



Empalme de series

Experiencia, avances y desafíos de la Encuesta Nacional de Empleo

Subdirección Técnica / Subdirección operativa

INE

05 de octubre de 2018

ine.cl

1. Objetivo del empalme
2. Empalme Encuesta Nacional de Empleo (ENE) 2006
3. Empalme Encuesta Casen 2006
4. Empalme Nueva Encuesta Nacional de Empleo 2010
5. Lecciones y desafíos para el empalme 2020

1. Objetivo del empalme
2. Empalme Encuesta Nacional de Empleo (ENE) 2006
3. Empalme Encuesta Casen 2006
4. Empalme Nueva Encuesta Nacional de Empleo 2010
5. Lecciones y desafíos para el empalme 2020

La implementación de una **metodología de empalme** busca darle continuidad y comparabilidad a series de datos que sufren un quiebre como consecuencia de cambios en:

- Año base para la construcción de índices
- Sistemas de clasificación
- Marco muestral
- Diseño muestral
- Conceptos y metodologías
- Cobertura y periodicidad de la medición
- Parámetros externos (proyecciones de población)

1. Objetivo del empalme
2. Empalme Encuesta Nacional de Empleo (ENE) 2006
3. Empalme Encuesta Casen 2006
4. Empalme Nueva Encuesta Nacional de Empleo 2010
5. Lecciones y desafíos para el empalme 2020

Antecedentes

- Hasta el año 2005, tanto el marco muestral como las proyecciones de población utilizados en la encuesta, estaban asociados al Censo del año 1992 .
- El levantamiento del Censo 2002 permitió contar con insumos para generar un **nuevo marco muestral** y **nuevas proyecciones de población** que reflejarían los cambios demográficos y estructurales del mercado laboral ocurridos entre 1992 y 2002.
- Se diseñó un plan que contemplaba las siguientes etapas



Metodología

Se eligió la metodología puesta en práctica en el **USA Bureau of Labor Statistics** basada en un **paper de Di Natale (1983)**, en esta metodología se **distribuyen las diferencias porcentuales** entre las dos mediciones, observadas al momento del cambio de muestras, de **manera lineal** durante un período definido previamente. De esta forma, se mantienen la tendencia y los ciclos de las series originales.

Las **series suavizadas** son básicamente **las series originales**, a las cuales se les aplica **un factor de ajuste** que se obtiene calculando la razón entre los **valores de los estimadores provenientes de una muestra u otra** en el momento de la **discontinuidad**.

$$\text{Factor de ajuste muestral (FM)} = \frac{\hat{y}_{MN} - \hat{y}_{MA}}{\hat{y}_{MA}}$$

\hat{y}_{MN} : estimación con muestra nueva (al momento de la discontinuidad)

\hat{y}_{MA} : estimación con muestra antigua (al momento de la discontinuidad)

$$\text{Factor de ajuste por proyecciones (FP)} = \frac{\hat{y}_{PN} - \hat{y}_{PA}}{\hat{y}_{PA}}$$

\hat{y}_{PN} : estimación con proyecciones nuevas (al momento de la discontinuidad)

\hat{y}_{PA} : estimación con proyecciones antiguas (al momento de la discontinuidad)

Empalme Encuesta Nacional de Empleo 2006

$$y_t^E = y_t^O * \left[FM * \frac{t}{T_m} + 1 \right] * \left[FP * \frac{t}{T_p} + 1 \right]$$

y_t^E : serie empalmada

y_t^O : serie original

FM : factor de ajuste por cambio de muestra

FP : factor de ajuste por cambio de proyección

T_m : período de ajuste por cambio de muestra

T_p : período de ajuste por cambio de proyección de población

Estrategia de implementación

A partir del trimestre móvil Noviembre, Diciembre 2005 y Enero 2006 se comenzó el proceso de recolección de datos con la **nueva muestra** y, paulatinamente, se incorporaron dichos datos a **la muestra nacional del empleo**. En el trimestre móvil Abril, Mayo y Junio 2006 culminó dicho proceso, demorando en total seis meses.

Para el trimestre móvil Abril, Mayo y Junio 2006 se incorporaron las nuevas proyecciones de población basadas en el Censo 2002 y en Estadísticas vitales.

Las estimaciones publicadas a partir del trimestre Mayo, Junio y Julio 2006, no son estrictamente comparables a la información similar publicada hasta esa fecha. Lo anterior se debe a que los cambios metodológicos introducen saltos o discontinuidades en los niveles de dichas variables, por tanto surgió la necesidad de realizar el empalme.

Empalme Encuesta Nacional de Empleo 2006

Período de inicio del empalme

Examinando la estimación poblacional total que entregaba la muestra ENE a lo largo de los últimos 10 años, y comparándola con las **estimaciones poblacionales** entregadas por modelos demográficos del INE-CELADE sobre la base Censo 1992, se observó que a partir del año **2000 las cifras diferían en más del 10%** (siendo mayores en la proyección INE-CELADE), diferencial que aumentaba progresivamente.

De acuerdo a esto se optó **por iniciar el período de empalme en el año 1998** cuando la diferencia alcanzó a 8,5%, siendo el máximo internacionalmente permitido 10%.

Específicamente, se fijó el **trimestre móvil Diciembre 1998, Enero y Febrero 1999** como **periodo inicial** para el ajuste por cambio de muestra y cambio de proyecciones para mantener el mismo punto de origen.

Cuadro N° 1

Proyección Población Censo 1992 y Estimación Población a partir de la Muestra de Empleo: 1996- 2002

	Unidades	1996	1997	1998	1999	2000	2001	2002
Muestra Empleo	Miles Personas	13.159	13.346	13.561	13.828	13.204	13.091	12.840
Proyección Población por Modelos Demográficos	Miles Personas	14.419	14.622	14.822	15.018	15.211	15.402	15.589
Diferencia porcentual	(%)	8,7	8,7	8,5	7,9	13,1	15,0	17,6

Empalme Encuesta Nacional de Empleo 2006

Cuadro N° 2

Esquema de incorporación rotativa de la nueva muestra de la Encuesta Nacional de Empleo (Número de secciones)

Año	2005						2006										Muestra del trimestre móvil	Porcentaje de nueva muestra
Mes	OCT	NOV	DIC	ENE		FEB		MAR		ABR		MAY		JUN				
Trimestre móvil	MA	MA	MA	MN	MA	MN	MA	MN	MA	MN	MN*	MN	MN*	MN	MN*			
OND 2005	1.196	1.190	1.196													3.582	0,0%	
NDE 2005		1.190	1.196	593	594											3.573	16,6%	
DEF 2006			1.196	593	594	593	595									3.571	33,2%	
EFM 2006				593	594	593	595	593	598							3.566	49,9%	
FMA 2006						593	595	593	598	593	593					3.565	66,5%	
MAM 2006								593	598	593	593	593	592			3.562	83,4%	
AMJ 2006										593	593	593	592	593	594	3.558	100,0%	

Incorporación 1/6 nueva muestra (Enero 2006)

Segunda entrevista para primer 1/6 de la nueva muestra (Abril 2006)

Nuevas proyecciones CENSO 2002

MA: Muestra antigua
MN: Muestra nueva
MN*: Muestra nueva segundas visitas

1. Cálculo del factor de ajuste por cambio de muestra (FM)

$$FM = \frac{\hat{y}_{MN}^{EFM06} - \hat{y}_{MA}^{EFM06}}{\hat{y}_{MA}^{EFM06}}$$

\hat{y}_{MN}^{EFM06} : estimación con muestra nueva en el trimestre EFM 2006

\hat{y}_{MA}^{EFM06} : estimación con muestra antigua en el trimestre EFM 2006

El periodo de suavizamiento producto del efecto de cambio de muestra es de 97 periodos hacia atrás.

2. Cálculo del factor de ajuste por cambio de proyecciones de población (FP)

$$FP = \frac{\hat{y}_{PN}^{AMJ06} - \hat{y}_{PA}^{AMJ06}}{\hat{y}_{PA}^{AMJ06}}$$

\hat{y}_{PN}^{AMJ06} : estimación con proyecciones de población nuevas en el trimestre AMJ 2006

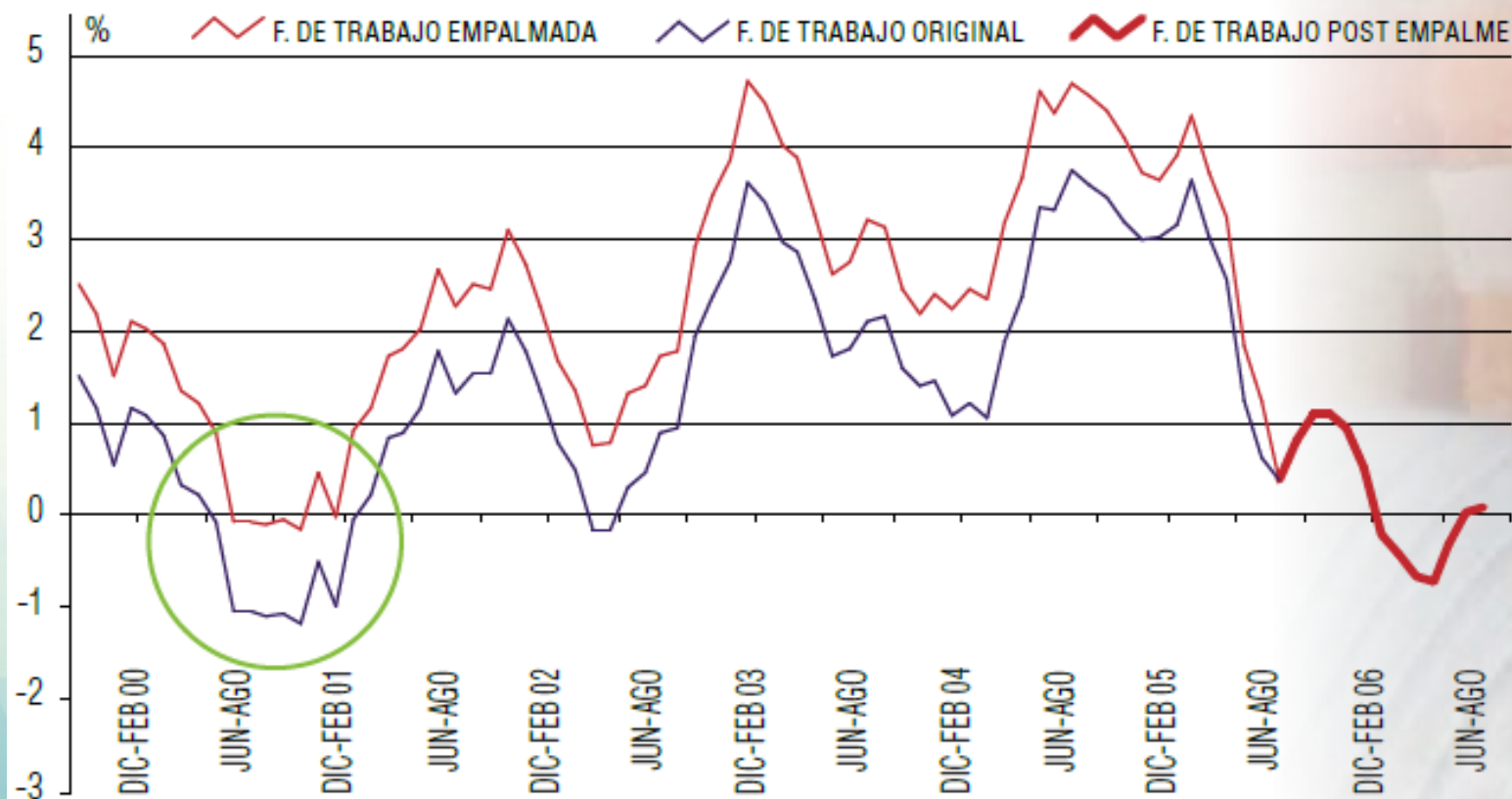
\hat{y}_{PA}^{AMJ06} : estimación con proyecciones de población antiguas en el trimestre AMJ2006

El periodo de suavizamiento producto del efecto de cambio de proyecciones de población es de 100 periodos hacia atrás.

Empalme Encuesta Nacional de Empleo 2006

Cambios en la fuerza de trabajo:

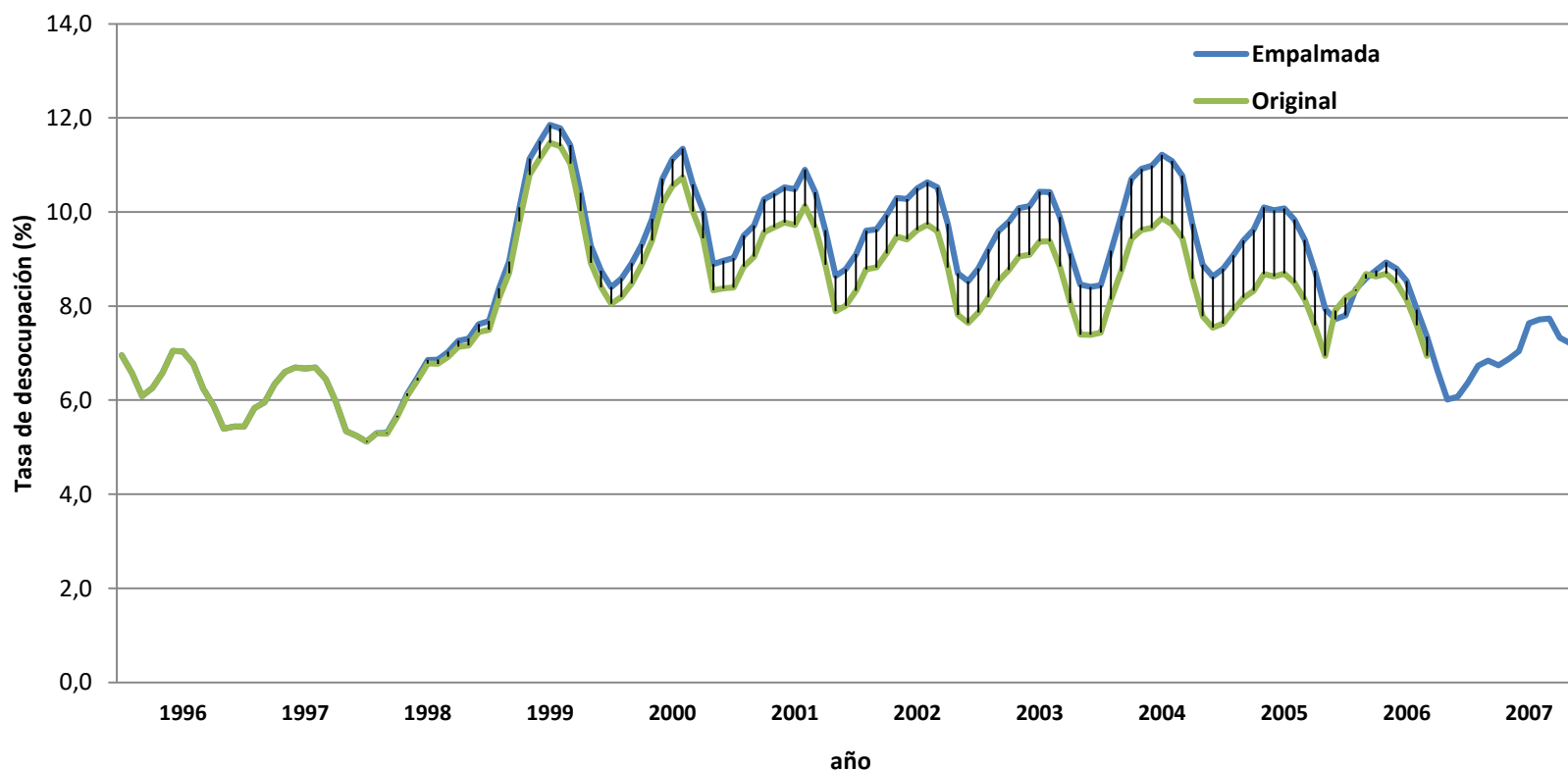
Fuerza de Trabajo: Serie Original, Empalmada y Post Empalme
2000 – 2006 (Tasa de variación 12 meses)



FUENTE: Instituto Nacional de Estadísticas.

Empalme Encuesta Nacional de Empleo 2006

Tasa de desocupación (Serie original y empalmada)
Total nacional: ambos sexos
Porcentaje



Para la aplicación de la **metodología de suavizamiento**, se decidió empalmar directamente las variables de nivel que miden la situación ocupacional, por ejemplo: **“Ocupados”, “Desocupados”, “Asalariados” y “Empleados por Cuenta Propia”**. Se optó por empalmar las series de ocupados y desocupados por separado para observar de manera independiente el comportamiento de ambos grupos. **En el año 1996, en cambio, se había aplicado el empalme a las tasas.**

La publicación de resultados del empalme se realizó durante el segundo trimestre de 2006 y su difusión implicó:

- Se generaron tabulados específicos de series empalmadas (un subconjunto de los tabulados que habitualmente publicaba el INE)
- No se modificaron las bases de datos, pues la metodología de empalme utilizada no afecta a los microdatos.
- Se publicó un Enfoque Estadístico titulado Macroeconomía de Empleo donde se analizó el cambio de niveles de los principales indicadores.

1. Objetivo del empalme
2. Empalme Encuesta Nacional de Empleo (ENE) 2006
3. Empalme Encuesta Casen 2006
4. Empalme Nueva Encuesta Nacional de Empleo 2010
5. Lecciones y desafíos para el empalme 2020

Marco Muestral

- 1996, 1998 y 2000: marco del Censo de Población y Vivienda del año 1992.
- 2003: marco desarrollado por el INE a partir de la información cartográfica del Pre-Censo 2002.
- 2006: marco “*Programa Integrado de Encuestas de Hogares*” (PIDEH) desarrollado por el INE a partir de la información cartográfica del Censo de Población y Vivienda del año 2002.

Proyecciones de población

- 1996, 1998, 2000 y 2003: Proyecciones basadas en Censo de 1992.
- 2006 en adelante: Proyecciones basadas en Censo de 2002.

Este cambio en las proyecciones de población genera un quiebre en la serie de estadísticas generadas a partir de la Encuesta Casen: pobreza, distribución del ingreso y otros indicadores sociales

Características del empalme

- Basado en cambio en las proyecciones de población.
- Cambios de muestra no implican necesidad de empalme.
- Aplicado a **nivel de microdatos**.
- El período de ajuste corresponde aproximadamente al punto central de la vigencia de las proyecciones de población originales (1996).

Ajuste del factor de expansión a nuevas proyecciones de población

$$a_{ij} = pr02_{ij} / pr92_{ij}$$

Donde:

a_{ij} : ajuste en la comuna i, zona j

$pr02_{ij}$: proyección de población en base al Censo 2002 en la comuna i, zona j

$pr92_{ij}$: proyección de población en base al Censo 1992 en la comuna i, zona j

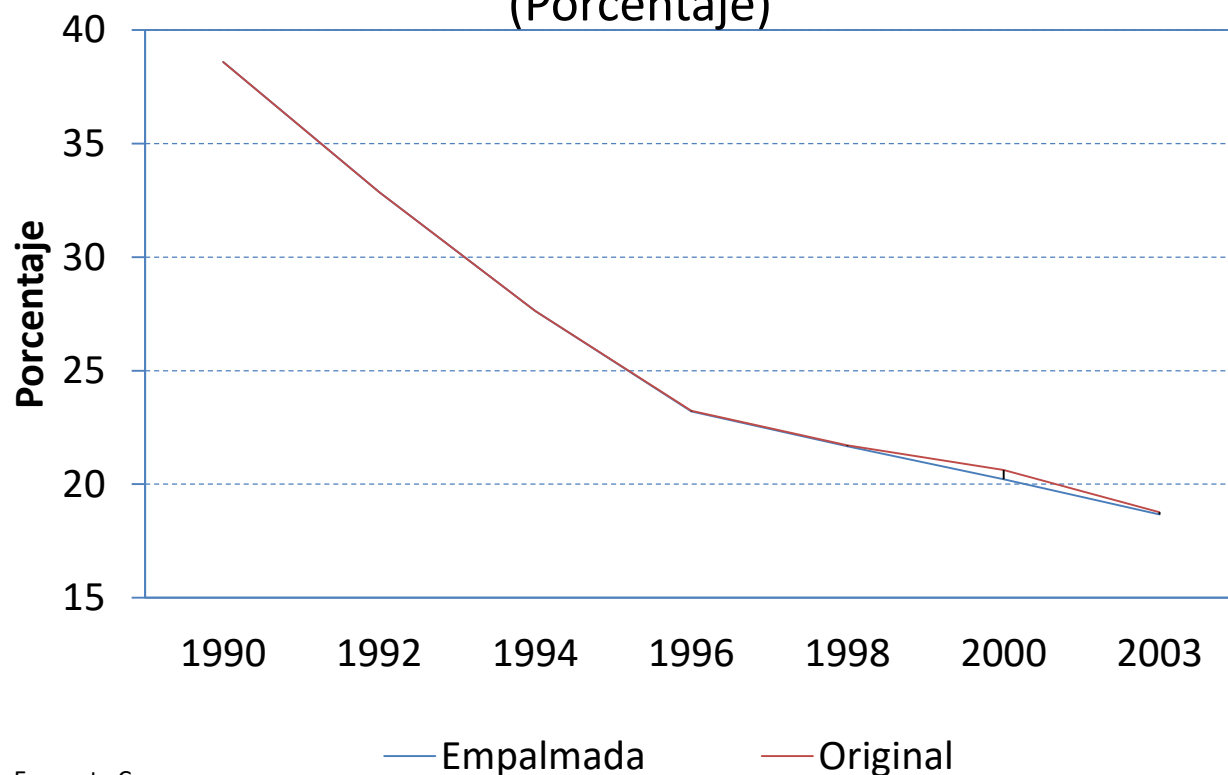
i: subíndice de comuna

j: subíndice de zona

Factor de expansión ajustado

$$FE_A_{ij} = FE_O_{ij} * a_{ij}$$

Serie original y empalmada Tasa de pobreza: Total país (Porcentaje)



Fuente: Encuesta Casen

La aplicación del empalme se realizó durante 2005, de manera que para la publicación de los primeros resultados basados en las nuevas proyecciones de población (Casen 2006), se disponía de información comparable.

Se recalcularon y publicaron los principales tabulados generados por la encuesta desde Casen 1996.

Se publicaron bases de datos con factores de expansión ajustados a las nuevas proyecciones, de manera que los usuarios podían generar series comparables de todos los indicadores y con todas las desagregaciones pertinentes.

1. Objetivo del empalme
2. Empalme Encuesta Nacional de Empleo (ENE) 2006
3. Empalme Encuesta Casen 2006
4. Empalme Nueva Encuesta Nacional de Empleo 2010
5. Lecciones y desafíos para el empalme 2020

La incorporación de estándares internacionales de medición implicó **cambios conceptuales y metodológicos** en la Encuesta Nacional de Empleo.

Cambios conceptuales

- Nueva conceptualización de la población ocupada, incorporando los criterios de 1) haber trabajado **al menos una hora en la semana** de referencia y 2) **recibir un ingreso por la actividad** realizada.
- Nueva conceptualización de población desocupada, incorporando los criterios 1) **búsqueda empleo en las últimas cuatro semanas** y 2) **disponibilidad para trabajar** en, al menos, las dos semanas siguientes a las de la entrevista.
- Clasificación de las personas fuera de la fuerza de trabajo, según razones de inactividad y según **su potencialidad de participar** en el mercado del trabajo.

Los anteriores alcances implicaron **la pérdida de comparabilidad**. El fenómeno que se está midiendo no es exactamente el mismo.

La experiencia internacional muestra que ante este tipo de cambios se requieren **metodologías o estrategias diferenciadas** que se basan en **los cambios de condición** de las personas según el instrumento que se aplique. Esta es la experiencia de España, Colombia y Ecuador.

Para la aplicación de esta metodología idealmente se requiere contar con un **levantamiento simultáneo o paralelo** que permita observar cambios en la condición de interés y luego modelar los determinantes de estos cambios

Ejemplo: España 2002, ante un cambio en la definición de desocupado

$$p = P(Y = 1 / \text{sexo}, \text{edad}) = \Phi(\beta_0 + \beta_1 \cdot \text{sexo} + \beta_2 \cdot \text{edad} + \beta_{12} \cdot \text{int})$$

donde denotamos por

$$Y = \begin{cases} 1 & \text{si parado con ambas definiciones} \\ 0 & \text{si no} \end{cases}$$

$$\text{edad} = \begin{cases} 1 & \text{si} < 25 \text{ años} \\ 0 & \text{si} \geq 25 \text{ años} \end{cases}$$

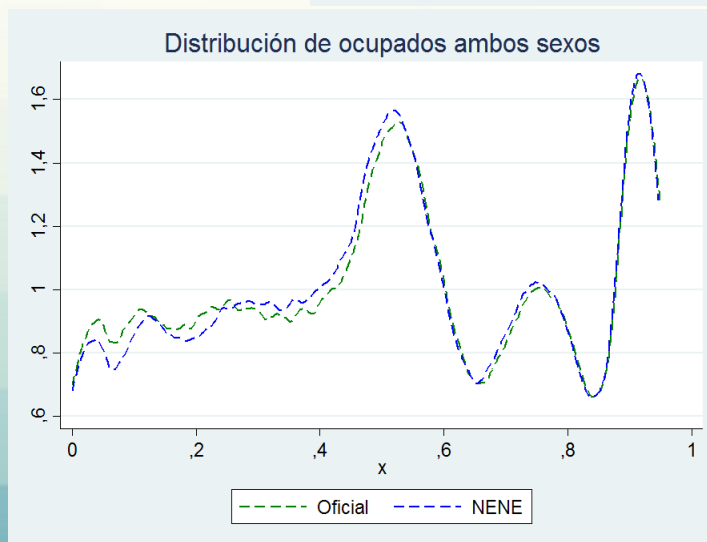
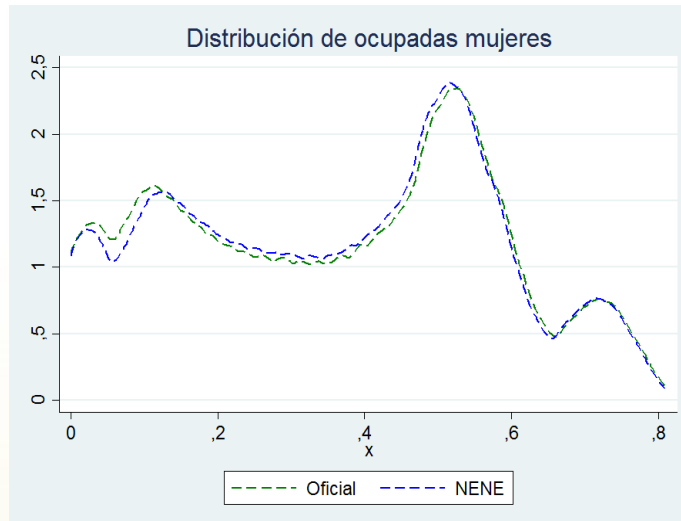
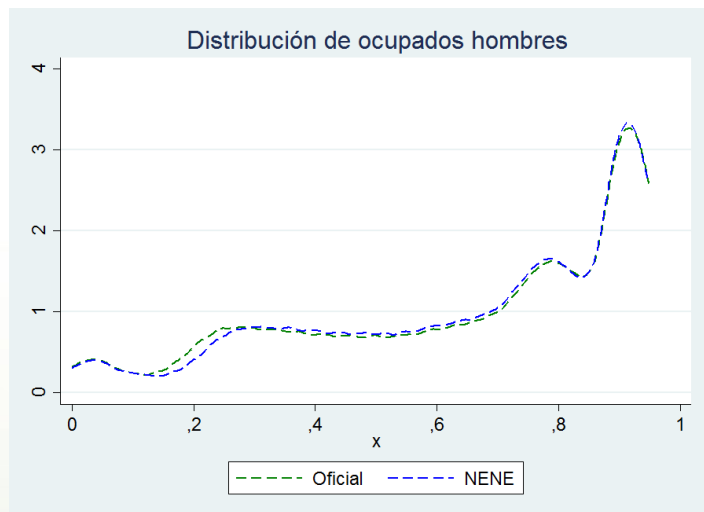
$$\text{sexo} = \begin{cases} 1 & \text{si varón} \\ 0 & \text{si mujer} \end{cases}$$

$$\text{int} = \text{edad} \cdot \text{sexo} = \begin{cases} 1 & \text{si varón} < 25 \text{ años} \\ 0 & \text{en otro caso} \end{cases}$$

p_i : es la probabilidad condicional de que el individuo i se encuentre desocupado

Empalme Nueva Encuesta Nacional de Empleo 2010

El contraste entre las k-densities, usando la base NENE (o paralela) y la oficial para el mismo período (del trimestre octubre – diciembre 2009) arroja las siguientes diferencias:



Cambios de diseño muestral


- Al entrar en vigencia la nueva ENE, también se registraron cambios en el diseño muestral y nuevas unidades primarias de muestreo en las zonas urbanas definidas como manzanas.

1. Objetivo del empalme
2. Empalme Encuesta Nacional de Empleo (ENE) 2006
3. Empalme Encuesta Casen 2006
4. Empalme Nueva Encuesta Nacional de Empleo 2010
5. Lecciones y desafíos para el empalme 2020

- Es relevante disponer de **estimaciones previas de los eventuales impactos** de los cambios a introducir → Institucionalmente se están realizando todos los esfuerzos operativos y técnicos para cumplir con esto.
- En el caso del cambio de marco y muestra que enfrentaremos ahora, se debe considerar que la vigencia del marco actual ha sido más prolongada que en el empalme anterior. Esto **impacta la estrategia para la introducción de la nueva muestra** → introducción progresiva de muestra proveniente de un nuevo marco o levantamiento paralelo.
- Las metodologías que implican cambios en los microdatos tienen la ventaja que generan bases que pueden generar datos empalmados directamente → **es distinto realizar un empalme por cambio de marco que por proyecciones de población.**
- **Existen alternativas metodológicas que permitirían abordar el empalme frente al cambio conceptual aplicado desde 2010.** La definición de esta metodología requiere un proceso de evaluación riguroso y detallado → En una siguiente sección se de la mesa se abordará en más detalle la propuesta metodológica.

Una estrategia de implementación en la incorporación de actualización de muestra ante cambio de marco:

Encuesta	Mes de levantamiento año 2020									
	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre
ENE [MM 2008] [PP 2014]	36.000	36.000	36.000	36.000	36.000	36.000	36.000	36.000	36.000	-
ENE [MM 2017] [PP 2017]	1/6	2/6	3/6	4/6	5/6	6/6	6/6	6/6	6/6	6/6



Levantamiento paralelo a un % de la muestra

Levantamiento paralelo al 100% de la muestra

Reprocesamiento debido al empalme

Lanzamiento oficial

MM: Marco muestral
PP: Proyección de población



Empalme de series

Experiencia, avances y desafíos de la Encuesta Nacional de Empleo

Subdirección Técnica / Subdirección operativa

INE

05 de octubre de 2018

ine.cl