

## “La EPH en los ‘90: una mirada desde el usuario” \*

Matías D. Cattaneo\*\*

### I. Introducción

---

*“All of these sources of information have, of course, various kinds of error for which allowance should be made, if only the source and extent of the error were known.”*

*Oskar Morgenstern, 1963.*

La información, tanto cuantitativa como cualitativa, juega un rol central en cualquier tipo de investigación. Si bien su participación resulta indispensable en todas las etapas que conforman el cuerpo de un proyecto, su mayor importancia relativa siempre ha recaído en la etapa de contrastación empírica. Lógicamente, la utilización de la información ha estado sujeta a fuertes debates en torno a aspectos de índole estadística, epistemológica o teórica, siendo, en general, muy difícil satisfacer simultáneamente los distintos estándares impuestos. Es habitual encontrar situaciones en las cuales, dentro del marco de un proyecto de investigación, la información a ser utilizada no respeta ciertos estándares estadísticos o no capta exactamente el suceso que la teoría intenta explicar o, quizás, no posee el sustento necesario para su utilización.

En las ciencias sociales, a diferencia de las ciencias naturales, la situación resulta más complicada aún, pues además de los problemas antes sintetizados, nos encontramos con uno mucho mayor: **la predominante disociación entre productores y usuarios de dicha información** [Morgenstern, 1963]. Este hecho no es menor, ya que ha sido el causante de grandes inconvenientes y retrasos –cuando no aboliciones– de proyectos que en su cuerpo formal resultaban sumamente atractivos [Wainerman y Sautu, 1997]. Dicho de otro modo, mientras que otros argumentos que han sido esgrimidos en relación a las diferencias fundamentales entre estos dos tipos de ciencia han podido ser relativizados [Nagel, 1961], tal característica intrínseca de las ciencias sociales continúa en el centro del debate.

Cabe entonces preguntarse cuán grave puede llegar a ser este inconveniente. Como punto de partida para responder este interrogante, podemos destacar dos de las principales causas de la existencia de la disociación antes mencionada:

1. *Estudios no controlados*. Al no existir la posibilidad de realizar experimentos controlados o de “laboratorio”, surge la necesidad de que alguien releve sistemáticamente la información más

---

\* El presente trabajo es una versión revisada y ampliada de la tesina de licenciatura del autor, bajo la dirección del Lic. Javier Lindenboim.

\*\* Licenciado en Economía (UBA). Docente y asistente de investigación, FCE, UBA. E-mail: mdcattan@econ.uba.ar

importante aún cuando no exista una demanda real; dicho de otra forma, al tratarse de sucesos sociales –donde el tiempo juega un rol central, ya que éstos son únicos e irrepetibles– es imperativo que los mismos sean relevados periódicamente más allá del interés de uno u otro investigador en particular. Esto da lugar a la existencia de instituciones tanto públicas como privadas que se encargan principalmente de su relevamiento.

2. *Costo*. En el campo de las ciencias sociales, el obtener datos exige realizar una serie de procesos que resultan en general extremadamente costosos, tanto por el tamaño de la muestra como por cuestiones geográficas, institucionales o sociales, siendo justamente éstos los principales condicionantes que explican la disociación antes mencionada.

Existen también al menos tres consecuencias claramente individualizables, donde es el usuario de la información quien se ve más afectado:

1. *Generalidad*. Como no es posible conocer con exactitud cuál será el uso que se le dará a la información, los productores se proponen captar sucesos “de amplio espectro”, al tiempo que intentan minimizar costos y la duplicación de esfuerzos. Con el avance de la ciencia y el afán de estudio más sistemático y preciso, muchas veces las fuentes actuales comienzan a quedar insuficientes u obsoletas.
2. *Asimetría de información*. Son sumamente extraños los casos en donde los usuarios cuentan con todo el conocimiento, tanto teórico como empírico en relación a la producción de las fuentes a estudiar. En general esta información se les presenta en forma discontinua y separada. Desde la teoría económica esto puede ser entendido como un problema de principal-agente, donde los usuarios delegan en los productores la tarea de recopilación de información, pero al no existir mecanismos de control –dado que los productores son claramente autónomos– muchas veces su utilización se hace sumamente difícil.
3. *Sesgos*. Existen innumerables distorsiones o sesgos (reales o potenciales): algunos de índole técnica o no controlable que reducen la robustez de la información captada; otros, en cambio, surgen justamente porque el productor intenta captar varios sucesos distintos –por ejemplo, orden, tiempo, formato, etc.–, alterando el resultado final e impidiendo la utilización parcial o total de la información.

Ahora bien, en este trabajo no nos centraremos puntualmente en el estudio sistemático de las causas y consecuencias relacionadas al uso (y abuso) de la información en las ciencias sociales. En cambio, intentaremos presentar, a modo de estudio de caso, un sucinto análisis –desde la perspectiva del usuario y para la década del noventa– de una de las fuentes secundarias más utilizadas en la República Argentina: la Encuesta Permanente a Hogares (EPH). Si bien existen, al menos, tres niveles de análisis en este sentido –metodología, cuestionario y bases–, en el presente nos focalizaremos principalmente en el estudio de las bases informáticas, las cuales compendian gran parte de la información disponible que entrega su productor, el Instituto de Estadística y Censos (INDEC), a los usuarios. En consecuencia, aquí sólo nos referiremos a la metodología y/o a los cuestionarios de la encuesta en aquellos momentos que resulte indispensable su

abordaje<sup>1</sup>. Debemos remarcar que nuestro interés radica únicamente en el análisis de la provisión de la información y su posterior uso por parte de los usuarios. De este modo, reconocemos que aún cuando algún fenómeno pudo haber sido captado correctamente, puede suceder que los datos entregados por el productor no nos permitan estudiarlo total o parcialmente.

Nuestro objetivo es claro: identificar algunos de los principales problemas a los que se enfrenta todo investigador (social) que intente utilizar esta fuente secundaria a la hora de iniciar algún tipo de investigación. Debe aclararse que de ningún modo nuestro análisis pretende ser exhaustivo; por el contrario, intenta la búsqueda del problema, estableciendo e individualizando algunos de los inconvenientes más comunes dentro del ámbito de la investigación.

Siendo nuestro objetivo encuadrar el estudio, consideramos apropiado como punto de partida, en primer lugar, analizar el marco conceptual de las encuestas de hogares, es decir, cuáles son las motivaciones y limitaciones que encierran las mismas en general; para luego, en segundo lugar, presentar el marco histórico de la EPH, es decir, cuáles fueron y son actualmente sus alcances. También intentaremos resumir el modo en que el usuario se encuentra con la información provista por INDEC, pues creemos que es justamente allí donde se inician los inconvenientes a la hora de su uso. Este desarrollo se encuentra en el Apartado II del presente trabajo.

Como siguiente paso proponemos analizar la estructura que el usuario recibe como fuente (secundaria) básica de información, es decir, intentaremos describir e interpretar los datos provistos. Así, en el Apartado III, se compendian todas las variables incluidas en las principales bases que el INDEC entrega, además de sugerir una metodología de compatibilización de los distintos formatos en que son presentadas. Se intentará mostrar también cuáles son las pérdidas<sup>2</sup> asociadas a dicha compatibilización, tanto cuando ésta es realizada para un año en particular como cuando se aplica para dos o más.

En el Apartado IV nos centramos en el estudio de la variable “pondera” existente en la base informática y que en general ha presentado varios problemas a los usuarios a lo largo del tiempo. Aquí la idea es doble. Por un lado, mostrar al lector un ejemplo de cómo las variables incluidas en las bases son construidas y por qué debe prestarse especial atención tanto a la metodología como a las implicancias de su utilización. En este sentido, intentaremos poner en evidencia por qué la compatibilización en el tiempo no siempre es posible. Por el otro, presentaremos dos métodos (uno de ellos comúnmente utilizado) que intentan solucionar los problemas previamente presentados.

Por último, se ensayan algunas reflexiones en donde la idea general es alertar al lector sobre los problemas en que puede incurrir un investigador toda vez que intente utilizar la EPH como fuente secundaria de información.

Además, se ha incluido un Anexo con todos los datos utilizados a lo largo de este trabajo. Allí podrán encontrarse todos los cuadros oficiales utilizados, además de aquellos que han sido especialmente construidos.

---

<sup>1</sup> Un análisis en esta línea puede ser encontrado en el artículo de Lindenboim y Serino, incluido en este Cuaderno.

<sup>2</sup> A lo largo de este trabajo, nos referiremos a “pérdidas” en un sentido especial. Buscamos simplemente remarcar cuáles son las alteraciones o mutaciones que sufren los datos, dado que éstos han sido procesados y entregados en un determinado formato.

## II. Una conceptualización histórica

---

### Las encuestas de hogares como unidad de análisis

Las encuestas de hogares son un instrumento estadístico inmerso en la problemática social. Existen, al menos, tres ejes temáticos claramente individualizables que merecen ser destacados a la hora de realizar cualquier análisis serio de este herramental: el técnico, el económico y el social. En toda encuesta puede observarse, en mayor o menor medida, la presencia e interacción de estos aspectos. Sin embargo, en la actualidad las encuestas de hogares –en sus diferentes formas– han representado un instrumento de análisis esencialmente de índole económica, siendo en realidad su justificación y prácticamente única motivación la necesidad de medir y captar con cierta periodicidad los cambios en las variables de mayor interés.

En cuanto a su implementación, si bien por un lado existen quienes ven a las encuestas como un reto, donde intentan lograr la mejor inferencia al menor costo, al tiempo que trabajando conjuntamente con especialistas de distintas disciplinas buscan alcanzar ciertos estándares internacionales de medición y precisión; por otro lado, podemos encontrar otros responsables que perciben a las encuestas como simple sustitutos de los restantes tipos de fuentes existentes: los censos y los registros<sup>3</sup>, donde su interés radica en obtener la información necesaria para dar sustento a sus intenciones en el plano político. Este conflicto de intereses es una de las características centrales de cualquier instrumento de medición social.

Muchas veces suele olvidarse que toda encuesta –como parte del sistema de estadísticas existente– posee como motivación primordial reflejar los deseos e intenciones de las personas, con el objeto de asegurar las correctas decisiones en el plano político, económico y social. Ahora bien, un atributo que debe tenerse en cuenta a la hora de relacionar a las encuestas con otras fuentes de información es su carácter cortoplacista, es decir, reflejan una situación dada en un momento puntual del tiempo y frente a ciertas condiciones particulares, siendo justamente esta característica la que exige que las mismas sean realizadas con cierta periodicidad.

Los orígenes de la medición pueden ser rastreados hasta los años treinta del siglo veinte, cuando se dieron a conocer los primeros desarrollos en materia de censos socio-económicos y de las Cuentas Nacionales [ONU, 1993; Müller, 1998; Cardinali, 1998]. En este contexto surgen las primeras encuestas con el objeto de mantener y actualizar los datos o informaciones más relevantes originados desde los instrumentos antes mencionados. Cabe destacar que el nacimiento de toda encuesta refiere a la necesidad de medir una o más variables de interés económico y/o social. Sin embargo, son muchos los casos en donde habiendo surgido originalmente con el objetivo de medir un suceso en particular, luego la misma se constituyó en un sistema de alto contenido social, totalmente independiente del instrumento o motivación que le diera origen<sup>4</sup>. Muchas son las posibles intenciones en el desarrollo de las encuestas a hogares. Pueden variar desde la simple medición de distintas poblaciones, hasta el complejo intento de captar los ingresos de distintos grupos o estratos socioeconómicos.

---

<sup>3</sup> Vale la pena destacar las principales características de cada tipo de fuente de información existente: desde el punto de vista temporal, las encuestas y los censos son periódicos, mientras que los registros son continuos; en cambio, desde el punto de vista del relevamiento, las encuestas son muestrales, mientras que los censos y registros son de carácter universal (vale aclarar que en algunas ocasiones los censos también pueden ser de índole muestral, al menos parcialmente).

<sup>4</sup> Por ejemplo, en un comienzo, en casi todo el mundo la medición de los precios estuvo integrada a otras encuestas, mientras que en la actualidad existen complejos sistemas autónomos e independientes.

Durante todo el siglo veinte, hemos asistido a una expansión y profundización del capitalismo como modelo imperante. En este marco, el Estado –como legitimador de dicho modelo– ha cobrado un rol mucho más activo, generando una serie de políticas desde la demanda y la oferta para asistir a los individuos que tienden a quedar excluidos del sistema en forma creciente. De este modo, resulta claro que el rol actual de las encuestas de hogares será, al menos en parte, el de monitorear este sistema. Vale la pena destacar que no siempre las encuestas de hogares han poseído este interés.

Mientras que en el pasado el objetivo de estas encuestas era captar la situación de la fuerza de trabajo en relación al proceso de apropiación y acumulación, en la actualidad, su objetivo es mucho más complejo: debe dar cuenta del triple mecanismo de provisión de servicios, trabajo e ingreso, donde los servicios –muchas veces sumamente difíciles de medir correctamente– son el eje central de la discusión. De este modo, una rápida mirada del siglo nos marca tres intenciones de las encuestas, claramente individualizables y relacionadas entre sí. En los años 30 y 40, proveían esencialmente información macroeconómica; en los años 50, 60 y 70, continuaron generando datos agregados, pero con mucho más contenido social; y en los años 90, se produjo un giro en el contenido tratado, brindando principalmente información de índole microeconómica [Beverley, 1998; Feres, 1998; Barreiros, 1998]. Este último tipo de encuesta de hogares son los más útiles para los *policy-makers* que buscan dimensionar y desarrollar las inversiones sociales, como parte del nuevo rol del Estado. En síntesis, nos encontramos frente a dos tipos completamente distintos:

- ♦ *Económica*. Dirigida a los hogares, con el objeto de detectar los trabajadores, los que quieren trabajar y los que tienen razones para no desearlo. Aquí se intenta medir la relación de estas personas con la actividad económica, formal e informal.
- ♦ *Socio-Política*. También focalizada en los hogares, pero con el objeto de determinar cuánto –el hogar como un todo y sus componentes individualmente– necesitan de esa compleja mercancía llamada “desarrollo” en término de servicios y prestaciones sociales. Este tipo de encuestas son las que comienzan a cobrar mayor importancia.

Como es lógico, estas encuestas de hogares van mucho más lejos que las llamadas encuestas con muestras de propósito múltiple, basando sus diferencias en la utilización de la información provista por cada una de ellas, y no en su diseño técnico [Medina, 1998; Barreiros, 1998].

Si bien las encuestas no deben ser consideradas como un fin en sí mismas, la realidad nos ha mostrado que muchas veces –principalmente en los países en desarrollo– su confección e implementación propone un reto por separado. Son muchos los casos en donde son olvidados sus objetivos principales como, por ejemplo, monitorear los aumentos en los niveles de vida o identificar las consecuencias derivadas de las políticas planeadas o implementadas. Parecería obvio que una estrategia basada en la inversión social necesita que se encuesten aquellos grupos sociales que efectivamente recibirían los beneficios –esta información sería muy útil tanto en los estudios de prefactibilidad como en la etapa de evaluación del proyecto–; sin embargo, en reiteradas ocasiones las encuestas son generalizadas y con un relativo bajo nivel de captación de grupos particulares<sup>5</sup>. Además, deberían definirse claramente nociones como “progreso” o “riqueza” a la

---

<sup>5</sup> Muchas veces las fallas son meramente estadísticas, en el sentido que al descomponer la encuesta se pierde toda validez en la inferencia.

hora de su preparación, ya que lógicamente estos términos no representan lo mismo para todos los individuos.

Finalmente, dentro de los aspectos técnicos, podemos identificar una serie de sesgos comunes a todas las encuestas. Los sesgos instrumentales relacionados con el cuestionario pueden ser de *punto de partida*, cuando el encuestado es inducido desde la misma pregunta; de *información*, cuando la persona desconoce cierta información relevante; del *entrevistador*, cuando éste, con o sin intención, condiciona la respuesta del encuestado; de *orden*, cuando la forma del cuestionario influye sobre el encuestado. Existen también sesgos no instrumentales, por ejemplo, *el estratégico*, cuando el encuestado busca su propio interés en la respuesta suministrada [Azqueta Oyarzún, 1994].

Otro aspecto interesante de las encuestas de hogares es su inserción a nivel internacional. Por un lado, se supone que éstas deberían ser lo más objetivas y comparables posible. Sin embargo, la realidad en los países en desarrollo parece ser otra. Aún resulta sumamente difícil compatibilizar las encuestas de distintos países de Latinoamérica. Por tal motivo, ese trabajo ha recaído en organismos de alcance internacional, como por ejemplo la CEPAL o la OIT [Feres, 1998]. Por otro lado, la necesidad de “internalización” de los datos originados desde las encuestas de hogares han exigido a sus productores una mayor transparencia y rigurosidad en la producción de los mismos, lo que obligó –y aún obliga– a la realización de constantes mejoras en su estructuración e implementación.

Por último, las encuestas son por definición de carácter muestral y periódico. Sin embargo, los países difieren en forma significativa en su implementación en el plano temporal. En la actualidad pueden hallarse ambos extremos: por un lado, se encuentra la Argentina –como se explica en detalle más abajo– donde la encuesta es periódica<sup>6</sup>; y por el otro se encuentra Uruguay, por ejemplo, donde la encuesta es continua, es decir, no implica ningún período de referencia. Como es obvio, estas discrepancias no hacen más que alejarnos de nuestro objetivo de lograr que los resultados de las encuestas sean más comparables entre los países.

## **La Encuesta Permanente de Hogares en Argentina**

En nuestro país, luego del Censo de Población y Vivienda de 1970, surge la necesidad de continuar con un programa de relevamiento periódico<sup>7</sup>. En consecuencia, la Encuesta Permanente de Hogares nace en 1972 como un programa nacional e intercensal que, desarrollado y coordinado por el INDEC –conjuntamente con las Direcciones Provinciales de Estadística (DPE) desde 1974–, intenta relevar una serie de aglomerados urbanos de todo el país con el objeto de obtener información socioeconómica actualizada, siendo su principal interés el de caracterizar al mercado laboral. De este modo, la EPH fue concebida como un instrumento estadístico que, a través del conocimiento de las características sociales y económicas de la población, contribuyera a la formulación y desarrollo de un sistema integrado de indicadores sociales. Su sustento consiste en caracterizar a la población en términos de su inserción socioeconómica, abordando la situación de los individuos y de los hogares, éstos últimos entendidos

---

<sup>6</sup> En la Argentina el período de referencia es una semana calendario, con excepción de los ingresos que refieren a todo el mes. Por otra parte, actualmente se plantea cambiar el régimen, con el objeto de reformular a la encuesta de forma continua, como ocurre en Uruguay o México.

<sup>7</sup> La idea era encontrar a un sucesor de la Encuesta de Empleo y Desempleo (EED), la cual estuvo en vigencia desde mediados de siglo.

como núcleo o vínculo entre los primeros. Si bien el objetivo de esta encuesta es el relevamiento de la vivienda, sus resultados permiten el estudio de los individuos por separado.

Con el fin de ampliar el marco original, la encuesta se ha referenciado globalmente a dos perspectivas analíticas centrales en la investigación social: la estructura social y las estrategias de vida; ambas incluidas en su eje conceptual básico, es decir, la inserción de la población en la producción social de bienes y servicios. A partir de esta premisa, pueden delimitarse dos campos principales de ocurrencia de los fenómenos: el mercado de trabajo y los circuitos de satisfacción de necesidades que están en estricta relación con los grandes ejes conceptuales y las áreas temáticas de la EPH. Así, mientras que en relación al mercado de trabajo la temática es la situación laboral, en relación a la satisfacción de necesidades, las temáticas son la educativa, de vivienda y de ingresos. Además, como se explicitará en párrafos posteriores, a lo largo de la década se fueron incluyendo –y recuperando– otras temáticas de índole más específica como por ejemplo los fenómenos migratorios [INDEC, 1999b].

Actualmente, la EPH se realiza dos veces al año –en mayo y octubre– en 28 aglomerados urbanos<sup>8</sup> que representan el 70% de la población urbana del país y el 98% de la población residente en centros de más de 100.000 habitantes. La captación se realiza a través de dos cuestionarios: uno familiar y otro individual. El primero de ellos permite la recolección de datos que caracterizan al hogar como unidad elemental de análisis, como por ejemplo cuestiones habitacionales, aspectos sanitarios y otros. A través del cuestionario individual, en cambio, se intenta captar las características personales, reduciendo la unidad de análisis a cada miembro del hogar seleccionado.

Como es lógico, cuando el INDEC construye las bases de datos para los usuarios utiliza la información de ambos cuestionarios, pero la separación entre unidades de análisis familiar e individual es preservada sólo en algunos de los tipos de bases disponibles para el usuario.

En relación al diseño muestral, puede destacarse que el plan de muestreo probabilístico es realizado en dos etapas y para cada dominio en particular. En la primera etapa, se seleccionan radios censales (Unidades de Primera Etapa - UPE), con una probabilidad proporcional a su tamaño, medido en cantidad de viviendas ocupadas. Luego, éstas son agrupadas en estratos, utilizando como variable de estratificación el porcentaje de jefes de hogar con educación primaria incompleta. En la segunda etapa, se seleccionan viviendas en forma sistemática, de manera de obtener una muestra autoponderada. Al utilizar un esquema de panel, una vez seleccionada una vivienda, ésta permanecerá cuatro ondas (dos años) en la encuesta. Así, el 75% de la muestra permanece constante entre dos ondas consecutivas.

En términos cuantitativos, la muestra comprende aproximadamente 36.000 viviendas pertenecientes a los principales aglomerados urbanos de todo el país, donde el Gran Buenos Aires (GBA) –Capital Federal más los partidos del conurbano– representa el 12% (4.500 hogares) y el resto urbano oscila entre 800 y 1.500 hogares por aglomerado. Debe destacarse que –con excepción de las características relacionadas con el ingreso– el período de referencia se limita a una semana calendario completa [IASI e INDEC, 1998].

La calidad de la encuesta queda determinada por tres elementos básicos: primero, el tema abordado; segundo, la oportunidad en la cual se encuentren disponibles los resultados; y tercero,

---

<sup>8</sup> Estas características no fueron así desde su implementación. Por ejemplo, existen períodos donde se realizaba la encuestas en tres ondas. También el número de los aglomerados incluidos fue aumentado con el tiempo (en referencia a esto, para la década del noventa, ver el Apartado III).

la precisión de los mismos. Así, la precisión de las estimaciones se explica tanto por el factor estadístico como por el factor humano, lo que claramente genera dos tipos de errores: muestrales y no muestrales. Los primeros son objeto del campo de la estadística y escapan a los objetivos del presente trabajo. Los segundos, en cambio, resultan ser en alguna medida uno de nuestros ejes de análisis. Entre ellos se encuentran los errores de cobertura, de no respuesta, de medición, de captación, de encuestador, de cuestionario, referentes a la forma de recolección de los datos, de codificación y de procesamiento.

En relación a la difusión, el INDEC ha presentado en distintos formatos la información obtenida –y procesada en muchos casos– a través de los cuestionarios: “Informaciones de Prensa”, tabulaciones básicas disponibles en papel, disquetes y bancos de datos propios del organismo, series de indicadores especiales y bases de datos. Estas últimas, que constituyen el eje central de nuestro análisis, son presentadas dentro de un “paquete informático”, donde se incluyen varias informaciones y/o herramientas que intentan asistir al usuario. Puntualmente, en una base tipo de la década del noventa puede encontrarse una serie de directorios donde cada uno de ellos refiere a un aglomerado en particular; dentro de ellos encontramos la base informática en formato “DBF”, una serie de tablas tabuladas con información referente al aglomerado, una tabla de errores muestrales, una tabla de población de referencia y –dependiendo del formato de que se trate– otras bases como por ejemplo la de hogares o anexos con información adicional de las personas encuestadas.

El INDEC, a lo largo de la década del noventa, ha incluido en estos paquetes tres formatos de bases de datos: R2, BU (base usuaria) y BUA (base usuaria ampliada) [INDEC, 1996; INDEC, 1998; INDEC 1999a]. El primero de ellos fue concebido originalmente como transitorio y su única misión era la de proveer información mientras se confeccionaba el definitivo, que se entregaría en el segundo de los formatos (esto en la práctica nunca sucedió). Por tal motivo, es presentado como un único archivo que nuclea tanto variables correspondientes a la encuesta personal como a la encuesta del hogar.

Por otro lado, el formato BU tiene por objetivo “permitir al usuario acceder a la mayor información posible y efectuar sus propios cruces de variables” [INDEC, 1999a], es decir, posee una “apertura” mucho mayor en el sentido de que tiene mayor información por encuestado. Pero además, incluye bases separadas, incorporando una correspondiente a la encuesta personal, otra a la encuesta del hogar y otra llamada “anexo” que aporta información adicional a las dos principales.

A partir de 1998, el INDEC comenzó a distribuir la información proveniente de la encuesta en un nuevo formato: Base Usuaria Ampliada. Esta base, al igual que la BU, también está formada por dos archivos: “hog\_bua.dbf” (que contiene la información de los hogares) y “per\_bua.dbf” (que contiene la información de las personas). Como rasgo diferenciador, se destaca la reincorporación del bloque de preguntas denominado Bloque de Migraciones y la inclusión de muchas más variables, las cuales al combinarse permiten la construcción de categorías antes inalcanzables.

La EPH ha estado sujeta a continuas mejoras, reformulaciones y pruebas. Si bien éstas comienzan desde sus mismos orígenes, dados los objetivos de este trabajo sólo resumiremos aquí las acontecidas en la década del noventa. En 1990, se desarrolló la temática relacionada con la precariedad laboral, incluyendo un módulo que se aplicó en sólo 16 aglomerados de ese año. En 1991, se diseñó un nuevo cuestionario básico. En 1992, se aplicó el nuevo cuestionario básico a una prueba exploratoria de campo. Además, se actualizaron los diseños muestrales utilizando la



información censal de 1991. En 1993, se profundizó la temática relacionada con la desocupación, aplicándose los cambios sólo en principio a los aglomerados de GBA, Córdoba y Rosario. En 1994, se desarrolló una metodología de subdivisión sistemática del área del GBA, lo que permite obtener indicadores separados para cuatro franjas diferenciadas del Conurbano Bonaerense. En 1995, se incorporaron nuevas temáticas al cuestionario básico (distinción del cuestionario por tamaño, condiciones para la obtención del subsidio de desempleo, etc.). Cabe aclarar que en esta oportunidad no se introdujeron cambios en la captación de la condición de actividad en el cuestionario básico habitual. En 1996, se retomó el desarrollo conceptual y el rediseño del cuestionario de la EPH, incluyendo ajustes en lo concerniente a la captación de la condición de actividad y del subempleo. En 1998, se realizó la primera prueba de encuesta telefónica. Por último, debe mencionarse la intención de pasar a un nuevo régimen de periodicidad (continua), el cual se espera sea implementado en los próximos años [IASI e INDEC, 1998].

### III. Las “bases” del problema

---

#### La compatibilización

El INDEC ha tendido a presentar en distintos formatos la información proveniente de la EPH, generando –tales cambios– variados inconvenientes para los usuarios a la hora de ensayar comparaciones, contrastar hipótesis y alcanzar conclusiones. Si bien es indiscutible que este organismo a logrado –y continua logrando– grandes mejoras en torno a las principales niveles de análisis de la EPH antes mencionados (metodología, cuestionarios y bases de datos), las mismas han sido graduales, cobrando un mayor protagonismo en la década de los noventa.

En particular, al analizar el progreso en dicho período de la forma de presentación de la información que los usuarios reciben, encontramos tres etapas claramente individualizables: 1990~1995, donde existen bases del tipo R2 y BU en forma conjunta; 1996~1997, donde sólo existen bases BU; y 1998~1999, donde sólo existen bases del tipo BUA. En consecuencia, y como veremos a continuación, a medida que avanzamos en el tiempo –y por ende en el formato– las distintas alternativas de proceso comienzan a incrementarse, así como el número de variables.

Para adentrarnos en el tema, se presenta el Cuadro 1 donde se identifica cada formato de presentación de la encuesta para cada aglomerado y año, lo que permite observar cómo han variado los formatos de presentación no sólo en el tiempo, sino también hacia adentro de los distintos años. Más precisamente, en el panel A puede verse toda la década del noventa para la onda de mayo, mientras que en el panel B se presenta la misma información para la onda de octubre. En el Cuadro 1, los asteriscos (\*\*\*) corresponden a los aglomerados que no se habían integrado a la EPH para el año en cuestión y la palabra “FALTA” implica que el INDEC no proveyó ninguna información para dicho año.

Puede deducirse fácilmente que un método de compatibilización entre R2, BU y BUA es una **condición necesaria** para cualquier investigación que intente analizar el conjunto de todos los aglomerados relevados por la EPH para la década de los noventa<sup>9</sup>. Más aún, cualquier intento de

---

<sup>9</sup> Quizás esta es una de las razones por las cuales muchos investigadores han optado sólo por utilizar la encuesta en el marco del GBA (donde la base tuvo el mismo formato para toda la década) y luego generalizar las conclusiones

análisis de la encuesta para años anteriores (o eventualmente posteriores) exigiría de un procedimiento equivalente al aquí presentado<sup>10</sup>.

Ahora bien, la compatibilización debe realizarse en dos sentidos. Por un lado, tenemos un procedimiento que podríamos denominar *diacrónico*, donde el objetivo sería empalmar las bases de datos a lo largo del tiempo. Por otro lado, se presenta un segundo procedimiento que podríamos llamar *sincrónico* donde el interés se centraría en unir las bases dentro cada año analizado para obtener una única base completa. Si bien no parecería ser necesario preestablecer un orden entre estos dos procedimientos, pensamos que la forma más sencilla para su implementación sería realizar en primer término el procedimiento sincrónico para luego implementar el diacrónico. El primero implica compatibilizar la forma de presentación con la que se encuentra el usuario, es decir, unir únicamente R2 y BU para los años comprendidos entre 1990 y 1995, como muestra el Cuadro 1. Por otro lado, un empalme diacrónico es por definición más complejo, ya que no sólo exige compatibilizar los datos, sino también implica tomar en cuenta temáticas que escapan a las bases como, por ejemplo, la metodología de captación o la situación estructural en el momento en que se realizó la encuesta.

**CUADRO 1 – Los formatos de las bases provistas por el INDEC para la década de los ‘90.**

Panel A – Onda Mayo.

	<b>AGLOMERADOS</b>	<b>1990</b>	<b>1991</b>	<b>1992</b>	<b>1993</b>	<b>1994</b>	<b>1995</b>	<b>1996</b>	<b>1997</b>	<b>1998</b>	<b>1999</b>
01	G.B.A.	BU(1)	BU	BU	BU	BU	BU	BU	BU	BUA	BUA
02	GRAN LA PLATA	BU	<b>R2</b>	BU	BU	BU	BU	BU	BU	BUA	BUA
03	BAHIA BLANCA	BU	<b>R2</b>	<b>R2</b>	<b>R2</b>	<b>R2</b>	BU	BU	BU	BUA	FALTA
04	GRAN ROSARIO	BU	<b>R2</b>	BU	BU	BU	BU	BU	BU	BUA	BUA
05	STA. FE Y STO. TOME	<b>R2</b>	<b>R2</b>	<b>R2</b>	BU	BU	BU	BU	BU	BUA	BUA
06	PARANA	<b>R2</b>	<b>R2</b>	BU	BU	BU	BU	BU	BU	BUA	BUA
07	POSADAS	BU	<b>R2</b>	BU(1)	<b>R2</b>	<b>R2</b>	<b>R2</b>	BU	BU	BUA	BUA
08	GRAN RESISTENCIA	<b>R2</b>	<b>R2</b>	<b>R2</b>	<b>R2</b>	BU	BU	BU	BU	BUA	BUA
09	COMODORO RIVADAVIA	FALTA	<b>R2</b>	BU	BU	BU	BU	BU	BU	BUA	BUA
10	GRAN MENDOZA	BU	BU	BU	BU	BU	BU	BU	BU	BUA	BUA
12	CORRIENTES	BU	<b>R2</b>	<b>R2</b>	<b>R2</b>	BU	BU	BU	BU	BUA	BUA
13	GRAN CORDOBA	BU	BU	BU	BU	BU	BU	BU	BU	BUA	BUA
14	CONCORDIA	**	**	**	**	**	BU	BU	BU	BUA	BUA
15	FORMOSA	FALTA	<b>R2</b>	<b>R2</b>	<b>R2</b>	<b>R2</b>	BU	BU	BU	BUA	BUA
17	NEUQUEN Y PLOTTIER	BU	BU	BU	BU	BU	BU	BU	BU	BUA	BUA
18	STGO DEL ESTERO Y LA BANDA	FALTA	<b>R2</b>	<b>R2</b>	<b>R2</b>	BU	BU	BU	BU	BUA	BUA
19	SAN S. JUJUY Y PALPALA	BU	BU	BU	BU	BU	BU	BU	BU	BUA	BUA
20	RIO GALEGOS	BU	BU	BU	BU	BU	BU	BU	BU	BUA	BUA
22	GRAN CATAMARCA	BU	<b>R2</b>	<b>R2</b>	BU	<b>R2</b>	BU	BU	BU	BUA	BUA
23	SALTA	BU	BU	BU	BU	BU	BU	BU	BU	BUA	BUA
25	LA RIOJA	<b>R2</b>	<b>R2</b>	<b>R2</b>	<b>R2</b>	BU	<b>R2</b>	BU	BU	BUA	BUA
26	SAN LUIS Y EL CHORRILLO	BU	<b>R2</b>	BU	BU	BU	BU	BU	BU	BUA	BUA
27	GRAN SAN JUAN	BU	<b>R2</b>	BU	<b>R2</b>	BU	BU	BU	BU	BUA	BUA
29	S. M. DE TUCUMAN Y TAFI VIEJO	BU	<b>R2</b>	BU	BU	BU	BU	BU	BU	BUA	BUA
30	SANTA ROSA Y TOAY	BU	BU	BU	BU	BU	BU	BU	BU	BUA	BUA
31	TIERRA DEL FUEGO	FALTA	<b>R2</b>	BU	BU	BU	BU	BU	BU	BUA	BUA
32	Cdad. Buenos Aires	BU	BU	BU	BU	BU	BU	BU	BU	BUA	BUA
33	PARTIDOS	BU	BU	BU	BU	BU	BU	BU	BU	BUA	BUA
34	MAR DEL PLATA	**	**	**	**	**	BU	BU	BU	BUA	BUA
35	ALTO VALLE DEL RIO NEGRO	**	**	**	**	**	**	**	**	BUA	BUA
36	RIO CUARTO	**	**	**	**	**	BU	BU	BU	BUA	BUA
37	INTERIOR DE MENDOZA	**	**	**	**	**	**	**	**	BUA	BUA

(1) Falta anexo.

al resto urbano, más allá de la validez que tienen tal generalizaciones. En relación a esto, ver el artículo de Lindenboim y Serino (2000) incluido en el Cuaderno 4 del CEPED.

<sup>10</sup> Para los años anteriores la dificultad reside en la disponibilidad efectiva de las bases en algún formato compatible.

Panel B – Onda Octubre.

AGLOMERADOS		1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
01	G.B.A.	BU	BU	BU	BU	BU	BU	BU	BU	BUA	BUA
02	GRAN LA PLATA	R2	BU	BU	BU	BU	BU	BU	BU	BUA	BUA
03	BAHIA BLANCA	R2	R2	R2	R2	R2	R2	BU	BU	BUA	FALTA
04	GRAN ROSARIO	R2	R2	BU	BU	BU	BU	BU	BU	BUA	BUA
05	STA. FE Y STO. TOME	R2	R2	BU	BU	BU	BU	BU	BU	BUA	BUA
06	PARANA	R2	R2	BU	BU	BU	BU	BU	BU	BUA	BUA
07	POSADAS	R2	R2	R2	R2	R2	R2	BU	BU	BUA	BUA
08	GRAN RESISTENCIA	R2	R2	R2	R2	BU	BU	BU	BU	BUA	BUA
09	COMODORO RIVADAVIA	FALTA	FALTA	BU	BU	BU	BU	BU	BU	BUA	BUA
10	GRAN MENDOZA	R2	BU	R2	BU	BU	BU	BU	BU	BUA	BUA
12	CORRIENTES	R2	R2	R2	R2	BU	BU	BU	BU	BUA	BUA
13	GRAN CORDOBA	BU	BU	BU	BU	BU	BU	FALTA	BU	BUA	BUA
14	CONCORDIA	**	**	**	**	**	BU	BU	BU	BUA	BUA
15	FORMOSA	R2	R2	R2	R2	R2	R2	BU	BU	BUA	BUA
17	NEUQUEN Y PLOTTIER	BU	BU	BU	BU	BU	BU	BU	BU	BUA	BUA
18	STGO DEL ESTERO Y LA BANDA	R2	R2	BU	R2	BU	BU	BU	BU	BUA	BUA
19	SAN S. JUJUY Y PALPALA	R2	BU	BU	BU	BU	BU	BU	BU	BUA	BUA
20	RIO GALLEGOS	BU	BU	BU	BU	BU	BU	BU	BU	BUA	BUA
22	GRAN CATAMARCA	R2	R2	R2	BU	BU	BU	BU	BU	BUA	BUA
23	SALTA	R2	BU	BU	BU	BU	BU	BU	BU	BUA	BUA
25	LA RIOJA	R2	R2	R2	R2	R2	BU	BU	BU	BUA	BUA
26	SAN LUIS Y EL CHORRILLO	R2	R2	BU	BU	BU	BU	BU	BU	BUA	BUA
27	GRAN SAN JUAN	R2	R2	BU	R2	BU	BU	BU	BU	BUA	BUA
29	S. M. DE TUCUMAN Y TAFI VIEJO	BU	BU	BU	BU	BU	BU	BU	BU	BUA	BUA
30	SANTA ROSA Y TOYA	R2	BU	BU	BU	BU	BU	BU	BU	BUA	BUA
31	TIERRA DEL FUEGO	R2	R2	BU	BU	BU	BU	BU	BU	BUA	BUA
32	Cdad. Buenos Aires	BU	BU	BU	BU	BU	BU	BU	BU	BUA	BUA
33	PARTIDOS	BU	BU	BU	BU	BU	BU	BU	BU	BUA	BUA
34	MAR DEL PLATA	**	**	**	**	**	BU	BU	BU	BUA	BUA
35	ALTO VALLE DEL RIO NEGRO	**	**	**	**	**	**	**	**	BUA	BUA
36	RIO CUARTO	**	**	**	**	**	BU	BU	BU	BUA	BUA
37	INTERIOR DE MENDOZA	**	**	**	**	**	**	**	**	BUA	BUA

FUENTE: Centro de Estudios sobre Población, Empleo y Desarrollo, Instituto de Investigaciones Económicas, UBA.

En este Apartado nos centraremos en la presentación del procedimiento sincrónico, realizando un breve tratamiento del procedimiento diacrónico sólo al final del mismo<sup>11</sup>. Antes de comenzar a describir el proceso de compatibilización propuesto debemos remarcar que a causa de la tecnología<sup>12</sup> todos los nombres de los campos de las bases serán de ocho dígitos o menos, aún cuando en los Cuadros presentados puedan ser más largos. El trabajo de compatibilización será facilitado por la presencia de subdivisiones en los instructivos<sup>13</sup>, donde puede ser encontrada una serie de bloques temáticos y dentro de ellos las variables relevantes. Por último, debemos aclarar que, como criterio metodológico, la compatibilización se realizará entorno a la base BU. Esta decisión se fundamenta en dos elementos: (a) la base usuaria es la más preponderante de la década; y (b) la base usuaria amplia guarda una estricta similitud a la BU en muchos aspectos. Ambos elementos conjuntamente facilitarán nuestro trabajo diacrónico. Así, primero, realizaremos el trabajo de compatibilización sincrónico que implica llevar la base R2 a un formato del tipo BU, dando origen a una base que llamaremos *Base Consolidada (BC)*. Luego, realizaremos el

11 Un ejemplo de los inconvenientes implícitos en este procedimiento es analizado en el siguiente apartado.

12 Como las primeras computadoras poseían sistemas operativos primitivos, el máximo de largo de un nombre era justamente 8 dígitos. El INDEC parece haber optado por mantener este tipo de metodología.

13 Estrictamente hablando, sólo las bases BU y BUA posee una división de variables por bloque temático, en lo concerniente a la R2, hemos optado por colocar sus variables respetando los bloques de las otra presentaciones.

trabajo de compatibilización diacrónico que demandará aproximar la BUA a este nuevo formato la BC.

*La compatibilización sincrónica.*

Si bien el paquete provisto por el INDEC incluye varias bases, datos operativos, técnicos y estadísticos, por cuestiones de espacio en este trabajo nos focalizaremos exclusivamente en los archivos de personas y hogares. El procedimiento será el siguiente: se presentan dos Cuadros de doble entrada (uno por cada unidad básica de análisis) donde cada fila corresponde a una variable y las columnas quedan divididas según la base de que se trate; además, en los casos que se requiera una compatibilización, se mostrarán los valores que las variables toman y al final del bloque se explicará cómo podría llevarse a cabo. Cabe remarcar que, de no realizarse ninguna aclaración, las variables en las distintas bases son perfectamente compatibles.

<b>ARCHIVO PERSONAS</b>			
<b>IDENTIFICACION DE LA BASE DE DATOS</b>			
BU		R2	
Onda	Mes de relevamiento	Onda	Onda correspondiente
Año	Año de relevamiento	Año	Año de la encuesta
Aglomerado	Centro urbano	Aglomerado	Aglomerado al que pertenece la persona
<b>IDENTIFICACION DE CADA PERSONA</b>			
BU		R2	
Cod	Código para individualizar a la persona	N/D	
Componente	Número de componente	N/D	
Acnoauto	Activos No autorrespondentes	N/D	
Realizada	Entrevista realizada	Respuesta	Entrevista realizada
Entrevista	Motivo de entrevista no realizada "1" Ausencia  "2" Rechazo	Codraz	Causa de no respuesta "0" Ausencia momentánea  "1" Ausencia temporal "2" Casa de fin de semana "3" Casa deshabilitada o demolida "4" Casa o departamento en construcción "5" Vivienda establecimiento "6" Rechazo Total "7" Rechazo Parcial "9" Otras Causas "F" Falta código
Pondera	Ponderador del registro	Pondera	Ponderador del registro
Aquí es necesario recodificar la variable Codraz, cambiando el valor 0 (cero) por 1 (uno) y los restantes valores por 2 (dos).			
<b>BLOQUE DE CONDICION DE ACTIVIDAD</b>			
BU		R2	
Estado	Estado ocupacional de la persona "1" Ocupado "2" Desocupado "3" Inactivo  "9" Desconocido	Estado	Estado ocupacional de la persona "1" Ocupado "2" Desocupado "3" Inactivo  "0" Sin especificar
Inactivo	Condición de inactividad	N/D	
N/D		P	Trabajó en la semana
Aquí es necesario recodificar la variable Estado, cambiando el valor 0 (cero) por el valor 9 (nueve).			

<b>BLOQUE CARACTERISTICAS DEMOGRÁFICAS</b>			
BU		R2	
Relacion	Relación de parentesco "1" Jefe "2" Cónyuge "3" Hijo/a "4" Hijo/a político "5" Hermano/a "6" Otros familiares "7" Servicios domésticos "8" Otros componentes	Paren	Relación de parentesco "1" Jefe "2" Cónyuge "3" Hijo/a "4" Hijo/a político "5" Hermano/a "6" Otros familiares "7" Servicios domésticos "8" Otros componentes <b>"9" Desconocido</b>
Edad	Años cumplidos	Edad	Años cumplidos
Sexo	Sexo	Sexo	Sexo
Civil	Estado civil	N/D	
Aquí no es posible compatibilizar. Por lo tanto al unir las bases debe recordarse que existe el valor 9 (nueve) proveniente de la R2.			
<b>BLOQUE DE OCUPADOS</b>			
BU		R2	
Ocupacion	Cantidad de ocupaciones <b>"1" una</b> <b>"2" dos o más</b>	Q	Cantidad de ocupaciones <b>".." número de ocupaciones</b>
Horastot	Total horas trabajadas <b>"1" de 1 a 19 hs.</b> <b>"2" de 20 a 29 hs.</b> <b>"3" de 30 a 40 hs.</b> <b>"4" de 41 a 45 hs.</b> <b>"5" de 46 a 61 hs.</b> <b>"6" de 62 hs. y más</b>  <b>"7" no trabajo en la semana</b>	R	Cantidad horas trabajadas <b>".." cantidad en números</b>  <b>"99" 99 horas o más</b> <b>"00" no trabajó en la semana</b>
Ingreso	Decil de ingreso de la ocupación principal	N/D	
Antigüed	Antigüedad en la ocupación	N/D	
Benef	Beneficios sociales	N/D	
N/D		S	Desea trabajar más horas
N/D		U	Busca otra ocupación
Aquí es necesario recodificar las variables Q y R, respetando los tramos existentes en la base BU.			
<b>BLOQUE DE DESOCUPADOS</b>			
BU		R2	
Tiempo	Tiempo de desempleo <b>"1" hasta 2 meses</b>  <b>"2" más de 2 meses hasta 6 meses</b>  <b>"3" más de 6 meses hasta 1 año</b>  <b>"4" más de 1 año</b>	V	Tiempo de desempleo <b>"0" de 1 a 20 días</b> <b>"1" de 21 a 1 mes</b> <b>"2" más de 1 mes a 2 meses</b>  <b>"3" más de 2 meses a 3 meses</b> <b>"4" más de 3 a 4 meses</b> <b>"5" más de 4 a 5 meses</b> <b>"6" mas de 5 a 6 meses</b>  <b>"7" más de 6 meses a 1 año</b>  <b>"8" más de 1 año</b>
Forma	Forma de desempleo	N/D	
Tipodes	Tipo de desempleo	N/D	
Aquí es necesario recodificar la variable V, respetando los tramos existentes en la base BU.			

## BLOQUE DE POBLACION ECONOMICAMENTE ACTIVA

BU		R2	
Categoría	Categoría ocupacional	T	Categoría ocupacional
Rama	Rama de actividad <b>"1" Actividades primarias</b>  <b>"2" Alimentos, bebidas y tabaco</b> <b>"3" Textiles, confecciones y calzado</b> <b>"4" Productos químicos y de la [...]</b> <b>"5" Productos metálicos, maquinarias y [...]</b> <b>"6" Otras industrias manufactureras</b>  <b>"7" Suministro de electricidad, luz y agua</b>  <b>"8" Construcción</b>  <b>"9" Comercio al por mayor</b> <b>"10" Comercio al por menor</b> <b>"11" Restaurantes y hoteles</b>  <b>"12" Transporte</b> <b>"13" Servicios conexos del transporte y [...]</b> <b>"14" Intermediación financiera</b> <b>"15" Actividades inmobiliarias, [...]</b>  <b>"16" Administración pública y defensa</b> <b>"17" Enseñanza</b> <b>"18" Servicios sociales y de salud</b> <b>"19" Otras actividades de servicios [...]</b> <b>"20" Servicios de reparación</b> <b>"21" Hogares privados con servicio [...]</b> <b>"22" Otros servicios personales</b>  <b>"89" Nuevos trabajadores</b>  <b>"99" Sin especificar</b>	Rama	Rama de actividad <b>"1" Agricultura, caza, silvicultura y pesca</b> <b>"2" Explotación de minas y canteras</b>  <b>"3" Industrias manufactureras</b>  <b>"4" Electricidad, gas y agua</b>  <b>"5" Construcción</b>  <b>"6" Comercio al por mayor y al por menor y restaurantes y hoteles.</b>  <b>"7" Transporte, almacenaje y comunicaciones</b> <b>"8" Establecimientos financieros, seguros [...]</b>  <b>"9" Servicios comunales, sociales y personales</b>  <b>"0" Actividades no bien especificadas y desconocidas</b>
Tamaño	Tamaño del establecimiento	N/D	
Tarea	Carácter y calificación	N/D	
N/D		X	Busca trabajar para
N/D		Y	Tuvo trabajo anterior

La variable rama es la más complicada de la década. Su compatibilización resulta sumamente difícil porque implica compatibilizar: (a) hacia dentro de un año (respetando la agrupación arriba presentada); y (b) a partir de 1994, algunas bases R2 presentaron una descripción alfabética en lugar de la aquí expuesta (la misma no ha sido incluida por motivos de espacio). Nótese que en este caso, la compatibilización se realiza en sentido inverso, pues es necesario respetar los valores provenientes de la base R2.

## BLOQUE DE INGRESOS

BU		R2	
Tipo	Tipo de ingreso.	Ingresos	Tipo de ingreso.
	"1" Total  <b>"2" No responde</b>  <b>"3" Sin ingresos</b>		"1" Tiene ingresos  <b>"9" Desconocido</b>  <b>"0" No tiene ingresos</b>
Fuente	Fuentes de ingresos	N/D	
Monto	<b>Decil del ingreso total individual</b>	Monto	<b>Monto total del ingreso percibido</b>
Especies	Tenencia de ingresos en especies	N/D	

Aquí es necesario recodificar la variable Ingresos, cambiando el valor 9 (nueve) por 2 (dos) y el valor 0 (cero) por 3 (tres). En relación a la variable Monto, debe recalcularse.

BLOQUE EDUCACIÓN			
BU		R2	
Alfabeto	Sabe leer y escribir	N/D	
Asiste	Asistencia a la educación formal	N/D	
Nivel	Máximo nivel alcanzado. <b>"10" Preescolar</b> <b>"11" Primario completo</b> <b>"12" Primario incompleto</b> <b>"21" Secundario no técnico completo</b> <b>"22" Secundario no técnico incompleto</b> <b>"31" Superior o Universitario completo</b> <b>"32" Superior o Universitario incompleto</b> <b>"41" Secundario técnico completo</b> <b>"42" Secundario técnico incompleto</b>	Nivel	Máximo nivel alcanzado. <b>"0" Preescolar o sin instrucción</b> <b>"1" Primario</b> <b>"2" Nacional</b> <b>"3" Comercial</b> <b>"4" Normal</b> <b>"5" Técnica</b> <b>"6" Otra enseñanza media</b> <b>"7" Superior</b> <b>"8" Universitaria</b> <b>"9" Desconocido</b>
		C_nivel	Completó nivel alcanzado. <b>"1" Si</b> <b>"2" No</b> <b>"8" Desconocido</b> <b>"9" sin instrucción</b>

Aquí es necesario generar una nueva variable a partir de las variables Nivel y C\_nivel, con el objeto de obtener una codificación equivalente a la existente en la BU.

### VARIABLES ELABORADAS POSTERIORMENTE

BU		R2	
Decif	Decil de ingreso familiar "0" Sin ingresos "12" Ingresos parciales y Ns/Nr <b>"13" Entrevista no realizada</b>	Decif	Decil de ingreso familiar "0" Sin ingresos "12" Ingresos parciales y Ns/Nr <b>"14" Hogar no entrevistado</b>
Deccf	Decil de ingreso per capita familiar "0" Sin ingresos "12" Ingresos parciales y Ns/Nr <b>"13" Entrevista no realizada</b>	Deccf	Decil de ingreso per capita familiar "0" Sin ingresos "12" Ingresos parciales y Ns/Nr <b>"14" Hogar no entrevistado</b>
Inghora	Ingreso horario de la ocupación principal	N/D	
P47hor	Ingreso total horario	N/D	
N/D		Decind	Decil de ingreso individual

Aquí es necesario recodificar las variables Decif y Deccf, cambiando el valor 14 (catorce) por 13 (trece).

### ARCHIVO HOGAR

#### IDENTIFICACION DE LA BASE DE DATOS

BU		R2	
Onda	Mes de relevamiento	N/D	
Año	Año de relevamiento	N/D	
Aglomerado	Centro urbano	N/D	

#### IDENTIFICACION DE CADA HOGAR

BU		R2	
Cod	Identificación	Codigo	Clave del hogar
Realizada	Entrevista realizada	Respuesta	Entrevista realizada
Razon	Razón de no entrevista	N/D	
Pondera	Factor de expansión	Pondera	Factor de expansión

#### BLOQUE DE CARACTERISTICAS DE LA VIVIENDA

BU		R2	
Tipoviv	Tipo de vivienda	N/D	
Habitacion	Habitaciones de uso exclusivo	N/D	
N/D		Cuartos	Número de cuartos
Agua	Servicio de agua corriente	N/D	

Bano	Servicio de baño	N/D	
Tenencia	Tenencia de la vivienda	Tenen	Tenencia
Material	Tipo de material	N/D	
<b>BLOQUE DE TAMANO DEL HOGAR Y GRUPOS DE POBLACION</b>			
BU		R2	
Pob_tot	Total de personas que habitan el hogar	Compo	Número de componentes
Ocupado	Ocupados que habitan el hogar	N/D	
Subocupado	Subocupados que habitan el hogar	N/D	
Desocupado	Desocupados que habitan el hogar	N/D	
Inactivo	Inactivos que habitan el hogar	N/D	
<b>VARIABLES ELABORADAS POSTERIORMENTE</b>			
BU		R2	
Men14	Menores de 14 años que habitan el hogar	N/D	
Percept	Perceptores de ingresos monetarios	N/D	
Decif	Decil de ingreso total Familiar (Idem personas)	Decif	Decil de ingreso total Familiar (Idem personas)
Deccf	Decil de ingreso per capita familiar (Idem base personas)	Deccf	Decil de ingreso per capita familiar (Idem base personas)
N/D		Iff	
N/D		Ipcf	

Una vez aplicado este procedimiento de unión, la compatibilización sincrónica queda completada, obteniendo dos bases para toda la década: 1990~1997, la Base Consolidada; y 1998~1999, la Base Usaria Ampliada.

#### *La compatibilización diacrónica.*

Nos encontramos en condiciones de empalmar las series temporalmente a partir de la Base Consolidada. Este procedimiento ahora es, sin duda, muy sencillo para todo el período en estudio, pues la BUA posee todos los campos existentes en la BC (además de muchos otros). Más aun, el mismo a quedado automáticamente realizado en términos de los datos existentes en ambas base. Aquí nuevamente el formato BC (originado a partir del BU) es el preponderante.

#### **Las pérdidas asociadas, tanto sincrónica como diacrónicamente.**

El proceso de compatibilización implica claramente la “pérdida” de varias variables y la recodificación de otras, que a su turno conlleva a la disminución en la especificidad de la información, pasando (como se mostró en la tabla anterior) de datos concretos y puntuales a rangos de valores que pueden o no coincidir a priori con aquellos que al investigador le hubiera interesado construir. Estos hechos se dan tanto en el procedimiento diacrónico como en el sincrónico.

Como consecuencia de la aplicación del procedimiento sincrónico podemos destacar las pérdidas más relevantes, separándolas en tres grandes grupos que a continuación se presentan<sup>14</sup>.

(1) Pérdidas de variables al construir la Base Consolidada:

(1.a) Prácticamente todos los datos correspondientes a bs hogares no estarán disponibles (R2).

<sup>14</sup> Entre paréntesis se informa la base que genera el inconveniente.



- (1.b) No es posible identificar a las personas intervinientes en la encuesta como elementos individuales ni identificar si la persona es autorrespondente (R2).
  - (1.c) No es posible conocer la condición de inactividad de los individuos (R2).
  - (1.d) No es posible establecer la causa por la que no trabajo en la semana de referencia (BU).
  - (1.e) No es posible individualizar el estado civil del respondente (R2).
  - (1.f) No es posible conocer el decil de ingreso de la ocupación principal, la antigüedad en la ocupación ni si posee algún beneficio social (R2).
  - (1.g) No es posible conocer si desea trabajar más horas ni si busca otra ocupación (BU).
  - (1.h) No es posible conocer la forma de desempleo ni el tipo de desempleo (R2).
  - (1.i) No es posible conocer el tamaño del establecimiento ni el carácter y calificación de la tarea (R2).
  - (1.j) No es posible conocer ni la fuente y se pierde la posibilidad de saber si la persona percibe ingresos en especies (R2);
  - (1.k) No es posible conocer si la persona sabe leer y escribir ni si recibe asistencia a la educación formal (R2).
  - (1.l) No es posible conocer el ingreso horario propio de la ocupación principal (R2).
- (2) Pérdidas de especificidad:
- (2.a) La causa de la no entrevista queda reducida a sólo dos posibilidades (BU).
  - (2.b) En la relación de parentesco aparece la categoría desconocido (R2).
  - (2.c) No es factible determinar el número exacto de ocupaciones (BU).
  - (2.d) La cantidad de horas trabajadas quedan reducidas a un rango preestablecido (BU).
  - (2.e) La subocupación ya no puede ser libremente calculada (BU). Sin embargo, debemos remarcar que puede ser aproximada, ya que la presentación BU incluye en la base “anexo” la variable Intensidad la cual incorpora ciertas categorías de subocupación.
  - (2.f) El tiempo de desempleo queda reducido a cuatro tramos (BU).
  - (2.g) En relación a las ramas de actividad el problema es sumamente serio. Sólo es posible preservar la codificación proveniente de la base R2.
- (3) Variables que pueden ser calculadas:
- (3.a) La variable monto de los ingresos (BU) puede ser construida a partir de la variable incluida en la base “anexo” llamada P47T, la cual muestra el ingreso total.
  - (3.b) La variable decil de ingresos (R2) pueden ser calculada desde la base.
  - (3.c) La variable ingreso total horario (R2) puede construirse a partir de las variables monto y R, ambas incluidas en la base R2.

Se han destacado las principales pérdidas que el procedimiento de compatibilización sincrónico implica para cualquier año dentro del intervalo de 1990 a 1995. Debemos aclarar que estas pérdidas sólo ocurrirían si nuestra intención es caracterizar a la totalidad de los aglomerados relevados. Nótese que de excluir aquellos aglomerados que presentaron la información en formato R2 o en el caso de iniciar el estudio en año 1996, la mayoría de las pérdidas serían evitadas. A continuación se comenta brevemente el otro tipo de pérdidas asociadas con la utilización de esta encuesta como generadora de series de tiempo.

En el empalme temporal, es decir, luego de aplicar el procedimiento diacrónico, nos encontramos con pérdidas muy importantes. La primera de ellas escapa a las discrepancias entre el formato R2, BU y BUA, porque tiene que ver con los aglomerados en cuestión. Puntualmente, como ya se mostró, en la búsqueda de homogeneidad para toda la década no deben ser incluidos en la base los aglomerados 14, 34, 35, 36 y 37 (Concordia, Mar del Plata, Alto Valle del Río Negro y Interior de Mendoza, respectivamente), ya que éstos se han ido incorporando a lo largo del período en estudio. Vale la pena mencionar que los aglomerados Alto Valle del Río Negro y Interior de Mendoza responden a una lógica de relevamiento sumamente diferente respecto de los restantes aglomerados urbanos, donde se intenta captar otros sucesos como, por ejemplo, la estacionalidad laboral. Por lo tanto, debemos ser extremadamente cuidadosos en caso de intentar una agregación total de la encuesta.

Otra pérdida importante la impone la misma BC, porque ésta es muy inferior en información a la BUA; por ejemplo, ya no es posible seguir a un individuo a lo largo de las cuatro ondas en que éste interviene, perdiendo todas las ventajas del estudio de datos de panel. Pueden darse también pérdidas exógenas, pero que todo investigador debe tomar en cuenta. Por ejemplo, a partir de la imposición del nuevo sistema de educación (EGB) se haría bastante difícil la comparación del nivel educativo de los individuos para toda la década.

Existe otro tipo de pérdidas mucho más grave. Toda vez que ha ocurrido un cambio metodológico en alguna fuente de información, la literatura ha intentado reflejar las implicancias de dichas variaciones a la hora de la comparación [Botta y Pok, 1987]. Un ejemplo de lo antedicho es el trabajo de Giusti y Lindenboim [1997] donde se analizan los cambios acaecidos en la medición de la tasa de actividad entre el censo de 1991 y el inmediato anterior. Resulta claro que entre dos momentos del tiempo es necesario identificar cuáles cambios en las variables a medir son de índole intrínseca y cuáles de índole exógena, o como afirman los autores para el caso de los Censos de Población y Vivienda que se llevaron a cabo en 1980 y 1991:

[...] el cambio en las tasas de actividad que los usuarios habrían de calcular entre 1980 y 1991 es "aparente" y derivan de la existencia de otros dos, uno "real" o "histórico" (el de los de la estructura del empleo) y otro "técnico" o "metodológico" (el aportado por las citadas modificaciones en el instrumento).

De este modo, si intentáramos empalmar las series de la EPH a lo largo de la década –o en cualquier otro intervalo temporal– deberíamos tener presente la posible existencia de cambios metodológicos o técnicos que pueden hacer que nuestras conclusiones se vean fuertemente afectadas, pudiendo también llevarnos a aseveraciones completamente equivocadas. En este trabajo ya hemos resumido algunos de los principales cambios que sufrió la EPH a lo largo de la década. Sin embargo, un análisis exhaustivo del tema requeriría una mayor y más detallada

información que permita analizar no sólo la encuesta como un todo, sino también cada aglomerado incluido en ella.

En relación a este tema, se presenta el siguiente Apartado en donde la variable analizada muestra fuertes cambios en su constitución a lo largo de los noventa. Más allá de tratarse de un mero ejemplo, puede resultar útil como ilustración de lo dañinos que pueden ser los cambios metodológicos u operativos. Sin embargo, dos salvedades deben realizarse. Primero, si bien la variable presenta cambios muy radicales, ésta es construida desde el INDEC, por lo que poco tiene que ver con problemas de captación o implementación de la encuesta. Segundo, si bien todos los datos son oficiales, las opiniones vertidas son de nuestra exclusiva responsabilidad, aceptando la posibilidad de que se haya incurrido en errores u omisiones que alteren sustancialmente el análisis.

#### **IV. Un ejemplo de uso: la variable “pondera”**

---

##### **Presentación del problema**

Sin duda ya debe resultar claro cuán complicado –además de útil– puede ser para un investigador social utilizar como fuente secundaria de información a la EPH (y en especial para la década del noventa). Su utilización implica necesariamente la realización de una serie de procesos y modificaciones en las bases, cuando no eliminaciones o sacrificios en términos analíticos y explicativos. Por ello, nuestra idea aquí es, a partir de todo el análisis precedente, mostrar un caso que no presente problemas sincrónicos pero que sí sea diacrónicamente conflictivo. Así, si bien ningún usuario encontraría problemas en compatibilizar las bases BU y R2, en lo concerniente a la variable pondera, la situación puede resultar muy distinta si el interés radicara en analizar la década utilizándola.

Como es lógico y natural, todo investigador intenta al menos describir la realidad. Por tal motivo, es también coherente que esta persona no se contente con los datos muestrales que la encuesta proporciona –o con las relaciones entre ellos–, sino que querrá inferir conclusiones o resultados buscando describir al universo representado. Además, justamente por tratarse de una encuesta y siguiendo las recomendaciones que el INDEC realiza en la materia:

por provenir de una muestra es necesario al analizar o procesar los datos utilizar la ponderación (campo PONDERA) que tiene cada persona u hogar de la base. Esta ponderación indica el número de personas u hogares del universo en estudio representado por cada persona u hogar de la muestra (corregidas por no respuesta).

INDEC, 1998, página 8.

El investigador decide utilizar la variable pondera<sup>15</sup>. Sin embargo, aquí se inician los inconvenientes: son los mismos responsables de la EPH y funcionarios del INDEC los que insisten en advertir que los resultados propios de la encuesta tienen validez sólo a nivel de tasas –y

---

<sup>15</sup> Esta variable es una de las más importantes de la encuesta, porque es justamente la encargada de expandir el registro; dicho de otra manera, nos informa a cuantas personas (hogares) representa cada persona (hogar) encuestada.

por ende las estructuras que éstas representan, pero que no deben ser utilizados los valores absolutos originados de aplicar el expansor (ponderador). De esta forma, nos encontramos ante la presencia de un ejemplo más de lo que sucede cuando se da una disociación entre productores y usuarios, pues de no conocer las recomendaciones del propio INDEC, quien quisiera utilizar los datos “expandidos” se encontraría con graves problemas.

Como ya se ha explicado, el ponderador se construye a partir del diseño de la muestra y de los datos censales, tratándose de una muestra “autoponderada”. Sin embargo, no siempre hay coincidencia entre la delimitación censal del aglomerado y la aplicada por la Encuesta, induciendo al usuario a cometer errores no deseados, tanto a un nivel analítico como explicativo.

Un sencillo ejemplo puede aclarar este punto. Tomando la encuesta<sup>16</sup> para toda la década y dividiéndola en GBA y Resto Urbano (donde se incluyen los restantes aglomerados revelados por la EPH)<sup>17</sup>, podemos rápidamente preguntarnos cuál es el universo alcanzado o, dicho de otro modo, cuál es según la encuesta la población para cada uno de los grupos en cada momento del tiempo. Los resultados son presentados en el Cuadro 2; además incluimos dos estimaciones independientes de población del INDEC (1996 y 1999)<sup>18</sup>, las cuales pueden servirnos como parte de la explicación de lo acontecido. Cabe remarcar que dentro del propio INDEC se han realizado varias proyecciones poblacionales a partir del Censo de 1991. En este trabajo se utilizan únicamente las últimas dos [INDEC, 1999c].

#### CUADRO 2- Población según grupo geográfico. Onda de Octubre.

PANEL A – Total País (EPH).

ANO	INDEC'96	INDEC'99	EPH. Base	INDEC'96-EPH	INDEC'99-EPH
1991	19.195.706	19.662.016	19.218.367	-22.661	443.649
1992	19.519.675	19.981.178	19.890.798	-371.123	90.380
1993	19.849.695	20.302.265	20.032.954	-183.259	269.311
1994	20.177.968	20.623.398	19.989.459	188.509	633.939
1995	20.503.884	20.943.940	21.365.957	-862.073	-422.017
1996	20.826.708	21.265.709	21.675.882 (1)	-849.174	-410.172
1997	21.148.888	21.588.572	22.001.888	-853.000	-413.316
1998	21.470.081	21.911.500	21.886.436	-416.355	25.064
1999	21.789.953	22.233.422	22.227.379 (2)	-437.426	6.043

PANEL B – GBA (EPH).

ANO	INDEC'96	INDEC'99	EPH. Base	INDEC'96-EPH	INDEC'99-EPH
1991	10.913.945	11.355.748	11.315.246	-401.301	40.502
1992	11.043.818	11.479.960	11.662.050	-618.232	-182.090
1993	11.175.131	11.602.556	11.690.471	-515.340	-87.915
1994	11.306.177	11.724.676	11.474.122	-167.945	250.554
1995	11.435.276	11.847.595	11.435.046	230	412.549
1996	11.562.569	11.973.008	11.561.652	917	411.356
1997	11.689.212	12.099.510	11.689.330	-118	410.180
1998	11.815.049	12.226.270	12.227.177	-412.128	-907
1999	11.939.934	12.352.412	12.352.050	-412.116	362

<sup>16</sup> Si bien nuestro análisis contempla únicamente la onda de octubre, resulta sumamente similar la situación en lo concerniente a la onda de mayo.

<sup>17</sup> Lógicamente, se han realizado los procesos de compatibilización por lo que no se incluyen ninguno de los “nuevos” aglomerados.

<sup>18</sup> Cabe mencionar que estas estimaciones son de uso restringido y que, en general, no se encuentran disponibles para uso del público, aunque sí para funcionarios e investigadores relacionados con este tema.

PANEL C – Resto Urbano (EPH).

ANO	INDEC'96	INDEC'99	EPH. Base	INDEC'96-EPH	INDEC'99-EPH
1991	8.281.761	8.306.269	7.903.121	378.640	403.148
1992	8.475.857	8.501.218	8.228.748	247.109	272.470
1993	8.674.564	8.699.709	8.342.483	332.081	357.226
1994	8.871.791	8.898.722	8.515.337	356.454	383.385
1995	9.068.608	9.096.345	9.930.911	-862.303	-834.566
1996	9.264.139	9.292.701	10.114.230 (1)	-850.091	-821.529
1997	9.459.676	9.489.062	10.312.558	-852.882	-823.496
1998	9.655.032	9.685.230	9.659.259	-4.227	25.971
1999	9.850.019	9.881.010	9.875.329 (2)	-25.310	5.681

(1) La base correspondiente a Córdoba para este año no se encuentra disponible. Por tal motivo, la estimación poblacional ha sido imputada como la semisuma entre el año 1995 y 1997.

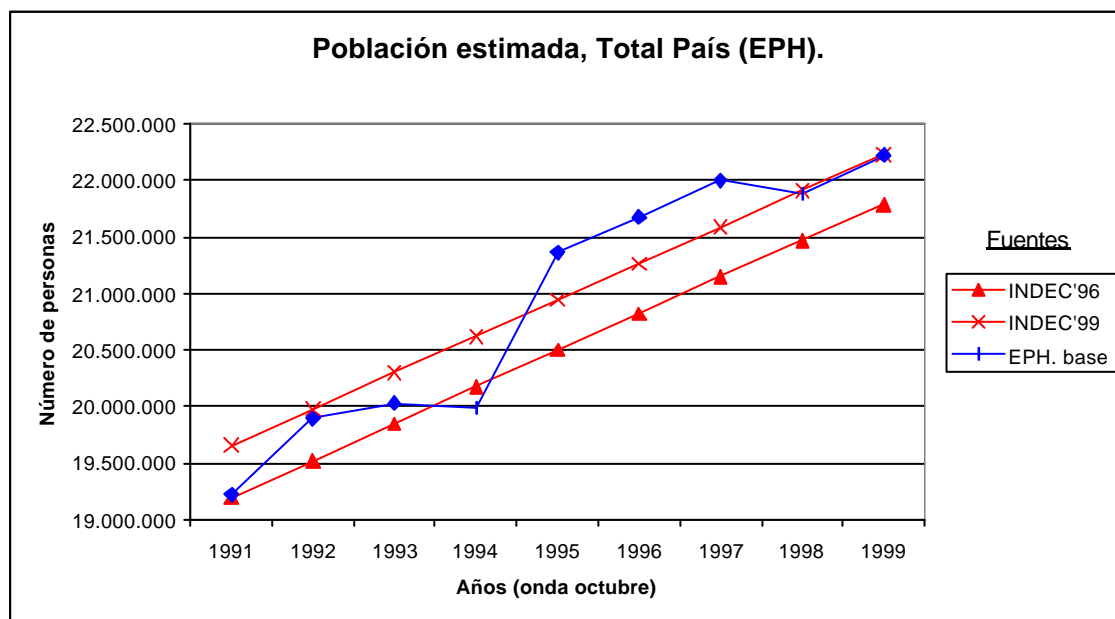
(2) El dato correspondiente al aglomerado de Bahía Blanca no se encuentra disponible. En este caso, el mismo ha sido imputado con un valor de 300.000 habitantes.

FUENTES: Elaboración propia en base a INDEC-EPH, INDEC -Dirección de Estadísticas Poblacionales.

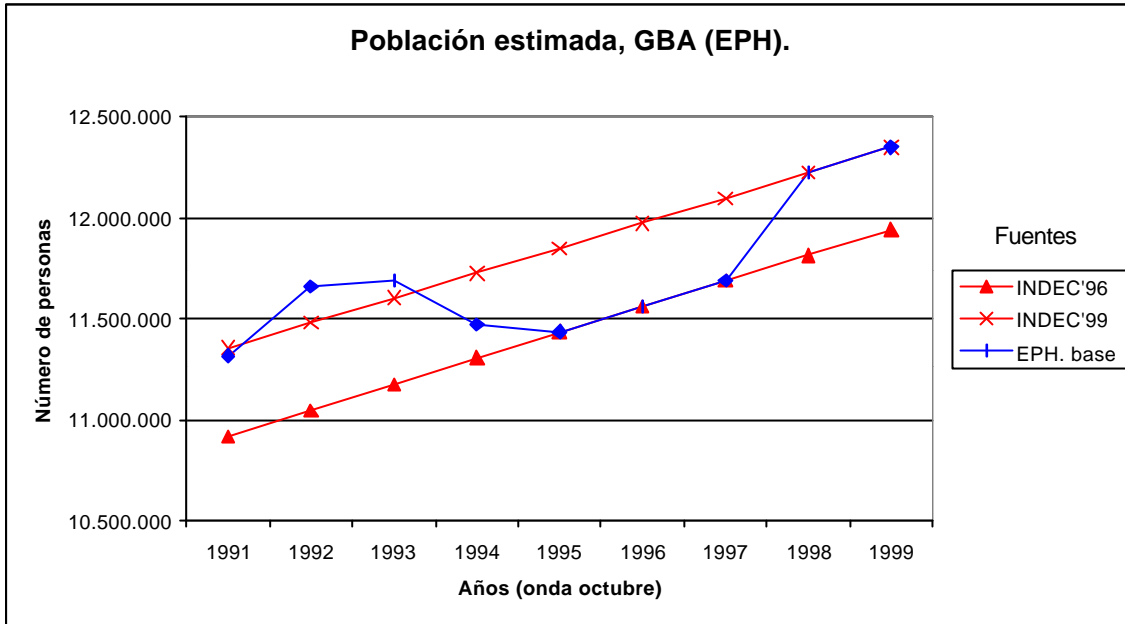
Una primer lectura nos permite apreciar la gran volatilidad existente en las estimaciones poblacionales desde la EPH respecto de las extrapolaciones censales. Asimismo, podemos remarcar que el aglomerado GBA es el que más ha visto modificada su nivel tendencial entre las extrapolaciones oficiales de 1996 y 1999. Nótese que conforme transcurre el tiempo, la brecha existente entre la EPH y las distintas extrapolaciones oficiales de población comienza a mermar en forma significativa. Si bien este hecho representa un aliciente en cuanto a la rigurosidad de los procesos llevados a cabo por el INDEC, nos confirma la volatilidad intrínseca dentro de la década.

Debemos realizar una aclaración: en el año 1996 se ha imputado el valor correspondiente al aglomerado de Gran Córdoba y en el año 1999 el valor correspondiente a Bahía Blanca también ha sido imputado, por lo que las oscilaciones presentadas en las variables “Total País” y “Resto Urbano” deben relativizarse levemente.

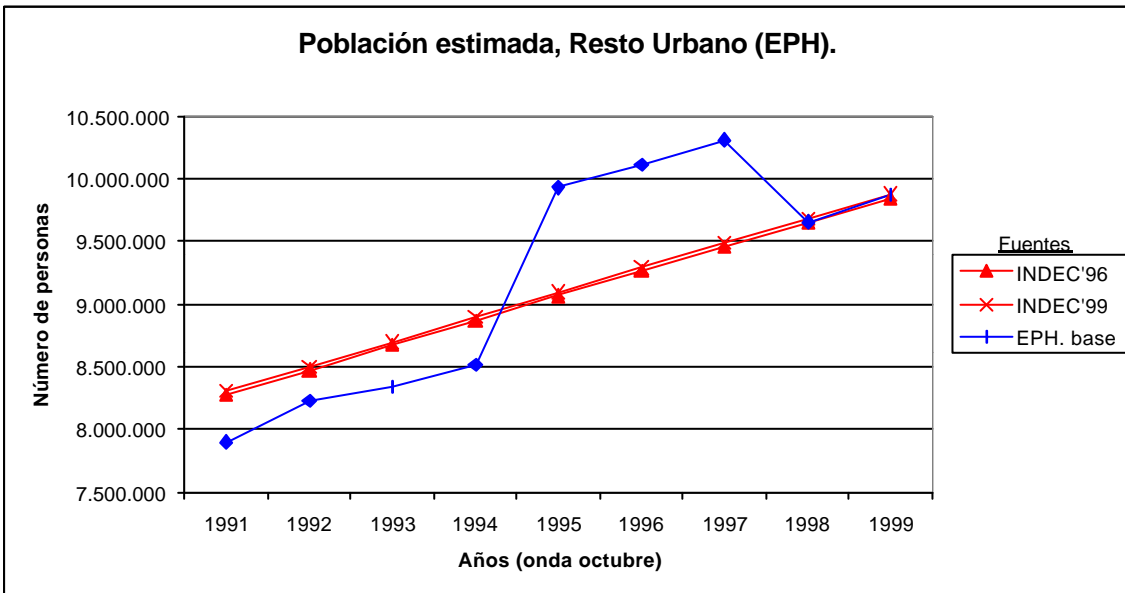
**FIGURA 1 – Población según grupo geográfico. Onda Octubre.**  
 PANEL A – Total País (EPH) (\*).



PANEL B – GBA. (EPH)



PANEL C – Resto Urbano (EPH) (\*).



(\*) La estimación para Gran Córdoba para el año 1996 ha sido imputada como la semisuma entre 1995 y 1997. La estimación para Bahía Blanca ha sido imputada con el valor de 300.000 habitantes.

FUENTES: Elaboración propia en base a INDEC-EPH, INDEC-Dirección de Estadísticas Poblacionales.

Mucho más claras son las conclusiones si observamos los datos gráficamente. En la Figura 1, se muestran los datos arriba expuestos, donde cada panel corresponde a un agrupamiento diferente. Las diferencias quedan expresadas sencillamente como la distancia geométrica entre las cruces, triángulos y cuadrados para cada año.

Como puede apreciarse, la conclusión es contundente: **bajo ninguna circunstancia puede utilizarse a la EPH como fuente secundaria confiable en lo que refiere a las estimaciones de población**; o lo que es lo mismo, para expresar cualquier conclusión en valores absolutos.

En relación al Resto Urbano, las estimaciones poblacionales derivadas de la EPH no sólo parecen erráticas, sino que tampoco se ajustan a ninguna de las estimaciones alternativas de población (salvo para el corto período 1998-1999). Debemos mencionar además, que justamente éstas resultan sumamente similares, lo que daría una impresión más acabada de su correcta representatividad respecto de los verdaderos valores poblacionales.

El caso del GBA parece mucho más interesante. En primer lugar, las dos proyecciones de población del INDEC difieren sustancialmente, donde la del año 1996 es llamativamente menor que la del año 1999. En segundo lugar, podemos identificar tres períodos bien definidos: entre 1991 y 1994, las estimaciones poblacionales de la EPH presentan también un comportamiento errático y no consistente con las proyecciones poblacionales; entre 1995 y 1997, coinciden perfectamente con la primera de las estimaciones de población del INDEC utilizadas en este trabajo, pero que por cierto sabemos ha sido reemplazada; en el tramo 1998-1999, la EPH presenta una congruencia perfecta pero con la nueva estimación independiente de población provista por el INDEC. Pero entonces, ¿existe una explicación racional de este suceso? ¿podemos atribuirle alguna (o toda la) responsabilidad a cambios intrínsecos a la encuesta?, o dicho de otra manera, si intentáramos compatibilizar estas bases, ¿estaríamos en presencia de un problema de orden diacrónico? A nuestro juicio, las repuestas a estas preguntas serían todas afirmativas.

En relación al resto Urbano, una primera aproximación nos muestra que las bases R2 –todas ellas provenientes de algún aglomerado del interior– poseen un factor de no-respuesta bastante importante, es decir, la encuesta tuvo un alto índice de personas u hogares enteros que no la respondieron. Por lo tanto, para el primer quinquenio en estudio, la escasa robustez de la muestra puede ser la explicación de por qué se ven tan afectadas las estimaciones poblacionales. En términos cuantitativos, el porcentaje de no respuesta sobre el total de encuestados fue 4.1%, 2.7%, 1.8%, 1.4% y 0.4% para los años comprendidos entre 1990 y 1994 respectivamente<sup>19</sup>; a partir de 1995 estos valores fueron reducidos a cero, dado que ya no existen más bases R2. De este modo, vemos cómo problemas en la presentación de la información pueden afectar significativamente un estudio como el presente. Con respecto a la segunda mitad de la década, este agregado continuó alejado de las estimaciones poblacionales, aunque ahora se coloca por encima de las mismas. Podría pensarse que la imputación realizada por el ente oficial, con el fin de corregir el factor de no-respuesta, es el causante de esta sobreestimación de población urbana. Finalmente, se logra alcanzar una homogeneidad entre las proyecciones censales y la EPH, pero sólo para los años 1998 y 1999.

En relación al GBA, el análisis puede también dividirse de acuerdo al comportamiento de la EPH a lo largo de la década. Hasta 1994 inclusive, se observa que las estimaciones poblacionales no se ajustan a ningún patrón preestablecido. Sin embargo, en el período 1995-1997, la encuesta estima una población casi exactamente igual a la estimación del INDEC de 1996. Luego, en 1998 y 1999, se produce un brusco salto que alcanza la nueva estimación de población de este

---

<sup>19</sup> Elaboración propia, en base a EPH-INDEC. Los cuadros no se incluyen por motivos de espacio.

organismo. La explicación de estos eventos podría ser como sigue: mientras que el Censo considera al GBA como una porción de territorio mayor que la EPH –la encuesta históricamente consideró sólo a la Capital Federal y a los 19 partidos del Conurbano Bonaerense, mientras que el censo siempre tomó una zona geográfica más grande– sus estimaciones se realizaron utilizando justamente la definición de la EPH, lo que se ejemplifica en la proyección de 1996. Sin embargo, fue la misma encuesta la que ha intentado desde mediados de la década ampliar su zona de captación con el objeto de ganar una mayor homogeneidad con los censos, dando origen a la necesidad de una nueva proyección (1999). Así, observamos como en 1995 la muestra de la EPH es ajustada con la vieja proyección censal<sup>20</sup> y sólo a partir de 1998 esta situación es corregida.

Más allá de las posibles explicaciones aquí esbozadas, lo llamativo de este análisis es la heterogeneidad en el comportamiento de una de las variables más importantes de la base suministrada a los usuarios. Además, como oportunamente remarcamos, este problema implica a la totalidad de las bases (es decir, no discrimina según formato) y por ende a las conclusiones derivadas.

### **Dos soluciones alternativas.**

Como ya hemos mencionado, el problema que nos presenta la variable “pondera” radica en su característica de “transformador” de datos muestrales en poblacionales. Si bien no existe objeción alguna a los datos obtenidos en términos relativos –esto es, las tasas– en cuanto a su precisión, tanto estadística como metodológica, resulta también evidente que los datos absolutos no pueden ser aceptados como verdaderos. Dada la gran insatisfacción que el análisis precedente genera, a continuación presentamos dos métodos alternativos<sup>21</sup> que buscan resolver dicho inconveniente. El primero de ellos puede ser llamado *Directo*, en el sentido de que utiliza los datos de la EPH y los datos del censo directamente y no realiza ninguna modificación sustantiva. El segundo de los métodos puede ser llamado *Indirecto*, dado que realiza ciertas alteraciones en la variable pondera.

Nótese que ambos métodos incluyen como supuesto básico la homogeneidad metodológica entre los valores censales y los valores de la EPH (este supuesto implica que las tasas calculadas desde una u otra fuente deben coincidir; el mismo es utilizado y explicado con mayor detalle en la demostración de la Proposición 3 que se encuentra al final de este Apartado), y el supuesto de constancia en las tasas (dado que el Censo de 1991 se realizó en mayo y nosotros utilizamos la EPH de la onda de octubre, este supuesto resulta sumamente importante). Además, como es lógico, la validez de los procedimientos dependerá primordialmente de la representatividad estadística y la consistencia que posean tanto las tasas calculadas como los datos de población iniciales.

#### *Método Directo de cálculo de valores poblacionales.*

El primero de estos métodos puede presentarse como sigue. Tomando como válidos los datos provenientes del INDEC contruidos a partir del Censo Nacional de Población, rápidamente pueden construirse los datos a nivel poblacional: simplemente deben aplicarse las tasas propias de la EPH (de actividad, de desempleo, por categoría ocupacional, etc.).

---

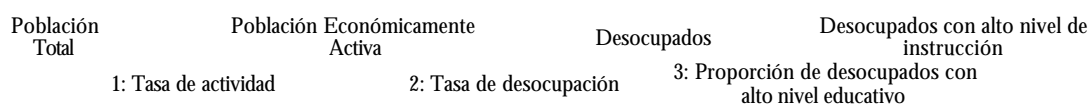
<sup>20</sup> Llamativamente, este acontecimiento coincide con los abruptos crecimientos permanentes de las tasas de actividad y de desocupación; hecho que plantea el interrogante de si dichos cambios pudieron influir de algún modo.

<sup>21</sup> Puede mostrarse fácilmente que ambos métodos resultan equivalentes desde una perspectiva analítica. Para la demostración formal, ver Proposición 3, al final de este apartado.



Aquí la idea es aplicar los valores que describen la estructura a los datos de nivel (absolutos). En el Esquema 1 se representa un ejemplo de este procedimiento. El mismo puede realizarse para distintos niveles de agregación en términos geográficos o de tasa, siempre y cuando se cumplan los supuestos antes mencionados.

**Esquema 1 – Método directo de cálculo de variables a nivel poblacional.**



Nótese que este método no implica mayores inconvenientes, pues su aplicación es sumamente fácil. No obstante ello, debemos aclarar que su aplicación en términos operativos puede resultar excesivamente complicada cuando, por ejemplo, se intenta alcanzar valores absolutos muy puntuales, es decir, cuando buscamos universos particulares, o quizás, cuando se manejan una gran cantidad de datos y/o con variables con diversa proporción de datos “sin especificar”.

Por todo ello, aquí intentaremos proponer un procedimiento alternativo más rápido en términos operativos.

*Método Indirecto de cálculo de valores poblacionales.*

La idea del segundo de los métodos es sumamente sencilla y puede explicarse del siguiente modo: pensando a los ponderadores como simples conversores de datos muestrales en datos de nivel poblacional, los mismos pueden ser corregidos con el objeto de utilizar directamente la base de EPH al momento de realizar inferencias de índole poblacional. Nuestro objetivo es encontrar un serie de vectores de “coeficientes de corrección” que, aplicados a los ponderadores originales, nos permitan construir los vectores de ponderación “corregidos”.

Dado que a cada hogar le corresponde un ponderador y que, a partir de éste, le es asignado uno a cada miembro, rápidamente podemos expresar simbólicamente la forma en que se alcanza la estimación poblacional desde la EPH.

$$pondera^t_j = \sum_{n=1}^N pondera^t_{n,j}$$

Claramente, la población en el momento t del aglomerado j será igual a la suma de los ponderadores (*pondera*) correspondientes a cada uno de los N registros de la base de datos de las personas –la fila de la base corresponde a un individuo–. Debemos remarcar que si miramos hacia dentro de los vectores de ponderación, los mismos varían según las características del hogar. Más aun, distintos estados (ocupado, desocupado, inactivo) pueden recibir distintos niveles de

ponderación media. En relación a esto, en el anexo se incluye el Cuadro A.4 donde se muestran algunas de estas diferencias. Si bien este hecho no afecta el presente análisis, debe ser considerado a la hora de delimitar el “universo” en estudio.

Ahora bien, si tomamos la proyección oficial del INDEC de población (ver en anexo, Cuadro A.2), observamos que para cada año y aglomerado existe una estimación independiente y propia a partir del censo de 1991 de población. Por consiguiente, podríamos construir el factor de corrección del siguiente modo (los datos correspondientes al denominador pueden encontrarse en el Cuadro A.3 del anexo):

$$\mathbf{a}_j^t = \frac{pc_j^t}{pondera_j^t}$$

donde  $pc$  significa proyecciones censales.

Este nuevo factor  $\alpha$  corresponde a la proporción de desvío que existe en la EPH respecto del censo. Cuanto más cercano al valor 1, menor será el desvío existente. Mientras que un  $\alpha$  menor que 1 implica una subestimación de la EPH, un  $\alpha$  mayor que 1 implica una sobreestimación de la encuesta. Al término de esta Sección se presenta la matriz completa de coeficientes  $\alpha$ ; obsérvese que éstos describen (en términos relativos) las desviaciones que gráficamente presentamos en la Sección anterior.

Una vez calculado el factor alfa, éste debe ser aplicado al vector de ponderadores originales con el objetivo de construir un nuevo vector de ponderadores corregidos. Una vez realizado este procedimiento, puede utilizarse la base de EPH directamente para el cálculo de valores poblacionales.

Nótese que este procedimiento de ningún modo modifica la estructura relativa dentro de un aglomerado, pues al tratarse de una transformación lineal, simplemente se ha “corrido” el eje de medición (ver Proposición 1 más abajo). Sin embargo, sí se ha alterado la estructura de la encuesta, dado que al modificar en forma sensible los ponderadores de cada aglomerado, se han modificado las participaciones relativas de los mismos en el todo (ver Proposición 2 más abajo).

Si bien esto no debería presentar ningún inconveniente, dado que las técnicas de muestro se realizan para cada aglomerado por separado, no puede ser olvidado ya que es de práctica común en el ámbito de la investigación agrupar aglomerados de acuerdo a distintos criterios u objetivos.

#### *Demostraciones de las principales proposiciones utilizadas.*

Sea  $\mathbf{a}_j^t$  el factor de corrección calculado por el método indirecto para el aglomerado  $j$  en el momento  $t$ .

Sea  $\mathbf{r}_j^t$  el vector columna transpuesto de ponderadores del aglomerado  $j$  para el momento  $t$ .

Sean  $\mathbf{q}_{i,j}^t$  y  $\mathbf{q}_{k,j}^t$  dos vectores columna binarios de los universos  $i$  y  $k$  respectivamente, ambos para el momento  $t$ ; donde 1 es asignado a quien pertenece y 0 a quien no pertenece (ej. [1, 0, 0, 1, ...]).

Sea  $pc_{k,j}^t$  el vector columna transpuesto de proyecciones censales independientes del universo  $k$  para el momento  $t$ .

**Proposición 1:** Aplicar el factor de corrección no modifica la estructura relativa intraglomerado.

**Demostración:**

La tasa  $i$  respecto de  $k$  se construye de la siguiente forma:

$$\mathbf{t}_{i/k;j}^t = \frac{\mathbf{a}_j^t \cdot \mathbf{r}_j^t \cdot \mathbf{q}_{i;j}^t}{\mathbf{a}_j^t \cdot \mathbf{r}_j^t \cdot \mathbf{q}_{k;j}^t} = \frac{\mathbf{r}_j^t \cdot \mathbf{q}_{i;j}^t}{\mathbf{r}_j^t \cdot \mathbf{q}_{k;j}^t}$$

Nótese que modificar los vectores de ponderación de ningún modo altera las participaciones relativas de la encuesta en un aglomerado.

**Proposición 2:** Aplicar el factor de corrección modifica la estructura relativa interaglomerado.

**Demostración:**

La tasa  $i$  respecto de  $k$  para los  $J$  (con  $J > 1$ ) aglomerados se construye de la siguiente forma:

$$\mathbf{t}_{i/k}^t = \frac{\sum_{j=1}^J \mathbf{a}_j^t \cdot \mathbf{r}_j^t \cdot \mathbf{q}_{i;j}^t}{\sum_{j=1}^J \mathbf{a}_j^t \cdot \mathbf{r}_j^t \cdot \mathbf{q}_{k;j}^t}$$

Nótese que modificar los vectores de ponderación afecta sensiblemente las participaciones relativas de los  $J$  aglomerados en el conjunto, salvo para el caso trivial en donde  $\mathbf{a}_1^t = \mathbf{a}_2^t = \dots = \mathbf{a}_J^t$ .

**Proposición 3:** Los métodos directo e indirecto resultan equivalentes.

**Demostración:**

$$pc_{k;j}^t \cdot \mathbf{t}_{i/k;j}^t = pc_{k;j}^t \cdot \frac{\mathbf{a}_j^t \cdot \mathbf{r}_j^t \cdot \mathbf{q}_{i;j}^t}{\mathbf{a}_j^t \cdot \mathbf{r}_j^t \cdot \mathbf{q}_{k;j}^t} = \frac{pc_{k;j}^t}{\mathbf{r}_j^t \cdot \mathbf{q}_{k;j}^t} \cdot \mathbf{r}_j^t \cdot \mathbf{q}_{i;j}^t = \frac{pc_{k;j}^t}{pondera_{k;j}^t} \cdot \mathbf{r}_j^t \cdot \mathbf{q}_{i;j}^t = \mathbf{a}_j^t \cdot \mathbf{r}_j^t \cdot \mathbf{q}_{i;j}^t$$

Nótese que el método directo implica al método indirecto implícitamente y viceversa. De regir el supuesto de homogeneidad metodológica entre las tasas calculadas desde el censo y las tasas calculadas desde la EPH, puede mostrarse fácilmente que  $\mathbf{a}_{k;j}^t = \mathbf{a}_j^t$ . Más aun, en la práctica sabemos que la característica  $k$  es efectivamente la población total y que luego se aplica a ésta las tasas de la EPH.

**CUADRO 3 – Factor de corrección de la variable pondera (EPH), según aglomerado y año. Onda de Octubre.**

<b>COD.</b>	<b>AGLOMERADOS</b>	<b>1991</b>	<b>1992</b>	<b>1993</b>	<b>1994</b>	<b>1995</b>	<b>1996</b>	<b>1997</b>	<b>1998</b>	<b>1999</b>
	<b>TOTAL (25 aglomerados)</b>	<b>1,023085</b>	<b>1,004544</b>	<b>1,013443</b>	<b>1,031714</b>	<b>1,020578</b>	<b>1,088896</b>	<b>1,021533</b>	<b>1,001145</b>	<b>1,013957</b>
1	<b>GRAN BUENOS AIRES</b>	<b>1,003579</b>	<b>0,984386</b>	<b>0,992480</b>	<b>1,021836</b>	<b>1,036078</b>	<b>1,035579</b>	<b>1,035090</b>	<b>0,999926</b>	<b>1,000029</b>
	<b>RESTO</b>	<b>1,051011</b>	<b>1,033112</b>	<b>1,042820</b>	<b>1,045023</b>	<b>1,001072</b>	<b>1,166259</b>	<b>1,004752</b>	<b>1,002689</b>	<b>1,031924</b>
2	GRAN LA PLATA	0,987313	0,952984	0,977843	0,971929	1,000701	1,002897	1,001823	1,001424	1,001347
3	BAHIA BLANCA	0,977548	0,961964	0,949252	0,945426	0,999537	0,999925	1,002059	1,000693	N/D
4	GRAN ROSARIO	0,983284	0,987621	0,978444	0,983959	1,000537	1,000816	0,999754	1,001348	1,000306
5	STA. FE Y STO. TOME	1,069495	1,102519	1,037795	1,104348	1,021296	1,024667	1,025956	1,024576	0,998938
6	PARANA	0,994705	1,033212	1,054242	1,083088	1,023076	1,023598	1,021982	1,024113	0,998652
7	POSADAS	1,035557	0,955766	0,972050	0,925959	0,998390	0,999084	1,004385	0,999881	1,001737
8	GRAN RESISTENCIA	0,994635	1,030273	1,041749	1,107555	0,989487	1,000664	1,002457	0,999946	1,000714
9	COMODORO RIVADAVIA	N/D	0,986577	1,034079	1,070052	0,997265	0,999621	1,006755	1,006675	0,997154
10	GRAN MENDOZA	1,007406	0,997902	1,012987	1,033799	1,001286	1,000344	0,999570	0,999496	1,000053
12	CORRIENTES	1,151125	1,191330	1,172474	1,195734	1,007733	0,996630	1,000322	1,000736	1,000920
13	GRAN CORDOBA	1,008938	1,019696	1,035355	0,999906	1,000161	N/D	0,999841	1,000262	0,999749
15	FORMOSA	1,033469	0,959423	1,036096	0,945303	1,000741	0,999456	0,999473	0,997270	1,004328
17	NEUQUEN Y PLOTTIER	1,086021	1,154374	1,128031	1,149957	0,999092	1,001708	1,001157	0,999087	1,000318
18	STGO DEL ESTERO Y LA BANDA	1,114441	1,107070	1,141683	1,207779	0,998906	1,000789	1,001164	0,999371	0,999973
19	SAN S. JUJUY Y PALPALA	1,144860	1,187889	1,180592	1,244670	0,997819	1,000001	1,006361	0,997864	1,002311
20	RIO GALLEGOS	1,098454	1,150816	1,214909	1,194780	0,999331	1,007829	1,000860	0,998184	1,003241
22	GRAN CATAMARCA	1,125752	1,125738	1,147546	1,144939	1,089265	1,086019	1,091773	1,001321	0,991177
23	SALTA	1,093078	1,078268	1,129486	1,056995	1,001785	1,000882	0,999138	0,999134	1,000110
25	LA RIOJA	1,000312	1,071117	1,046190	1,069275	0,997024	1,008322	1,006423	1,002656	1,001276
26	SAN LUIS Y EL CHORRILLO	1,090126	1,047110	1,074633	1,105689	0,992825	0,997351	0,999441	0,998154	1,001914
27	GRAN SAN JUAN	1,109329	1,093873	1,083595	1,055970	1,001406	1,003537	1,004251	1,002213	1,004248
29	S. M. DE TUCUMAN Y TAFI VIEJO	1,075519	1,054824	1,075830	1,096900	0,977793	1,003956	1,007112	1,007864	1,003446
30	SANTA ROSA Y TOAY	1,034655	1,035277	1,060792	1,085632	0,993876	1,004060	0,999669	0,997670	1,000077
31	TIERRA DEL FUEGO	1,075131	1,066699	1,123468	1,069722	1,006740	0,988446	1,010710	1,002209	0,997847

Fuente: Elaboración propia en base a INDEC-EPH, INDEC-Dirección de Estadísticas Poblacionales y Centro de Estudios sobre Población, Empleo y Desarrollo, UBA.

## V. Reflexiones finales

---

Utilizar a la Encuesta Permanente de Hogares como fuente secundaria de información con el objetivo de caracterizar al total de aglomerados urbanos impone sin duda un reto. Más aún, el simple intento de usarla para el estudio particular de algún centro urbano implicaría tomar en cuenta ciertas precauciones que los usuarios muchas veces por apuro o descuido suelen no considerar.

A lo largo de este trabajo, hemos estudiado distintos aspectos de la EPH (y de las encuestas de hogares en general) con el objetivo de destacar los principales problemas metodológicos u operativos con que pueden encontrarse los usuarios al momento de utilizar esta fuente secundaria de información.

En una primer instancia, decidimos recurrir a una conceptualización de índole teórica para mostrar cómo las encuestas de hogares deben ser, además de cuáles son sus objetivos, alcances y limitaciones. Allí rápidamente nos dimos cuenta que toda encuesta posee al menos dos características que la diferencia de los otros tipos de fuentes. Por un lado, su motivación y uso, es decir, pudimos apreciar cómo han cambiado significativamente a lo largo de la historia, respondiendo a distintos modelos de acumulación e intereses de diferentes actores sociales. Por otro lado, su característica técnica, dado que ésta es en última instancia la razón central de por qué una encuesta resulta útil o no.

Luego, nos acercamos a la EPH en la Argentina para la década de los noventa, analizándola desde una perspectiva histórica y conceptual. Pudimos destacar dos aspectos centrales en esa Sección. Primero, presentamos las principales características técnicas, metodológicas y operativas que posee la encuesta y que –si bien no fueron objeto de análisis en extenso de este trabajo– merecerían un estudio mucho más detallado. Segundo, observamos que desde sus comienzos el INDEC intentó mejorar y ampliar en distintos aspectos a la EPH. Así, pudimos destacar varios cambios de sumo interés a lo largo de la década del noventa y otros que se prevé sean implementados en breve, como por ejemplo la reformulación de la encuesta a fin de hacerla de carácter continua. Vale la pena destacar el profundo interés que presentó el organismo productor (así como sus integrantes) por mejorar en forma continua a este herramental.

Una vez encuadrado nuestro objeto de análisis (la EPH), mostramos algunos ejemplos de cómo la disociación entre productores y usuarios, predominante en las fuentes de información de índole social, puede afectar seriamente el uso de la información dada, claro está, la forma de su presentación. Llamativamente, pudimos constatar que, aún cuando los sucesos quizás fueron captados, éstos no se encuentran disponibles para los usuarios dada la forma en que la información fue provista. Observamos que la existencia de distintas formas de presentación generan fuertes limitaciones. Dentro de dicho Apartado, se propuso una solución de índole metodológica-operativa para la homogeneización de los datos presentados por el INDEC. La misma poseía un doble propósito, por un lado, documentar los procedimientos que habitualmente se llevan a cabo previos a la utilización de la encuesta y, por el otro, mostrar las pérdidas que este procedimiento –y por lo tanto la disociación que le dio origen– implica. Concluimos que sería injusto culpar solamente a la presentación de tipo R2, ya que en muchas ocasiones las limitaciones provenían del formato BU, por ejemplo, mientras que en las bases R2 el investigador gozaba del dato concreto, en las bases BU sólo existen intervalos, constituyéndose en un ejemplo de pérdida de especificidad. Es claro que esto nada tiene que ver con la metodología y/o implementación de la encuesta, sino que aquí discutimos únicamente su presentación.

Por último, decidimos incluir el estudio de una de las variables más importantes de la encuesta con el objetivo de mostrar cómo los usuarios, de no estar prevenidos, pueden incurrir en serios problemas al utilizar las bases provistas. Pudimos remarcar dos conclusiones esenciales: (a) no es correcto utilizar los valores absolutos internos de la encuesta, tal como siempre alertó el INDEC; y (b) existen distintas alternativas para subsanar estos inconvenientes. En particular, aquí se presentaron dos de ellas. La primera, llamada “método directo”, utilizaba los datos poblacionales (proyectados) censales para luego aplicarles directamente las tasas provenientes de la EPH. Aun cuando este método pueda parecer sumamente sencillo, si pretendiéramos alcanzar aquellos valores poblacionales que responden a universos muy particulares el procedimiento puede volverse tedioso y de difícil implementación. Por tal motivo, se propuso una segunda solución, llamada “método indirecto” donde su aplicación implica la modificación sustantiva del vector de ponderadores de la base original para luego utilizarlo en la construcción de los distintos indicadores y valores poblacionales. Esta alternativa resulta sumamente sencilla de aplicar en términos operativos ya que requiere una única corrección de la encuesta. También se mostró que ambos métodos pueden resultar equivalentes.

En conclusión, este documento de índole metodológica, además de presentar un sucinto análisis de las encuestas de hogares en general y de la EPH en particular, intentó mostrar dos casos claros de problemas para los usuarios en la presentación de la información, y cuáles podrían ser las soluciones, tomando en cuenta las pérdidas que éstas implicarían. Este trabajo, por lo tanto, realiza un doble aporte. Por un lado, muestra los problemas existentes en la década de los noventa en lo que a los usuarios de la EPH refiere y, por el otro, nos permite inferir que algunos de los principales inconvenientes parecerían desaparecer conforme avanza la década de los noventa. El primero de ellos (referido a la compatibilización sincrónica) es solucionado a partir de la creación de una única base de datos, la Base Consolidada, común para todos los aglomerados del país. El segundo de ellos (referido a la construcción de valores absolutos estimados a partir de la EPH) también es resuelto a partir de 1998, tal como lo mostramos en la primera Sección del Apartado IV. Si bien esto resulta sumamente alentador para el futuro, no debemos desconocer la existencia de otros problemas, en su mayoría referidos a la compatibilización diacrónica.

Sólo resta mencionar que, si bien este trabajo analiza únicamente la década de los noventa, todo lo aquí esbozado puede servir como punto de partida para futuras investigaciones. En particular, los aspectos metodológicos y de implementación de la encuesta deberían recibir una atención especial, pues han sido poco estudiados en el pasado y resultan esenciales para el correcto usufructo de este herramental.

## **VI. Bibliografía**

---

Alsina, S., C. Badler, C. Puigsubirá, M. Vitelleschi y N. Arnesi (1998); "Datos perdidos en encuestas de hogares. Un aporte metodológico"; Primera reunión sobre estadística pública del Instituto Interamericano de Estadística.

Azqueta Oyarzún, D. (1994); "Valoración económica del medio ambiente"; McGraw Hill.

Barreiros, Ive (1998); "The household survey and the current development model"; Primera reunión sobre estadística pública del Instituto Interamericano de Estadística.

Beverly, C. (1998); "Measuring and analyzing labour market outcomes of education: the crucial role of labor force and household surveys and their design"; Primera reunión sobre estadística pública del Instituto Interamericano de Estadística.

Botta, A. y C. Pok (1987); "Lo censal y lo intercensal: hacia la integración conceptual de las estadísticas; en Los Censos del 90: características económicas de la población", INDEC-CENEP-CELADE.

Cardinali, A. (1998); "Encuesta a hogares y cuentas nacionales: la estadística social básica en la producción de información económica"; Primera reunión sobre estadística pública del Instituto Interamericano de Estadística.

CEUR, Centro de Estudios Urbanos y Regionales, (1984); "Cuestiones metodológicas relativas a la investigación sociodemográfica basada en censos y encuestas de hogares"; Cuaderno del CEUR núm. 12.

Elizalde M. L., C. Pok, A. Botta y J. M. Villareal (1974); "Encuesta permanente de hogares: marco teórico-metodológico de la investigación temática", INDEC.

Feres, J. (1998); "Hacia un sistema integrado de encuestas de hogares en los países de América Latina"; Primera reunión sobre estadística pública del Instituto Interamericano de Estadística.

Giusti, A. y J. Lindenboim (1997); "Cambio "técnico" en el censo de 1991: cuantificación de su efecto en las tasas de actividad y evaluación de los resultados sobre las características económicas de la población"; presentado en las IV Jornadas Argentinas de Estudios de Población, AEPA.

IASI e INDEC (1998); "Encuesta a hogares: reformulación de la encuesta permanente de hogares de Argentina"; Primera reunión sobre estadística pública del Instituto Interamericano de Estadística.

INDEC (1996); "Encuesta Permanente de Hogares. Códigos utilizados en la base de datos. Proceso preliminar -R2-". INDEC, instructivo de la Encuesta Permanente de Hogares

INDEC (1998); "Encuesta Permanente de Hogares. Base Usuaria Ampliada (BUA)". INDEC, instructivo de la Encuesta Permanente de Hogares.

INDEC (1999a); "Encuesta Permanente de Hogares. Base Usuaria". INDEC, instructivo de la Encuesta Permanente de Hogares.

INDEC (1999b); "Marco teórico y metodológico de la investigación temática. Encuesta Permanente de Hogares". INDEC, informes de prensa de la Encuesta Permanente de Hogares, período 1989/1995.

INDEC (1999c); "Estimaciones de la población de los aglomerados urbanos que cubre la Encuesta Permanente de Hogares período 1991-2000. Reseña metodológica y resultados". Documento de Trabajo EPH4, Dirección de Estadísticas Poblacionales.

Lindenboim, Javier (con la colaboración de Leandro Serino) (2000). "Mercados de Trabajo Urbanos en la Argentina de los 90". Cuadernos del CEPED 5, Facultad de Ciencias Económicas, UBA.

Medina, Fernando (1998); "Tamaño óptimo de muestra en encuestas de propósitos múltiples"; Primera reunión sobre estadística pública del Instituto Interamericano de Estadística.

- Morgenstern, O. (1963); "On the accuracy of economic observations"; Princeton University Press.
- Müller, A (1998); "Economía Descriptiva"; Catálogos.
- Nagel, E. (1961); "La estructura de la ciencia"; Editorial Paidós.
- ONU (1993); "Sistema de Cuentas Nacionales 1993"; ONU.
- Salvia, A. y colaboradores (1998); "Aproximaciones al tema de la desigualdad social a partir de la EPH. Un estudio de la evolución y distribución del ingreso en el Gran Buenos Aires: 1980-1996". Primera reunión sobre estadística pública del Instituto Interamericano de Estadística.
- Wainerman, C. y R. Sautu (compiladoras) (1997); "La trastienda de la investigación"; Editorial de Belgrano.

## **VII. Anexo: Cuadros estadísticos**

---

A continuación se adjuntan los principales Cuadros que se utilizaron a lo largo de este trabajo. Los mismos han sido levemente modificados con el único fin de lograr una mayor claridad expositiva.

Los primeros dos (A.1 y A.2) son oficiales y corresponden a las proyecciones de población por aglomerado para la década a partir de los datos del Censo de 1991. Únicamente se han eliminado los aglomerados que no se utilizaron en este trabajo. Nótese que mientras los datos del año 1991 corresponden a los verdaderos valores de población, en los años subsiguientes se presentan los datos correspondientes a proyecciones censales, es decir, a estimaciones del verdadero valor poblacional.

El tercer Cuadro (A.3) incluye las estimaciones de población calculadas independientemente a partir de las bases de EPH. Este Cuadro es de elaboración propia a partir de las bases oficiales. Aquí también han sido eliminados los aglomerados no utilizados en el análisis. Obsérvese que aquí todos los datos corresponden a estimaciones a partir de la muestra de la encuesta.

Por último, se incluye el Cuadro A.4 donde se analizan los ponderadores de los principales aglomerados para los años impares de la década de los noventa. En el mismo se incluye la cantidad de casos distintos (cantidad), el valor mínimo (Min), el valor máximo (Max), el valor medio (Med) y el desvío estándar (Des). En este Cuadro se han cruzado dichos datos por estado de ocupación (ocupado, desocupado, inactivo) para los aglomerados seleccionados.

El objetivo de este último Cuadro es mostrar la gran heterogeneidad existente en el vector de ponderadores. Más aun, podemos destacar algunos comportamientos sumamente interesantes, por ejemplo, nótese que el valor medio de los ponderadores de los desocupados es menor que el valor medio de los ponderadores de los otros estados de ocupación. También el desvío varía sensiblemente según el aglomerado en estudio. Por último, sólo para mencionar algunos casos interesantes, obsérvese que los inactivos reciben ponderadores en promedio mayores.



**(A.1) Proyecciones censales del INDEC, 1996<sup>22</sup>. Onda de Octubre.**

AGLOMERADOS	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
<b>TOTAL (25 aglomerados)</b>	<b>19.195.706</b>	<b>19.519.675</b>	<b>19.849.695</b>	<b>20.179.968</b>	<b>20.503.886</b>	<b>20.826.706</b>	<b>21.148.888</b>	<b>21.470.081</b>	<b>21.789.953</b>	<b>22.108.157</b>
<b>1 GRAN BUENOS AIRES</b>	<b>10.913.945</b>	<b>11.043.818</b>	<b>11.175.131</b>	<b>11.308.177</b>	<b>11.435.278</b>	<b>11.562.569</b>	<b>11.689.212</b>	<b>11.815.049</b>	<b>11.939.934</b>	<b>12.063.718</b>
CAP. FEDERAL	2.955.951	2.963.898	2.971.734	2.978.970	2.985.144	2.990.318	2.994.833	2.998.858	3.001.760	3.004.118
PARTIDOS G.B.A.	7.957.995	8.079.920	8.203.397	8.327.207	8.450.132	8.572.252	8.694.379	8.816.393	8.938.174	9.059.602
<b>RESTO</b>	<b>8.281.761</b>	<b>8.475.857</b>	<b>8.674.564</b>	<b>8.871.791</b>	<b>9.068.608</b>	<b>9.264.139</b>	<b>9.459.676</b>	<b>9.655.032</b>	<b>9.850.019</b>	<b>10.044.439</b>
2 GRAN LA PLATA	635.744	644.548	653.464	662.405	671.281	680.100	688.918	697.729	706.523	715.292
3 GRAN BAHIA BLANCA	266.312	270.484	274.709	278.945	283.151	287.329	291.508	295.683	299.849	304.004
4 GRAN ROSARIO	1.123.123	1.141.032	1.159.170	1.177.356	1.195.412	1.213.350	1.231.289	1.249.211	1.267.099	1.284.936
5 SANTA FE	399.473	406.810	414.240	421.690	429.087	436.435	443.784	451.126	458.454	465.760
6 PARANA	209.224	214.258	219.357	224.469	229.545	234.587	239.630	244.668	249.697	254.710
7 POSADAS	215.208	222.611	230.109	237.626	245.090	252.505	259.920	267.328	274.722	282.095
8 GRAN RESISTENCIA	296.703	304.902	313.206	321.532	329.799	338.011	346.224	354.430	362.620	370.786
9 COMODORO RIVADAVIA	127.825	131.016	134.248	137.489	140.706	143.902	147.099	150.292	153.480	156.658
10 GRAN MENDOZA	781.764	799.207	816.873	834.586	852.172	869.643	887.116	904.572	921.995	939.367
12 CORRIENTES	263.547	272.419	281.404	290.414	299.359	308.248	317.133	326.012	334.874	343.710
13 GRAN CORDOBA	1.179.578	1.201.144	1.222.984	1.244.882	1.266.624	1.288.224	1.309.825	1.331.406	1.352.946	1.374.423
15 FORMOSA	151.712	157.702	163.769	169.851	175.890	181.890	187.890	193.884	199.867	205.833
17 NEUQUEN	189.436	199.014	208.715	218.441	228.099	237.693	247.288	256.873	266.441	275.981
18 SANTIAGO DEL ESTERO	265.454	272.428	279.490	286.571	293.601	300.586	307.571	314.549	321.514	328.459
19 SAN SALV. DE JUJUY	222.019	229.061	236.192	243.342	250.441	257.493	264.546	271.593	278.626	285.638
20 RIO GALLEGOS	65.543	68.156	70.804	73.457	76.093	78.711	81.330	83.945	86.556	89.159
22 GRAN CATAMARCA	123.980	127.777	131.622	135.477	139.305	143.108	146.910	150.710	154.502	158.283
23 SALTA	375.160	387.138	399.269	411.432	423.508	435.506	447.504	459.491	471.455	483.384
25 LA RIOJA	106.520	110.762	115.058	119.365	123.642	127.891	132.139	136.384	140.621	144.848
26 SAN LUIS	115.927	120.752	125.638	130.536	135.400	140.233	145.065	149.893	154.712	159.516
27 GRAN SAN JUAN	355.146	360.411	366.744	371.090	376.399	381.672	386.946	392.215	397.474	402.718
29 GRAN TUCUMAN	661.187	675.192	690.490	705.829	721.058	736.187	751.318	766.434	781.521	796.565
30 SANTA ROSA	82.281	85.212	88.200	91.195	94.170	97.125	100.080	103.032	105.979	108.917
31 USHUAIA Y R.GRANDE	68.895	73.821	78.809	83.811	88.776	93.710	98.643	103.572	108.492	113.397

Fuente: INDEC, Programa de Análisis DemoGráfico. Versión revisada del D.T.62bis. a partir de las áreas de cobertura definidas por el Censo de Población de 1991.

<sup>22</sup> Se han eliminado los aglomerados que no fueron utilizados en el presente estudio con el objeto de facilitar la lectura del cuadro.

**(A.2) Proyecciones censales del INDEC, 1999<sup>23</sup>. Onda de Octubre.**

AGLOMERADOS	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
<b>TOTAL (25 aglomerados)</b>	<b>19.662.016</b>	<b>19.981.178</b>	<b>20.302.265</b>	<b>20.623.398</b>	<b>20.943.940</b>	<b>21.265.709</b>	<b>21.588.572</b>	<b>21.911.500</b>	<b>22.233.422</b>	<b>22.553.973</b>
<b>1 GRAN BUENOS AIRES</b>	<b>11.355.748</b>	<b>11.479.960</b>	<b>11.602.556</b>	<b>11.724.676</b>	<b>11.847.595</b>	<b>11.973.008</b>	<b>12.099.510</b>	<b>12.226.270</b>	<b>12.352.412</b>	<b>12.477.770</b>
CAP. FEDERAL	2.955.951	2.963.898	2.971.734	2.978.970	2.985.144	2.990.318	2.994.833	2.998.656	3.001.760	3.004.116
PARTIDOS G.B.A.	8.399.797	8.516.062	8.630.822	8.745.706	8.862.451	8.982.690	9.104.677	9.227.614	9.350.652	9.473.654
<b>RESTO</b>	<b>8.306.269</b>	<b>8.501.218</b>	<b>8.699.709</b>	<b>8.898.722</b>	<b>9.096.345</b>	<b>9.292.701</b>	<b>9.489.062</b>	<b>9.685.230</b>	<b>9.881.010</b>	<b>10.076.203</b>
2 GRAN LA PLATA	635.744	644.548	653.464	662.405	671.281	680.100	688.918	697.729	706.523	715.292
3 GRAN BAHIA BLANCA	266.312	270.484	274.709	278.945	283.151	287.329	291.508	295.683	299.849	304.004
4 GRAN ROSARIO	1.123.123	1.141.032	1.159.170	1.177.356	1.195.412	1.213.350	1.231.289	1.249.211	1.267.099	1.284.936
5 SANTA FE	408.930	416.440	424.046	431.672	439.243	446.766	454.288	461.804	469.306	476.785
6 PARANA	214.170	219.324	224.542	229.775	234.972	240.133	245.295	250.452	255.600	260.732
7 POSADAS	215.208	222.611	230.109	237.626	245.090	252.505	259.920	267.328	274.722	282.095
8 GRAN RESISTENCIA	296.703	304.902	313.206	321.532	329.799	338.011	346.224	354.430	362.620	370.786
9 COMODORO RIVADAVIA	127.825	131.016	134.248	137.489	140.706	143.902	147.099	150.292	153.480	156.658
10 GRAN MENDOZA	781.764	799.207	816.873	834.586	852.172	869.643	887.116	904.572	921.995	939.367
12 CORRIENTES	263.547	272.419	281.404	290.414	299.359	308.246	317.133	326.012	334.874	343.710
13 GRAN CORDOBA	1.179.578	1.201.144	1.222.984	1.244.882	1.266.624	1.288.224	1.309.825	1.331.406	1.352.946	1.374.423
15 FORMOSA	151.712	157.702	163.769	169.851	175.890	181.890	187.890	193.884	199.867	205.833
17 NEUQUEN	189.436	199.014	208.715	218.441	228.099	237.693	247.288	256.873	266.441	275.981
18 SANTIAGO DEL ESTERO	265.454	272.428	279.490	286.571	293.601	300.586	307.571	314.549	321.514	328.459
19 SAN SALV. DE JUJUY	222.019	229.061	236.192	243.342	250.441	257.493	264.546	271.593	278.626	285.638
20 RIO GALLEGOS	65.543	68.156	70.804	73.457	76.093	78.711	81.330	83.945	86.556	89.159
22 GRAN CATAMARCA	134.103	138.441	142.780	147.119	151.458	155.796	160.127	164.444	168.738	173.002
23 SALTA	375.160	387.138	399.269	411.432	423.508	435.506	447.504	459.491	471.455	483.384
25 LA RIOJA	106.520	110.762	115.058	119.365	123.642	127.891	132.139	136.384	140.621	144.846
26 SAN LUIS	115.927	120.752	125.638	130.536	135.400	140.233	145.065	149.893	154.712	159.516
27 GRAN SAN JUAN	355.146	360.411	365.744	371.090	376.399	381.672	386.946	392.215	397.474	402.718
29 GRAN TUCUMAN	661.187	675.192	690.490	705.829	721.058	736.187	751.318	766.434	781.521	796.565
30 SANTA ROSA	82.261	85.212	88.200	91.195	94.170	97.125	100.080	103.032	105.979	108.917
31 USHUAIA Y R.GRANDE	68.895	73.821	78.809	83.811	88.776	93.710	98.643	103.572	108.492	113.397

Fuente: INDEC, Programa de Análisis Demográfico. Versión revisada del D.T.62bis. a partir de las áreas de cobertura definidas por el Censo de Población de 1991.

<sup>23</sup> Se han eliminado los aglomerados que no fueron utilizados en el presente estudio con el objeto de facilitar la lectura del cuadro.

**(A.3) Total de población urbana según EPH-INDEC. Onda de Octubre.**

AGLOMERADOS	1991	1992	1993	1994	1995	1996	1997	1998	1999
<b>TOTAL (25 aglomerados)</b>	<b>19.218.367</b>	<b>19.890.798</b>	<b>20.032.954</b>	<b>19.989.459</b>	<b>20.521.652</b>	<b>19.529.606</b>	<b>21.133.512</b>	<b>21.886.436</b>	<b>21.927.379</b>
<b>1 GRAN BUENOS AIRES</b>	<b>11.315.246</b>	<b>11.662.050</b>	<b>11.690.471</b>	<b>11.474.122</b>	<b>11.435.046</b>	<b>11.561.652</b>	<b>11.689.330</b>	<b>12.227.177</b>	<b>12.352.050</b>
<b>RESTO</b>	<b>7.903.121</b>	<b>8.228.748</b>	<b>8.342.483</b>	<b>8.515.337</b>	<b>9.086.606</b>	<b>7.967.954</b>	<b>9.444.182</b>	<b>9.659.259</b>	<b>9.575.329</b>
2 GRAN LA PLATA	643.913	676.347	668.271	681.536	670.811	678.135	687.665	696.737	705.573
3 GRAN BAHIA BLANCA	272.429	281.179	289.395	295.047	283.282	287.351	290.909	295.478	N/D
4 GRAN ROSARIO	1.142.216	1.155.334	1.184.707	1.196.550	1.194.770	1.212.361	1.231.592	1.247.529	1.266.712
5 SANTA FE	382.358	377.717	408.603	390.884	430.084	436.011	442.795	450.727	469.805
6 PARANA	215.310	212.274	212.989	212.148	229.672	234.597	240.019	244.555	255.945
7 POSADAS	207.819	232.914	236.725	256.627	245.485	252.736	258.785	267.360	274.246
8 GRAN RESISTENCIA	298.303	295.943	300.654	290.308	333.303	337.787	345.376	354.449	362.361
9 COMODORO RIVADAVIA	N/D	132.799	129.824	128.488	141.092	143.957	146.112	149.296	153.918
10 GRAN MENDOZA	776.017	800.888	806.400	807.300	851.078	869.344	887.497	905.028	921.946
12 CORRIENTES	228.947	228.668	240.009	242.875	297.062	309.288	317.031	325.772	334.566
13 GRAN CORDOBA	1.169.129	1.177.943	1.181.221	1.244.999	1.266.420	N/D	1.310.033	1.331.057	1.353.285
15 FORMOSA	146.799	164.372	158.063	179.679	175.760	181.989	187.989	194.415	199.006
17 NEUQUEN	174.431	172.400	185.026	189.956	228.306	237.288	247.002	257.108	266.356
18 SANTIAGO DEL ESTERO	238.195	246.080	244.805	237.271	293.923	300.349	307.213	314.747	321.523
19 SAN SALV. DE JUJUY	193.927	192.830	200.062	195.507	250.988	257.493	262.874	272.174	277.983
20 RIO GALLEGOS	59.668	59.224	58.279	61.482	76.144	78.100	81.260	84.098	86.276
22 GRAN CATAMARCA	119.123	122.978	124.422	128.495	139.046	143.456	146.667	164.227	170.240
23 SALTA	343.214	359.037	353.496	389.247	422.754	435.122	447.890	459.889	471.403
25 LA RIOJA	106.487	103.408	109.978	111.632	124.011	126.835	131.296	136.023	140.442
26 SAN LUIS	106.343	115.319	116.912	118.059	136.379	140.605	145.146	150.170	154.416
27 GRAN SAN JUAN	320.145	329.482	337.528	351.421	375.870	380.327	385.308	391.349	395.793
29 GRAN TUCUMAN	614.761	640.099	641.821	643.476	737.434	733.286	746.012	760.454	778.837
30 SANTA ROSA	79.506	82.308	83.145	84.002	94.750	96.732	100.113	103.273	105.971
31 USHUAIA Y R.GRANDE	64.081	69.205	70.148	78.348	88.182	94.805	97.598	103.344	108.726

Fuente: Elaboración propia en base a EPH-INDEC.

**(A.4) Principales características de la variable pondera (EPH-INDEC), según aglomerado<sup>24</sup> y estado de ocupación. Onda de Octubre.**

AGLOMERADO	ESTADO	PONDERADORES																								
		1991					1993					1995					1997					1999				
		Cant dad	Min	Max	Med	Des	Cant dad	Min	Max	Med	Des	Cant dad	Min	Max	Med	Des	Cant dad	Min	Max	Med	Des	Cant dad	Min	Max	Med	Des
1	G.B.A. Ocupados	4174	819	1553	1048	103	4453	845	1519	1027	88	4215	869	2478	991	80	4396	904	1232	1027	64	4580	895	1315	1055	73
	Desocupados	235	819	1479	1033	93	482	885	1338	1012	77	893	869	1392	981	71	742	904	1232	1016	60	793	895	1315	1050	70
	Inactivos	6435	819	1553	1041	96	6497	845	1519	1020	86	6492	869	2478	983	76	6280	904	1232	1022	60	6386	895	1315	1047	67
2	GRAN LA PLATA Ocupados	1315	169	223	178	10	1311	172	238	189	19	1363	160	189	168	7	1480	161	181	171	5	758	281	381	330	27
	Desocupados	99	169	194	177	8	93	172	238	188	20	337	160	189	167	7	256	161	181	170	5	126	281	381	324	26
	Inactivos	2211	169	223	177	10	2139	172	238	188	18	2305	160	189	167	7	2285	161	181	171	5	1261	281	381	329	27
4	GRAN ROSARIO Ocupados	1452	277	308	287	8	1399	279	325	295	13	1434	257	313	269	16	1299	270	594	323	70	856	438	616	486	53
	Desocupados	151	277	308	286	8	188	279	325	293	12	322	257	313	265	11	206	270	594	309	58	176	438	616	476	48
	Inactivos	2384	277	308	286	8	2441	279	325	294	12	2699	257	313	268	14	2344	270	594	319	64	1581	438	616	485	54
5	SANTA FE Ocupados	1573	87	91	89	1	1533	89	96	90	2	1366	94	107	98	4	1354	97	116	104	5	1106	90	285	133	55
	Desocupados	198	87	91	89	2	269	89	96	91	2	357	94	107	97	4	264	97	116	103	3	173	90	285	131	52
	Inactivos	2525	87	91	89	1	2712	89	96	91	2	2688	94	107	97	3	2662	97	116	103	5	2305	90	285	130	53
10	GRAN MENDOZA Ocupados	1397	209	209	209	0	1366	210	210	210	0	1392	214	214	214	0	1350	233	233	233	0	1422	225	260	238	10
	Desocupados	66	209	209	209	0	66	210	210	210	0	100	214	214	214	0	87	233	233	233	0	105	225	260	236	9
	Inactivos	2250	209	209	209	0	2408	210	210	210	0	2485	214	214	214	0	2372	233	233	233	0	2363	225	260	237	9
13	GRAN CORDOBA Ocupados	1626	251	271	257	4	1634	253	285	259	5	1338	275	387	305	14	1239	268	494	370	97	1169	394	440	409	16
	Desocupados	93	251	271	258	4	120	253	285	258	5	256	275	387	303	14	238	268	494	369	96	185	394	440	408	15
	Inactivos	2826	251	271	257	4	2817	253	285	258	5	2562	275	387	305	14	2061	268	494	371	97	1958	394	440	408	15
19	SAN S. DE JUJUY Ocupados	1249	49	56	50	2	1261	47	52	48	1	1309	47	98	54	15	789	97	120	99	5	805	93	112	97	4
	Desocupados	42	49	56	50	2	87	47	52	48	2	187	47	98	53	15	146	97	120	98	3	156	93	112	96	3
	Inactivos	2617	49	56	50	2	2798	47	52	48	1	3117	47	98	55	16	1726	97	120	99	5	1924	93	112	96	4
23	SALTA Ocupados	1679	62	89	66	4	1781	64	72	65	1	963	128	151	135	6	1015	135	174	144	10	1145	139	163	143	7
	Desocupados	93	62	85	65	0	204	64	72	65	1	180	128	151	135	5	170	135	174	143	9	189	139	163	142	5
	Inactivos	3462	62	89	66	4	3480	64	72	65	1	1986	128	151	135	5	1942	135	174	143	10	1966	139	163	143	7
27	GRAN SAN JUAN Ocupados	1256	80	97	83	4	1400	78	95	82	4	1368	82	105	87	5	1446	89	96	91	2	1410	91	111	95	5
	Desocupados	94	80	97	83	5	90	78	95	81	3	222	82	105	87	5	137	89	96	91	1	160	91	111	94	3
	Inactivos	2501	80	97	83	4	2645	78	95	82	4	2713	82	105	87	5	2654	89	96	91	2	2605	91	111	95	4
29	GRAN TUCUMAN Ocupados	1042	185	218	197	11	1105	182	205	191	7	894	233	274	241	12	976	229	266	246	10	1020	227	329	250	25
	Desocupados	137	185	218	194	10	149	182	205	190	7	214	233	274	239	8	181	229	266	243	9	195	227	329	247	23
	Inactivos	1954	185	218	196	10	2114	182	205	190	7	1956	233	274	241	12	1884	229	266	245	10	1907	227	329	249	25

Fuente: Elaboración propia en base a EPH-INDEC.

<sup>24</sup> Se seleccionaron aquellos aglomerados con más de 400.000 habitantes.