstico por tamaño del establecimiento, según fuente y etapa. Comparación respecto a nivel agregado.



Fuente: elaboración propia en base a MTEySS e INDEC.

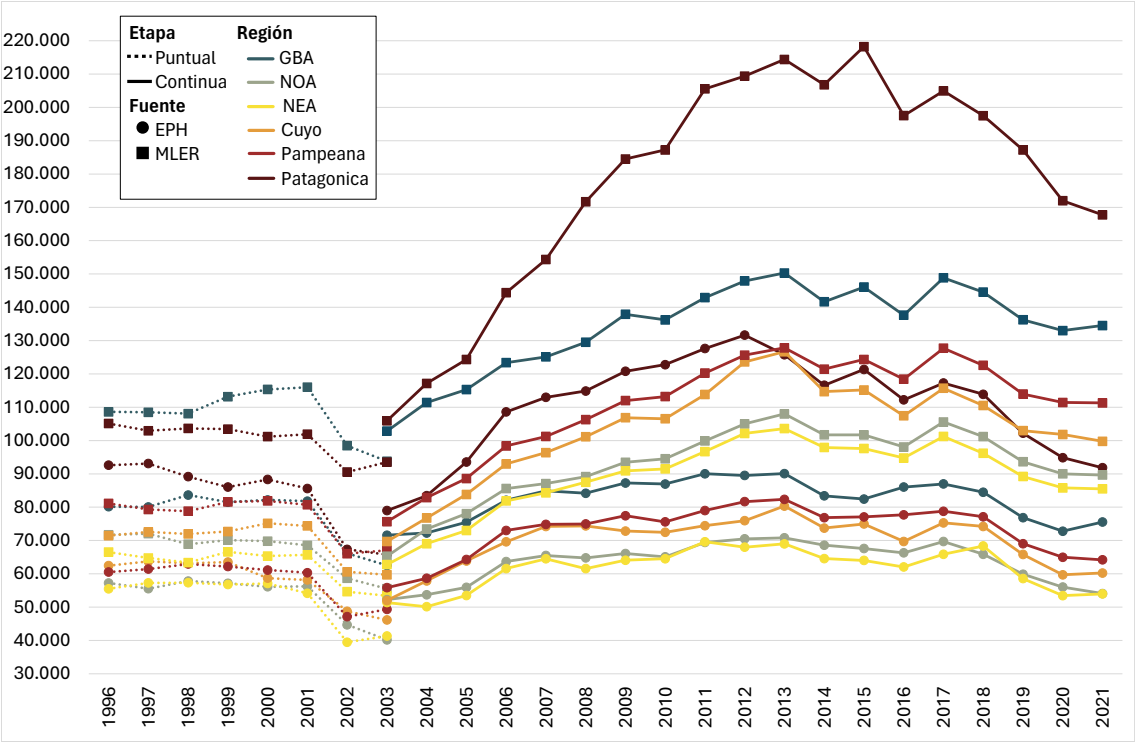
Al analizar las brechas presentes por categoría, nuevamente el patrón presente es el que corresponde mayor nivel de ingresos con menor ratio Br_t. La categoría de establecimientos grandes (+200) aparece con peor captación que el nivel general a lo largo de todo el período[24], mientras que el resto de las categorías aparecen siempre con mejor captación que la del nivel agregado. A su vez en esta categoría de mayor tamaño de establecimiento y mayor nivel de ingresos, la captación de ingresos medida a través del ratio Br_t resulta en una diferencia promedio respecto al nivel general en ambas etapas es 5 pp. menor. De las que están por encima del nivel general, le sigue la categoría de entre 51 y 200 personas con 8.3 pp. mayor promedio en puntual y 8.7 pp. mayor promedio en continua en continua. La categoría de establecimientos de entre 11 y 50 personas presenta Br_t en promedio 18,4 pp. mayor en la etapa puntual y 13,7 pp. Br_t mayor en la etapa

[24] La excepción del año 2001, en donde todas las categorías presentan mejor captación que el promedio agregado, puede atribuirse a las observaciones presentes en el nivel general que son luego desechados para este caso dada la imposibilidad de reconstruir la información acerca del tamaño del establecimiento, tal como fuera descripto.

continua. Finalmente, la categoría de establecimientos de 10 personas o menos presenta Br_t promedio del orden de 31 pp. mayor respecto al nivel general en etapa puntual y 22,5 pp. mayor en la etapa continua. **Nuevamente, a categorías de mayor nivel de ingresos (mayor tamaño) se corresponde una peor captación de estos en la encuesta**.

En términos de ubicación geográfica, el Gráfico 11 muestra la evolución del nivel de ingresos en las 6 regiones seleccionadas, para cada fuente de información.

Gráfico 11. Ingresos de asalariados registrados sin servicio doméstico de EPH y MLER, por región. Promedios anuales en $ constantes de nov-21.

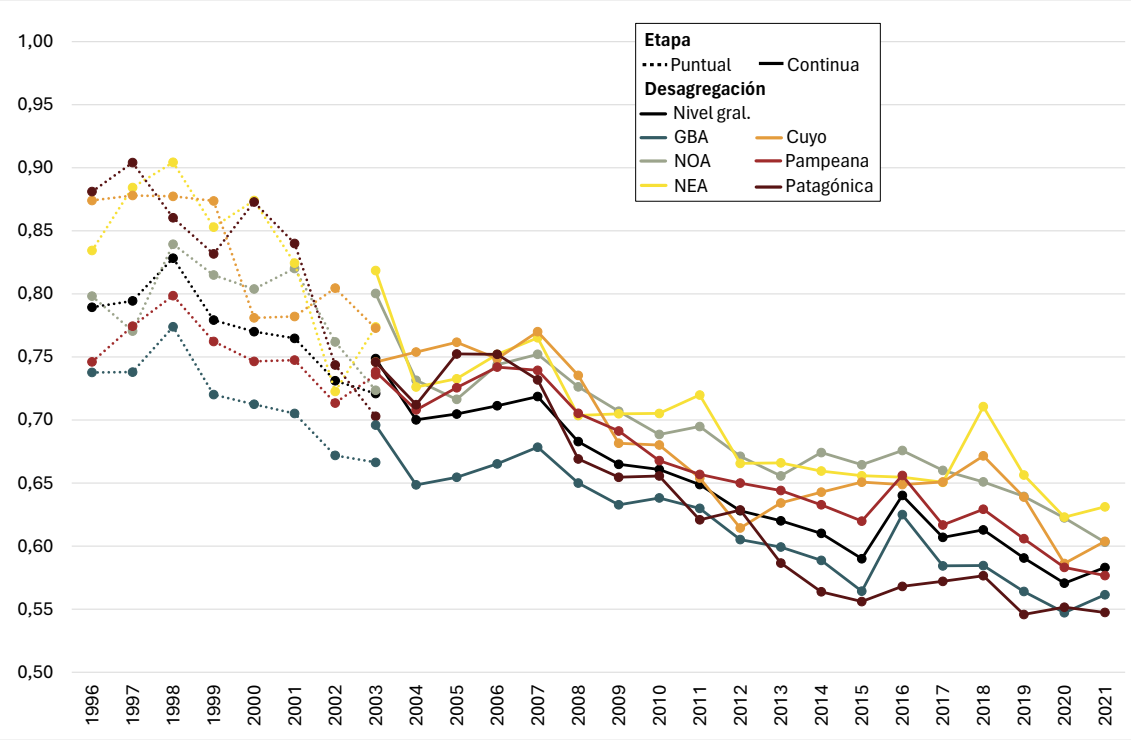


Fuente: elaboración propia en base a MTEySS e INDEC.

Como es sabido, los salarios muestran, en ambas fuentes, un nivel diferencial por regiones. En la etapa puntual, las regiones de GBA y la Patagonia resaltan particularmente como las regiones de mayores ingresos. En un escalafón intermedio, alrededor de un 20% inferior son los ingresos que presenta la región Pampeana, y similares en un último nivel los de las regiones de Cuyo, y el norte argentino. En la etapa continua resaltan dos cuestiones en torno a este ordenamiento. En primer lugar, el notable crecimiento de los ingresos promedio en la región patagónica, superando con holgura al GBA, resultando casi un 25% mayor para el último período disponible. La segunda es el acercamiento en niveles de la región de Cuyo a la Pampeana, conformando un más marcado segundo

escalafón (tercero si Patagonia pasa a ser el primero en soledad y GBA el segundo), diferenciado del último correspondiente al NOA y el NEA.

Gráfico 12. Ratio entre el salario promedio de EPH y MLER para asalariados registrados sin servicio doméstico por región, según fuente y etapa. Comparación respecto a nivel agregado.



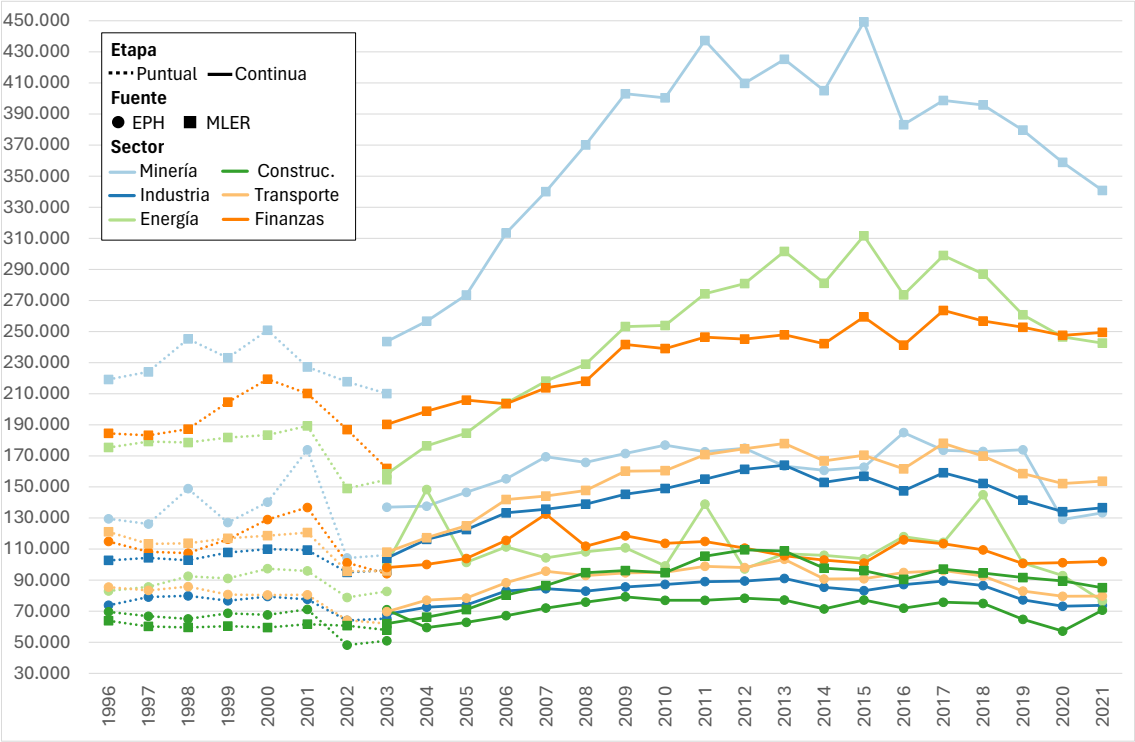
Fuente: elaboración propia en base a MTEySS e INDEC.

El Gráfico 12 refleja cómo, nuevamente, a mayor nivel de ingresos se corresponde un menor coeficiente Br_t -mayor brecha entre fuentes-. Si bien las dos regiones de peor captación son las resaltadas previamente como las de mayor nivel de ingresos, sorprende que la región Patagónica muestre una menor ratio que GBA recién en 2013, cuando su nivel salarial según MLER es mayor desde 2003. Incluso la captación en esta región respecto al nivel general es inicialmente mejor (Br_t promedio entre 1996 y 2001 de 0,87 contra 0,77 de nivel general en el mismo subperíodo), y luego evidenciando un progresivo deterioro: mantiene una relativa paridad entre 2002 y 2007, cambio de etapa mediante, para luego pasar de un Br_t de 0,68 para Patagonia contra un 0,66 del nivel agregado, en el último dato disponible el nivel general se ubica en 0,58 y Patagonia en 0,54. De hecho, resalta que el signo del coeficiente de brecha promedio para esta región respecto de la brecha del nivel general cambia entre etapas: es 5,7 pp. mayor en puntual -mejor captación que el nivel general- y 1,6 pp. menor en la etapa continua -peor captación-. Para GBA, en cambio, las diferencias se mantienen entre etapas: 5,7 pp. menor en promedio

para la etapa puntual y 3 pp. menor para la etapa continua. La región Pampeana también cambia de signo: presenta una diferencia respecto al nivel agregado de 1,9 pp. menor en la etapa puntual y 1,5 pp. mayor en la etapa continua. Cuyo, por su parte, se ubica en 5,6 pp. puntual y 3 pp. continua, superior en ambos casos. NEA en 6,2 pp. y 4,8 pp. Finalmente, NOA en 1,9 pp. y 4,1 pp. **A grandes rasgos, a mayor nivel de ingresos en la información registral para una región en un año determinado se corresponde una peor captación de estos en la encuesta, y viceversa**.

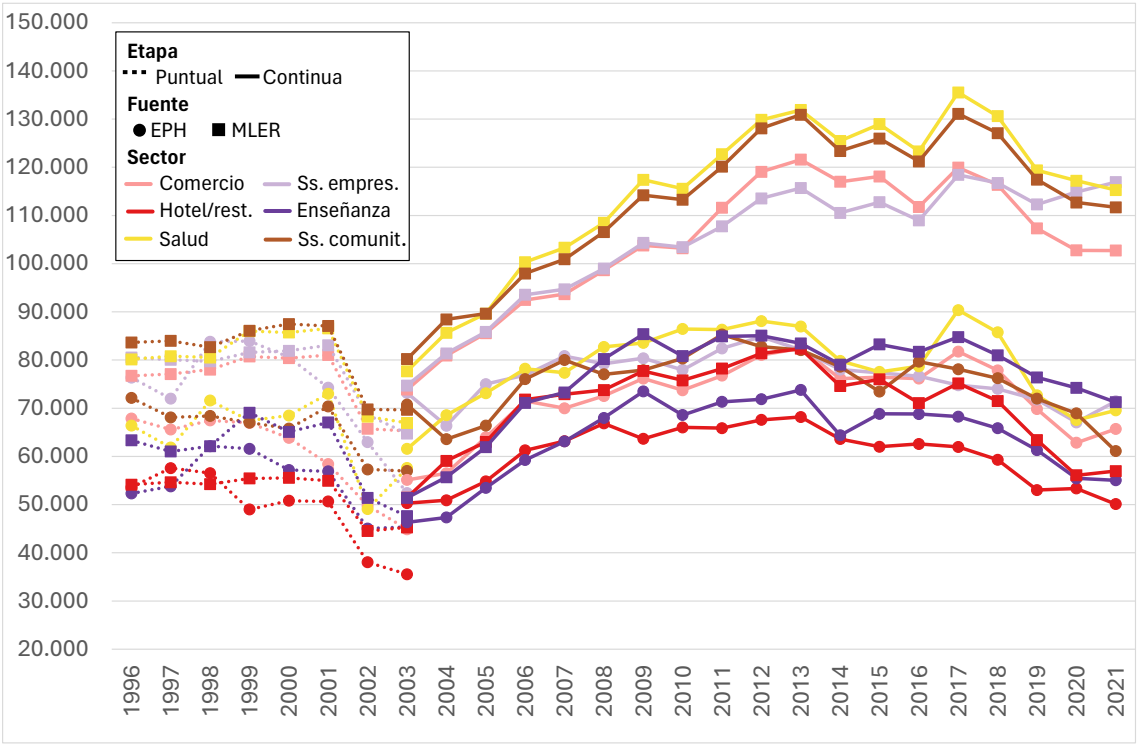
Finalmente, la última dimensión socio ocupacional en términos del puesto laboral a revisar es el sector de la actividad. Dada la gran cantidad de categorías a visualizar, para su presentación en gráficos se dividen en dos partes. El Gráfico 13 contiene los niveles de ingresos para los sectores de Minería, Industria, Energía, Construcción, Transporte y Finanzas. El Gráfico 14, los restantes Comercio, Hotelería y restaurantes, Salud, Enseñanza, Servicios empresariales y Servicios comunitarios.

Gráfico 13. Ingresos de asalariados registrados sin servicio doméstico de EPH y MLER, por sector productivo (parte 1). Promedios anuales en $ constantes de nov-21.



Fuente: elaboración propia en base a MTEySS e INDEC.

Gráfico 14. Ingresos de asalariados registrados sin servicio doméstico de EPH y MLER, por sector productivo (parte 2). Promedios anuales en $ constantes de nov-21.

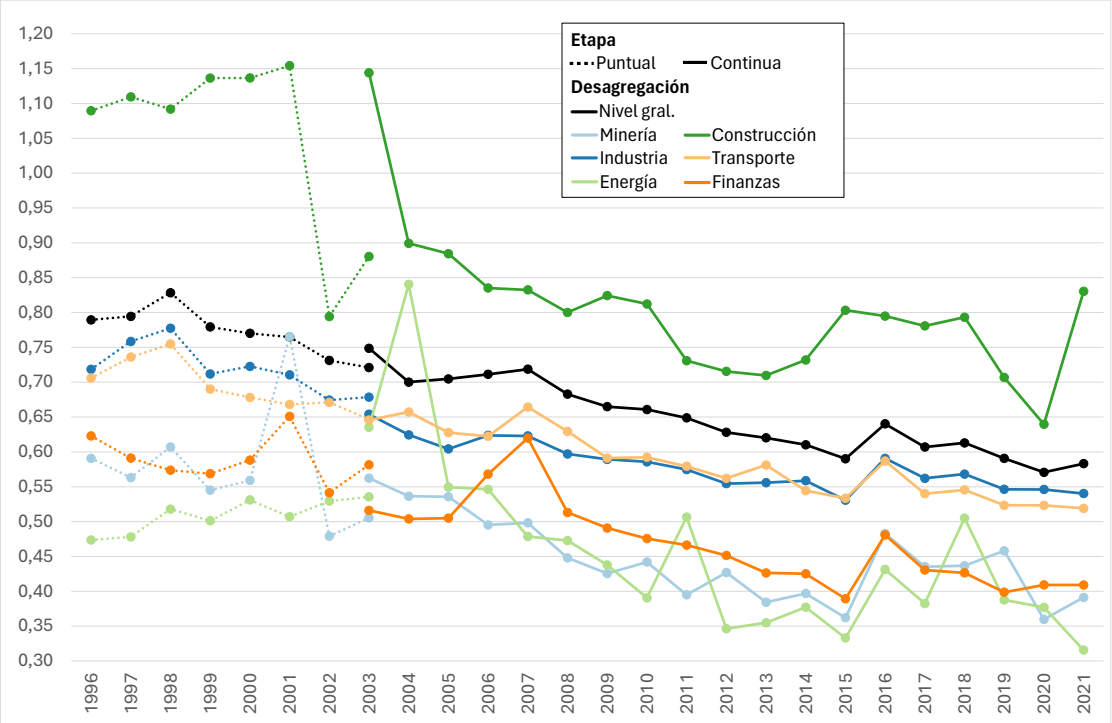


Fuente: elaboración propia en base a MTEySS e INDEC.

A grandes rasgos -con la excepción de Construcción-, los sectores de mayores ingresos se ubican en el Gráfico 13, y los de menores en el Gráfico 14. En el primero, resaltan Minería, Energía y Finanzas como los sectores con un ingreso superior a la media -particularmente, el primero durante la etapa continua-. En segundo orden aparecen Transporte e Industria. Luego, los superiores del Gráfico 14: Comercio, Salud, Servicios comunitarios y Servicios empresariales. Enseñanza y Hotelería/restaurantes, por su parte, componen un grupo de más bajos ingresos en ambas fuentes. Construcción no termina ubicándose ni en el tercer ni el cuarto escalafón en tanto se asimila a los niveles de ingreso de este último en la etapa puntual, pero escala hacia niveles más cercanos a los del tercer escalafón en la etapa continua.

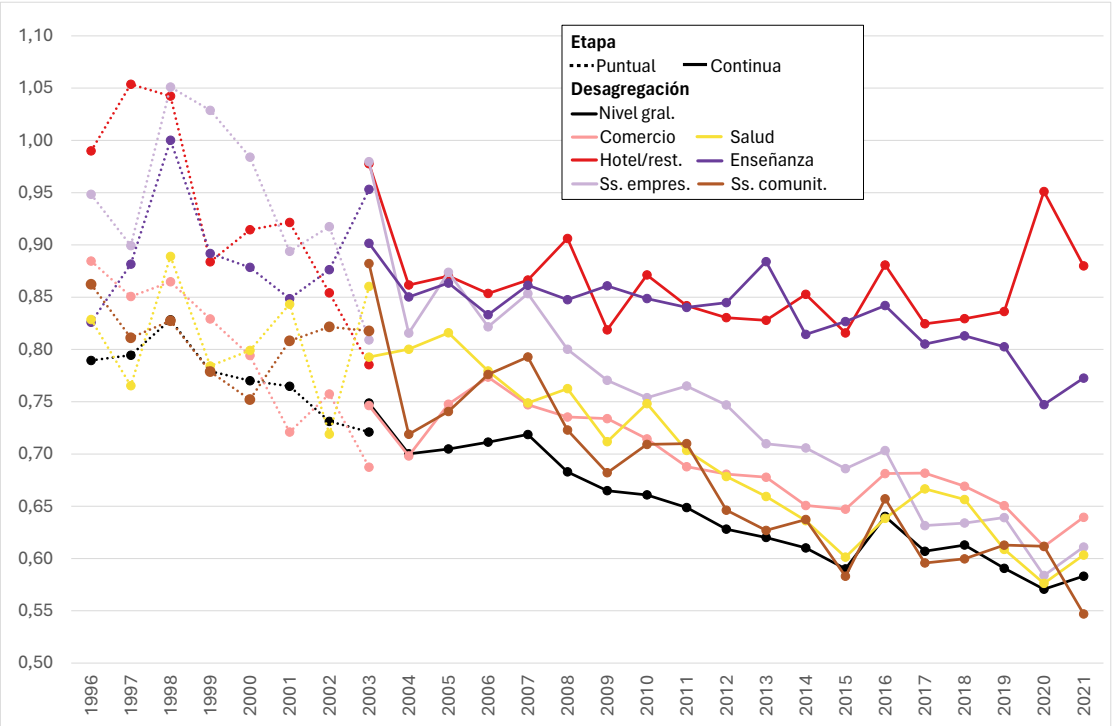
A tono con la división empleada para una mejor visualización, los Gráficos 15 y 16 presentan las brechas correspondientes a estos sectores, y sus correspondientes brechas entre MLER y EPH.

Gráfico 15. Ratio entre el salario promedio de EPH y MLER para asalariados registrados sin servicio doméstico por sector, según fuente y etapa (parte 1). Comparación respecto a nivel agregado.



Fuente: elaboración propia en base a MTEySS e INDEC.

Gráfico 16. Ratio entre el salario promedio de EPH y MLER para asalariados registrados sin servicio doméstico por sector, según fuente y etapa (parte 2). Comparación respecto a nivel agregado.



Fuente: elaboración propia en base a MTEySS e INDEC.

Minería, Energía y Finanzas, sectores de mayor nivel de salario, registran para ambos períodos un ratio Br_t notablemente menor que el promedio agregado. Minería, por su parte, en promedio 19 pp. menor, estable entre ambas etapas. Energía 26,3 pp. menor promedio en puntual y 19 pp. menor promedio en continua. Finanzas 18,3 pp. menor en puntual y 17,8 pp. en continua.

El segundo grupo clasificado, donde se ubican Industria y Transporte, muestra de forma estable -sin la volatilidad del primer grupo- variaciones en Br_t similares a los evidenciados por el nivel general en ambas etapas, aunque siempre en un nivel algo inferior -de mayor brecha-: en promedio, Industria 5,3 pp. inferior al nivel general en puntual y 6,6 pp. inferior en continua; Transporte 7,8 pp. inferior en la etapa puntual y 6,4 pp. inferior en la etapa continua.

El tercer grupo en términos de nivel de ingresos (Comercio, Salud, Servicios Comunitarios y Servicios Empresariales) aparece durante todo el período con niveles de brecha menores a los del nivel general -mayor Br_t-, a excepción de algunos pocos casos en años de etapa puntual (Salud en 1997 y 2002, Servicios Comunitarios en el 2000, y Comercio en 2001 y 2003). Resalta también la elevada sobre volatilidad y la menor brecha durante la etapa puntual, -particularmente en los sectores de servicios, en muchos años resultando $Br_t > 1$-, o sea, sobre captación en EPH respecto de MLER, mayores ingresos nominales (sin considerar niveles teóricos devenidos de los aportes patronales). En promedio, Comercio por encima 2,6 pp. en la etapa puntual y 4,6 pp. en la etapa continua; Salud por encima 3,8 pp. en la etapa puntual y 4,8 pp. en la etapa continua; Ss. Empresariales por encima 16,9 pp. en la etapa puntual y 9,4 pp. en la etapa continua (entre puntas es notable la reducción: de 16,9 pp. por encima del nivel general en 2005 a 1,5 pp. en 2021); y Ss. Comunitarios por encima 3,7 pp. en la etapa puntual y 2,9 pp. en la etapa continua

Finalmente, sectores de nivel de ingresos típicamente bajos para los asalariados registrados privados como Enseñanza, Hotelería/restaurantes y también Construcción aparecen con la menor brecha relativa hacia la actualidad, al mismo tiempo que la misma muestra una fuerte reducción desde los niveles de sobre captación ($Br_t > 1$) que acostumbraban a mostrar en la etapa continua. Enseñanza promedia 12,2 pp. por encima del nivel general en la etapa puntual (Br_t promedio en 0,89 en esta etapa) y 18,7 pp. por encima en la continua (Br_t promedio en 0,83); Hotelería y Restaurantes 15,8 pp.

promedio de menor brecha en la etapa puntual (Br_t promedio en 0,93) y 21,6 pp. en la continua (Br_t promedio en 0,86); y finalmente Construcción 27,7 pp. por encima promedio en la etapa puntual (Br_t promedio en 1,04) y 15,6 pp. promedio por encima en la continua (Br_t promedio en 0,80).

En esta dimensión es particularmente notable la tendencia general en donde a mayor ingreso en niveles, peor captación -menor ratio Br_t-. Se reflejan con fidelidad los distintos escalafones de ingresos según sector como determinantes del nivel y la evolución de la captación de ingresos en la encuesta.

4.4. La brecha depende del nivel de ingresos, pero existen matices

A modo de resumen de lo revisado dimensión por dimensión, un sucinto resumen puede obtenerse reduciendo la dimensión temporal, a partir de comparaciones entre períodos extremos[25] dentro de cada etapa. Esto es lo que se expone en los Cuadros 2 y 3.

Cuadro 2. Niveles de ingreso y brechas entre puntas de la etapa puntual (1996-2002).

			Promedios						
			1996			2002			
Tipo	Dimensión	Edad/género	EPH	MLER	Br	EPH	MLER	Br	ΔBr
Nivel general	Nivel general	Nivel general	74.329	94.150	0,7895	59.744	81.709	0,7312	-0,0583
Atributos personales	Género	Varón	79.508	101.320	0,7847	64.980	89.354	0,7272	-0,0575
		Mujer	62.421	75.683	0,8248	50.379	65.931	0,7641	-0,0607
	Edad	-25	53.644	58.554	0,9162	40.524	45.130	0,8979	-0,0182
		25-45	79.443	97.662	0,8134	62.678	82.741	0,7575	-0,0559
		+45	81.586	114.197	0,7144	65.852	99.840	0,6596	-0,0549
Atributos del puesto de trabajo	Tamaño del establecimiento	-10	59.635	54.653	1,0911	47.767	44.726	1,0680	-0,0231
		11-50	71.344	70.242	1,0157	53.372	60.976	0,8753	-0,1404
		51-200	80.399	92.709	0,8672	63.738	81.778	0,7794	-0,0878
		+200	99.204	137.339	0,7223	84.266	118.399	0,7117	-0,0106
	Región	GBA	80.131	108.616	0,7377	66.153	98.449	0,6720	-0,0658
		Pampeana	60.525	81.123	0,7461	47.120	66.038	0,7135	-0,0326
		NOA	57.240	71.708	0,7982	44.691	58.652	0,7620	-0,0363
		NEA	55.508	66.521	0,8344	39.476	54.629	0,7226	-0,1118
		Cuyo	62.452	71.448	0,8741	48.758	60.600	0,8046	-0,0695

[25] Se excluye de esta comparación el año 2003, al tratarse del año de cambio de versión en la EPH y no contar con la información completa para ese período en ninguna de las dos versiones.

		Patagonia	92.608	105.102	0,8811	67.350	90.558	0,7437	-0,1374
	Sector productivo (letra)	Minería	129.477	219.170	0,5908	104.267	217.707	0,4789	-0,1118
		Industria	73.834	102.767	0,7185	64.095	95.049	0,6743	-0,0441
		Energía	83.065	175.399	0,4736	78.881	149.008	0,5294	0,0558
		Construcción	69.627	63.908	1,0895	48.221	60.712	0,7942	-0,2952
		Comercio	67.868	76.723	0,8846	49.785	65.727	0,7574	-0,1271
		Hotel/rest	53.595	54.140	0,9899	38.028	44.522	0,8541	-0,1358
		Transporte	85.498	121.107	0,7060	64.231	95.717	0,6711	-0,0349
		Finanzas	114.913	184.494	0,6229	101.181	186.885	0,5414	-0,0814
		Ss. Empres.	76.367	80.524	0,9484	62.956	68.620	0,9175	-0,0309
		Enseñanza	52.316	63.342	0,8259	45.043	51.401	0,8763	0,0504
		Salud	66.415	80.146	0,8287	49.073	68.253	0,7190	-0,1097
		Ss. Comunit.	72.134	83.655	0,8623	57.309	69.750	0,8216	-0,0407

Fuente: elaboración propia en base a MTEySS e INDEC.

Cuadro 3. Niveles de ingreso y brechas entre puntas de la etapa continua (2004-2021).

			Promedios						
			2004			2021			
Tipo	Dimensión	Edad/género	EPH	MLER	Br	EPH	MLER	Br	ΔBr
Nivel general	Nivel general	Nivel general	67.778	96.803	0,7002	71.224	122.142	0,5831	-0,1170
Atributos personales	Género	Varón	72.792	105.194	0,6920	76.081	132.431	0,5745	-0,1175
		Mujer	57.774	78.399	0,7369	62.026	101.831	0,6091	-0,1278
	Edad	-25	47.286	56.650	0,8347	50.820	68.237	0,7447	-0,0900
		25-45	68.433	96.969	0,7057	71.104	117.024	0,6076	-0,0981
		+45	78.657	121.746	0,6461	78.051	149.474	0,5222	-0,1239
Atributos del puesto de trabajo	Tamaño del establecimiento	-10	58.210	59.672	0,9755	56.561	72.505	0,7801	-0,1954
		11-50	62.107	75.475	0,8229	64.579	93.322	0,6920	-0,1309
		51-200	73.816	93.511	0,7894	74.747	111.297	0,6716	-0,1178
		+200	90.396	137.119	0,6593	92.362	167.591	0,5511	-0,1081
	Región	GBA	72.257	111.410	0,6486	75.533	134.527	0,5615	-0,0871
		Pampeana	58.663	82.857	0,7080	64.190	111.300	0,5767	-0,1313
		NOA	53.720	73.436	0,7315	54.076	89.659	0,6031	-0,1284
		NEA	50.133	69.036	0,7262	53.974	85.510	0,6312	-0,0950
		Cuyo	57.865	76.756	0,7539	60.236	99.762	0,6038	-0,1501
		Patagonia	83.442	117.153	0,7122	91.835	167.764	0,5474	-0,1648
	Sector productivo (letra)	Minería	137.643	256.619	0,5364	133.354	340.832	0,3913	-0,1451
		Industria	68.123	104.154	0,6541	73.796	136.604	0,5402	-0,1138
		Energía	148.301	176.451	0,8405	76.552	242.579	0,3156	-0,5249
		Construcción	59.504	66.169	0,8993	70.702	85.145	0,8304	-0,0689

		Comercio	56.501	80.929	0,6982	65.683	102.718	0,6394	-0,0587
		Hotel/rest	50.901	59.076	0,8616	50.114	56.950	0,8800	0,0184
		Transporte	77.131	117.349	0,6573	79.789	153.718	0,5191	-0,1382
		Finanzas	100.120	198.715	0,5038	102.029	249.482	0,4090	-0,0949
		Ss. Empres.	66.372	81.372	0,8157	71.441	116.916	0,6110	-0,2046
		Enseñanza	47.336	55.674	0,8502	55.047	71.249	0,7726	-0,0776
		Salud	68.549	85.660	0,8002	69.551	115.274	0,6034	-0,1969
		Ss. Comunit.	63.566	88.426	0,7189	61.095	111.693	0,5470	-0,1719

Fuente: elaboración propia en base a MTEySS e INDEC.

Surgen aquí varias cuestiones a resaltar. En primer lugar, **para cada característica se cumple el hecho estilizado señalado en la literatura, que vincula mayor nivel de ingresos con peor captación en la encuesta respecto a los registros administrativos**. Tanto para atributos personales como del puesto de trabajo, el grupo con mayor nivel de ingresos es también aquel que presenta menor ratio Br_t. En términos de género, los varones promedian mayor ingreso que las mujeres, y la encuesta lo capta menos respecto al registro. En edad, los mayores de 45 hacen lo propio respecto a las dos categorías de menor edad. Es así también en tamaño del establecimiento, donde los asalariados de establecimientos de más de 200 personas presentan los mayores niveles de ingresos y también la mayor brecha respecto a las otras categorías. En regiones, si bien quizás la dimensión menos dispersa, se repite también el patrón: CABA y Patagonia son, en forma rotativa, las regiones con mayor nivel de ingresos y peor captación de estos en las encuestas. Finalmente, en sectores productivos la tendencia se ve claramente confirmada: los sectores de mayores ingresos como Minería, Finanzas y Energía presentan para todo el período analizado los peores niveles de captación de ingresos en EPH.

Esta primera observación no debe pasar por alto **que todos los atributos revisados -tanto personales como del puesto de trabajo- resultan relevantes para determinar distinto nivel de captación de los ingresos**. Aunque inicialmente de a una dimensión a la vez y sin considerar la relevancia de múltiples dimensiones en simultáneo, sobre lo que se profundizará en el Capítulo 5, **en cada característica revisada existe una notable dispersión en los niveles de captación del ingreso**. En indeleble vinculación con lo anterior -dispersión guiada por el nivel de ingresos-, pero también dando indicios de **matices en distintas características** más allá de su nivel. El ejemplo más nítido se ve en la característica más homogénea en estos términos, la región: tras el fuerte crecimiento de ingresos en la región patagónica desde inicios de la etapa continua, para el año 2012 la misma refleja un ratio Br_t=0,63. Para el mismo año, GBA marca un Br_t=0,61, 2 pp.

inferior al de Patagonia -peor captación-, con un nivel de ingreso 30% inferior en MLER, y 32% inferior en EPH. Al siguiente año, los ingresos promedio para Patagonia según MLER ascienden un 2,3%, mientras que para la encesta descienden un 4,3%; consecuentemente empeorando su captación de ingresos hasta un ratio Br_t=0,59. Como contracara, en GBA ascienden 1,6% en MLER y 0,6% en EPH, mostrando ahora una peor captación relativa a sí mismo, pero mejor respecto a la que presenta Patagonia en el mismo período (Br_t=0,60).

En resumidas cuentas, **el nivel de ingresos parece ser el principal determinante del nivel de captación de la encuesta, pero no el único: existen fricciones particulares en esta captación, dadas tanto por las características del individuo como del puesto de trabajo**.

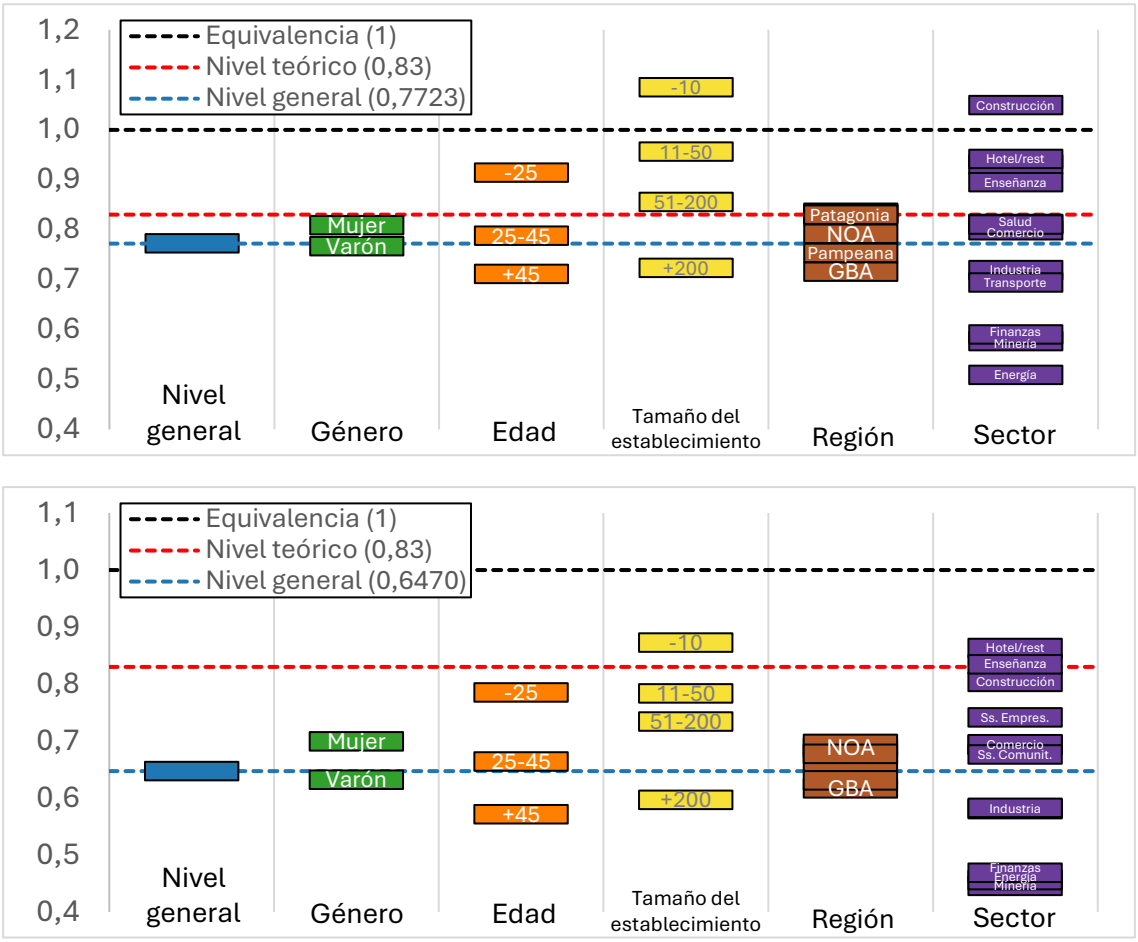
4.5. Hallazgos desprendidos del análisis individual de las dimensiones

Más allá de la concordancia del análisis con los hechos estilizados revisados en la literatura, de los Cuadros 2 y 3 se desprenden también nuevas lecturas que vale la pena resaltar. La etapa puntual, caracterizada por un buen nivel de captación de la encuesta respecto a la información que registran las fuentes administrativas, no posee un comportamiento homogéneo a través de las distintas dimensiones revisadas. Esto es, **la etapa puntual no encuentra a todas las dimensiones alrededor del Br_t=0,83 teórico[26], sino que encuentra algunas por encima y otras por debajo. De modo que, en el promedio agregado, para la etapa puntual el ingreso captado por la encuesta aparenta ubicarse en niveles acordes a los teóricos desprendidos de los registros administrativos; pero la desagregación revela la heterogeneidad de captaciones detrás de este promedio agregado, donde se contienen categorías subcaptadas compensadas por otras sobrecaptadas** -donde el nivel de ingreso en EPH es mayor al de MLER-. Esto resulta contrario a la lectura habitual revisada para inicios del nuevo milenio (Alaimo y Sosa Escudero, 2000; Roca y Pena, 2001), en donde se caracteriza a la subcaptación de ingresos como un fenómeno propio de perceptores de ingresos de fuente como remuneraciones al capital, rentas y trabajo cuentapropista; y no al trabajo

[26] Si bien, como fuera revisado, el nivel teórico no se ubica exactamente en 0,83 dado los cambios en las alícuotas de aportes personales entre 2000 y 2003 para la etapa puntual y entre 2003 y 2007 para la etapa la continua; al ser el valor de referencia en la mayoría de los años en ambas etapas, por simplicidad se toma este nivel de forma genérica.

asalariado. Así se visualiza en el Gráfico 17 se presentan las brechas promedio para cada categoría devenida de cada dimensión comparada entre fuentes.

Gráfico 17. Brecha de ingresos promedio para cada categoría dentro de cada dimensión. Etapa puntual (1996-2003; panel superior) y etapa continua (2003-2021; panel inferior).



Fuente: elaboración propia en base a MTEySS e INDEC.

Particularmente para la etapa puntual, como visualiza el panel superior, para todas las dimensiones revisadas se respeta el patrón en donde la peor captación se da en la categoría de mayor nivel de ingresos. Como anverso, la mejor captación (en realidad, sobre captación, no necesariamente "mejor", sí de mayor ratio Br_t) se da en la categoría de menor nivel de ingresos, en casi todos los casos por encima del nivel teórico. Incluso, en algunas características existe más de una categoría que registra sobrecaptación en EPH. Como resultante, **el promedio agregado de los ingresos de EPH se ubica en niveles similares a los que indican teóricamente los aportes personales, aparentando una buena captación que esconde esta heterogeneidad**. Cabe resaltar que sí varía la dispersión entre captaciones dentro de cada categoría, destacando la dimensión de región

como la más homogénea en estos términos. En la etapa continua, en líneas generales esta dispersión se mantiene, aunque con un peor nivel de captación tanto en el nivel general como en todas las dimensiones revisadas.

Finalmente, cabe una revisión del cambio en los niveles de brecha entre etapas en cada dimensión. Más allá de las diferencias metodológicas entre etapas que no hacen que sea cualitativamente comparables, este acercamiento puede brindar nociones acerca de en donde se dieron los más marcados empeoramientos relativos de captación entre una etapa y otra.

Cuadro 4. Brechas promedio por etapas y diferencias respecto al nivel general.

			Promedios		Diferencias	
Tipo	**Dimensión**	**Edad/género**	**Puntual**	**Continua**	**entre etapas**	**s/nivel gral**
Nivel general	**Nivel general**	**Nivel general**	0,7797	0,6414	-0,1383	-
Atributos personales	**Género**	**Varón**	0,7739	0,6273	-0,1466	**-0,0084**
		Mujer	0,8163	0,6916	-0,1247	0,0136
	Edad	**-25**	0,9156	0,7781	-0,1375	0,0007
		25-45	0,7962	0,6584	-0,1378	0,0004
		+45	0,7147	0,5666	-0,1481	**-0,0099**
Atributos del puesto de trabajo	**Tamaño del establecimiento**	**-10**	1,0964	0,8654	-0,2310	**-0,0927**
		11-50	0,9638	0,7769	-0,1869	-0,0486
		51-200	0,8579	0,7293	-0,1286	0,0096
		+200	0,7333	0,5899	-0,1434	-0,0052
	Región	**GBA**	0,7228	0,6124	-0,1104	0,0278
		Pampeana	0,8013	0,6821	-0,1192	0,0190
		NOA	0,8425	0,6880	-0,1545	-0,0162
		NEA	0,8387	0,6738	-0,1649	-0,0267
		Cuyo	0,7556	0,6584	-0,0972	0,0411
		Patagonia	0,8477	0,6248	-0,2229	**-0,0847**
	Sector productivo (letra)	**Minería**	0,5870	0,4394	-0,1476	-0,0094
		Industria	0,7248	0,5764	-0,1484	-0,0102
		Energía	0,5054	0,4462	-0,0592	0,0790
		Construcción	1,0732	0,7847	-0,2884	**-0,1502**
		Comercio	0,8145	0,6905	-0,1241	0,0142
		Hotel/rest	0,9514	0,8566	-0,0948	0,0434
		Transporte	0,7006	0,5791	-0,1215	0,0167
		Finanzas	0,5910	0,4661	-0,1249	0,0133
		Ss. Empres.	0,9604	0,7281	-0,2323	-0,0941
		Enseñanza	0,8861	0,8310	-0,0551	0,0832
		Salud	0,8041	0,6886	-0,1154	0,0228

		Ss. Comunit.	0,8088	0,6650	-0,1438	-0,0055

Fuente: elaboración propia en base a MTEySS e INDEC.

En el Cuadro 4 se aprecia que el empeoramiento relativo entre etapas se da de distinta manera entre tipos de atributos. Para atributos personales, la captación en la etapa continua empeora más en las categorías de típicamente mayor nivel de ingresos respecto a la etapa puntual. En términos de región, el peor empeoramiento relativo entre etapas se da en la región Patagonia, probablemente explicado por el notable cambio en el nivel de ingresos en esta región, ya revisado. Obviando este caso, el más marcado empeoramiento relativo de la captación sí se da en las regiones de menor nivel de ingresos, las del norte argentino. Para tamaño del establecimiento o sector productivo, restantes atributos del puesto de trabajo, se da a la inversa respecto a los atributos personales: el mayor empeoramiento relativo de la captación aparece en las categorías de menor nivel de ingresos: establecimientos de menos de 10 personas y construcción, respectivamente. De modo que, considerando estas variaciones entre los promedios de cada etapa, **los atributos personales parecen tener un desempeño distinto al de los atributos del puesto de trabajo en el marco del deterioro general de la captación de la encuesta. Mientras que en los primeros se profundiza el impacto diferencial de la captación de ingresos según su nivel, en los segundos se revieren: empeora relativamente más la captación de los ingresos de bajo nivel**. Esta observación resulta interesante en el marco de encontrar los matices que les dan las características socioocupacionales a la captación de ingresos más allá de su nivel, y también como insumo para afinar una explicación acerca del revisado deterioro en la captación del ingreso de EPH.

5. EL ANÁLISIS DE LAS DIFERENCIAS ENTRE EPH Y MLER A PARTIR DE UN MODELO PARAMÉTRICO

El análisis de a una característica a la vez puede perder de vista la existencia de distintas composiciones en las variables que no están siendo consideradas para las unidades observadas. De este modo, por ejemplo, lo referido al sector construcción puede estar -en mayor o menor grado- compensado a la masculinización de los trabajadores en ese sector. Podría suceder también que parte de las diferencias en el nivel de captación entre regiones responda a la distinta composición de edades entre las mismas. A fin de cuentas, de a una característica a la vez no se realiza de forma estricta una comparación entre comparables, sino que entre distintas composiciones se combinan los efectos del resto de las características no consideradas, todas relevantes -como fuera resumido en la última sección del capítulo anterior- para la captación de ingresos. Es por esto que cobra sentido intentar un análisis multidimensional, contemplando más de una característica a la vez para evaluar con mayor precisión la influencia de cada dimensión en la captación de ingresos de la encuesta. Al contener cada unidad observada -en ambas fuentes- más de una dimensión a la vez, la combinación de estas permite construir grupos socio ocupacionales compuestos por individuos razonablemente parecidos tanto en términos de atributos individuales como del puesto laboral.

Así, en una primera sección del capítulo se detalla la creación de grupos socioocupacionales a partir de la disponibilidad de información en el emparejamiento entre fuentes realizado. En una segunda, se revisa la vinculación entre el nivel de ingresos en cada fuente y la correspondiente brecha que se deriva entre estos. En la tercera y última sección se expone la configuración de distintas configuraciones paramétricas para modelizar la subcaptación, y se exponen los resultados obtenidos.

5.1. La unidad en grupos socioocupacionales

En este punto, considerar las 5 dimensiones exploradas en el capítulo anterior (género, edad, tamaño del establecimiento, región y sector productivo) enfrenta un problema metodológico: el insuficiente tamaño de la muestra. Sucede que la desagregación del sector productivo imposibilita la captación de casos para cada posible combinación. En ocasiones a causa de su combinación con la variable de ubicación

(regiones que no poseen empleo en cierto sector productivo), pero también en términos de género (sectores masculinizados y feminizados). La escasez relativa de casos se da principalmente en la encuesta, pero también sucede en los registros administrativos. Es por esto por lo que se opta por descartar sector productivo a la hora de construir los grupos socio ocupacionales.

De esta manera, es posible emparejar entre fuentes: 2 categorías por género, 3 categorías por edad, 4 categorías por tamaño de establecimiento empleador y 6 por región. Se obtienen así 144 distintos grupos para cada período temporal. De existir datos en cada combinación, quedan construidas bases de datos de 1.152 observaciones emparejadas entre EPH y MLER para la etapa puntual (8 años) y 2.736 para la continua (19 años). Un resumen de las combinaciones posibles obtenidas para cada período temporal puede observarse en el Cuadro 5.

Cuadro 5. Combinaciones de dimensiones emparejables posibles.

	Región	GBA		Pampeana		NOA		NEA		Cuyo		Patagonia	
Tam. Establ.	Edad/género	Varón	Mujer	Varón	Mujer	Varón	Mujer	Varón	Mujer	Varón	Mujer	Varón	Mujer
Hasta 10	-25	1	2	3	4	5	6	7	8	9	10	11	12
	25-45	13	14	15	16	17	18	19	20	21	22	23	24
	+45	25	26	27	28	29	30	31	32	33	34	35	36
Entre 11 y 50	-25	37	38	39	40	41	42	43	44	45	46	47	48
	25-45	49	50	51	52	53	54	55	56	57	58	59	60
	+45	61	62	63	64	65	66	67	68	69	70	71	72
Entre 51 y 200	-25	73	74	75	76	77	78	79	80	81	82	83	84
	25-45	85	86	87	88	89	90	91	92	93	94	95	96
	+45	97	98	99	100	101	102	103	104	105	106	107	108
Más de 200	-25	109	110	111	112	113	114	115	116	117	118	119	120
	25-45	121	122	123	124	125	126	127	128	129	130	131	132
	+45	133	134	135	136	137	138	139	140	141	142	143	144

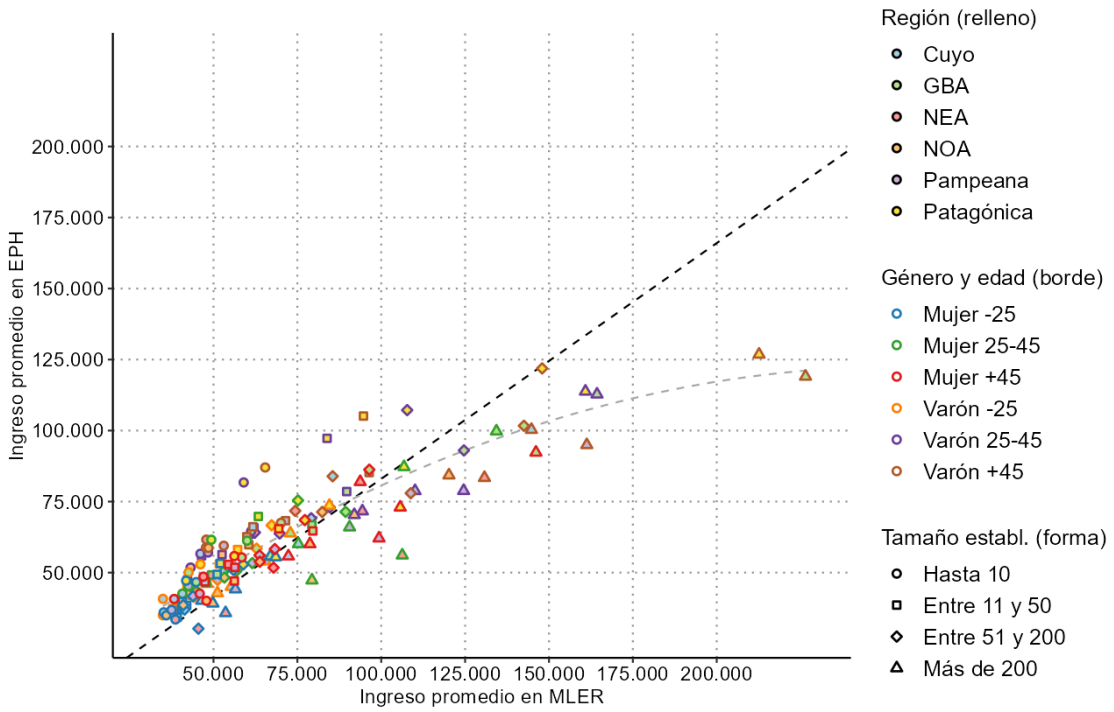
Fuente: elaboración propia en base a MTEySS e INDEC.

5.2. Un vistazo dentro de los grupos socioocupacionales

A 144 combinaciones posibles resulta problemático repetir el análisis realizado en la sección anterior, donde se revisaba para cada categoría disponible por dimensión el ingreso promedio del año en cada fuente y en cada etapa. Es por esto por lo que se opta por repetir la metodología elegida para reflejar la información en los Cuadros 2, 3 y 4, a modo de cierre del anterior capítulo, y tomar promedios de los valores anuales separados por etapa.

Adicionalmente, se combinan las dimensiones de edad y género en una sola de respectivas 6 categorías (Mujeres y varones menores de 25 años, de entre 25 y 45 años y mayores de 45 años). De este modo, es posible visualizar la ubicación relativa de cada uno de los 144 grupos disponibles en lo que respecta a nivel de ingresos en ambas fuentes en un gráfico de dispersión. Esto es, reflejando las (ahora) 3 dimensiones reducidas en tres distintos aspectos de cada punto de la dispersión: el color de relleno como la región, el color del borde como la dimensión que reduce género y edad, y la forma del punto para diferenciar asalariados de establecimientos de diferentes tamaños.

Gráfico 18. Promedio de ingresos anuales para la etapa puntual (1996-2003) por grupo y fuente, con curva de regresión cuadrática (línea punteada gris) y nivel teórico de referencia (línea punteada negra). En pesos de nov-21.

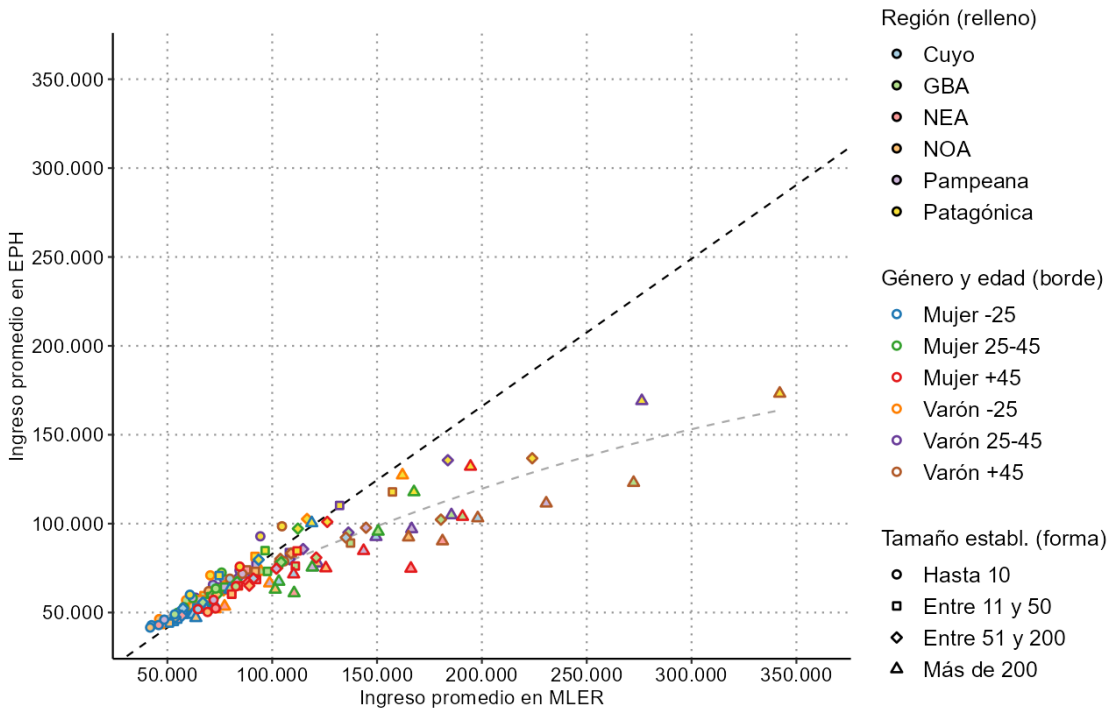


Fuente: elaboración propia en base a MTEySS e INDEC.

El Gráfico 18 destaca -del mismo modo que el capítulo anterior, pero ahora al combinar cuatro características a la vez- que a bajos ingresos se dan emparejamientos por encima del nivel teórico determinado en el ratio teórico promedio de 0,83. A medida que se avanza en nivel salarial según MLER, la distancia entre fuentes se va ensanchando notablemente. Dicho de otro modo, a mayor nivel salarial en la fuente de registro empeora la captación de la encuesta o disminuye el coeficiente Br_t que contempla ambos promedios. De la nube de puntos cercana al origen, por ejemplo, aparecen puntos como mujeres menores de 25 años de establecimientos de menos de 10 personas de la región

pampeana, donde el ingreso promedio para esta etapa marca un Br_t=1,02; por encima del nivel teórico y también por encima de la equivalencia entre fuentes. En el otro extremo, hombres mayores de 45 años de establecimientos de más de 200 personas de la región GBA registran un Br_t=0,53. O sea, una captación de ingresos sustancialmente menor.

Gráfico 19. Promedio de ingresos anuales para la etapa continua (2003-2021) por grupo y fuente, con curva de regresión cuadrática (línea punteada gris) y nivel teórico de referencia (línea punteada negra). En pesos de nov-21.



Fuente: elaboración propia en base a MTEySS e INDEC.

Para la etapa continua, como muestra el Gráfico 19, la relación observada previamente se agudiza aún más. La dispersión de los puntos alrededor de la curva teórica se reduce, y prácticamente desaparecen los casos de grupos por encima de la equivalencia entre fuentes. Al tomar los mismos grupos ejemplo que para la etapa continua, se revela que las mujeres menores de 25 años de la región pampeana de establecimientos de hasta 10 personas registran un coeficiente Br_t= 0,95. Del otro lado, hombres del GBA mayores de 45 años de establecimientos de más de 200 personas muestran un Br_t=0,45.

De este modo, en esta exploración en múltiples dimensiones se refuerzan dos cuestiones ya arribadas en el análisis unidimensional del capítulo anterior. En primer lugar, el nivel de ingresos aparece como el como principal determinante de la captación, donde a mayor nivel de ingresos en las fuentes registrales menor porción de esos ingresos

se captan en la encuesta. No obstante, tanto las dimensiones personales como del puesto de trabajo aplican matices a este nivel de captación. En segundo lugar, la relevancia del notorio y generalizado empeoramiento relativo de la captación en términos temporales o entre etapas de la encuesta, aún con la consideración de estas cuatro dimensiones emparejables y analizando grupos con mayor nivel de comparabilidad entre sí.

5.3. Una aproximación paramétrica para contemplar todas las dimensiones

En este punto, al trabajar con datos donde se cruzan distintas dimensiones en simultáneo, resulta conveniente implementar un modelo de regresión que permita considerar efectos *ceteris paribus* entre variables. Esto es, para obtener -más allá de la vinculación con el ingreso referida- la acción más "pura" de cada característica considerada sobre nuestra variable de interés. De este modo, se plantea un ejercicio de regresión multivariada a través del método de mínimos cuadrados ordinarios (MCO) *pooleado* temporalmente, en donde las unidades observadas son los grupos sociolaborales construidos con las 4 características detalladas en la sección anterior (género, edad, tamaño del establecimiento y región) para cada período temporal.

Se determina como variable dependiente el ingreso promedio de EPH para cada grupo y período temporal, variable que es buscada explicar en función del ingreso en MLER. A su vez, dada la relación no lineal apreciada en la sección anterior, resulta relevante la inclusión de la interacción del ingreso de MLER consigo misma, de modo de poder obtener una parábola en lugar de una recta. Esto permitirá, adicionalmente, la evaluación de los efectos de cada variable independiente en distintos lugares de la distribución del ingreso de MLER, permitiendo incluir la dimensión del nivel de ingresos en el análisis de forma transversal a los grupos creados y no a partir del nivel que muestran las características de estos. Por último, dada la variable independiente y nuestro interés en evaluar los matices que aportan las dimensiones consideradas a un nivel de captación dado en mayor medida por el ingreso de MLER, se incluyen interacciones de cada dimensión considerada con la variable dependiente de los ingresos en MLER.

Se evalúa el modelo en tres etapas distintas, a partir de una inclusión progresiva de parámetros. En una primera instancia, se consideran sólo efectos fijos e interacciones temporales, sin controles de otro tipo, como se muestra en la Ecuación 2:

$$WEPH_{i,t} = \alpha + \beta_1 WMLER_{i,t} + \beta_2 WMLER_{i,t}^2 + \theta_t + \beta_3 (WMLER_{i,t}\theta_t) + \epsilon_{i,t} \quad \textbf{(2)}$$

Donde $WEPH_{i,t}$ es el salario medio de EPH del grupo i en el año t, así como $WMLER_{i,t}$ es el salario medio para MLER. A su vez, $\epsilon_{i,t}$ es un término de error con propiedades estándar y θ_t, por su parte, un efecto fijo para cada período temporal (año). En una segunda etapa, se agregan los controles e interacciones de cada una de las dimensiones contempladas en el armado de los grupos socio ocupacionales, como muestra la Ecuación 3:

$$WEPH_{i,t} = \alpha + \beta_1 WMLER_{i,t} + \beta_2 WMLER_{i,t}^2 + \theta_t + \beta_3 (WMLER_{i,t}\theta_t) + x_i\gamma + \beta_4 (WMLER_{i,t}x_i) + \epsilon_{i,t} \quad \textbf{(3)}$$

Donde x_i es un vector de variables explicativas que contiene las características de cada observación, tanto en términos individuales (género y rango de edad) como las vinculadas al puesto laboral (tamaño del establecimiento y región).

Finalmente, la tercera especificación contempla además una interacción cúbica, en busca de una mejor modelización del efecto del salario de MLER en la captación en EPH. Así lo muestra la Ecuación 4:

$$WEPH_{i,t} = \alpha + \beta_1 WMLER_{i,t} + \beta_2 WMLER_{i,t}^2 + \theta_t + \beta_3 (WMLER_{i,t}\theta_t) + x_i\gamma + \beta_4 (WMLER_{i,t}x_i) + \beta_5 WMLER_{i,t}^3 + \epsilon_{i,t} \quad \textbf{(4)}$$

Adicionalmente, se consideran ponderadores analíticos para tener en cuenta la representatividad de cada grupo en términos cuantitativos, a través de las cantidades ponderadas derivadas de la MLER.

Tal como fuera descripto en la sección metodológica, los modelos se estiman y calculan de manera diferenciada entre etapas de la información puntual (1996-2003) y continua (2003-2021), dadas las diferencias cualitativas entre los datos entre ambos períodos temporales.

En los cuadros 6 y 7 se exponen los resultados de las estimaciones realizadas por especificación para cada etapa, presentando los parámetros estimados para todos los términos vinculados a $WMLER$. Esto es, obviando los controles de cada categórica que funcionan para aislar el efecto de las diferencias en niveles de ingreso -en tanto captan solamente los distintos niveles en $WEPH$-.

Cuadro 6. Resultados de estimaciones por especificación. Etapa puntual.

	(2)	(3)	(4)
WMLER	0.7754***	0.4228***	-0.0833
WMLER^2	-0,000001***	-0,000002***	-0,000003*

	WMLER^3	-	-	-0,0000000***
Año*WMLER	1997	-0,0292	-0,0331	-0,0317
	1998	0,3020	0,0298	0,0315
	1999	-0,1010	0,0000	-0,0007
	2000	0,0309	0,0514**	0,0497**
	2001	0,0161	0,0326	0,0327
	2002	-0,0664**	-0,0519**	-0,0305
	2003	-0,0944***	-0,0918***	-0,0646**
Género*WMLER	Mujer	-	0,0007	0,0377
Edad*WMLER	25-45	-	0,01029*	0,0331
	45	-	0,1076	0,0169
Tamaño del establecimiento*WMLER	10-50	-	0,1166*	0,089
	51-200	-	0,1939***	0,0946
	200	-	0,3110***	0,1536*
Región*WMLER	Pamp.	-	-0,1636***	-0,1221***
	Cuyo	-	-0,0742*	-0,0096
	NOA	-	-0,1359***	-0,0508
	NEA	-	-0,1553***	-0,0804
	Patag.	-	-0,0184	-0,0186
	Constante	18.044,13	37.504,45	51.567,49
	N	1.150	1.150	1.150
	R2	0,8244	0,8926	0,8935

Nota: *** significativo al 99%, ** al 95%, * al 90%.

Fuente: elaboración propia.

Cuadro 7. Resultados de estimaciones por especificación. Etapa continua.

		(2)	(3)	(4)
	WMLER	0,4802***	0,4817***	0,3165***
	WMLER^2	-0.0000002***	-0.0000003***	0.0000009***
	WMLER^3	-	-	-0.00000000***
Año*WMLER	2004	0,0008	0,0039	-0,0021
	2005	-0,0133	-0,0125	-0,0217
	2006	0,0100	0,0055	-0,0091
	2007	0,0116	0,0068	-0,0095
	2008	-0,0076	-0,0193	-0,0377*
	2009	-0,0309	-0,0346*	-0,0553**
	2010	-0,0127	-0,0185	-0,0390*
	2011	-0,0156	-0,0242	-0,0453**
	2012	-0,0338	-0,0374*	-0,0607***
	2013	-0,0755**	-0,0768***	-0,1002***
	2014	-0,0863**	-0,0951***	-0,1169***
	2015	-0,0961***	-0,1024***	-0,1228***
	2016	-0,0041	-0,0117	-0,0320
	2017	-0,0254	-0,0213	-0,0437**

	2018	-0,0499	-0,0479**	-0,0697***
	2019	-0,0841**	-0,0874***	-0,1064***
	2020	-0,0531	-0,0511**	-0,0685***
	2021	-0,0784**	-0,0712***	-0,0880***
Género*WMLER	Mujer	-	0,0519***	0,0662***
Edad*WMLER	25-45	-	-0,0714***	-0,0871***
	45	-	-0,1074***	-0,1328***
Tamaño del establecimiento*WMLER	10-50	-	0,0285	0,0221
	51-200	-	0,0255	0,0048
	+200	-	0,0195	-0,202
Región*WMLER	Pamp.	-	0,0027	0,0095
	Cuyo	-	0,0160	0,0299**
	NOA	-	0,0028	0,0241
	NEA	-	-0,517***	-0,0339*
	Patag.	-	0,1218***	0,1189***
	Constante	28.126,31***	25.759,09***	30.6044,16***
	N	2.735	2.735	2.735
	R2	0,8455	0,9438	0,9441

Nota: *** significativo al 99%, ** al 95%, * al 90%.

Fuente: elaboración propia.

Los Cuadros 6 y 7 destacan, para ambas etapas, la relevancia de todas las dimensiones contempladas en la captación de ingresos de la EPH, dada por la significatividad estadística de los parámetros estimados en al menos una categoría para todos los casos. La única excepción aparece en las interacciones de tamaño del establecimiento para la etapa continua (no así en la etapa puntual). Esta significatividad en las interacciones marca la relevancia de las categorías para $WMLER$ como variable independiente y, por ende, para la captación de ingresos de EPH como porción del ingreso de registros tributarios del SIPA.

Dada la inclusión de interacciones cuadráticas y cúbicas, los coeficientes estimados no pueden ser interpretados como efectos marginales de manera lineal. Esto es, no representan el efecto del cambio de una unidad sobre la variable dependiente. De este modo, los efectos marginales varían en función de la distribución no condicional del ingreso en $WMLER$ (en términos de grupos socio ocupacionales, unidad observacional en los modelos, aunque ponderando por las cantidades de cada grupo), lo cual permite estimaciones para distintos niveles de ingresos. Así, se opta por estimar los efectos marginales los percentiles 10, 50 y 90, en busca de evaluar impactos diferenciales en tres niveles distintos de ingresos: bajo, mediana y alto, respectivamente.

Cuadro 8. Efectos marginales estimados según percentil y especificación. Etapa puntual.

		(2) – sólo tiempo			(3) - todo			(4) – todo + cúbica		
		p10	p50	p90	p10	p50	p90	p10	p50	p90
Año	1996	0,6545***	0,5737***	0,3331***	0,4422***	0,3268***	-0,0166	0,2160**	0,2910***	0,1848*
	1997	0,6253***	0,5445***	0,3039***	0,4091***	0,2937***	-0,0497	0,1844*	0,2593***	0,1531
	1998	0,6847***	0,6039***	0,3634***	0,4720***	0,3566***	0,0131	0,2475**	0,3225***	0,2163**
	1999	0,6444***	0,5636***	0,3230***	0,4422***	0,3268***	-0,0166	0,2154**	0,2903***	0,1842*
	2000	0,6854***	0,6046***	0,3640***	0,4936***	0,3783***	0,0348	0,2657***	0,3407***	0,2345**
	2001	0,6706***	0,5898***	0,3492***	0,4748***	0,3594***	0,0159	0,2487**	0,3236***	0,2175**
	2002	0,5881***	0,5073***	0,2667***	0,3903***	0,2750***	-0,0685	0,1855**	0,2605***	0,1543
	2003	0,5601***	0,4793***	0,2387***	0,3504***	0,2351***	-0,1084	0,1514*	0,2264***	0,1202
Género	Varón	-	-	-	0,4413***	0,3259***	-0,0176	0,2084**	0,2834***	0,1772*
	Mujer	-	-	-	0,4420***	0,3266***	-0,0168	0,2461***	0,3211***	0,2149*
Edad	-25	-	-	-	0,3556***	0,2403**	-0,1032	0,1966*	0,2716***	0,1654
	25-45	-	-	-	0,4585***	0,3432***	-0,0003	0,2297**	0,3047***	0,1985**
	45	-	-	-	0,4633***	0,3479***	0,0045	0,2135**	0,2885***	0,1823**
Tamaño del establecimiento	-10	-	-	-	0,2658***	0,1504	-0,1930*	0,126	0,2010**	0,0948
	10-50	-	-	-	0,3824***	0,2670***	-0,0764	0,2150**	0,2900***	0,1838
	51-200	-	-	-	0,4597***	0,3443***	0,0009	0,2207**	0,2957***	0,1895**
	200	-	-	-	0,5768***	0,4614***	0,1180**	0,2796**	0,3546***	0,2484***
Región	GBA	-	-	-	0,5111***	0,3958***	0,0523	0,2648**	0,3398***	0,2336***
	Pamp.	-	-	-	0,3475***	0,2321***	-0,1113	0,1427	0,2177***	0,1115
	Cuyo	-	-	-	0,4370***	0,3216***	-0,0219	0,2552***	0,3302***	0,2240*
	NOA	-	-	-	0,3752***	0,2598***	-0,0836	0,2140**	0,2890***	0,1828
	NEA	-	-	-	0,3558***	0,2404***	-0,103	0,1844**	0,2594***	0,1532
	Patag.	-	-	-	0,4927***	0,3774***	0,0339	0,2462**	0,3212***	0,2150**

Nota: *** significativo al 99%, ** al 95%, * al 90%.

Fuente: elaboración propia.

Cuadro 9. Efectos marginales estimados según percentil y especificación. Etapa continua.

		(2) – sólo tiempo			(3) - todo			(4) – todo + cúbica		
		p10	p50	p90	p10	p50	p90	p10	p50	p90
Año	2003	0.4509***	0.4325***	0.3949***	0.4201***	0.3988***	0.3553***	0.3617***	0.4011***	0.4259***
	2004	0.4518***	0.4334***	0.3958***	0.4240***	0.4027***	0.3592***	0.3596***	0.3990***	0.4238***
	2005	0.4376***	0.4192***	0.3816***	0.4075***	0.3863***	0.3428***	0.3399***	0.3794***	0.4042***
	2006	0.4610***	0.4426***	0.4050***	0.4256***	0.4043***	0.3608***	0.3526***	0.3920***	0.4168***
	2007	0.4625***	0.4441***	0.4065***	0.4269***	0.4057***	0.3621***	0.3522***	0.3916***	0.4164***
	2008	0.4433***	0.4250***	0.3874***	0.4008***	0.3795***	0.3360***	0.3239***	0.3634***	0.3882***
	2009	0.4200***	0.4017***	0.3640***	0.3855***	0.3643***	0.3207***	0.3063***	0.3458***	0.3706***
	2010	0.4382***	0.4199***	0.3822***	0.4016***	0.3803***	0.3368***	0.3227***	0.3621***	0.3869***

	2011	0.4353***	0.4169***	0.3793***	0.3959***	0.3746***	0.3311***	0.3163***	0.3558***	0.3806***
	2012	0.4171***	0.3987***	0.3611***	0.3827***	0.3614***	0.3179***	0.3010***	0.3404***	0.3652***
	2013	0.3754***	0.3571***	0.3195***	0.3433***	0.3220***	0.2785***	0.2615***	0.3009***	0.3257***
	2014	0.3646***	0.3462***	0.3086***	0.3249***	0.3037***	0.2602***	0.2447***	0.2842***	0.3090***
	2015	0.3548***	0.3364***	0.2988***	0.3177***	0.2964***	0.2529***	0.2389***	0.2783***	0.3031***
	2016	0.4468***	0.4284***	0.3908***	0.4084***	0.3871***	0.3436***	0.3296***	0.3691***	0.3939***
	2017	0.4256***	0.4072***	0.3696***	0.3988***	0.3776***	0.3340***	0.3180***	0.3574***	0.3822***
	2018	0.4010***	0.3827***	0.3450***	0.3722***	0.3509***	0.3074***	0.2919***	0.3314***	0.3562***
	2019	0.3668***	0.3484***	0.3108***	0.3327***	0.3115***	0.2679***	0.2552***	0.2947***	0.3195***
	2020	0.3978***	0.3795***	0.3418***	0.3689***	0.3477***	0.3042***	0.2931***	0.3326***	0.3574***
	2021	0.3725***	0.3542***	0.3166***	0.3488***	0.3276***	0.2841***	0.2737***	0.3131***	0.3379***
Género	Varón	-	-	-	0.3638***	0.3425***	0.2990***	0.2821***	0.3215***	0.3463***
	Mujer	-	-	-	0.4157***	0.3944***	0.3509***	0.3483***	0.3877***	0.4125***
Edad	-25	-	-	-	0.4516***	0.4304***	0.3868***	0.3906***	0.4301***	0.4549***
	25-45	-	-	-	0.3802***	0.3589***	0.3154***	0.3035***	0.3430***	0.3677***
	45	-	-	-	0.3443***	0.3230***	0.2795***	0.2578***	0.2973***	0.3221***
Tamaño del establecimiento	-10	-	-	-	0.3618***	0.3405***	0.2970***	0.3051***	0.3446***	0.3694***
	10-50	-	-	-	0.3903***	0.3690***	0.3255***	0.3272***	0.3666***	0.3914***
	51-200	-	-	-	0.3873***	0.3660***	0.3225***	0.3099***	0.3493***	0.3741***
	200	-	-	-	0.3813***	0.3600***	0.3165***	0.2850***	0.3244***	0.3492***
Región	GBA	-	-	-	0.3734***	0.3521***	0.3086***	0.2915***	0.3309***	0.3557***
	Pamp.	-	-	-	0.3761***	0.3549***	0.3113***	0.3010***	0.3404***	0.3652***
	Cuyo	-	-	-	0.3894***	0.3682***	0.3246***	0.3214***	0.3608***	0.3856***
	NOA	-	-	-	0.3762***	0.3550***	0.3114***	0.3155***	0.3550***	0.3798***
	NEA	-	-	-	0.3216***	0.3004***	0.2569***	0.2575***	0.2970***	0.3218***
	Patag.	-	-	-	0.4952***	0.4740***	0.4304***	0.4104***	0.4498***	0.4746***

Nota: *** significativo al 99%, ** al 95%, * al 90%.

Fuente: elaboración propia.

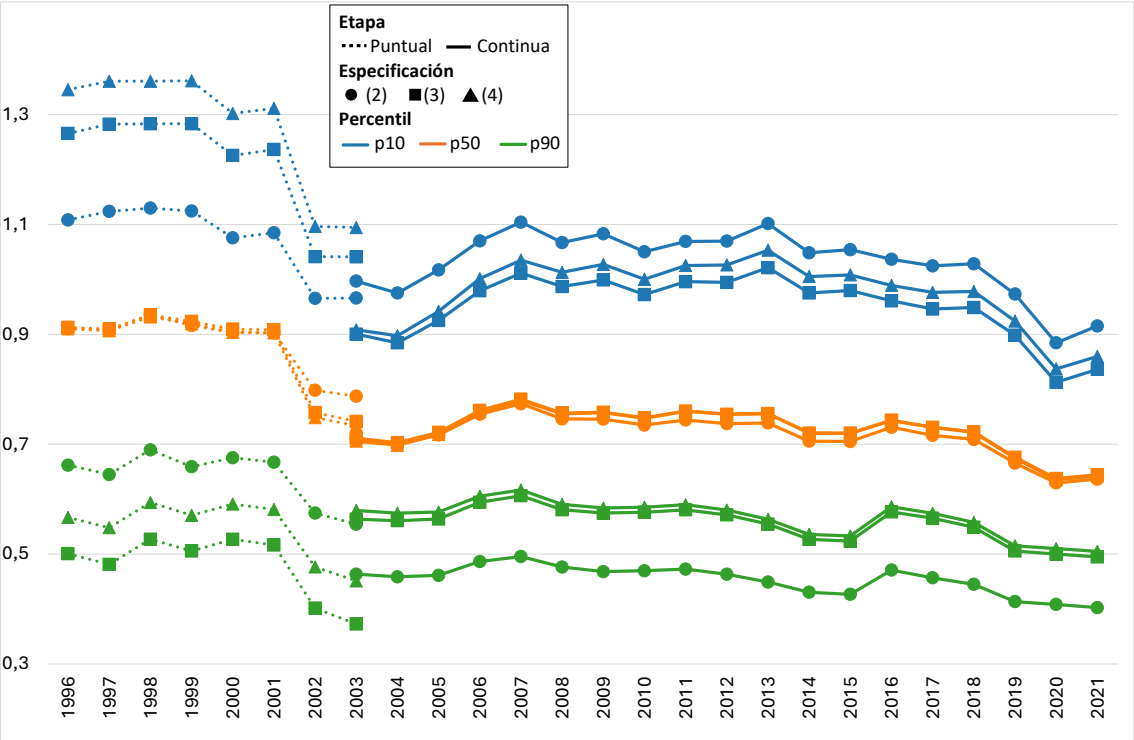
Las estimaciones presentadas en el Cuadro 8 y el Cuadro 9 reflejan distintos impactos en la captación según nivel de ingresos. Esto es, la proporción que se capta en $WEPH$ de un aumento de \$1.- en $WMLER$, *ceteris paribus* el resto de los regresores. En general, se tratan de traslados relativamente bajos, lejanos a los aproximadamente 0,83 que supondrían una captación "perfecta" según nivel teórico. Concretamente, la mayoría de los efectos marginales estimados se encuentran entren 0,3 y 0,6.

Por otro lado, se observa mayor distancia entre los efectos marginales de p90 y p10 en la etapa puntual respecto a la etapa continua. Esto habla de una homogeneización en la captación de ingresos en la encuesta según nivel de ingresos entre etapas. En términos temporales intra-etapas, en ambas se refleja una disminución en la captación a períodos temporales más actuales respecto de anteriores. Sólo alguna reversión en esta

tendencia hacia el año 1999 en la etapa puntual -para luego caer abruptamente en el año 2000- y entre 2005 y 2007 en la etapa continua.

Para interpretar las brechas entre $WEPH$ y $WMLER$ de manera directa, es decir, en forma de coeficientes similares a los ratios Br_t empleados en el capítulo anterior, es necesario predecir la variable dependiente. De este modo, en base a los efectos marginales ya revisados, se predice un $WEPH$ para distintos niveles de $WMLER$, y entre ambos se obtiene un ratio Br_t estimado. De esta manera, se apunta a dimensionar en términos de magnitudes cuantitativas el grado de influencia del nivel de ingresos a su captación, pero en una comparación paramétrica entre comparables. Se alimenta al modelo con distintos niveles de la distribución no condicional de $WMLER$ -p10, p50, p90- y se obtienen $WEPH$ estimados para cada uno, en cada año. Esto es lo que se visualiza, para las tres especificaciones del modelo, en el Gráfico 20.

Gráfico 20. Brechas estimadas entre fuentes, por percentil de ingresos y especificación de modelo. 1996-2021.



Fuente: elaboración propia.

En el Gráfico 20 se muestran las brechas símil los ratios Br_t predichas por el modelo presentado, en sus 3 distintas especificaciones. A grandes rasgos el modelo indica, tal como fue revisado en el capítulo anterior, que la captación del ingreso en la década de 1990 resulta mucho más heterogénea de lo que se suele caracterizar. Es de este modo a

partir de una sobre captación en los bajos ingresos: para el percentil 10 se estiman siempre mayores ingresos en términos absolutos en la encuesta respecto a la información registral. La captación en la mediana, por su parte, aparece en esta etapa algo por encima del nivel teórico determinado, alrededor de Br_t=0,9. En el percentil 90, en cambio, se estima una pronunciada subcaptación, de alrededor de Br_t=0,6. De este modo, como ya fuera descrito, se configura en promedio una apariencia de correcta captación, en torno al nivel teórico de Br_t=0,83, pero que esconde estas heterogeneidades.

Luego, en la etapa continua y a lo largo del siglo XXI, el modelo estima que la captación se redujo para toda la distribución. Sin embargo, no con la misma intensidad: lo hace de manera más suave cuanto más alto el nivel de ingresos. Así, en el movimiento de largo plazo se observa una mayor disminución en el coeficiente entre fuentes para el percentil 10, un movimiento intermedio respecto a los percentiles de los extremos en la mediana, y una relativa mayor estabilidad en un bajo nivel de captación para el percentil 90. Sin embargo, resaltan matices entre las tres especificaciones elegidas, que se reflejan con claridad en los distintos saltos y variaciones entre etapas, dándose de forma distinta según la especificación elegida. Esto agrega un relativo tamiz a la observación acerca del mayor empeoramiento relativo de la captación en los bajos ingresos como movimiento de largo plazo. Por ejemplo, la especificación (4) marca un pronunciado cambio hacia abajo para el percentil 10 entre etapa puntual y continua, y lo contrario para el percentil 90, resultando la etapa continua de esta estimación mucho más homogénea que la puntual.

Esta observación de trayectoria diferencial de la captación según nivel de ingresos refuerza la conclusión abordada al final del capítulo anterior en torno al mayor empeoramiento relativo en la captación de las categorías de bajos ingresos -en este caso, de los bajos ingresos controlando por cada dimensión disponible-. La sobrecaptación -o subdeclaración al SIPA (INDEC, 2006)- en los bajos ingresos, característica de la etapa puntual, merma hacia el nuevo milenio y la etapa continua. Paradójicamente, este acercamiento entre fuentes en los ingresos captados para los primeros percentiles de la distribución podría significar un deterioro de la relación agregada: con su correcta captación, deja de oficiar de "contrapeso" de la mala captación de los altos ingresos -lo que dotaba a la década de 1990 de apariencia de captación correcta-, que ante la relativa estabilidad en la -mala- captación de los percentiles superiores, redunda en un deterioro agregado de la captación de ingresos de la encuesta.

6. CONCLUSIONES, DISCUSIÓN Y TRAZOS A FUTURO

Como fuera revisado, la literatura especializada se ha enfrentado al problema de la subcaptación de ingresos desde los orígenes del empleo de las encuestas como fuente de información. La existencia de este fenómeno ha sido tratada en búsquedas de despejar o aminorar el efecto de los sesgos que produce para estimaciones derivadas como distribución del ingreso o pobreza. Son minoritarios los trabajos donde se ha tratado el estudio del fenómeno de forma exhaustiva en busca de mejorar su caracterización. Este trabajo se enmarca en este grupo, particularmente a causa de la creciente relevancia que muestra en los últimos años, apuntando a comprender mejor sus características y determinaciones para a su vez brindar mayor precisión a futuras correcciones.

Con esta intención, a partir de la disponibilidad de información en MLER se adecúa la EPH en todas sus dimensiones posibles para realizar un emparejamiento de información de asalariados registrados privados. A partir de este, se analiza la evolución cuantitativa de la subcaptación de ingresos en la encuesta, tanto en términos agregados como desagregando por distintas dimensiones en términos individuales como considerando distintas dimensiones de conjunto.

Del presente trabajo se desprenden múltiples conclusiones, presentes tanto en el Capítulo 4 como en el Capítulo 5. En primer lugar, como es documentado en la literatura, el nivel de ingresos determina buena parte de la brecha. En parte, por la existencia de tributos que deterioran la paridad entre fuentes por ensanchar la distancia entre bruto y neto, pero también por falta de respuesta o no captación en términos muestrales. En este sentido, se refuerzan los hallazgos de Camelo (1998) y Groppa (2009) en torno a marcar los diferenciales de captación para asalariados en función de su ubicación en la distribución del ingreso más allá de la fuente de estos.

En segundo lugar, las características de los individuos encuestados son relevantes para la captación de los ingresos. Es cierto que la brecha es menor para cada categoría de menor nivel de ingresos: mujer en género, menores de 25 años en edad, establecimientos de tamaño menor a 10 personas, regiones del norte del país y Construcción en sector productivo. Sin embargo, tanto características personales como características del puesto laboral muestran comportamientos diferenciales incluso comparando a similares niveles salariales de la fuente de registro, lo que revela los matices aportados por estas

dimensiones a la captación del ingreso (más allá de que gran parte esta captación está dada por el nivel de los salarios). Este hallazgo refuerza lo hallado por Groppa (2009) en términos de diferenciales de captación por región, así como lo hallado por INDEC (2006), donde se documenta la relevancia de incluir sectores productivos a la hora de evaluar captación de ingresos en EPH. Por el contrario, Camelo (1998) no les otorga relevancia a los sectores productivos en esta clave, en tanto encuentra que sus diferenciales de captación están en realidad determinados por sus diferentes niveles de ingreso.

En tercer lugar, se propone una caracterización distinta acerca de la captación de ingresos en la etapa puntual y principios de la etapa continua. Al contrario de cómo se postula en trabajos donde se supone o se estima la ausencia de subcaptación en asalariados hacia fines de la década de 1990 (Alaimo y Sosa Escudero, 2000; Roca y Pena, 2001), la dispersión de captación entre categorías revisada en este trabajo indica que los promedios agregados cercanos a los niveles teóricos esconden en realidad sobrecaptación de categorías ingresos que compensa la subcaptación de estratos de altos ingresos, acompañando otros trabajos que demuestran la debilidad de la encuesta para captar altos ingresos, incluso en esa etapa de apariencia ausencia de subcaptación (Alvaredo, 2010). Nuevamente, esto refuerza lo expuesto en INDEC (2006), donde se documenta evidencia de subdeclaración al SIJyP -sobrecaptación en EPH- relevante para sectores productivos como Construcción o Comercio para fines de la década de 1990.

En cuarto lugar y en estrecha vinculación con lo anterior, el movimiento de largo plazo en el período estudiado -tanto en el análisis de a una dimensión como en el modelo paramétrico- sugiere que el deterioro de la captación agregada es también conducido, al menos parcialmente, por los cambios en la captación en los bajos estratos de ingresos. Esto es, su mejor captación en términos de "acercarse" al nivel de brecha teórica entre fuentes, reduciendo la subdeclaración al SIPA, paradójicamente empeora el coeficiente de captación general. Esto es, en tanto la baja captación de altos ingresos continua relativamente estable, deteriorándose levemente, pero sin compensar la baja en el coeficiente de captación de los bajos ingresos. Cabe considerar aquí, de todas formas, dos cuestiones: por un lado, el menor peso relativo de los salarios de los deciles inferiores sobre el promedio agregado, tamizando el impacto sobre el ensanchamiento agregado de la brecha revisado al inicio. Por otro lado, que el cambio suscitado en 2003 en la modalidad de pago de las Asignaciones Familiares -mencionado en la sección 2.2- puede

estar influyendo sobre este fenómeno, abonando a la reversión de la sobrecaptación de bajos ingresos.

A partir de esta lectura se propone una nueva interpretación -al menos parcial- de la creciente relevancia del fenómeno. Podría pensarse que, junto con el incremento de la tasa de registro del empleo dado a principios de la década de 2000, se conjugó un blanqueo de los ingresos que se encontraban parcialmente no registrados (no declarados al SIPA), particularmente presentes en estratos de bajos ingresos y en las características del empleo de este tipo. Coloquialmente, un proceso de "blanqueo" del salario en gris. De ser así, este proceso es el que desencadena lo contraintuitivo de que, ante un registro de mayor fidelidad entre encuesta y registros administrativos, un nivel salarial más fiel al real en los registros administrativos, la captación agregada de la encuesta empeore.

Para obtener un panorama completo, podemos sintetizar algunas de las otras posibles explicaciones del crecimiento de la subcaptación de salarios en los últimos años de esta manera:

a) El crecimiento de modalidades de vivienda por fuera del alcance de la muestra de la encuesta (en barrios cerrados o predios conocidos coloquialmente como *countries*) deterioró la captación de los ingresos a partir de la no captación de individuos de ingresos típicamente altos, que se trasladan a estas modalidades de vivienda inaccesibles para las encuestas públicas.
b) La elevada variación de la nominalidad, en donde los ingresos pueden actualizarse con mayor frecuencia, conlleva en declaraciones de ingresos erróneas -inferiores- a la encuesta, ya sea de forma voluntaria o involuntaria por olvido o mero redondeo. Tal como se referencia generalmente le sucede a los no asalariados, dada la mayor variabilidad en sus ingresos (Beccaria, 1998). Altimir (1986) ya planteaba la nominalidad como causante del problema de captación de ingresos durante la década de1980. La apariencia de captación correcta en conjunción con la estabilidad nominal de la década de 1990 favorece también esta hipótesis.
c) Una progresiva desconfianza en las estadísticas públicas y en los institutos garantes de las mismas a partir de la intervención del INDEC de 2007 (Minoldo y Born, 2009).

d) Una creciente incidencia en la relevancia de tributos de asalariados que ensanchan la relación entre salario bruto y salario neto. Particularmente, del Impuesto a las Ganancias[27].

A la luz de las conclusiones ofrecidas en este trabajo, se deriva que la hipótesis que se ve relativamente más perjudicada es la a), vinculada a la creciente ausencia de captación perceptores de salarios de alto nivel por el alcance urbano de la encuesta. Esto es, dado que la misma vincula el deterioro directamente con la subcaptación de altos ingresos, hecho que no parece estar a la vista en el análisis cuantitativo realizado. Adicionalmente, según lo revisado en la sección 2.2 en torno a la cambiante composición de los asalariados que tributan este impuesto, la hipótesis d) no parece sustentar el constante rezago más allá de un ensanchamiento en nivel en el mismo en períodos particularmente desde 2011, dados los aumentos en la tasa efectiva y en proporción de asalariados tributantes de IIGG.

De este modo, como sobre las otras hipótesis no existen razones para descartarlas a la luz del análisis realizado, en conjunción con la nueva hipótesis planteada en este trabajo puede pensarse el fenómeno como, al menos en alguna medida, una combinación de los cuatro efectos a la vez: retroceso del salario en gris hacia principios de la década del 2000, creciente inflación y desconfianza en las estadísticas públicas a partir de 2007, y cambios en la relevancia y alcance del Impuesto a las Ganancias para los asalariados. Así, se desprenden temas futuros de investigación cuyo objetivo sea refinar la influencia de algunas o todas estas hipótesis como causas del fenómeno.

Finalmente, cabe contemplar que los hallazgos para el período comprendido por este trabajo pueden emplearse en futuras correcciones, en combinación con precisas correcciones de corte transversal como la de Albina *et al*. (2024), donde además se encuentran considerados perfiles laborales más allá del universo de asalariados registrados. Así, potencialmente refinar indicadores socioeconómicos como los de distribución de ingreso y de pobreza, en donde la subcaptación de ingresos impacta de forma asimétrica -con mayores sesgos subestimando desigualdad respecto a los que puede introducir exagerando la pobreza- a partir de la mayor relevancia del fenómeno en altos ingresos. En una corrección inter temporal de largo plazo, se debe tomar en cuenta el

[27] Acerca de las fluctuaciones de la tasa efectiva y sobre la cantidad de asalariados contribuyentes en este impuesto, ver MECON (2023)

movimiento aquí observado en la reversión de la sobrecaptación de los ingresos para los primeros percentiles, así como también la persistencia de una importante subcaptación de ingresos -pero relativamente más estable en el tiempo- en los percentiles más altos. De modo sucinto puede intuirse que, a la idea guía de la corrección en función de indicadores de pobreza sobreestimados e indicadores de desigualdad subestimados, los hallazgos de este trabajo agregarían un tamiz para fines de la década de 1990, donde la consideración de la sobrecaptación de bajos ingresos contrapesa esta tendencia tanto en términos de pobreza como de distribución del ingreso.

Bibliografía

Albina, A., Laguinge, L., Gasparini, G., Tornarolli, L., Cruces, G. &Afonso, S. (2024). *Ajustando la Imagen de la distribución del ingreso en Argentina: Encuestas y Registros Administrativos*. Documentos de Trabajo del CEDLAS N°336, CEDLAS-UNLP, La Plata, Buenos Aires.

Altimir, O. (1975). *Estimaciones de la distribución del ingreso en América Latina por medio de encuestas de hogares y censos de población*. Sede de la CEPAL en Santiago (Estudios e Investigaciones), Naciones Unidas, Comisión Económica para América Latina y el Caribe (CEPAL).

Altimir, O. (1979). *La dimensión de la pobreza en América Latina*. Cuadernos de la CEPAL, Santiago de Chile, Publicación de las Naciones Unidas.

Altimir, O. (1986). *Estimaciones de la distribución del ingreso en la Argentina, 1953-1980*. Desarrollo Económico, Vol. 25, N° 100, enero-marzo, pp.521-566.

Alvaredo, F. (2010). "The Rich in Argentina over the Twentieth Century, 1932–2004". En A. B. Atkinson & T. Piketty (ed.) *Top Incomes: A Global Perspective (cap. 6)*. PSE-Ecole d'économie de Paris.

Amarante, V. (2013). *Income inequality in Latin America. Data challenges and availability from a comparative perspective*. Políticas Sociales 6200, Naciones Unidas Comisión Económica para América Latina y el Caribe (CEPAL).

Arakaki, A. (2016). *Cambios recientes en la Encuesta Permanente de Hogares. Análisis de la actualización de las proyecciones poblacionales y la muestra maestra.* Documento de Trabajo del CEPED N° 22, CEPED-IIE-FCE-UBA, Buenos Aires, septiembre.

Atkinson, A.B. (2007). "Measuring Top Incomes: Methodological Issues". En A. B. Atkinson & T. Piketty (ed.) *Top Incomes over the 20th Century (cap. 2)*. PSE-Ecole d'économie de Paris.

Beccaria, L. (1998). *Criterios operativos de las encuestas de hogares y la medición de los ingresos*. CEPAL, Segundo Taller Regional sobre Medición del Ingreso en las Encuestas de Hogares, pp. 87-96. Buenos Aires, noviembre.

Bertranou, F., Cetrángolo, O., Grushka, C. & Casanova, L. (2011). *Encrucijadas en la seguridad social Argentina*. Buenos Aires: CEPAL y Oficina Internacional del Trabajo.

Camelo, H. (1998). *Subdeclaración de ingresos medios en las encuestas de hogares según quintiles de hogares y fuente del ingreso*. Segundo Taller Regional sobre Medición del Ingreso en las Encuestas de Hogares, pp. 459-473. Buenos Aires, noviembre.

Cattaneo, M. (2001). "La EPH en los '90: una mirada desde el usuario". En Lindenboim, J. (comp.) *Crisis y metamorfosis del mercado de trabajo. Parte 2: aportes metodológicos y otras evidencias*. Cuaderno del CEPED N° 5, CEPED-IIE-FCE-UBA, Buenos Aires.

Centro de Investigación y Formación de la República Argentina (CIFRA) (2009). *La evolución del sistema previsional argentino*. Documentos de Trabajo de CIFRA N°2, noviembre.

Comisión Económica para América Latina y el Caribe (CEPAL) (2018). *Medición de la pobreza por ingresos. Actualización metodológica y resultados*. Metodologías de la CEPAL, Santiago de Chile, diciembre.

Felcman, D., Kidyba, S. & Ruffo, H. (2004). *Medición del ingreso laboral: Ajustes a los datos de la Enuesta Permanente de Hogares para el análisis de la distribución del ingreso (1993-2002)*. 14° Taller regional del MECOVI - Imputación de datos en las encuestas de hogares: los procedimientos metodológicos y sus implicaciones, pp. 21-47.

Gasparini, L. (1998). "Incidencia distributiva del sistema impositivo argentino". En FIEL (ed.) *La reforma tributaria en la Argentina (cap. 12)*. Fundación de Investigaciones Económicas Latinoamericanas.

Gasparini, L. (1999). "Desigualdad en la distribución del ingreso y bienestar. Estimaciones para la Argentina". En FIEL (ed.) *La distribución del ingreso en la Argentina (secc. II)*. Fundación de Investigaciones Económicas Latinoamericanas.

Gasparini, L, Cicowiecz, M. & Sosa Escudero, W. (2013). *Pobreza y desigualdad en América Latina: conceptos, herramientas y aplicaciones*. UNLP, CEDLAS, Editorial Temas, febrero.

Gasparini, L. & Sosa Escudero, W. (1999). *Bienestar y distribución del ingreso en la Argentina, 1980-1998*. Económica, Vol. 45, N°3, pp. 183–217.

Gasparini, L., Marchionni, M. & Sosa Escudero, W. (2000*). La distribución del ingreso en Argentina y la Provincia de Buenos Aires*. Cuaderno de Economía Nº 49, Ministerio de Economía de la Provincia de Buenos Aires, marzo.

Gasparini, L. (2024). *Desigualdad*. Argendata. Fundar.

Gómez Sabaíni, J. & Rossignolo, D. (2014). *La tributación sobre las altas rentas en América Latina*. Serie Estudios y Perspectivas N°13, Oficina de la CEPAL en Montevideo. Impreso en Naciones Unidas, Santiago de Chile.

Groppa, O. (2009). *Un método para corregir la subdeclaración de ingresos y gasto de consumo a partir de la Encuesta Nacional de Gasto de los Hogares 2004-2005*. XLIV reunión anual de AAEP. Mendoza, noviembre.

Heckman, J. (1979). *Sample Selection Bias as a Specification Error*. Econometrica, Vol. 47, N° 1, pp. 153-161.

Instituto Nacional de Estadísticas y Censos (INDEC) (2003). *La nueva Encueste Permanente de Hogares de Argentina. 2003.* INDEC, Secretaría de Política Económica - Ministerio de Economía, Buenos Aires.

Instituto Nacional de Estadísticas y Censos (INDEC) (2006). *Generación del ingreso e insumo de mano de obra. Fuentes, métodos y estimaciones. Años 1993–2005*. INDEC, Buenos Aires.

Instituto Nacional de Estadísticas y Censos (INDEC) (2024). *Diseño de registro y estructura para las bases preliminares Hogar y Personas*. INDEC, Buenos Aires, noviembre.

Jiménez, J.P. & Rossignolo, D. (2019). *Concentración del ingreso y desigualdad en América Latina. El caso argentino*. Centro de Estudios para el Cambio Esctructural, enero.

Kennedy, D., Romero, A., Suárez, F. & Sánchez, M.A. (2024). *Fuentes de información sobre salarios en Argentina. Parte I: Metodología y resultados de las series oficiales actuales.* Documento de Trabajo N° 30, CEPED (FCE) - IIEP (UBA/CONICET). Buenos Aires, febrero.

Kuznets, S. (1953). *Shares of Upper Income Groups in Income and Savings*. National Bureau of Economic Research, NBER.

Llach, J.J. & y Montoya, S. (1999). *En pos de la equidad. La pobreza y la distribución del ingreso en el Área Metropolitana de Buenos Aires: diagnóstico y alternativas de políticas.*

Ministerio de Economía de la Nación (MECON) (2023). *Informe sobre el Impuesto a las Ganancias del personal en relación de dependencia. Período 1999-2023*. Secretaría de Hacienda, Dirección Nacional de Investigaciones y Análisis Fiscal, Ministerio de Economía, Argentina.

Minoldo, S., & Born, D. (2019). *Claroscuros: 9 años de datos bajo sospecha (EPH en Argentina entre 2007 y 2015)*. Estudios Sociológicos Editora.

Müller, A. (2023). *Medición de la pobreza: cuestiones conceptuales y de implementación - el caso de la Argentina*. Documento de Trabajo CESPA N°63, CESPA-FCE-UBA, Buenos Aires, octubre.

Observatorio de Empleo y el Dinámica Empresarial (OEDE) (2022). *Muestra longitudinal de empleo registrado. Datos a 2021*. Ministerio de Trabajo, Empleo y Seguridad Social, abril.

Okner, B. (1975). *Taxes and income: A microunit analysis*. Review of Income and Wealth, International Association for Research in Income and Wealth, vol. 21(3), pp. 279-299, September.

Oranización de las Naciones Unidas (ONU) (2009). *Clasificación Industrial Internacional Uniforme de todas las actividades económicas (CIIU) revisión 4.* Departamento de Asuntos Económicos y Sociales, División de Estadística, Informes estadísticos Serie M, No. 4.

Roca, E. & Pena, H. (2001). *La declaración de ingresos en las encuestas de hogares*. Congreso Nacional de Estudios del Trabajo, ASET. Buenos Aires, 1, 2 y 3 de agosto de 2001.

Salvia, A. & Donza, E. (1999). *Problemas de medición y sesgos de estimación derivados de la no respuesta a las preguntas de ingresos de la EPH (1990-1998)*. Revista Estudios del Trabajo, N° 18, ASET.

Sánchez, M. A., Pacífico, L. & Kennedy, D. (2016). *La Participación asalariada en el ingreso y su composición según vínculo laboral. Argentina 1993-2015. Fuentes de información, metodologías y alternativas de estimación*. Documento de Trabajo del CEPED N° 21, CEPED-IIE-FCE-UBA, Buenos Aires, septiembre.

Santiere, J.J., Gómez Sabaíni, J.C. & Rossignolo, D. (2000). *Impacto de los impuestos sobre la distribución del ingreso en Argentina en 1997*. Banco Mundial, SPEyR, Ministerio de Economía.

Sosa Escudero, W. & Alaimo, V. (2000). “La economía oculta en la Argentina: evidencia basada en encuestas de gasto”. En FIEL (ed.) *La Economía Oculta en la Argentina (cap. II.2)*. Fundación de Investigaciones Económicas Latinoamericanas.

Tornarolli, L., Gasparini, L. & Schteingart, D. (2024). *Salarios e ingresos*. Argendata. Fundar.

Trombetta, M. & Cabezón Cruz, J. (2020). *Brecha salarial de género en la estructura productiva argentina.* Documentos de Trabajo del CEP XXI N°2, Centro de Estudios para la Producción-Ministerio de Desarrollo Productivo de la Nación, julio.

Villatoro, P. (2015). *Ajuste de los ingresos de las encuestas a las Cuentas Nacionales. Una revisión de la literatura.* Serie estudios estadísticos N° 91, CEPAL, Santiago de Chile, abril.

Anexo I: Criterios operativos para la compatibilización entre EPH y MLER

Para empardar con la población objetivo de la MLER, en las EPH de la etapa tanto puntual como continua la muestra se limita a ocupados (ESTADO == 1), asalariados (P17==3 en EPHP, CAT_OCUP==3 en EPHC), en condición registrada ((BENEF2 == (11, 12, 14)) en EPHP; (pp07h == 1 | pp07i == 1) en EPHC), sin beneficiarios de planes (P18A != 1 en EPHP a partir de la onda de octubre del 2000, pp07e != 1 en EPHC), sin empleados de la administración pública (P18B != 1 en EPHP, pp04a != 1 en EPHC) ni servicio doméstico (RAMA !=21 en EPHP, vía CAES/CNO en EPHC).

La identificación del servicio doméstico combinando ambos clasificadores de efectúa de esta manera:

a. Para las bases de microdatos del período 2004-2011, con PP04B_COD = 9500 o, si no se cuenta con información de esta, si los primeros dos dígitos de PP04D_COD = 55.

b. Para el lapso 2012-2015, si la variable PP04B_CAES = 9700 o, si no se cuenta con información de esta, si los primeros dos dígitos de PP04D_COD = 55.

c. Para el período 2016-2022, si la variable PP04B_COD = 9700 o, si no se cuenta con información de esta, si los primeros dos dígitos de PP04D_COD = 55.