

NOTA TÉCNICA

ESTIMACIÓN DE SERIES DE SALARIOS REGIONALES EN CHILE

ESTIMATION OF REGIONAL WAGE SERIES IN CHILE

JORGE DRESDNER*
CARLOS SANHUEZA

Resumen

En este artículo se presentan series corregidas de salarios mensuales con periodicidad trimestral para trece regiones en Chile en el período diciembre 1994-diciembre 2004. La información se origina de la Superintendencia de Administradoras de Fondos de Pensiones, pero ha sido corregida por sesgos de truncación y censuramiento. El análisis comparativo entre series corregidas y sin corregir indica que éstas tienen un comportamiento diferenciado desde el punto de vista estadístico, y que las series sin corregir no constituyen un buen predictor de las corregidas. Por ello, para el análisis de salarios regionales es conveniente utilizar las series corregidas.

Palabras clave: Salarios regionales, estimación, truncamiento, censura.

Abstract

Monthly wage series with quarterly periodicity between December 1994 and December 2004 for thirteen regions in Chile are presented in this paper. These series are based on the information recollected by the Superintendencia de Fondos de Pensiones (Pensions Fund Superintendence), but are corrected for truncation and censoring bias. The comparative analysis of corrected and non-corrected series indicates that the series show a statistically differentiated behavior. The non-corrected wage series are not good predictors of the corrected series. Thus, to develop regional wage analysis, based on this information source, it is preferable to use the corrected series presented in this paper.

Key words: Regional wages, estimation, truncation, censored.

JEL Classification: J31, R11, R23.

* Depto. Economía, Universidad de Concepción, Casilla 160-C, Concepción, Chile.
Fono: 56 41/220 42 00, fax 56 41/ 220 45 91, jdresdne@udec.cl. Correspondencia a J. Dresdner.

1. INTRODUCCIÓN

La investigación en Economía Regional se encuentra poco desarrollada en Chile. La disponibilidad de información es sin duda una de las limitaciones que enfrenta el análisis empírico sobre las economías regionales. Una de las áreas que congrega una gran cantidad de esfuerzo (relativo) de investigación es el área de convergencia de productos, ingresos y salarios entre regiones. En una revisión sobre investigación en economía regional en Chile Ochoa (2007) menciona diez estudios distintos sobre el tema en el lapso de diez años. Esto corresponde a aproximadamente al 66% de todos los trabajos en investigación identificados en esta revisión.

La mayoría de los trabajos de convergencia se concentran en el análisis de la convergencia de los productos per cápita regionales. Este es el caso de Morandé, *et al.*, 1997; Fuentes, 1996; Aroca y Claps, 1997; Calisto y Riffo, 1999; Aroca y Bosch, 2000; Oyarzún y Araya, 2001; Díaz y Meller, 2004; Duncan y Fuentes, 2006, y Pincheira, 2006. Existen también algunos trabajos que abordan la convergencia de ingresos per cápita regionales, basado en datos de corte transversal para algunos años. Estos son Díaz y Herrera, 1999; Anríquez y Fuentes, 2001; Díaz y Meller, 2004; Duncan y Fuentes, 2006, y Soto y Torche, 2004. Finalmente, se identifican sólo dos trabajos que consideran la convergencia regional de salarios. Estos son Díaz y Meller, 2004, y Tramón y Dresdner, 2004.

El estudio sobre la convergencia de salarios es un área de interés no sólo desde la perspectiva de la Economía Regional, sino también desde la perspectiva del funcionamiento del mercado laboral. La forma y rapidez con que los salarios regionales se ajustan a desequilibrios en los mercados locales entrega información relevante sobre el funcionamiento del “mercado laboral global”. Por ello, el análisis de convergencia de salarios parece justificado no sólo como una perspectiva adicional de la convergencia regional, sino además como un área de investigación de interés *per se*.

El análisis de salarios regionales de Díaz y Meller (2004) está basado en la información de la encuesta que realiza periódicamente la Asociación Chilena de Seguridad (ACHS) a las empresas socias sobre, entre otras cosas, remuneraciones de sus empleados. Los investigadores han utilizado información sobre remuneraciones promedio por empresa. Sin embargo, esta información sufre de dos limitaciones importantes. Primero, la información recogida por la encuesta no asigna las remuneraciones de acuerdo a la región donde viven los empleados que reciben la remuneración. Muchas veces la empresa informante (casa matriz) se ubica en una localidad distinta a la que residen los trabajadores de los cuales informa. Esto implica que los investigadores han debido desarrollar un procedimiento (basado en las regiones de exportación de las empresas) para asignar la información de salarios a distintas regiones. Sin embargo, la información disponible para realizar esta asignación no es completa, por lo que existen potenciales errores de medición en la información utilizada. Segundo, la cobertura de la encuesta es limitada, porque cubre sólo un subconjunto del universo de empresas/trabajadores que debería cubrir. Y esta limitación no se refiere sólo a la ausencia de los trabajadores que laboran en microempresas o empresas informales, sino además a la representación parcial de la información

al interior del sector formal de empresas¹. En algunas regiones existen sectores de actividad económica que no cuentan con empresas en la muestra, lo cual obliga a los autores a introducir supuestos fuertes en los cálculos de las series de salarios regionales, que pueden distorsionar en forma importante las series calculadas. Tramón y Dresdner (2004) infieren la evolución de los salarios (ingresos laborales) regionales a partir de la información de ingreso imponible promedio de los cotizantes previsionales recogida por la Superintendencia de Aseguradoras de Fondos de Pensiones (SAFP). Esta información se encuentra con periodicidad trimestral desde el año 1981. Sin embargo, a raíz de la baja representatividad que tenía la información inicialmente, por la baja cobertura que tuvo el sistema de capitalización privado de pensiones en sus comienzos, se utilizan las cifras a partir del año 1989, donde los niveles de cobertura son importantes². Esta información, a diferencia de la base de la ACHS, proviene de la información individual que entregan las empresas o las personas naturales a su Aseguradora de Fondo de Pensiones (AFP), y que constituye la base para el cálculo de la cotización previsional que debe realizar periódicamente. Las limitaciones que tiene esta información son dos³. Primero, está basada en las personas que participan en el sistema de AFPs. Existe un segmento importante de trabajadores que no pertenecen a este sistema. Además, existe un porcentaje importante de personas que están inscritas en alguna AFP, pero que no cotizan. Es muy probable que las características de las personas que no cotizan sean diferentes a las personas que cotizan. En general, la incidencia de no cotizantes es mayor entre sectores de ingresos bajos, sectores informales, y trabajadores por cuenta propia. Lo mismo ocurre entre las personas que no participan del sistema. El hecho que la información de la SAFP no represente de buena forma estos sectores de la fuerza laboral que, en general, tienden a tener bajos ingresos, puede introducir un sesgo de truncación (por abajo) en la inferencia sobre la serie de salarios basada en una muestra con las características de la muestra de la serie de la SAFP. Segundo, el sistema de AFPs incluye un tope de imposibilidad para los trabajadores de muy altos ingresos. Ello significa que el trabajador que percibe ingresos por sobre este tope imponible, en la base de datos se le asigna el ingreso que corresponde a este tope, introduciendo un problema de censuramiento (por arriba) en los datos. De tal forma que la información sobre salarios regionales basada en la base de la SAFP se encuentra truncada (por abajo) y censurada (por arriba). Tramón y Dresdner recogen e intentan resolver la segunda de estas limitaciones, pero no la primera.

¹ Fuera de la ACHS existen dos mutuales de seguridad adicionales que entregan servicios de seguro a las empresas. Las empresas que contratan con estas otras instituciones no se encuentran representadas en la base de la ACHS. Díaz y Meller (2004) informan que la muestra utilizada corresponde a 8,7% de los trabajadores, y que en regiones distintas a la Metropolitana la representación es 5% o menor.

² A diciembre del 2003, el número de cotizantes alcanza la cifra de 2.982.805 personas, siendo la fuerza laboral registrada por INE de 5.578.081, es decir, 53,47% del total de la fuerza laboral.

³ Existe una tercera limitación a esta fuente de información, pero que es compartida por todas las fuentes: Se basa en la información que entrega el declarante. Existen diversas partidas de ingresos que no son declarados o que no forman parte de la base para el cálculo de las cotizaciones/pensiones.

Sin perjuicio de lo antedicho, la información de la SAFP aparece como una base de información preferible para el análisis de salarios regionales, en comparación con las alternativas (ACHS, CASEN) existentes para el análisis de ingresos laborales en Chile, desde al menos dos perspectivas. Primero, constituye una base de información pública que se genera en forma regular y con una periodicidad trimestral. Segundo, tiene un alto grado de representatividad y éste ha tendido a aumentar con el tiempo en la medida que más trabajadores se han incorporado y cotizan en el sistema de AFPs. Sin embargo, para poder realizar inferencias basada en esta información, en forma relativamente confiable, es recomendable resolver los problemas de truncación y censuramiento discutidos. Al menos es necesario conocer en qué medida y en qué sentido estos problemas pueden afectar las inferencias basadas en esta fuente de información.

En este artículo proponemos una metodología y calculamos series de salarios regionales basados en la información de la SAFP corregidos por los potenciales sesgos de truncación y censuramiento para el período 1989-2004. Adicionalmente, evaluamos estadísticamente las diferencias que surgen entre las series de salarios original (sin corregir) de las series de salarios corregida por estos sesgos. Para estos efectos, en la próxima sección presentamos la metodología de corrección de las series por truncación y censuramiento. Luego presentamos las series obtenidas de este procedimiento y realizamos diversas pruebas estadísticas sobre las diferencias entre series corregidas y sin corregir. Finalmente resumimos las conclusiones. En un apéndice se entregan los detalles de la metodología utilizada y en un anexo se presentan las series corregidas obtenidas.

2. METODOLOGÍA

El presente trabajo pretende recuperar los salarios promedio insesgados a partir de una muestra censurada (por arriba) y truncada (por abajo). Para realizar esto la metodología general es la siguiente: Primero, se deriva la expresión de la esperanza incondicional de una variable que sufre de estas limitaciones. La expresión obtenida es una función de las funciones de densidad y densidad acumulada evaluadas en distintos puntos, de los niveles de truncamiento y censura, y de los valores de la esperanza y varianza de esta variable sin truncar ni censurar. Utilizando información externa es posible estimar los niveles de truncamiento y censura, y resolver para los valores sin truncar ni censurar de la esperanza y varianza de las series (series corregidas). Luego, estas series pueden ser empíricamente evaluadas. Posteriormente, los valores obtenidos se comparan, por distintos procedimientos estadísticos, con las series sin corregir y se discute el efecto que tiene el truncamiento y la censura sobre el comportamiento de las series de salarios regionales. A continuación se detalla la metodología empleada.

Asuma que existe una variable latente continua⁴, $y^* [0, \infty] \sim N(\mu, \sigma^2)$ y una variable observada, y , censurada por arriba y truncada por abajo, donde los niveles de censura y truncamiento son c_1 y c_2 respectivamente. Formalmente,

⁴ Para nuestros propósitos $y^* = \log(\text{salario nominal})$, donde el salario nominal se mide en moneda corriente.

$$(1) \quad Y = y^*, \text{ si } c_2 \leq y^* < c_1$$

$$Y = c_1 \text{ si } y^* \geq c_1$$

donde: $0 < c_2 < c_1 < \infty$.

La esperanza de y^* es (ver apéndice técnico):

$$(2) \quad E(y^*) = (1 - \Phi(\alpha_1))c_1 + (\Phi(\alpha_1))[\mu + \sigma\lambda(\alpha_1, \alpha_2)]$$

donde Φ y ϕ son las funciones de probabilidad acumulada y densidad normal estándar respectivamente,

$$(3) \quad \lambda(\alpha_1, \alpha_2) \equiv \frac{\phi(\alpha_2) - \phi(\alpha_1)}{\Phi(\alpha_1, \alpha_2)}$$

es la función hazard para una distribución doblemente truncada, y donde:

$$(4) \quad \alpha_1 \equiv \frac{c_1 - \mu}{\sigma}$$

$$(5) \quad \alpha_2 \equiv \frac{c_2 - \mu}{\sigma}$$

Con información externa, y utilizando el sistema de ecuaciones conformado por las ecuaciones (2) y (4), o alternativamente (2) y (5), se puede resolver para las variables μ y σ , que son la esperanza y desviación estándar de la serie sin truncar ni censurar. Una vez obtenidos estos resultados, y recordando que la variable y representa el logaritmo del salario nominal, la esperanza de la serie sin truncar ni censurar se recupera de:

$$(6) \quad E(e^y) = e^{\mu + \frac{\sigma^2}{2}}$$

La información requerida para resolver μ y σ del sistema de ecuaciones (2) y (4) es información sobre los puntos de censura y truncamiento, c_1 y c_2 respectivamente; la esperanza de la variable limitada, $E(y^*)$; y la evaluación de las funciones de densidad y de probabilidad acumulada normal estándar en los puntos críticos α_1 y α_2 . El punto de censura c_1 se obtiene del valor del tope imponible para el sistema de las AFP. Este tope es de 60 unidades de fomento (U.F.) Para el punto de truncación c_2 utilizamos el valor del ingreso mínimo legal (“salario mínimo”) vigente en cada trimestre del período. La esperanza de la variable limitada es simplemente el logaritmo del promedio del ingreso imponible mensual regional publicado por la SAFP. La probabilidad de estar bajo el tope imponible, $\Phi(\alpha_1)$, y la probabilidad de estar bajo el salario mínimo, $\Phi(\alpha_2)$, se estimó a partir de la información contenida en las encuestas de Caracterización Socioeconómica Nacional, CASEN (Ministerio de Planificación y Coordinación, 1993-2003). Se calcularon estas frecuencias acumuladas para cada región para los años 1994, 1998, 2000, y 2003, utilizando la información de las muestras de las encuestas CASEN

para estos años⁵. Se asumió que la frecuencia acumulada era válida para un año calendario. Dado que la encuesta CASEN se aplica normalmente en noviembre de cada año, se asumió que la medición era válida desde diciembre del año en que se aplicaba la encuesta hasta octubre del año siguiente. Para los años 1995, 1996, 1997, 1999, 2001, y 2002 se utilizó la interpolación de las frecuencias del año anterior y posterior más próximos donde existía una estimación a partir de las encuestas CASEN. Los valores α_1 y α_2 , para cada año, se calcularon utilizando la inversa de la distribución normal estándar para la probabilidad evaluada en cada punto crítico, $\Phi(\alpha_1)$ y $\Phi(\alpha_2)$. Finalmente, utilizando estas estimaciones de α_1 y α_2 se calcularon las densidades requeridas evaluadas en cada punto crítico⁶.

Cabe indicar que la elección del nivel de ingreso mínimo legal, como punto de truncación, no está libre de controversias. Ello porque pueden existir trabajadores con remuneraciones ligeramente superiores al mínimo que no se encuentren cotizando en el sistema de pensiones⁷. De igual forma, es posible pensar en encontrar trabajadores cuyos ingresos declarados son inferiores al mínimo legal y que se encuentran cotizando en el sistema de pensiones. En cualquiera de los dos casos, el resultado puede ser una distorsión en los cálculos de la probabilidad $\Phi(\alpha_2)$. Sin embargo, no existe un punto de truncación evidente, alternativo al ingreso mínimo legal, que pueda ser utilizado en los cálculos. Por ello, en la sección de resultados, hemos optado por evaluar el impacto que tienen cambios en la probabilidad estimada de $\Phi(\alpha_2)$ sobre los salarios estimados.

Para evaluar la importancia de la corrección realizada a las series de salarios regionales se comparó la serie original (sin corregir) con la serie corregida. El enfoque fue utilizar procedimientos estadísticos alternativos que permitieran evaluar si la serie corregida constituye un buen predictor de la serie sin corregir. Si esta hipótesis no se puede rechazar, entonces la corrección de las series regionales no aparece como demasiado importante. En caso contrario de que la hipótesis se rechace, entonces cobra importancia el uso de las series corregidas para el análisis regional de los salarios.

3. RESULTADOS

En el anexo estadístico se presentan los resultados de las estimaciones de las series de salarios regionales corregidos para Chile para el período diciembre 1994-diciembre 2004⁸.

⁵ La encuesta CASEN existe para el año 1996. Sin embargo, esta información no fue utilizada porque la medición de salarios que por hora se obtuvo no era exactamente comparable con la de otros años.

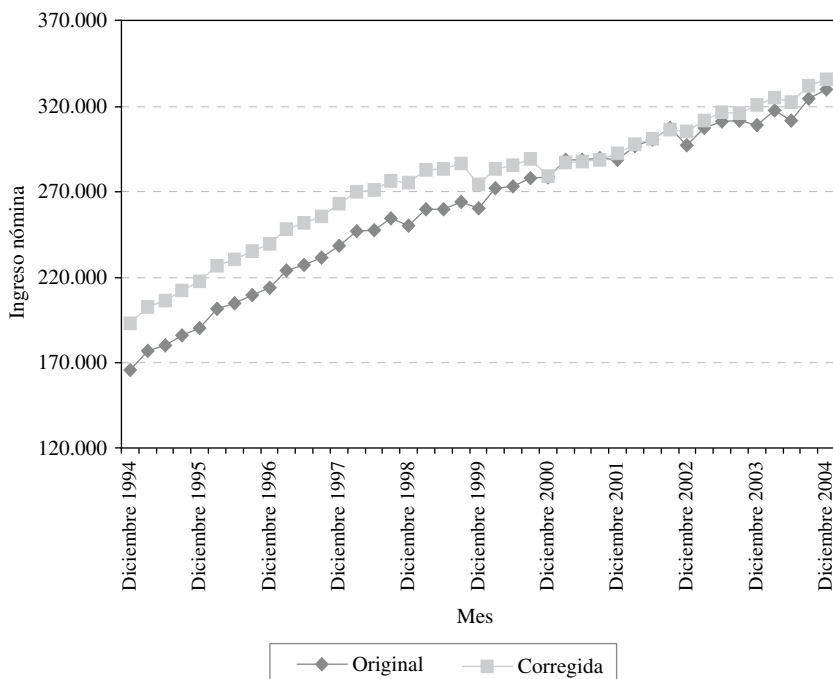
⁶ Una alternativa a la encuesta CASEN podría haber sido utilizar la información de la Encuesta Suplementaria de Ingresos del Instituto Nacional de Estadísticas. Esta opción se desechó porque esta información se encuentra disponible con regularidad sólo a partir del año 1995 y porque esta encuesta no permite estimar los porcentajes de población bajo el salario mínimo.

⁷ Debemos esta observación a un árbitro anónimo.

⁸ Se estiman las series de salarios para trece regiones, de acuerdo a la división política administrativa antigua. No es posible estimar las series de salario para las dos nuevas regiones, Región de Arica y Parinacota, Región de Los Ríos, por falta de series estadísticas de salarios para ello. La base de datos corregida puede bajarse directamente de <http://www.udec.cl/~jdrsdne/salarioscorregidos.rar>

En el Gráfico 1 se muestra la serie promedio de salarios regionales corregidos y sin corregir en el período diciembre 1994-diciembre 2004.

GRÁFICO 1
SERIE DE SALARIOS REGIONAL PROMEDIO CORREGIDA Y SIN CORREGIR
PARA EL PERÍODO DICIEMBRE 1994-DICIEMBRE 2004



Fuente: Cálculos propios y SAFP (1993-2003).

Se observa que en el período diciembre 1994-septiembre 1999 las series evolucionan en forma paralela, con un diferencial a favor de la serie corregida. En diciembre 1999 se observa un primer quiebre, que paulatinamente reduce este diferencial hasta ser prácticamente nulo en diciembre 2000. A partir de este período, las dos series evolucionan en forma similar y con niveles muy próximos. Esta diferencia en la evolución de las dos series es producto de que la serie corregida incorpora el cambio en las probabilidades de estar sobre el tope imponible y bajo el salario mínimo que se produce dentro del período analizado y la serie original no lo incorpora⁹. Sin embargo, si calculamos un estadístico

⁹ Este cambio en las probabilidades seguramente está asociado a los efectos que tuvo la Crisis Asiática sobre los salarios en Chile en este período (Marinakos, 2006). Estadísticamente esto se refleja en los resultados en un aumento en la probabilidad de estar bajo el salario

que normalmente se utiliza para evaluar la relación entre dos variables, como es el coeficiente de correlación entre las dos series, no detectamos este cambio en la evolución de la serie corregida. El coeficiente de correlación para la serie promedio es de 98,85% y para todas las series de salarios regionales es sumamente alto (ver Cuadro 1). Esto levanta la pregunta de cómo evaluar de mejor forma la relación estadística entre las dos series.

CUADRO 1
COEFICIENTES DE CORRELACIÓN Y RESULTADOS DE PRUEBAS DE COEFICIENTES
UNITARIOS PARA SERIES DE TRECE REGIONES
DICIEMBRE 1994-DICIEMBRE 2004

Región	Coef. Corr.	β	Desv. Estándar
I	0,9549	0,9810	0,0140
II	0,9992	0,9786*	0,0020
III	0,9934	1,0056	0,0051
IV	0,9556	1,0514*	0,0118
V	0,9912	0,9376*	0,0076
VI	0,9635	0,9970	0,0132
VII	0,9678	1,0344*	0,0098
VIII	0,9205	1,0155	0,0126
TX	0,9709	1,0078	0,0160
X	0,9893	0,9274*	0,0075
XI	0,9470	0,8368*	0,0097
XII	0,9976	0,8750*	0,0014
XIII	0,9761	0,8302*	0,0052
Promedio	0,9885	0,9507*	0,0067

Fuente: Cálculos propios.

Notas: Coef Corr. = Coeficiente de Correlación; β = parámetro estimado; Desv Estándar = desviación estándar; * = hipótesis nula de $\beta = 1$, significativa al 99% de confianza.

Para abordar esta pregunta se utilizaron diversos procedimientos estadísticos alternativos. Primero, se realizó una regresión de la serie de salarios sin corregir contra la serie corregida (sin intercepto), y se aplicó una prueba de coeficiente unitario. La idea es que si regresionamos una serie sobre sí misma, no sólo el coeficiente de determinación será igual a uno, sino que además, el coeficiente de la variable “independiente” será exactamente uno. En el caso de dos series distintas, esta hipótesis prueba si cambios en una variable, en promedio, se reflejan completamente en la otra. En este sentido, una serie sería una “buena” predictora de la otra.

mínimo y una reducción en la probabilidad de estar sobre el tope imponible. Al comparar los salarios regionales corregidos propuesto por Tramón y Dredsner (2004) con las series sin corregir, el quiebre en la evolución de las series corregida y original a partir del año 1999 no se observa, producto de que su ajuste no considera ajustar los salarios por la truncación por abajo.

En el Cuadro 1 se muestran los coeficientes de correlación y los resultados de las pruebas de coeficientes unitarios para la regresión de la serie sin corregir contra la serie corregida en el período diciembre 1994-diciembre 2004 para las series de salarios de las trece regiones y para el promedio total. Para la regresión de las series de salarios promedio la hipótesis nula fue rechazada al 99% de confianza (ver Cuadro 1). Ello sugiere la existencia de un sesgo al usar la serie original como predictora de la corregida con un coeficiente unitario.

Al considerar los resultados para cada una de las series regionales por separado, observamos que para todas las series los coeficientes de correlación son altos. Estos fluctúan entre 92,1% y 99,9%. Series de salarios con mayor porcentaje de correlación que el promedio son las de las Regiones II, III, V y XII. Los coeficientes estimados de una regresión de la serie de salarios sin corregir contra la serie de salarios corregida (sin intercepto) para cada una de las regiones son todos cercanos a uno. Sin embargo, al realizar una prueba de coeficiente unitario, ésta es rechazada al 99% de confianza en ocho de los trece casos. Se observa además que los coeficientes estimados de las Regiones XI, XII y XIII son los que más tienden a subestimar el coeficiente esperado para una serie insesgada.

Una segunda forma utilizada para evaluar las propiedades predictivas de las series corregidas en relación a las sin corregir fue calcular una serie de indicadores de capacidad predictiva de la serie de salarios corregida sobre la serie de salarios sin corregir. Estos son indicadores que normalmente se utilizan en la literatura para evaluar predicciones. Estos indicadores se calcularon para las series de salario promedio y para cada una de las regiones, y se presentan en el Cuadro 2.

Primero analicemos los resultados para las series de salarios (corregida y sin corregir) promedio. Se observa en este cuadro que la desviación promedio absoluta (DPA) de la serie promedio es de poco más de 15 mil pesos. Para aquilatar la importancia de esta diferencia se debería tener presente que el valor inicial de la serie sin corregir en diciembre de 1994 era de 165.681 pesos y su valor final en diciembre de 2004 era de 392.907 pesos. Una medida que relativiza la desviación de las series corregida y sin corregir es la desviación promedio absoluta porcentual (DPAP). Esta desviación es de 6,0%. Una medida que pondera relativamente más desviaciones grandes que la DPA es la raíz del error cuadrático medio (RECM). Esta medida es algo superior al DPA, y alcanza sobre 18 mil pesos. Finalmente, presentamos la medida de desigualdad de Theil (MDT), cuyo valor debería ser cero en caso de ajuste perfecto. El valor obtenido es relativamente bajo para la serie promedio. Sin embargo, ello es de esperar dado que se trata de una predicción de una serie transformada sobre sí misma. Más interesante es la descomposición que se realiza de esta medida, que permite descomponer las diferencias encontradas entre la serie sin corregir y corregida en proporciones de sesgo, varianza y covarianza. Las dos primeras fuentes son errores de predicción sistemáticos y la tercera es de error no sistemático. Generalmente para un predictor acertado es preferible que las dos primeras fuentes de error sean mínimas, dejando las diferencias por la tercera fuente, que son las diferencias no reductibles mediante correcciones al modelo. En el cuadro se presenta esta descomposición en términos porcentuales. Se observa que la mayor diferencia se encuentra en el componente de sesgo, y luego el componente de varianza. Es decir, de acuerdo a estos resultados, la serie de salarios corregida promedio no es una adecuada predictora de la serie sin corregir.

CUADRO 2
INDICADORES DE LA CAPACIDAD PREDICTIVA DE LAS SERIES DE SALARIOS
CORREGIDAS SOBRE LAS SERIES SIN CORREGIR.
SERIES REGIONALES Y SERIE PROMEDIO NACIONAL

Región	DPA	DPAP	RECM	MDT	CS_MDT	CV_MDT	CC_MDT
I	21.346	8,1%	25.176	0,0456	7,4%	78,2%	14,4%
II	8.758	2,5%	9.530	0,0125	84,1%	9,6%	6,3%
III	8.913	3,1%	9.779	0,0162	0,4%	75,0%	24,5%
IV	17.482	7,3%	21.104	0,0441	24,0%	50,9%	25,1%
V	18.140	7,4%	21.118	0,0407	72,2%	22,2%	5,5%
VI	17.989	8,0%	19.463	0,0419	1,3%	82,2%	16,5%
VII	11.081	5,3%	13.923	0,0347	16,0%	47,1%	36,9%
VIII	18.335	7,6%	19.664	0,0401	1,7%	48,4%	50,0%
IX	20.339	9,7%	22.190	0,0505	0,0%	87,1%	12,9%
X	19.102	8,6%	21.273	0,0449	80,2%	11,5%	8,3%
XI	46.773	14,9%	53.714	0,0948	75,4%	12,1%	12,6%
XII	48.870	14,3%	49.569	0,0771	97,1%	2,4%	0,5%
XIII	61.415	17,5%	62.292	0,0944	97,1%	0,0%	2,8%
Promedio	15.113	6,0%	18.098	0,0334	66,9%	21,2%	11,9%

Fuente: Cálculos propios.

Notas: DPA = Desviación Promedio Absoluta; DPAP = Desviación Promedio Absoluta Porcentual; RECM = Raíz del Error Cuadrático Medio; MDT = Medida de Desigualdad de Theil; CS_MDT = Componente de Sesgo de MDT; CV_MDT = Componente de Varianza de MDT; CC_MDT = Componente de Covarianza de MDT.

Si analizamos los resultados por región, observamos que en general estos son coincidentes para distintos indicadores. Las regiones que aparecen con series corregidas que predicen de mejor forma la serie original (sin corregir) son las Regiones II y III. En general, estas dos regiones muestran valores para la mayoría de los indicadores que denotan una mayor capacidad predictiva. Una excepción para la serie de la II Región lo constituye, sin embargo, el componente de sesgo de la medida de desigualdad de Theil (CS_MDT), que es uno de los más altos en este caso. En cambio las regiones que en forma consistente aparecen con series corregidas que constituyen pobres predictores de la serie sin corregir son las Regiones XI, XII y XIII. La descomposición de la medida de Theil indica que el componente de sesgo es cuantitativamente preponderante en seis de las trece regiones y que el componente de la varianza en otras seis regiones. Por otra parte, el componente de la covarianza es el componente porcentual más importante en sólo una región, la VIII Región. Es decir, en general, los componentes sistemáticos del error son muy importantes. Ello sugiere que las series sin corrección no son buenas predictoras de las series corregidas.

Una tercera forma de evaluar la relación estadística de las dos series es medir si la diferencia entre ambas tiene raíz unitaria. Si una de las series es una buena predictora de la otra, salvo por un error aleatorio, entonces esta diferencia debería estar integrada de orden cero. Esto equivale a realizar una prueba de cointegración de la serie de salarios original sobre la serie de salarios corregido con un vector de cointegración $[1, -1]$. Para ello se utilizan las pruebas de

Dickey Fuller y Dickey Fuller aumentado (Charemza y Deadman, 1992). Los resultados cualitativos no cambiaron entre estas dos pruebas, por lo que en el Cuadro 3 se muestran los resultados sólo para la prueba de Dickey Fuller con y sin variables determinísticas para los residuos de todas las series de salarios regionales y las series de salarios promedio.

CUADRO 3
RESULTADOS DE PRUEBA DE DICKEY FULLER PARA DIFERENCIA
DE SERIES DE SALARIOS SIN CORREGIR Y CORREGIDAS PARA
LAS REGIONES Y EL PROMEDIO NACIONAL

Región	S/constante	C/constante	C/constante y tendencia	C/constante y mudas estacionales	C/constante, tendencia y mudas estacionales
I	-1,89	-1,77	-1,08	-1,98	-0,76
II	-1,22	-2,42	-2,92	-2,11	-2,50
III	-1,24	-1,49	-2,18	-1,31	-1,42
IV	0,86	-0,32	-1,89	-0,33	-1,69
V	2,76**	0,31	-2,66	0,37	-2,09
VI	-0,54	-0,79	-1,37	-0,69	-1,25
VII	0,94	0,11	-1,96	0,20	-1,87
VIII	-0,52	-0,79	-1,78	-0,75	-1,64
IX	-0,99	-1,53	-0,89	-1,65	-0,70
X	-1,41	-1,56	-1,36	-1,49	-1,21
XI	0,08	-1,28	-1,72	-1,17	-1,56
XII	0,68	-1,84	-1,85	-1,60	-1,06
XIII	-0,07	-1,70	-1,69	-1,48	-1,48
Promedio	-1,49	-1,24	-1,62	-1,16	-1,44
Nivel significancia	Valores críticos				
5%	-1,95	-2,95	-3,54	-2,95	-3,543
1%	-2,63	-3,63	-4,24	-3,63	-4,241

Fuente: Cálculos propios.

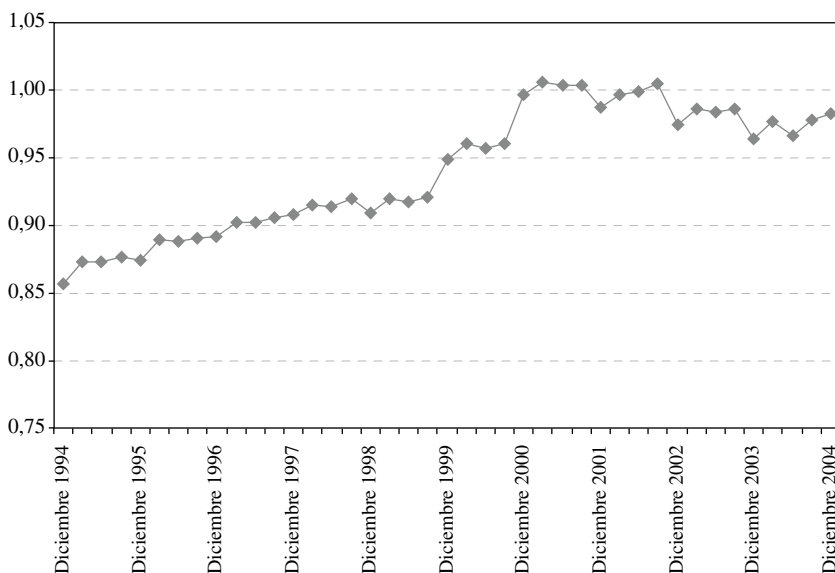
Notas: * significancia al 5%; ** significancia al 1%.

En todos los casos, salvo para la prueba sin variable determinística para los residuos de la serie de salario de la V Región, los resultados de la prueba de Dickey Fuller no rechazan la hipótesis nula de raíz unitaria. Ello sugiere que los residuos no se encuentran integrados de orden cero. Vale decir, en general la serie sin corregir no es un buen predictor de la serie corregida.

En el Gráfico 2 se muestra la evolución de la razón entre la serie de salarios promedio sin corregir y corregida en el período diciembre 1994-diciembre 2004.

GRÁFICO 2

EVOLUCIÓN DEL SALARIO RELATIVO ENTRE SERIE DE SALARIOS PROMEDIO SIN CORREGIR Y CORREGIDA EN EL PERÍODO DICIEMBRE 1994-DICIEMBRE 2004



Fuente: Cálculos en base a estimación propia y SAFP (1993-2003).

Se observa que la razón entre la serie promedio original y corregida está lejos de comportarse como ruido blanco. Esta claramente tiene deriva positiva en el primer período, hasta al menos diciembre 2000, corroborando los resultados anteriormente expuestos que indican que existe un sesgo importante al usar una serie como predictora de la otra. Si analizamos los gráficos de las razones de las series de salarios regionales (no mostrados aquí), éstas muestran en general la existencia de “derivadas” o “desplazamientos de niveles”, salvo para los de las series de salarios de las Regiones XI y quizás XII, que tienen un perfil más parecido a ruido blanco.

En Tramón y Dresdner (2004) se calculan series de salarios regionales corregidas basadas en las estadísticas de la SAFP, con un método similar al utilizado en esta nota. Sin embargo, a diferencia de las series corregidas presentadas aquí, que corrigen tanto el sesgo de censura por arriba como el de truncación por abajo, Tramón y Dresdner sólo corrigen el sesgo de censura por arriba. Para aquilatar la diferencia entre ambas series se calcularon los mismos indicadores presentados en el Cuadro 2 para las series corregidas por Tramón y Dresdner y las series corregidas en este artículo. La idea de esta comparación es ver si las series tienen un comportamiento distinto o no. Si el comportamiento es similar, entonces la corrección adicional propuesta en este artículo no aparece como demasiado importante empíricamente para el período muestral considerado. Pero si por otra parte el comportamiento entre ambas series es disímil, entonces la corrección adicional

propuesta aquí aparece como relevante para el período muestral considerado. En el Cuadro 4 se presentan los resultados obtenidos de esta comparación¹⁰.

CUADRO 4
INDICADORES DE LA CAPACIDAD PREDICTIVA DE LAS SERIES DE SALARIOS
CORREGIDAS POR CENSURA POR ARRIBA Y TRUNCACIÓN POR ABAJO
SOBRE LAS SERIES CORREGIDAS DE DRESDNER Y TRAMÓN.
SERIES REGIONALES Y SERIE PROMEDIO NACIONAL

Región	DPA	DPAP	RECM	MDT	CS_MDT	CV_MDT	CC_MDT
I	22.158	8,6%	26.150	0,0570	21,6%	65,8%	12,6%
II	6.168	1,9%	6.783	0,0109	71,8%	17,9%	10,3%
III	7.119	2,6%	9.034	0,0182	0,9%	68,7%	30,4%
IV	9.795	4,3%	14.574	0,0374	11,8%	38,2%	50,0%
V	20.179	8,4%	21.257	0,0496	89,8%	7,3%	2,9%
VI	15.080	7,1%	16.598	0,0431	23,2%	55,9%	20,9%
VII	5.048	2,5%	7.777	0,0240	7,4%	20,4%	72,2%
VIII	14.355	6,2%	16.257	0,0403	2,4%	19,9%	77,6%
IX	18.393	9,3%	21.194	0,0603	7,6%	79,2%	13,2%
X	17.970	8,6%	20.351	0,0539	76,7%	18,6%	4,7%
XI	32.355	12,0%	35.571	0,0801	82,2%	1,8%	16,0%
XII	41.997	13,1%	42.770	0,0804	96,3%	3,2%	0,5%
XIII	55.327	16,9%	56.026	0,1026	97,4%	0,4%	2,2%
Promedio	9.956	4,0%	10.829	0,0240	17,3%	48,3%	34,4%

Fuente: Cálculos propios.

Notas: DPA = Desviación Promedio Absoluta; DPAP = Desviación Promedio Absoluta Porcentual; RECM = Raíz del Error Cuadrático Medio; MDT = Medida de Desigualdad de Theil; CS_MDT = Componente de Sesgo de MDT; CV_MDT = Componente de Varianza de MDT; CC_MDT = Componente de Covarianza de MDT.

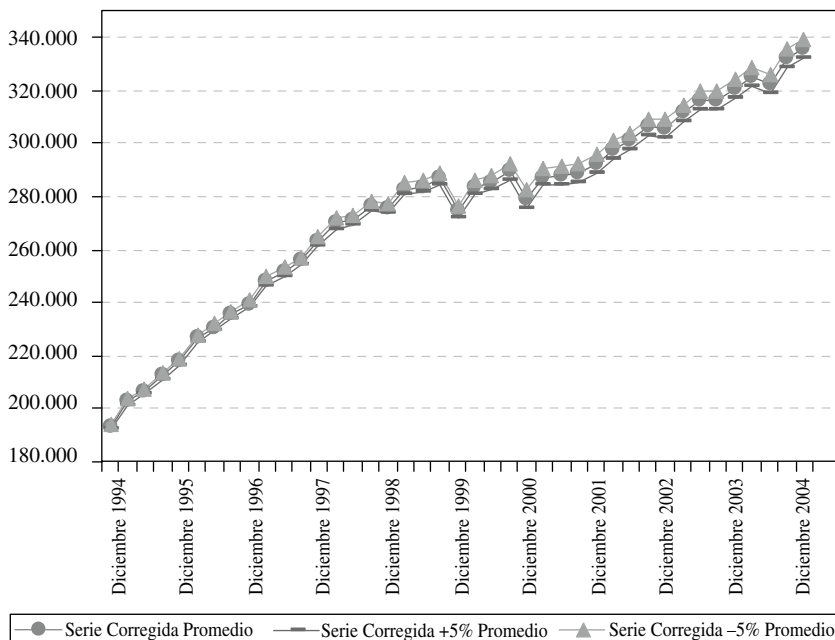
En general, se observa que la disparidad entre las series mostradas en el Cuadro 4 se reduce en comparación con las mostradas en el Cuadro 2 en la mayoría de los casos. Esto básicamente muestra que parte de la diferencia entre las series originales y las series corregidas mostradas en este artículo son producto de la corrección por sesgo de censura por arriba. Sin embargo, se observa que persisten diferencias importantes entre las series en el Cuadro 4. La desviación promedio absoluta (DPA) para la serie promedio nacional es de casi 10 mil pesos, que en términos porcentuales corresponde a un 4% (DPAP). La raíz del error cuadrático medio de esta serie es superior a los 10 mil pesos. Se observa una reducción en el componente de sesgo de la medida de desigualdad de Theil (CS_MDT), pero aumenta el componente de varianza de esta medida (CV_MDT). Pero quizás más importante, al considerar las series regionales,

¹⁰ Cabe mencionar que el período muestral para el cual se encontraban disponibles las series corregidas de Tramón y Dresdner son más cortas que las presentadas en este artículo. Los cálculos mostrados en el Cuadro 3 comprenden el período muestral diciembre 1994-diciembre 2001, por lo que los resultados no son estrictamente comparables con los del Cuadro 2.

no se observa un patrón común de mejoramiento en la capacidad predictiva de las series corregidas en todas las regiones. Es más, en algunas regiones, los indicadores de predicción incluso son menos precisos, sugiriendo que para esas regiones el ajuste logrado, cuando sólo se considera el sesgo de censura por arriba, es claramente insuficiente. Es decir, el análisis comparativo entre las series de Tramón Dresdner y las presentadas en este artículo indica que parece altamente relevante introducir la corrección del sesgo de truncación por abajo a las series de salarios regionales, como lo hacemos aquí.

Como se discutió en la sección de metodología, el uso del ingreso mínimo legal como punto de truncación no está libre de controversias. Por ello, sensibilizamos los resultados obtenidos a cambios en la probabilidad de encontrarse bajo o sobre el punto de truncación. Probamos como variaban los salarios corregidos a cambios en esta probabilidad. Los resultados obtenidos indican que los salarios estimados son muy poco sensibles a variaciones en esta probabilidad. Por ejemplo, para la serie de salarios promedio, un aumento (o disminución) importante de 5% en la probabilidad de encontrarse bajo (o sobre) el punto de truncación definido por el ingreso mínimo generó un cambio promedio de 0,88% en la serie de salarios y una variación máxima del salario corregido en 1,15% (ver Gráfico 3). El mayor

GRÁFICO 3
RELACIÓN ENTRE LA SERIE DE SALARIOS PROMEDIO CORREGIDA
Y LAS SERIES CORREGIDAS CON SENSIBILIZACIÓN DE LA
PROBABILIDAD DEL PUNTO DE TRUNCACIÓN EN +5% Y -5%

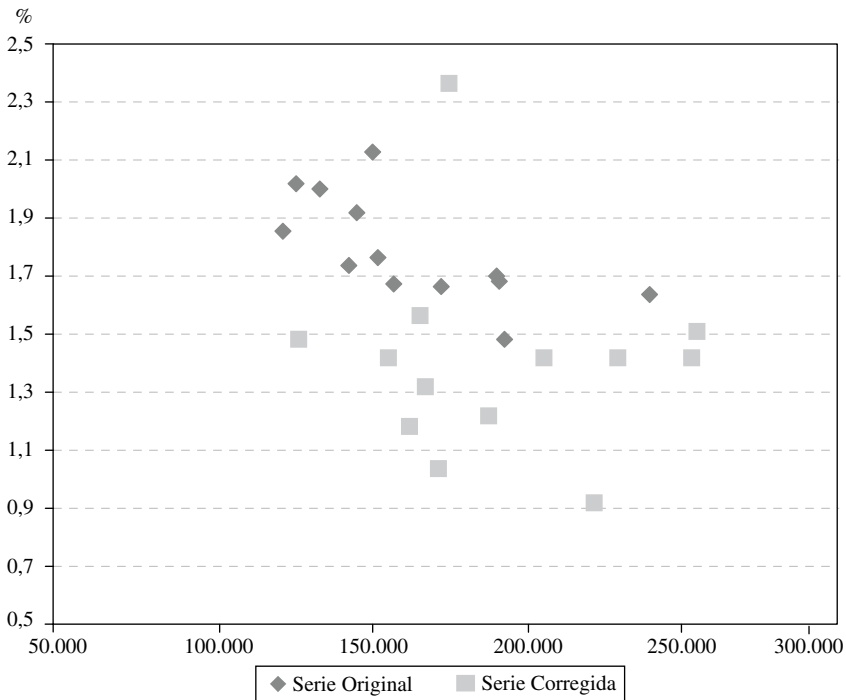


Fuente: Cálculos en base a estimación propia y SAFP (1993-2003).

impacto percibido en los salarios regionales estimados para algún período por un cambio de 5% en esta probabilidad fue de 2%. Es decir, en general el impacto de cambios en esta probabilidad sobre los salarios estimados fue bajo. Esto entrega mayor confianza al uso de las series estimadas, aún en el caso de que el punto de truncación haya sido escogido en forma imprecisa.

Finalmente, otra forma de evaluar el efecto que tiene la corrección de las series de salarios regionales es en términos del impacto que tiene este cambio para el análisis económico. Para describir heurísticamente este potencial impacto, en el Gráfico 4 presentamos la tasa de cambio de los salarios regionales entre diciembre 1994 y diciembre 2004 y el nivel de salarios regionales en diciembre 1994, tanto para las series sin corregir y corregidas. Este es un típico gráfico que se utiliza en el análisis de convergencia de los salarios. Normalmente se acostumbra a buscar una relación lineal entre estas dos variables. En el caso de ser negativa, esta relación indicaría que existe convergencia (incondicional) en los salarios regionales.

GRÁFICO 4
 RELACIÓN ENTRE EL SALARIO PROMEDIO EN EL AÑO 1994 Y LA TASA DE CRECIMIENTO PROMEDIO ANUAL DE LOS SALARIOS ENTRE DICIEMBRE 1994 Y DICIEMBRE 2004 PARA LAS SERIES ORIGINALES Y CORREGIDAS DE SALARIOS REGIONALES



Fuente: Cálculos en base a estimación propia y SAFP (1993-2003).

El gráfico sugiere que en el caso de las series de salarios sin corregir (series originales), sería posible encontrar una relación negativa entre las dos variables graficadas. Sin embargo, al considerar las series corregidas, la relación entre los distintos salarios regionales se transforma en bastante más difusa. No es claro de que exista una relación negativa, y claramente existe una observación que se escapa del conjunto de puntos (la relación para la Región XIII). Entonces, la corrección de las series probablemente en este caso tendrá implicancias para las conclusiones económicas que se pueda obtener del análisis de éstas.

4. CONCLUSIONES

En esta nota se han calculado series de salarios regionales corregidas basadas en la información de ingresos laborales promedio de la base de datos de la Superintendencia de Asociación de Fondo de Pensiones (SAFP). Esta base de datos constituye una fuente de información sobre ingresos laborales con desagregación regional continua, y con un alto grado de representatividad. Sin embargo, esta información sufre de sesgos de truncación y censuramiento, como se describe en este trabajo, que generan aprensiones sobre el uso de esta información en forma confiable para el análisis regional. En este trabajo se ha diseñado una metodología que utilizando información externa permite corregir las limitaciones que sufre esta información. Esta metodología se ha aplicado a las series para cada una de las regiones, obteniéndose series de salarios corregidas para el período diciembre 1994-diciembre 2004.

Se ha evaluado la capacidad predictiva de la serie de salarios no corregida sobre la serie de salarios con corrección para todas las series regionales y para la serie promedio nacional. Los resultados indican que en general las series sin corregir y corregidas, a pesar de que a primera vista parecen ser similares, tienen un comportamiento estadístico diferenciado. Se muestra que este resultado también es válido cuando se comparan las series corregidas en este artículo con las series corregidas presentadas en Tramón y Dresdner (2004). Las series no corregidas, en general, no constituyen buenos predictores de las series con corrección. Especialmente relevante parece ser la existencia de shocks económicos, como es el caso de la crisis asiática, para alterar la relación entre las series corregidas y sin corregir. La excepción podría ser las series de salarios para la III Región. Sin embargo, que en este caso la serie corregida haya funcionado relativamente bien para predecir la serie sin corregir en el pasado no garantiza que va a mantener estas propiedades en el futuro. Por ello, para realizar los análisis de salarios a nivel regional es preferible utilizar las series corregidas. Todo ello sugiere que el uso de las series de salarios sin corregir puede inducir potencialmente a conclusiones erróneas en el análisis económico. Esta impresión se ve reafirmada al comparar los resultados que se obtienen con series corregidas y sin corregir para un análisis gráfico de convergencia de salarios.

Adicionalmente, las relaciones entre las distintas series regionales no muestran un patrón común. Existen series que divergen para un subperíodo muestral y en otro no. Otras divergen a lo largo de todo el período. En otros casos, existe un desplazamiento en la relación en algún momento en el tiempo, etc. Por ende, no es fácil establecer reglas de ajuste para las series regionales sin corregir. Parece entonces que, para fines de análisis de salarios regionales, es recomendable realizar una corrección a las series de salarios de la SAFP, como el que se presenta en este trabajo.

5. REFERENCIAS

- Anríquez, G. y Fuentes, R. (2001). "Convergencia de Producto e Ingreso de las Regiones en Chile: Una Interpretación", en *Convergencia Económica e Integración: La Experiencia en Europa y América Latina*, eds. Mancha, T. y Sotelsek, D. (Madrid) Ediciones Pirámide.
- Aroca, P. y Bosch, M. (2000). "Crecimiento, Convergencia y Espacio en las Regiones Chilenas: 1960-1998", *Estudios de Economía*, vol. 27; 199-224.
- Aroca, P. y Claps, D. (1997). "Regional Convergence Analysis of Chilean Economy between 1960 and 1996", *Instituto Nacional de Estadísticas*, pp. 9.
- Calisto, N. y Riffo, L. (1999). "Crecimiento Económico Regional en Chile: 1985-1998", *Estadística y Economía* N° 17, (enero), pp. 51-74.
- Charemza Wojciech y Derek Deadman (1992). "*New Directions in Econometric Practice. General to Specific Modelling, Cointegration and Vector Autoregression*", Edward Elgar Publishing Limited, Cambridge
- Díaz, L. y Herrera, N. (1999). "Desigualdad de Ingresos y Bienestar 1990-1996 Análisis Comparativo Desde un Enfoque Nacional/Regional". *Documentos MIDEPLAN*, Santiago.
- Díaz, R. y Meller, P. (2004). "Crecimiento Económico Regional en Chile: ¿Convergencia?", *Documentos de Trabajo, Serie Economía* CEA, N° 180.
- Duncan, R. y Fuentes, R. (2006). "Regional Convergence in Chile: New Test, Old Results", *Cuadernos de Economía* N° 127; 81-112 (mayo).
- Fuentes, R. (1996). "¿Convergen las Regiones en Chile? Una Interpretación". En *Análisis Empírico del Crecimiento Económico Chileno*, eds. Morandé, F. y Vergara, R. Centro de Estudios Públicos / ILADES-Georgetown; 179-196.
- Marinakís, A. (2006). "Rigidez de Salarios en Chile", *Revista de la CEPAL* N° 90 (diciembre); 135-150, Santiago.
- Ministerio de Planificación y Coordinación, *Encuesta CASEN*, 1993-2003, <http://www.mideplan.cl/casen> .
- Morandé, F.; Soto, R. y Pincheira, P. (1997). "Águiles, la Tortuga y el Crecimiento Regional en Chile", en *Análisis Empírico del Crecimiento en Chile*, eds. Morandé, F. y Vergara, R. Centro de Estudios Públicos / ILADES. Georgetown University, Santiago.
- Ochoa, M. (2007). Presentación realizada en el *Taller sobre Investigación Económica Regional* organizado por el Banco Central de Chile el 30 de marzo del 2007.
- Oyarzún, C. y Araya, I. (2001). "Long Run Dynamics of Regional Growth in Chile", *Estudios de Economía*, vol. 28, N° 1; 69-78.
- Pincheira, P. (2006). "Convergence and Long Run Uncertainty", *Documentos de Trabajo del Banco Central* N° 391 (Diciembre), pp. 33.
- Soto, R. y Torche, A. (2004). "Spatial Inequality, Migration and Economic Growth in Chile", *Cuadernos de Economía*, vol. 41; 401-424.
- Superintendencia de Administradoras de Fondos de Pensiones (1993-2003). *Boletín Estadístico*.
- Tramón, L. y Dresdner, J. (2004). "¿Convergen los Salarios Regionales en Chile?". *Economía y Administración*, Universidad de Concepción, Año XLI, N° 62; 7-31.

APÉNDICE TÉCNICO

VALOR ESPERADO DE UNA DISTRIBUCIÓN NORMAL TRUNCADA POR ABAJO Y CENSURADA POR ARRIBA

La esperanza de una variable distribuida normal doblemente truncada es:

$$\begin{aligned} E(y^* | c_1 > y^* > c_2) &= \int_{c_2}^{c_1} y^* \frac{f(y^*)}{\text{Prob}(c_1 > y^* > c_2)} \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{c_2}^{c_1} y^* \frac{e^{-\frac{1}{2}\left(\frac{y^*-\mu}{\sigma}\right)^2} dy^*}{\text{Prob}(c_1 > y^* > c_2)} \end{aligned}$$

donde se ha utilizado la definición de probabilidad condicional.

Definiendo $z = \frac{y^* - \mu}{\sigma} \Rightarrow y^* = \mu + \sigma z$ y $dy^* = \sigma dz$ y aplicando obtenemos:

$$\begin{aligned} E(y_i^* | c_1 > y_i^* > c_2) &= \frac{\mu}{\sqrt{2\pi}} \int_{c_2}^{c_1} \frac{e^{-z^2/2} dz}{\text{Prob}(c_1 > y_i^* > c_2)} \\ &+ \frac{\sigma}{\sqrt{2\pi}} \int_{c_2}^{c_1} \frac{ze^{-z^2/2} dz}{\text{Prob}(c_1 > y_i^* > c_2)} \end{aligned}$$

Ahora

$$\begin{aligned} \text{Prob}(c_1 > y^* > c_2) &= 1 - \text{Prob}(y^* > c_1) - \text{Prob}(y^* < c_2) = \Phi(\alpha_1) - \Phi(\alpha_2) \\ &= \frac{1}{\sqrt{2\pi}} \int_{c_2}^{c_1} e^{-z^2/2} dz \equiv \Phi(\alpha_1, \alpha_2) \end{aligned}$$

donde $\alpha_1 \equiv \frac{c_1 - \mu}{\sigma}$, $\alpha_2 \equiv \frac{c_2 - \mu}{\sigma}$.

Por tanto el primer término de la esperanza es igual a μ . El segundo término:

$$\begin{aligned} \frac{\sigma}{\sqrt{2\pi}} \int_{c_2}^{c_1} \frac{ze^{-z^2/2} dz}{\text{Prob}(c_1 > y^* > c_2)} &= \frac{\sigma}{\sqrt{2\pi}} \frac{\left[-e^{-z^2/2} \right]_{c_2}^{c_1}}{\Phi(\alpha_1, \alpha_2)} = \frac{\frac{\sigma}{\sqrt{2\pi}} e^{-c_2^2/2} - \frac{\sigma}{\sqrt{2\pi}} e^{-c_1^2/2}}{\Phi(\alpha_1, \alpha_2)} \\ &= \frac{\sigma(\phi(\alpha_2) - \phi(\alpha_1))}{\Phi(\alpha_1, \alpha_2)} \\ &\equiv \sigma\lambda(\alpha_1, \alpha_2) \end{aligned}$$

Entonces, juntando términos obtenemos:

$$E(y^* | c_1 > y^* > c_2) = \mu + \sigma\lambda(\alpha_1, \alpha_2)$$

La esperanza de un modelo truncado por abajo y censurado por arriba queda entonces:

$$\begin{aligned} E(y^*) &= (1 - \Phi(\alpha_1))E(y^* | y^* = c_1) + \Phi(\alpha_1)E(y^* | c_1 > y^* > c_2) \\ &= (1 - \Phi(\alpha_1))c_1 + (\Phi(\alpha_1))[\mu + \sigma\lambda(\alpha_1, \alpha_2)] \end{aligned}$$

