

Avance proyecto de investigacion

Juan Gabriel Juaraa

28/11/2020

Introducción

El proyecto de investigación busca indagar en las características de la distribución del ingreso de los trabajadores durante los últimos años, prestando especial atención a las características que dicha distribución asume desde un análisis sectorial del retorno educativo. En particular se buscará evidencia de si existe un diferencial específico estructural de ingresos para los trabajadores de sectores productivos determinados y si ese diferencial es sensible o no al nivel educativo de los trabajadores. Se procura realizar el análisis sectorial a un nivel de desagregación tal que permita identificar ramas de actividad con comportamientos laborales específicos. ~~Se evaluará en este sentido una posible homologación entre conjuntos sectoriales del CAES según lo relevado en la EPH y la estructura de ramas que organiza la negociación sindical en la Argentina.~~

Interrogantes

Objetivos

Hipótesis

Antecedentes

Revisar Citas y Referencias

Adriana Marshall en un trabajo pionero de 1975 titulado “Mercado de trabajo y crecimiento de los salarios en la Argentina” señaló que si bien el comportamiento de los salarios seguía entre 1950 y 1970 un patrón similar relacionado a las fuerzas intervinientes en un mercado laboral global - esto es, la presión general en el mercado de trabajo del nivel de desempleo y precariedad del trabajo -, se observaba una creciente diferenciación de los niveles salariales por actividad y también a nivel interindustrial. Desde el periodo analizado por Marshall han transcurrido importantes transformaciones en la economía argentina, pero el análisis de distribución del ingreso entre los trabajadores desde una perspectiva sectorial ha perdido centralidad. El campo académico se ha centrado con frecuencia en la tipología de los sectores formales y los informales o precarios como divisiones del mercado de trabajo, y en el nivel educativo de los trabajadores como atributo individual explicativo. Estas categorías de análisis han demostrado tener relevancia, Benza y Calvi (2006) demostraron el comportamiento diferencial negativo de los asalariados precarios especialmente a partir de mediados de la década de 1990, aunque con un breve periodo previo en sentido contrario. En cuanto a la importancia del nivel educativo, Groisman (2008) ha aportado evidencia de una asociación negativa entre el bajo nivel educativo del jefe de hogar y los ingresos de los hogares, especialmente afectados por la inestabilidad y variabilidad de los ingresos. Sin embargo, Beccaria (2006) ha examinado los retornos de la educación sobre los ingresos de los trabajadores a la luz del creciente nivel educativo promedio de la fuerza laboral argentina entre 1974 y 2004, y a partir de dicho examen ha propuesto matizar o al menos complementar la explicación educativa del nivel de los salarios con otras dimensiones posibles. Beccaria en particular hace notar que el peso de las negociaciones sindicales pudo haber hecho disminuir la diferenciación de los salarios de los trabajadores en base a su nivel educativo, al menos durante los primeros años de la década del 2000. Esta línea de exploración señalada por Beccaria resulta coincidente con los planteos citados de Marshall, debido a la importancia de la estructura de

negociaciones paritaria por rama de actividad que caracteriza al sistema sindical argentino (Trajtemberg, Medwid., & Senén González, 2010). A su vez el análisis sectorial del mercado de trabajo permite revisar los problemas entorno a las estrategias nacionales de desarrollo y la cuestión de la heterogeneidad estructural. Si bien esta relación temática ha sido considerada, por ejemplo en el trabajo de Salvia y Vera (2010), allí la dimensión de rama de actividad fue considerada sólo en tanto variable control y de forma demasiado agregada como para que sea posible obtener información sobre la relación entre dicha variable, el nivel educativo y el ingreso de los trabajadores.

Marco Teórico

Metodología y Fuentes

Preparación de las bases

A partir de las bases de microdatos de la Encuesta Permanente de Hogares publicadas por el INDEC para cada uno de los cuatro trimestres del año 2017 se procede a unificarlas en una base de datos única. Luego se selecciona de la base de datos única a aquellos individuos que pertenecen a hogares que fueron entrevistados en un sólo trimestre del año de referencia. Para que la base de datos única posea una distribución uniforme de casos a través de los cuatro trimestres se seleccionaron 2,137 hogares de cada uno: en el caso del tercer trimestre el valor corresponde al total de hogares que sólo fueron entrevistados en ese trimestre, para el resto de los trimestres se selecciona al azar una muestra de 2137 casos sobre el total de hogares entrevistados únicamente en el trimestre.

A continuación se expresa el código utilizado:

```
#carga de bases
eph2017 <- list()
eph2017[[1]] <- readRDS("~/Documentos/rProyectos/seminarioPhilipps/eph_2017_1.RDS")
eph2017[[2]] <- readRDS("~/Documentos/rProyectos/seminarioPhilipps/eph_2017_2.RDS")
eph2017[[3]] <- readRDS("~/Documentos/rProyectos/seminarioPhilipps/eph_2017_3.RDS")
eph2017[[4]] <- readRDS("~/Documentos/rProyectos/seminarioPhilipps/eph_2017_4.RDS")
#suma de bases
eph2017 <- bind_rows(eph2017[[1]][[5]], eph2017[[2]][[5]], eph2017[[3]][[5]], eph2017[[4]][[5]])

#creo una variable "idhogar" que une el CODUSU y el NRO de hogar en una sola linea de caracteres
eph2017 <- eph2017 %>%
  mutate(idhogar = paste0(CODUSU, NRO_HOGAR))

# creo una tabla que expresa en qué trimestres aparece cada hogar
tablahogares <- eph2017 %>%
  select(idhogar, TRIMESTRE) %>%
  #agrupo primero por idhogar y dps por trimestre
  pivot_wider(names_from = TRIMESTRE, values_from = TRIMESTRE, values_fn = list(TRIMESTRE = is.vector))

# elimino na y transformo los valores en "verdadero" / "falso" segun el hogar aparezca en el trimestre
tablahogares[2:5] <- !is.na(tablahogares[2:5])

# creo variable = a la suma de veces que aparece el hogar en el año
tablahogares <- tablahogares %>%
  mutate( ntrims = tablahogares$`1`+tablahogares$`2`+tablahogares$`3`+tablahogares$`4`)

# guardo la seleccion realizada
tablahogares %>%
  write_csv(., path = "tablahogares.csv")
```

Se observa en la tabla siguiente que se registran aproximadamente 18,500 hogares de forma homogénea para cada trimestre, sin embargo los hogares que fueron entrevistados una sola vez durante el año de referencia se distribuyen desigualmente entre los cuatro trimestres. Esto se debe al efecto que produce la aplicación del criterio de una sola observación durante el año calendario sobre el esquema de rotación de la muestra panel de la EPH. **Evaluar de qué forma específica se produce esta distribución**

Tabla 1. Cantidad de hogares entrevistados por trimestre. Año 2017

```

tablahogares <- read_csv("tablahogares.csv")

tablahogares %>%
  group_by(ntrims) %>%
  summarise(t1 = sum(`1`),
            t2 = sum(`2`),
            t3 = sum(`3`),
            t4 = sum(`4`)) %>%

  t(.) %>%
  as.data.frame() %>%
  .[2:5,1:2] %>%
  mutate(porcentaje = 100*V1/(V1+V2)) %>%
  knitr::kable(.,row.names = T ,digits = 2, format = "latex",
               col.names = c("Hogares entrevistados una vez",
                             "Hogares entrevistados dos veces",
                             "Porcentaje de Hogares entrevistados una vez sobre el total"))

kableExtra::kable_styling(latex_options = c("scale_down"))

```

	Hogares entrevistados una vez	Hogares entrevistados dos veces	Porcentaje de Hogares entrevistados una vez sobre el total
1	6240	12237	33.77
2	2265	16606	12.00
3	2137	16669	11.36
4	6493	12300	34.55

Como se observa para obtener una base de datos de hogares anual con igual representación de cada trimestre y sin repetición de hogares no se puede seleccionar más de 2137 hogares en cada trimestre, límite determinado por el valor del tercer trimestre.

```

# selecciono al azar para cada trimestre solo los hogares que aparecen una sola vez durante el año de o
set.seed(25) # determino el seed del RNG a fin de replicabilidad
eph1 <- tablahogares %>%
  filter(`1` == T & ntrims == 1) %>%
  sample_n(tbl = .,size = 2137,replace = F)
eph2 <- tablahogares %>%
  filter(`2` == T & ntrims == 1 ) %>%
  sample_n(tbl = .,size = 2137,replace = F)
eph3 <- tablahogares %>%
  filter(`3` == T & ntrims == 1) %>%
  sample_n(tbl = .,size = 2137,replace = F)
eph4 <- tablahogares %>%
  filter(`4` == T & ntrims == 1) %>%
  sample_n(tbl = .,size = 2137,replace = F)

# creo una tabla con la lista de hogares seleccionados para cada trim
tablahogares <- bind_rows(eph1,eph2,eph3,eph4)

# Recorto el total de individuos registrados por la EPH en 2017 según la selección de hogares precedent

```

```
eph2017 <- semi_join(eph2017, tablahogares)
write_rds(eph2017, path = "eph2017.RDS", compress = "gz")
```

La base de datos única resultante tiene `{r} nrow(eph2017)` observaciones individuales que se distribuyen de la siguiente manera en los trimestres:

Tabla 2. Distribución de observaciones individuales según trimestre en la Base de Datos Única conformada. Año 2017

```
eph2017 %>%
  group_by(TRIMESTRE) %>%
  summarise("Observaciones individuales" = n())
```

```
## # A tibble: 4 x 2
##   TRIMESTRE `Observaciones individuales`
##       <int>                <int>
## 1         1                6608
## 2         2                5878
## 3         3                5838
## 4         4                6458
```

A continuación se procedió a seleccionar del total de casos a la población específica que es objeto de estudio, es decir, personas de 18 años y más que se hallaban ocupados y encuadran en la categoría de “obreros” al momento de la entrevista. **PREGUNTAR: ¿y si en vez de seleccionar N hogares de cada trimestre, selecciono N individuos de 18 años y , ocupados y obreros?**

```
eph2017 <- eph2017 %>%
  filter(CH06 >= 18 & ESTADO == 1 & CAT_OCUP == 3)
```

Realizada la selección por edad, condición de actividad y categoría ocupacional se obtienen `{r} nrow(eph2017)` observaciones individuales distribuidas de la siguiente manera entre los cuatro trimestres.

Tabla 3. Distribución de observaciones individuales de sujetos de 18 años y más, ocupados y de categoría ocupacional obrera. Año 2017

```
eph2017 %>%
  group_by(TRIMESTRE) %>%
  summarise("Observaciones individuales" = n())
```

```
## # A tibble: 4 x 2
##   TRIMESTRE `Observaciones individuales`
##       <int>                <int>
## 1         1                1971
## 2         2                2008
## 3         3                1996
## 4         4                2068
```

Se procedió además a aplicar tres funciones de etiquetado a las variables: `organize_labels`, `organize_caes` y `organize_cno` del paquete “eph” (**ver referenciar**). La primera aplica las etiquetas descriptivas de cada variable según el diseño de registro de la EPH. Las dos restantes trabajan sobre los datos correspondientes a los ítems PP04B_COD y PP04D_COD de las bases de microdatos de la EPH, correspondientes a los códigos de cuatro dígitos del Clasificador de Actividades Económicas para Encuestas Sociodemográficas del Mercosur (CAES) y al Clasificador Nacional de Ocupaciones 2001 (CNO). En ambos casos las funciones interpretan los códigos de clasificación para generar nuevas variables con los valores correspondientes: a clase, sección y división para CAES, y categoría, jerarquía, tecnología y calificación para CNO.

```
eph2017 <- eph2017 %>%
  organize_labels()
```

```
eph2017 <- eph2017 %>%  
  organize_caes()  
  
eph2017 <- eph2017 %>%  
  organize_cno()
```

Resultados Principales

Conclusiones y reflexiones finales

Bibliografía

Anexo 1