

Reunión de Discusión N° 184

Fecha: 08/07/2008

Hs.: 16

La Encuesta Permanente de Hogares: Fuente de datos socioeconómicos de Argentina

Carolina Piselli*

1. Introducción

El presente trabajo forma parte de uno mayor “Diferenciación salarial y Distribución del ingreso”, cuyo objetivo principal es el estudio de la relación entre la dinámica del mercado laboral y la dinámica de desigualdad del ingreso, teniendo como ejes de análisis, el temporal y el espacial. Así, al comenzar con el manejo de las bases de datos durante el periodo de interés, 1990-2003 y para el total país, se inició una etapa de investigación que permitió vislumbrar problemas a resolver en cuanto a los ejes de análisis propuestos. De este modo, surgió la necesidad de compilar y ordenar información, muchas veces dispersa y no siempre homogénea.

Por lo tanto, el objetivo de esta parte del trabajo es realizar una exposición que resuma características relevantes de la Encuesta Permanente de Hogares (EPH), fuente de datos de este trabajo y principal fuente intercensal de datos socioeconómicos de la República Argentina.

Durante el periodo examinado se observan importantes cambios en la EPH, como por ejemplo la existencia de formatos diferentes de bases en distintos años y aglomerados, cambios en el diseño muestral, modificaciones en las variables captadoras de ingresos, en la codificación de variables. Se trata siempre de aspectos importantes a tener en cuenta si se busca trabajar con datos comparables.

En síntesis, este capítulo brinda un compendio de información sobre la EPH y el mismo se organiza de la siguiente manera: En la sección 2 se exponen las características

* Se agradece, la colaboración brindada por Juan Carlos Cid, Sandra Duclós y Cynthia Pok sobre temas metodológicos de la Encuesta; los comentarios de Jorgelina Mena y Eugenio Martínez acerca de aspectos estadísticos y a Jorge Paz, su colaboración hizo posible la materialización de este trabajo.

generales de la Encuesta. Las secciones 3 y 4 se dedican al diseño muestral y sus cambios, mientras que en la 5 se cuenta con información en cuanto a su disponibilidad en soporte informático. Por último en la sección 6 se caracteriza a la EPH continua y se realizan algunos comentarios.

2. Aspectos generales de la Encuesta Permanente de Hogares¹

2.1 Presentación

La EPH es un programa nacional e intercensal, que realiza el Instituto Nacional de Estadísticas y Censos (INDEC) desde 1972 y en forma conjunta con las Direcciones Provinciales de Estadística (DPE) desde 1974.

Sus objetivos son conocer y caracterizar a la población desde su inserción socioeconómica. Pretende por lo tanto echar luz sobre la situación socioeconómica y demográfica de las personas y especialmente de los hogares, por ser estos los núcleos básicos de convivencia en donde los individuos se asocian e interactúan con el resto de la estructura social.

A fin de cumplir con los objetivos establecidos, la EPH establece como temática un conjunto de dimensiones básicas que pretenden caracterizar los siguientes ejes conceptuales en cuanto a la población: Desde un punto de vista demográfico, por su participación en la producción de bienes y servicios (características ocupacionales y de migración), y por su participación en la distribución del producto social (características habitacionales, educacionales, del ingreso y migratorias).

En cuanto a la cobertura geográfica, los aglomerados de la Encuesta se fueron incorporando paulatinamente desde 1974, en el Cuadro 1 del Anexo² se presenta un detalle de las fechas de incorporación al programa de los distintos aglomerados del país³. Por tratarse de un tema relevante, el siguiente apartado se dedica a profundizar acerca de la cobertura geográfica, temporal y temática de la Encuesta.

2.2 Cobertura

La Encuesta se realiza en aglomerados urbanos⁴, en su mayoría con más de 100.000 habitantes. Teniendo en cuenta los datos poblacionales del Censo 1991, la muestra representaba al 70% de la población urbana en 1995 y al 71,1% en 2002, al llegar a un total de 31 aglomerados urbanos relevados.

La población urbana en Argentina representa un 87,1% de la población total, así el conjunto de aglomerados muestreados representa un 61,9% de dicha población. Por lo tanto, los resultados obtenidos de la Encuesta estarán siempre limitados a la población urbana.

La situación socioeconómica de poblaciones rurales, no puede conocerse con la EPH, a excepción del aglomerado Alto Valle del Río Negro. También existe una Encuesta realizada por el Banco Mundial en áreas rurales de Argentina⁵.

¹ La información que se expone en esta sección, se puede profundizar consultando: INDEC (1995, 1998a, 1999, 2000, 2002a, 2002b).

² A partir de ahora, los cuadros que se encuentren en el Anexo serán indicados con Anx a continuación del número del cuadro y con Ap si se trata de cuadros del Apéndice.

³ Sin embargo, la información en soporte magnético, en general, no se encuentra disponible desde los años de incorporación. Acerca de este tema consultar la Sección 5.

⁴ En el Cuadro 2Anx se presenta información acerca de este aspecto.

⁵ (Cfr.) Gasparini (2007), pág. 3.

Si nos enfocamos en el periodo de relevamiento 1990-2003, se observa que existen dos tipos de EPH, la puntual (EPHP) vigente desde el inicio del Programa hasta Mayo de 2003 y la EPH continua (EPHC) que comienza en el año 2003⁶.

La EPHP concentraba la observación de la muestra en una semana (la tercera del mes) y se realizaba en dos momentos del año, Mayo y Octubre, los cuales se conocen como la Primera y la Segunda onda respectivamente.

Cabe agregar que para 1998 y 1999 existen relevamientos especiales en el mes de agosto, onda reducida, las cuales sólo son comparables entre ellas, ya que la selección muestral difiere de la realizada para las otras ondas, debido a no querer perturbar el esquema de rotación⁷ de la EPHP, INDEC (1998a).

Por otro lado, como ya fue mencionado, los aspectos socioeconómicos de los hogares son la temática que permite abordar el conjunto de variables relevadas por la EPH. Si bien los aspectos considerados por la Encuesta se han mantenido a lo largo del tiempo, la forma de captación de la información se fue perfeccionando. En algunos casos se aprovechó la realización de la Encuesta para llevar a cabo relevamientos especiales, anexando módulos a los cuestionarios básicos, como por ejemplo: Módulo de actividad física y deportiva de la población, relevado en Mayo 2000 para el 50% de la muestra de Gran Buenos Aires; 100% de las muestras de Gran Rosario, Gran Córdoba y Salta.

Encuesta de Victimización física, relevado en Mayo 2000 para el 100% de la muestra de Neuquén-Plottier.

Base de datos anexa de Pobreza, desde Mayo 2001.

Información sobre Planes de Empleo, relevado desde Octubre 2002.

Modulo Programa Jefas/ Jefes de Hogar, relevado en Octubre 2002.

De esta manera, teniendo en este momento una idea básica de la Encuesta argentina, aspectos técnicos que hacen a la muestra, se desarrollan en la siguiente sección.

3. Diseño muestral de la EPH puntual

La muestra de la EPH es estratificada y bietápica. Los hogares que serán encuestados se seleccionan de la siguiente manera:

Dentro de cada aglomerado se selecciona una cantidad de radios censales o subdivisiones de los mismos en adelante “áreas”.

Luego se listan todas las viviendas particulares de las áreas seleccionadas y finalmente se realiza una selección aleatoria de viviendas.

De esta manera los hogares⁸ que habitan las viviendas seleccionadas constituyen el conjunto de hogares a encuestar.

La EPHP renueva el conjunto de hogares a encuestar por cuartos, es decir, el 25% de la muestra se renueva de semestre a semestre, por lo tanto, sólo, un máximo teórico de 25% de la muestra puede ser seguido a lo largo de cuatro semestres.

⁶ En este apartado se hace referencia a la EPH puntual o tradicional. Aspectos generales de la EPH continua se pueden consultar en la Sección 6.

⁷ Acerca del esquema de rotación de la EPH puntual ver el apartado 3.

⁸ Se debe señalar que es posible encontrar más de un hogar residiendo en una vivienda.

Poco se puede decir sobre el marco muestral de la EPH, dado que INDEC considera que todo aquello que contribuiría a caracterizar, en profundidad, dicho marco se encuentra bajo secreto estadístico. Sin embargo, se brinda información sobre dos aspectos relevantes, el error muestral y el cálculo del ponderador de la muestra, ambos temas se exponen en las siguientes secciones.

3.1 Errores de muestreo

Realizar estimaciones con información procedente de la EPH implica tener en cuenta que existirá un error muestral, dado que se trata de una muestra. En INDEC (2002b) y en otras de sus publicaciones se presenta la metodología para calcular una medida de error muestral, el desvío standard (*DS*) y los intervalos de confianza de la estimación, distinguiendo si se trata de estimaciones para *bases de Aglomerado* o para la *base Total* de la EPH.

Para bases de Aglomerado

En primer lugar se explica el uso de una Tabla de errores de muestreo⁹ para el caso de estimaciones de subtotaes poblacionales como por ejemplo mujeres ocupadas entre 15 y 19 años. En este caso si la estimación es $\hat{x} = 75.638$ mujeres, se buscaría en la Tabla el valor más cercano, 80.000 y se tiene que el $DS=9.047$ con el cual se pueden obtener intervalos de confianza y también calcular el coeficiente de variación, $CV = \frac{DS}{\hat{x}}$.

Se podrían calcular intervalos del 99%; 95%; 90% y 67% de confianza, que implican que la estimación estaría contenida entre el valor absoluto de 3; 2; 1,64 y 1 vez el *DS*. Para el ejemplo, el intervalo de confianza del 90% para el total de mujeres ocupadas entre 15 y 19 años, estaría entre 80.000- 1,64 * 9.047 y 80.000 + 1,64 * 9.047.

Mientras el *CV* es 11,3%, aquí INDEC advierte que estimaciones con *CV* superiores al 10% deben ser tratadas con cierta cautela.

Sin embargo, cuando la Tabla resulta limitada, se puede calcular el *DS* con la siguiente formulación:

$DS = \sqrt{K-1 \cdot \frac{\hat{x} * (N - \hat{x})}{N}}$, donde $K = \frac{N}{n}$, siendo *N*= estimación de total de personas en la onda y *n*= total muestral de personas en la onda.

Para base Total EPH

Cuando se trabaja con el total de los aglomerados, base total, se aplica una metodología basada en la Función de Varianza Generalizada, consiste en calcular los desvíos para una gran cantidad de estimaciones y luego encontrar un modelo que permita al usuario calcular el desvío de cualquier otra estimación¹⁰.

En el caso del cálculo de desvíos de estimaciones de un subtotal poblacional, el modelo es:

⁹ La Tabla se puede ver en Cuadro 3Anx.

¹⁰ INDEC (2002b), pág. 22.

$CV^2 = a + b * \frac{1}{\hat{x}}$, donde \hat{x} es el subtotal poblacional estimado y CV^2 es el coeficiente de variación al cuadrado, reemplazando por su formulación $CV^2 = \frac{DS^2}{\hat{x}^2}$, despejando y aplicando raíz cuadrada se llega a:

$DS = \sqrt{a \hat{x} + b \hat{x}}$, los parámetros para la base total toman los valores $a = 0$ y $b = 910,471$. Por lo tanto, la fórmula a aplicar es $DS = \sqrt{910,471 \hat{x}}$.

Cuando se trata de la estimación de un cociente de dos subtotales poblacionales (indicadores expresados como tasas), el modelo resultante es:

$DS = \sqrt{\frac{a * \hat{t}(1-\hat{t})}{\hat{y}}}$, donde $\hat{t} = \frac{\hat{x}}{\hat{y}}$; \hat{x} e \hat{y} son subtotales poblacionales. El parámetro $a = 870,987$.

En ambos casos una vez obtenidos los DS de las estimaciones se pueden calcular los intervalos de confianza.

Se resalta que en el caso de trabajar con el total de la muestra, los modelos permiten calcular los DS de estimaciones puntuales y cocientes poblacionales pero no de promedios.

3.2 Expansión Muestral

La EPH posee una variable denominada PONDERA¹¹, representa los factores de expansión a utilizar al realizar estimaciones a partir de la muestra.

La variable de expansión está compuesta por tres variables, las cuales se describen a continuación:

- 1) Inversa de la probabilidad con que fue seleccionada cada vivienda en la muestra, F .
- 2) Corrector por no respuesta a nivel de estrato, C_h . En su construcción se emplea información del cuestionario familiar sobre las causas por las que no pudo realizarse la entrevista: Ausencia, rechazo u otras caracterizan a una *vivienda encuestable* (ve); Deshabitada, demolida, en construcción, fin de semana, vivienda usada como establecimiento, problema de listado, identifican a una *vivienda no encuestable* (vne). Usando esa información se calcula C_h como:

$$C_h = \frac{(ve + vr)_{estrato h}}{vr_{estrato h}}$$

total de viviendas a encuestar se extraen vne , se tiene la suma de ve y vr .

Usando estos dos primeros componentes se calcula para cada hogar (Base hogar) e individuo (Base personas) w_i , por lo tanto, $w_i = F * C_h$ donde i es hogar o individuo.

¹¹ El Cuadro 4Ap muestra la ubicación de la variable en los distintos formatos de bases. Se respeta el uso de mayúsculas en los nombres de las variables por la EPH. Para el caso de la EPHC ver Sección 6

3) Ajuste demográfico¹², se utilizan proyecciones demográficas de población elaboradas por INDEC a nivel de total de población de aglomerados, se construye

$D = \frac{P}{\hat{P}_{eph}}$, donde P = proyección demográfica de población para el aglomerado y

$\hat{P}_{eph} = \sum_1^{n_r} w_i$ es la estimación de la población del aglomerado a partir de la muestra y

n_r es el total de registros de la base personas.

De este modo, finalmente se llega a que la variable de expansión es el producto de los tres componentes descriptos, por lo tanto:

$$\text{PONDERA} = F * C_h * D$$

El uso de la variable de ponderación muestral es recomendable para el cálculo de los estimadores estadísticos descriptivos, sin embargo, en el caso de estimaciones econométricas existen opiniones contrapuestas acerca del beneficio de utilizar o no dichas variables¹³.

4. Cambios en los diseños muestrales de la EPH

El cambio fundamental en el diseño muestral de la EPH es el que se produce durante el año 2003 y es por el cual se debe distinguir entre EPHP y EPHC¹⁴.

En esta parte del capítulo se señalarán las modificaciones de muestra que tuvieron lugar durante los años 1990-mayo 2003, por lo tanto siempre se considera el contexto de EPHP. En este periodo se cambiaron los diseños y tamaños de muestra en distintos aglomerados, por lo que se considera importante brindar información agregada sobre estos temas, objetivo desarrollado en los siguientes apartados.

4.1 Cambios en los diseños muestrales

En el Cuadro 4Anx se señalan aquellos aglomerados que sufrieron modificaciones en su diseño muestral y la fecha en que se produjeron.

Para los aglomerados que sufrieron cambio total, generalmente el cambio viene acompañado con una modificación en la codificación usada para matching, CODUSU¹⁵. Este es el caso de Neuquén, que al incorporar la localidad de Plottier en 1995 comienza con una nueva codificación para dicha variable.

Tal situación también es afrontada por Santa Rosa-Toay y La Rioja en 1998; Formosa y Tierra del Fuego en 1999; Gran San Juan y San Luis-El Chorrillo en 2000, en todos los casos las publicaciones con novedades editadas por INDEC advierten sobre la nueva codificación.

¹² El ajuste demográfico se efectúa en Gran Buenos Aires desde Mayo de 1991 y en los aglomerados del interior se inició desde la segunda onda, Octubre de 1995.

¹³ (Cfr.) Korn y Graubard (1999), pp. 180-81.

¹⁴ En relación al diseño muestral de la EPHC, la sección 6 trata el tema y la sección inmediata anterior lo hace respecto a la EPHP.

¹⁵ Sin embargo, la variable, recién se incorpora en el diseño de base BUA, consultar el Cuadro 4Ap.

4.2 Ampliación de tamaño muestral de Gran Buenos Aires (GBA)

A partir de mayo de 1998, se realiza una ampliación del área de relevamiento en el aglomerado GBA, se incorporan a la muestra, áreas urbanas de los partidos de Pilar; General Rodríguez; Escobar; San Vicente; Presidente Perón, Cañuelas y Marcos Paz; incluidos en el aglomerado definido por el Censo de Población 1991.

Por tal motivo, INDEC (1998a) advierte que pueden existir problemas de comparabilidad entre la muestra del aglomerado relevada hasta octubre de 1997 y la nueva. Así, la base de datos de GBA incorpora una variable que permite identificar si las personas u hogares corresponden o no a estas nuevas áreas, AREASNUE¹⁶.

Por lo tanto, teniendo en cuenta esta modificación, se establece la validez de las estimaciones si corresponden a las siguientes áreas geográficas:

GBA (incluyendo las nuevas áreas)

GBA (sin incluir las nuevas áreas, es decir el dominio relevado hasta octubre de 1997)

Partidos (incluyendo nuevas áreas)

Partidos (sin incluir las nuevas áreas)

Ciudad de Buenos Aires (ex Capital Federal y no afectada por esta extensión geográfica)

4.3 Reducción de tamaño muestral

En el Cuadro 5Anx se presentan los aglomerados que resultaron afectados por reducción muestral como también el porcentaje de reducción durante el periodo 1990-mayo 2003.

La reducción al igual que el cambio en el diseño muestral produce alteraciones en el uso de la variable matching, aunque en este caso sólo disminuye el porcentaje de población que se puede emparejar entre las distintas ondas.

Así por ejemplo, la reducción que sufrió el aglomerado de La Plata en 1999 se debió a fin de lograr un adecuado manejo de campo y esta alteración produce que entre octubre 1998 y mayo de 1999 sólo se pueda emparejar el 37% de la población mientras que con la onda octubre 1999 nuevamente se podrá tener el 75% de la población emparejada.

Los aglomerados de San Luis-El Chorrillo y Posadas se encuentran en igual situación en mayo de 2000, salvo que el primero afronta un cambio en el diseño muestral en la onda inmediata posterior.

5. La disponibilidad de la EPH puntual en soporte informático

El formato en que se presenta la EPH en soporte informático durante el periodo Mayo1990-2003 no es del todo homogéneo. Así, la información muestral de la Encuesta está disponible en: Planilla de vuelco preliminar (R2), Base Usuarios/Usuaría (BU), Base de Trabajo (BT) y Base Usuaría Ampliada (BUA).

En los Cuadro 1Ap, 2Ap y 3Ap, se detalla el tipo de base con que se cuenta en cada instante del periodo bajo análisis y en el Cuadro 4Ap, se muestra en forma resumida la estructura de cada una de las bases, por lo tanto, sólo se exponen las variables más importantes.

A continuación, se sintetiza la información acerca de los tipos de bases y se detalla sobre algunas de las variables consideradas relevantes para el estudio de la diferenciación salarial y de la distribución del ingreso.

¹⁶ Sobre la localización de la variable en la base de datos BUA se puede consultar el Cuadro 4Ap.

R2

Es una planilla que se llena con antelación al análisis y codificación de los cuestionarios y se lleva a cabo inmediatamente después de finalizada la Encuesta, se la conoce como “Planilla de vuelco preliminar”.

Esta base cuenta con información a nivel de hogar e individual en un mismo archivo y comparte muchas variables con la BU.

Además de las variables expuestas en el Cuadro 4Ap, hay otras variables que permiten conocer algo más sobre un posible estado de subocupación o de desempleo. Sin embargo, el principal problema surge respecto a las variables de ingreso¹⁷, sólo se cuenta una variable que representa el total de ingresos percibidos, MONTO.

Respecto a las fuentes de ingreso, no existe manera de determinar cuales son las fuentes que contribuyen con los ingresos del individuo, ya que no existe una distinción entre fuente laboral y otras fuentes.¹⁸

BU

Más detallada que la anterior, esta base cuenta con dos archivos de información, uno a nivel hogar y el otro individual. Como se observa en el Cuadro 2Ap, las variables se exponen siguiendo una estructura de bloques.

Se debe señalar que en esta base se dispone de dos variables adicionales que no están contenidas en R2: Una releva información que permite caracterizar la actividad laboral desarrollada por los individuos, TAREA, y otra permite determinar distintas combinaciones posibles de fuentes de ingreso, FUENTE.

BT

Se trata de una base más completa que la BU y junto a la BUA son las que guardan más semejanzas en cuanto a su estructura. Si bien se mantiene la presentación de variables haciendo referencia a individuos y hogares en archivos separados, existe alguna diferencia respecto a las variables contenidas en cada bloque de información.

BUA

A partir de Mayo de 1998¹⁹, el equipo técnico encargado de la EPH resuelve implementar la Base Usuaría Ampliada, la cual presenta información con un mayor nivel de desagregación, por ejemplo, en cuanto a la información de horas trabajadas, la BUA muestra datos de horas extras y totales dedicadas a la ocupación principal, secundaria y otras ocupaciones. También se cuenta con variables construidas como por ejemplo RAMA, FUENTE, contenidas en BU y BT. Por otro lado, a fin de caracterizar la ocupación que realiza el individuo, la BUA ya no presenta la variable TAREA, la cual es reemplazada por la variable P20²⁰.

Una incorporación muy importante en cuanto a la *identificación* del individuo y del hogar, es la de la variable CODUSU, número de identificación único dentro de cada aglomerado. A cada hogar se le asigna este código y a todos los individuos de un mismo hogar le corresponde el mismo, la identificación personal se puede realizar con la consideración auxiliar del número de componente del hogar, COMPONENTE. Esta

¹⁷ Siempre teniendo en cuenta el objetivo general que es la guía de este capítulo.

¹⁸ En el siguiente punto de esta sección se puede consultar sobre la codificación de algunas variables relevantes en R2 y en los otros formatos de bases, especialmente sobre las variables de ingreso.

¹⁹ Una versión aproximada de la BUA (BUAa) se presenta posteriormente para ondas anteriores a Mayo 1998, por pedido de los usuarios, con el objeto de contar con información más homogénea entre ondas. En las BUAa se cuenta con la variable CODUSU, por lo tanto es posible realizar matching.

²⁰ A fin de realizar comparaciones en cuanto a la codificación de esta y otras variables, consultar la siguiente sección.

incorporación permite relacionar un hogar y sus integrantes a lo largo de cuatro ondas de la EPH (matching), accionar imposible de realizar para toda base distinta de la BUA.

5.1 Codificación de variables relevantes

Se ha observado que el formato de la EPH tradicional fue en evolución y ciertas diferencias ya han sido remarcadas. En este apartado, se hace referencia especial a los cambios en nombres de variables y codificaciones de aspectos que ayudan a caracterizar la actividad laboral de los individuos y sus ingresos.

En el Cuadro 5Ap se presenta la codificación sobre categoría ocupacional; rama de actividad; caracterización de la tarea u ocupación e ingresos de las personas en cada uno de los formatos anteriormente especificados.

Categoría Ocupacional

Quien se clasifica como “patrón o empleador”; “trabajador por su cuenta”; “obrero o empleado” o “trabajador sin salario” es el entrevistado, se observa que a pesar que esta variable cambió dos veces de nomenclatura ha conservado su codificación.

Rama de Actividad

Esta variable modifica su nombre y codificación al pasar de las bases R2 a BU, sin embargo, en las bases BT y BUA se incorpora una nueva variable, para clasificar la rama del establecimiento en el cual trabaja o trabajó el individuo, P18, cuya codificación corresponde a la establecida por CIU (Clasificación Internacional Industrial Uniforme), Revisión 2 y 3²¹.

Caracterización de la tarea u ocupación

La variable que permite caracterizar la actividad ocupacional de las personas, TAREA, se incorpora a partir de la BU, esta variable contenida también en BT siempre tendrá igual codificación. Se debe resaltar que tanto en BT como BUA se cuenta con otra variable para tener en cuenta la actividad ocupacional, P20.

La codificación de P20 en BT sigue el Clasificador de Ocupaciones de la EPH, el cual consiste en un código bidimensional, que tiene en cuenta la *finalidad* del trabajo (carácter) y la *complejidad* de la tarea (calificación). En tanto que para BUA, la codificación, consiste en un código tridimensional, que tiene en cuenta el Carácter general de la tarea; el Carácter específico (*Jerarquía y Tecnología*) y la Calificación (*complejidad*), esto queda establecido por el Clasificador Nacional de Ocupaciones. En el Cuadro 6Ap se muestra la codificación de esta variable.

Variables de ingresos

Respecto a las variables captadoras de los datos de ingreso se observa una constante evolución, mientras en R2 se cuenta con el monto total del ingreso percibido MONTO, en BU se relevan dos variables, expresadas en unidades monetarias por hora, INGHORA, que representa el ingreso horario de la ocupación principal y P47HOR, ingreso total horario.

Mientras que en R2 no se dispone de información sobre la fuente generadora de ingresos, en BU ya se cuenta con la variable FUENTE, la cual sintetiza un conjunto de combinaciones posibles de las mismas

²¹ En INDEC (1995) se recomienda tener en cuenta la fecha de cambio en la clasificación de la variable P18, ya que en algunos aglomerados se realiza mediante CIU-Rev.2 y en otros con CIU-Rev.3 en una misma onda.

Las bases BT y BUA presentan con mayor detalle la conformación de los ingresos personales, en el Cuadro 1 se realiza una descripción resumida de las variables de ingreso en estas bases.

Cuadro 1: Variable y fuente de Ingreso BT-BUA

Variable		BT (hasta octubre 1994)	BT (a partir de Mayo 1995) y BUA
		Descripción: Ingreso obtenido como/por:	
<u>Fuentes laborales</u>			
P471	P47_1	Obrero o empleado	Asalariado
	P47_2		Bonificación o gratificaciones no habituales de los asalariados
P472	P47_3	Trabajador por cuenta propia	Cuenta Propia
P473	P47_4	Utilidades y beneficios de Patrón	Ganancia de patrón (incluye sueldo asignado)
<u>Otras fuentes</u>			
P475	P48_1	Jubilación o Pensión	
P474	P48_2	Alquileres, rentas, intereses y dividendos	Alquileres, rentas o intereses
	P48_3		Utilidades, beneficios o dividendos
	P48_4		Seguro de desempleo
	P48_5		Indemnización por despido
	P48_6		Beca de estudio
	P48_7		Cuota de alimentos
	P48_8		Aporte de Personas que no viven en el hogar
P476	P48_9	Otros ingresos	
P47T	P47T	Monto Ingresos Totales	

Fuente: Elaboración propia en base a INDEC (s.f) y Ocaranza et al (s.f)

Por lo tanto, es posible contar con la desagregación por fuentes de ingresos siempre y cuando las bases sean BT o BUA y el individuo realice la declaración pertinente.

En el cuadro se puede ver, que a partir de Mayo 1995, la BT presenta una nueva tipificación en la desagregación de ingresos por las fuentes que le han dado origen, dicha desagregación se mantiene en la BUA.

6. Acerca de la EPH continua: Breves consideraciones

A partir de enero 2003 el Programa EPH comienza una etapa de modificaciones, las reformulaciones abarcan tres grandes aspectos *temáticos*, *muestrales* y *organizativos*:

Temáticos: Rediseño de cuestionarios para reflejar más adecuadamente el mercado laboral.

Muestrales: Rediseño de muestra para la captación continua de la información que permite mayor frecuencia de presentación de resultados.

Organizativos: Ajuste de la modalidad de trabajo y procesos informáticos a los diseños temáticos y muestrales establecidos.

Aquí se hace referencia a los dos primeros, destacando los principales cambios²².

²² (Cfr.) INDEC (2003).

Cuestionario rediseñado

Los cambios estuvieron orientados a la reformulación de las preguntas del cuestionario individual, se pasó de preguntas breves de lectura textual a otro que cuenta con preguntas auxiliares, con mayor autoexplicación que garantiza una mayor homogeneidad.

A la EPH tradicional la conformaban dos cuestionarios, uno para la vivienda y el hogar y otro individual que se aplicaba a toda la población sin límite de edad, la EPHC consta de tres cuestionarios, uno para la vivienda, otro para el hogar y el individual para cada una de las personas de 10 años y más que constituyen el hogar.

Las reformas realizadas tienen como objetivo captar formas ocultas de ocupación y de desempleo como también modalidades especiales de empleo (planes de empleo, pasantías) y mayor información sobre las condiciones de trabajo. Así mismo, se incorporaron variables sobre cobertura médica, división familiar de tareas domésticas y se modificaron algunas sobre migración.

En resumen, las modificaciones están orientadas a producir una captación más adecuada de los rasgos estructurales de la realidad económico-social y su dinámica.

Captación continua de la información

La EPHP se relevaba dos veces por año, en la tercera semana de los meses de mayo y octubre, las cuales constituían la primera y segunda onda respectivamente. Bajo la nueva modalidad, la muestra está distribuida a lo largo de cada uno de los cuatro trimestres del año, por lo tanto se trata de una Encuesta “continua” que permite dar estimaciones trimestrales, semestrales y anuales.

El muestreo mantiene su característica de bietápico y se desarrolla de igual manera que para la EPH tradicional, sin embargo, la gran diferencia es que las áreas seleccionadas se distribuyen a lo largo de las 12 semanas del trimestre, por lo tanto cada área tiene asignada una semana de referencia dentro del trimestre, la semana de referencia se mantiene de trimestre a trimestre y de año a año.

Actualmente el tamaño muestral por trimestre es de 17.000 hogares pertenecientes a los aglomerados relevados por la antigua EPH, el objetivo es llegar a relevar 25.000 hogares por trimestre y 100.000 por año una vez que se pueda realizar una ampliación geográfica.²³

La asignación de la muestra dentro de cada aglomerado se realizó tratando de que cada vivienda tenga la misma probabilidad de ser seleccionada. Una restricción importante que se respetó, es un tamaño mínimo de muestra para los aglomerados más pequeños, de manera que la agregación de la muestra permita dar estimaciones semestrales con niveles de precisión similares a la EPH puntual.

Por otro lado, para aquellos casos donde el tamaño de la muestra continua resultó menor al de la puntual, se decidió mantener la cantidad de áreas y se redujo la cantidad de viviendas seleccionadas dentro de cada una de ellas.

Un concepto nuevo e importante es el dominio de estimación, se trata de cualquier subdivisión de la población acerca de la cual se puede dar información numérica de precisión conocida, en el Cuadro 2 se dan los detalles.

El esquema de rotación de la EPH continua es un poco más complejo que el de la puntual. Las viviendas de un área que ingresan a la muestra son encuestadas en dos trimestres consecutivos, se retiran los siguientes dos trimestres y se vuelven a encuestar en dos trimestres siguientes, en el mes y semana que se les ha asignado.

²³ Actualmente se trabaja en el nuevo Marco Muestral Nacional, por lo tanto las limitaciones de la ampliación de la cobertura geográfica son metodológicas, reformulación del marco muestral, por un lado, pero también existen restricciones presupuestarias, INDEC (2003).

Cuadro 2: Dominios de estimación, EPH continua

Dominios de estimación	Períodos de estimación		
	Trimestre	Semestre	Año
Total de aglomerados agrupados	X	X	X
Conjunto de aglomerados agrupados por regiones estadísticas	X	X	X
Conjunto de aglomerados con 500.000 y más habitantes	X	X	X
Conjunto de aglomerados con menos de 500.000 habitantes	X	X	X
Cada uno de los aglomerados con 500.000 y más habitantes	X	X	X
Cada uno de los aglomerados con menos de 500.000 habitantes		X	X

Fuente: INDEC (2003), pág. 18.

Así, en caso de que el objetivo sea el análisis a lo largo del tiempo, un hogar puede ser seguido a lo largo de un año y medio, sin embargo sólo se observa el hogar en los dos trimestres del año que corresponda realizar la entrevista y no se tiene un seguimiento continuo del hogar a lo largo del año y medio.

Respecto a la expansión muestral, las bases correspondientes a la EPH continua, además de contar con la variable PONDERA²⁴, tienen nuevas variables de ponderación, las cuales incorporan una corrección por no respuesta de ingresos.

Dicha corrección se realiza asignando a los no respondientes el comportamiento de los respondientes por estrato de la muestra. Por lo tanto, para el tratamiento de ingresos y pobreza, se deben tener en cuenta los ponderadores signados en el Cuadro 3.

Cuadro 3: Nuevos ponderadores de la EPH continua

Aplicable al tratamiento del:	Variable de ponderación
Ingreso total individual	PONDII
Ingreso de la ocupación principal	PONDIO
Ingresos total familiar y per cápita familiar de los hogares	PONDIH
Ingresos total familiar y per cápita familiar de los individuos	PONDIHP

Fuente: Elaboración propia en base a INDEC (2004)

Así, INDEC resalta que se pueden realizar estimaciones con los ponderadores tradicionales, implementados desde la EPH puntual, conociendo que en este caso se omite de corregir la no respuesta o a partir de la EPH continua, se pueden implementar los nuevos ponderadores.

En la EPH continua, tanto los deciles de ingresos calculados en la base como la clasificación de la condición de pobreza e indigencia, se realizan con los ponderadores que incluyen la corrección por no respuesta de ingresos.

7. Comentarios Finales y Aplicaciones

En este capítulo se hizo un resumen de las características de la EPH atendiendo principalmente a las variables de ingreso y otras que contribuyen al objetivo general de este trabajo: Indagar sobre el efecto de la diferenciación salarial sobre la distribución del ingreso.

La observación en profundidad de la disponibilidad de formatos de la Encuesta, conlleva a reconsiderar los ejes temporal y espacial bajo análisis. Así, se elige trabajar

²⁴ Variable de expansión de la EPH puntual, aspectos sobre la misma se presentan en la Sección 3.2 de este capítulo.

con las bases BT, para el periodo comprendido entre mayo de 1990 y octubre de 1994 y las bases BUA, mayo 1995-2003, ya que resultan las más apropiadas para alcanzar los objetivos planteados.

La división temporal obedece fundamentalmente a la incorporación de la variable CODUSU en la BUA, la que permite la realización de matching para la confección de paneles de datos.

Los Cuadros 7Ap y 8Ap presentan la información de la continuidad en la disponibilidad de datos en el periodo bajo estudio.

Por un lado, el análisis puntual, se propone la realización de ocho cortes transversales, mayo 1990; mayo 1992; octubre 1994 tomando como fuente de datos la BT y la primera o segunda onda de 1995; 1997; 1999; 2001 y la primera de 2003 considerando la BUA.

Para el análisis temporal del periodo mayo 1990-octubre 1994, solo podría analizarse con Paneles de cohortes. Capital Federal y Partidos son los únicos aglomerados para los cuales se dispone de información a largo del periodo completo.

En el Cuadro 7Ap se puede observar como se van incorporando las bases BT con el tiempo, es decir que al extender el análisis espacial se pierde en el análisis temporal, por ejemplo entre 1991 y 1992, 5 son los aglomerados del interior que cuentan con la BT, mientras entre 1993-1994 ya llegan a ser 15.

Tomando bajo consideración la información que aporta este capítulo, se pensó trabajar del siguiente modo: analizar el periodo 1990-2003 para el aglomerado Gran Buenos Aires, es decir Capital Federal y Partidos; mientras que para el periodo 1995-2003 se incorporará un análisis de aglomerados del interior sujeto a la adecuada disponibilidad de las bases de datos.

Dos propuestas para el análisis del interior se dejan entrever en los Cuadros 8 y 9Ap, por un lado teniendo en cuenta la preparación de paneles, se podrían conformar 4 entre la primera onda de 1995 y segunda onda de 2002 ó entre la segunda onda de 1995 y primera onda de 2003. Esta última alternativa sería más apropiada, si se tiene en cuenta la fecha desde que empieza a regir el ajuste demográfico para el interior (Octubre 1995) y además se cuenta con un total de 17 aglomerados comunes a lo largo de todo el periodo, sin embargo no se podría incorporar el aglomerado de Santa Fe y en ninguna de las dos opciones se cuenta con el aglomerado de Gran Córdoba, el cual representa el 4,1% de la población urbana argentina.

En resumen, las propuestas son:

Paneles	Propuesta 1		Propuesta 2	
	Periodo	Nº de aglomerados	Periodo	Nº de aglomerados
I	1995-1/1996-2	18	1995-2/1997-1	26
II	1997-1/1998-2	23	1997-2/1999-1	21
III	1999-1/2000-2	25	1999-2/2001-1	25
IV	2001-1/2002_2	27	2001-2/2003_1	26
Aglomerados comunes		13		17

Seleccionar una de las alternativas propuestas para el interior es de suma importancia y se debe tener en cuenta que están en juego dos aspectos relevantes, incorporar un mayor número de aglomerados significará trabajar con una muestra más representativa del periodo, sin embargo, se pierde comparabilidad entre los distintos periodos.

El conjunto de aglomerados seleccionados para el periodo 1990-2003, será objeto del análisis estadístico descriptivo del siguiente capítulo, principalmente nos enfocaremos en el mercado laboral.

Referencias:

- GASPARINI, L. (2007): Monitoring the Socio-Economic Conditions in Argentina 1992-2006. CEDLAS. Universidad Nacional de la Plata.
- INDEC (s.f): Encuesta Permanente de Hogares. Diseño de Registro y Estructura para las bases de trabajo (BT) Hogar y Personas. Dirección de Encuesta Permanente de Hogares. Argentina.
- INDEC (1995): Encuesta Permanente de Hogares. Base Usuaría Ampliada (BUA) Ondas Octubre y Mayo '95.
- INDEC (1998a): EPH. Novedades de la Onda. Agosto y Octubre 1998. Argentina.
- INDEC (1998b): Clasificador Nacional de Ocupaciones. Revisión Mayo 1998
- INDEC (1999): EPH. Novedades de la Onda. Mayo 1999. Argentina.
- INDEC (2000): EPH. Novedades de la Onda. Mayo y Octubre 2000. Argentina.
- INDEC (2002a): EPH. Novedades de la Onda. Octubre 2002. Argentina.
- INDEC (2002b): Base Usuaría Ampliada de E. P. H (BUA). Octubre 2002 (Versión para Bases de Aglomerado y Base del Total EPH).
- INDEC (2003): La Nueva Encuesta Permanente de Hogares de Argentina.2003 en <http://www.indec.mecon.gov.ar>.
- INDEC (2004): Encuesta Permanente de Hogares Continua. Diseño de Registro y Estructura para las bases preliminares. Segundo semestre 2003.
- INDEC- LAUPHAN, W. (2000): Utilización de clasificadores de ocupación en las bases de datos de la Encuesta Permanente de Hogares. Coord. Cynthia Pok.
- KORN, E. L. y B. I. GRAUBARD (1999): Analysis of Health Surveys. Wiley Series in Probability and Statistics. USA
- OCARANZA, A., IGLESIAS, C. y CARDINALI, A. (s.f): Los ingresos en la Encuesta Permanente d Hogares. Dirección de Ocupación e Ingresos. Subsecretaría de Programación Macroeconómica. Ministerio de Economía.