Documento No. 32

Promedios Paramétricos: Su Selección y Empleo en la Determinación de Índices de Precios

por

Víctor M. Guerrero

Enero, 1981.

Las ideas contenidas en el presente ensayo son responsabilidad exclusiva del autor y no reflejan la posición del Banco de México, S.A.

PROMEDIOS PARAMETRICOS: SU SELECCION Y EMPLEO EN LA DETERMINACION DE INDICES DE PRECIOS

por

Víctor M. Guerrero*

I. INTRODUCCION

Para hacer válidos algunos de los supuestos clásicos de los modelos lineales, Box y Cox (1964) sugieren la siguiente transformación potencia

$$X^{(\lambda)} = \begin{cases} (X^\lambda - 1)/\lambda & \text{si } \lambda \neq 0 \\ \log(X) & \text{si } \lambda = 0 \end{cases} \quad (1.1)$$

que es aplicable cuando $X > 0$. La transformación inversa correspondiente viene dada por

$$X^{-(\lambda)} = \begin{cases} (\lambda X + 1)^{1/\lambda} & \text{si } \lambda \neq 0 \\ \exp(X) & \text{si } \lambda = 0 \end{cases} \quad (1.2)$$

para $X > -1/\lambda$, según $\lambda < 0$.

En este trabajo se hará uso de (1.1) y (1.2) para definir el promedio paramétrico (1.3) con base en una muestra de n observaciones

$$\bar{x}^{(\lambda)} = \left(\frac{1}{n} \sum_{i=1}^n X_i^{(\lambda)} \right)^{-(\lambda)} \quad (1.3)$$

* Se desea hacer patente el reconocimiento a la colaboración y apoyo de Gabriel Vera, quien ayudó a formular el problema y siempre estuvo dispuesto a dar sugerencias para mejorar este trabajo. De igual manera, la participación de Odette Barrón fue muy valiosa para llevar acabo la parte empírica, que requirió del uso de computadora. Así mismo, la parte mecanográfica del documento fue realizada de manera eficiente por Ana María Gutiérrez.

el cual es equivalente al promedio generalizado (1.4) que ya se estudió ampliamente (c. f. Norris, 1976).

$$(\lambda)_{\bar{x}} = \begin{cases} \left(\frac{1}{n} \sum_{i=1}^n X_i^\lambda \right)^{1/\lambda} & \text{si } \lambda \neq 0 \\ \left(\prod_{i=1}^n X_i \right)^{1/n} & \text{si } \lambda = 0 \end{cases} \quad (1.4)$$

Debido a que $X^{(\lambda)}$ es una función monótona creciente de λ , se demuestra fácilmente que $^{(\lambda_1)}\bar{X} \leq ^{(\lambda_2)}\bar{X}$ si $\lambda_1 < \lambda_2$ (con estricta igualdad cuando todos los valores promediados son iguales). En particular se puede observar que cuando λ toma los valores -1, 0 y 1, se obtiene la media armónica, geométrica y aritmética, respectivamente. De igual manera, cuando $\lambda \rightarrow \infty$ se observa que $^{(\lambda)}\bar{X} \rightarrow X_{(n)}$, en donde $X_{(n)}$ denota al mayor estadístico de orden. Esto último se deduce de

$$\lim_{\lambda \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n (X_i/X_{(n)})^\lambda \in [1/n, 1], \quad \text{ya que } X_i \leq X_{(n)} \text{ para toda } i$$

y, por consiguiente

$$\lim_{\lambda \rightarrow \infty} \log(^{(\lambda)}\bar{x}/X_{(n)}) = \lim_{\lambda \rightarrow \infty} \frac{1}{\lambda} \log \left[\frac{1}{n} \sum_{i=1}^n (X_i/X_{(n)})^\lambda \right] = 0.$$

Asimismo, cuando $\lambda \rightarrow -\infty$, se tiene $^{(\lambda)}\bar{x} \rightarrow X_{(1)}$, en donde $X_{(1)}$ es el menor estadístico de orden. De las propiedades anteriores se deduce que cualquier medida de tendencia central debe ser expresable como un promedio paramétrico, para alguna λ en particular.

II. SELECCION DE UN PROMEDIO ADECUADO

Si no existe más información que la contenida en la muestra, lo más conveniente es dejar que los datos mismos sugieran el tipo de promedio que se les debe aplicar y, por ende, el valor de λ más adecuado. Así, el problema de seleccionar el promedio paramétrico más apropiado para un conjunto particular de datos consiste entonces en determinar el valor correspondiente de λ .

Aquí se partirá de las siguientes bases:

- a) el uso de promedios como medidas del centro de una distribución es adecuado solamente si existe una tendencia central real en los datos;
- b) la media aritmética es un promedio adecuado cuando no existen observaciones aberrantes^{1/} y la distribución de frecuencias es aproximadamente simétrica y unimodal;
- c) la fórmula $^{(\lambda)}\bar{x}$ es una media aritmética en la escala transformada, expresada en la escala original mediante la transformación inversa.

Con estas bases, λ deberá ser elegida con el criterio de que, en la escala transformada, la distribución de las $X_i^{(\lambda)}$, $i=1, \dots, n$, sea simétrica. Se ha demostrado en la práctica que el uso de la transformación de Box y Cox para "normalizar" una distribución, si bien no la vuelve estrictamente normal, sí logra al menos volverla simétrica (véase Draper y Cox, 1969 o Guerrero y Johnson, 1979). De aquí que se proponga el método de máxima verosimilitud de Box y Cox para estimar el parámetro λ . Este método evita la inclusión de criterios subjetivos para seleccionar el promedio apropiado y permite que los datos per se, se indiquen qué promedio debe usarse. Nótese, sin embargo, que es necesario suponer, adicionalmente, que no

^{1/} El término observación aberrante pretende incluir a todos aquellos datos tan peculiares o extremos que bien podrían considerarse como ajenos al conjunto de datos bajo estudio.

existen observaciones aberrantes aunque, de hecho, el método es robusto contra este tipo de observaciones (lo cual significa que tales observaciones casi no afectan los resultados).

La manera en que el método funciona es la siguiente. Considérese una muestra aleatoria $X_1, \dots, X_n$ de una distribución absolutamente continua, cuya función de densidad se concentra en $(0, \infty)$. Para obtener el estimador de máxima verosimilitud de λ , se supondrá tentativamente que $X_i \sim N(\mu_0, \sigma_0^2)$ $i = 1, \dots, n$, para algún vector de parámetros $\theta_0 = (\mu_0, \sigma_0^2, \lambda_0)'$.

La razón por la que se pone como tentativo este supuesto es que $X_i > 0$ implica que $X_i \sim N(\mu_0, \sigma_0^2, \lambda_0)$ de acuerdo con $\lambda_0 < 0$; así, para $\lambda_0 \neq 0$, X_i está acotada y, hablando estrictamente, no puede tener una distribución normal.

Esta suposición conduce a utilizar la siguiente función de verosimilitud

$$\ln(\theta | X_1, \dots, X_n) = (2\pi)^{-n/2} \sigma^{-n} \exp\left[-\sum_{i=1}^n (X_i^{(\lambda)} - \mu)^2 / 2\sigma^2\right] \prod_{i=1}^n X_i^{\lambda-1} \quad (2.1)$$

la cual se puede maximizar en dos etapas. Primera: se fija un valor de λ y se obtienen los estimadores máximo-verosímiles

$$\hat{\mu}(\lambda) = \frac{1}{n} \sum_{i=1}^n X_i^{(\lambda)} \quad \text{y} \quad \hat{\sigma}^2(\lambda) = \frac{1}{n} \sum_{i=1}^n (X_i^{(\lambda)} - \hat{\mu}(\lambda))^2 \quad (2.2)$$

que producen la siguiente función de log-verosimilitud maximizada

$$\ln \max^{(\lambda)} = \text{constante} - \frac{n}{2} \log [\hat{\sigma}^2(\lambda)] + (\lambda-1) \sum_{i=1}^n \log (X_i), \quad (2.3)$$

segunda: se hace variar λ para obtener diferentes valores de $\ln \max^{(\lambda)}$. Aquel valor de λ que

proporcione el máximo de la función (2.3) será el estimador $\hat{\lambda}$; por consiguiente, el

estimador máximo-verosímil de θ será

$$\hat{\theta} = (\hat{\mu}(\hat{\lambda}), \hat{\sigma}^2(\hat{\lambda}), \hat{\lambda})'. \quad (2.4)$$

Como Box y Cox sugieren, es conveniente graficar la función de log-verosimilitud maximizada contra λ , a fin de encontrar el valor $\hat{\lambda}$ y su correspondiente intervalo de $100(1-\alpha)\%$ de confianza, el cual se obtiene, para muestras grandes, de la relación

$$g_{n \max}(\hat{\lambda}) - g_{n \max}(\lambda) < \frac{1}{2} X^2_1(\alpha) \quad (2.5)$$

con $\Pr [X^2_1 > X^2_1(\alpha)] = \alpha$ y $X^2_1 \sim$ Ji-cuadrada con 1 grado de libertad.

Cabe aclarar que al seleccionar de esta manera el valor de λ , se determina no sólo el promedio paramétrico más adecuado, sino también una medida de dispersión asociada con este promedio.

III. CORRECCION POR SESGO

Varios autores (e. g. Hoyle, 1973 y Patterson, 1966) han indicado que al aplicar una transformación, hacer estimaciones en la escala transformada y expresar dichas estimaciones en las unidades originales (mediante la transformación inversa) se introduce un sesgo que puede, en muchas ocasiones, ser importante.

Debido a que los promedios paramétricos funcionan precisamente de la forma señalada, resulta necesario encontrar una medida del sesgo introducido, para poder corregirlo. En particular, debido a la monotonidad de (1.1) se tiene que el promedio paramétrico (1.3) es un

estimador de la mediana de la variable original, de tal forma que mientras mayor asimetría exista en la variable original, mayor será el sesgo que se introduzca al estimar $E(X)$ con $^{(\lambda)}\bar{X}$. La corrección se puede obtener mediante las siguientes aproximaciones producidas por una expansión en serie de Taylor de $X^{(\lambda)}$ alrededor de $E(X)$

$$\begin{aligned} E(X^{(\lambda)}) &\simeq [E(X)]^{(\lambda)} + (\lambda-1) [E(X)]^{\lambda-2} \text{Var}(X)/2 \\ \text{Var}(X^{(\lambda)}) &\simeq [E(X)]^{2\lambda-2} \text{Var}(X). \end{aligned} \quad (3.1)$$

De estas expresiones se tiene que

$$E(X) \simeq c(\lambda) [E(X^{(\lambda)})]^{-\lambda} \quad (3.2)$$

$$\text{Var}(X) \simeq [E(X)]^{2-2\lambda} \text{Var}(X^{(\lambda)}). \quad (3.3)$$

donde el factor de corrección $c(\lambda)$ que aparece en (3.2), está dado por

$$c(\lambda) = \begin{cases} \left\{ \left[1 + \sqrt{1 - 2\lambda(\lambda-1) [1 + \lambda E(X^{(\lambda)})]^2 \text{Var}(X^{(\lambda)})} \right] / 2 \right\}^{1/\lambda} & \text{si } \lambda \neq 0 \\ \exp\{\text{Var}[\log(X)]/2\} & \text{si } \lambda = 0 \end{cases} \quad (3.4)$$

Estimaciones de $E(X)$ y $\text{Var}(X)$ pueden obtenerse sustituyendo las estimaciones $\hat{\lambda}$, $\hat{\mu}$ y $\frac{n}{(n-1)} \hat{\sigma}^2$ por los correspondientes valores poblacionales λ , $E(X^{(\lambda)})$ y $\text{Var}(X^{(\lambda)})$, donde $\hat{\lambda}$, $\hat{\mu}$ y $\hat{\sigma}^2$ se obtienen como se indicó en la sección anterior.

Como un indicador de la bondad de las aproximaciones (3.2) y (3.3), considérese el caso en que $\lambda_0 = 0$, lo cual significa que $\log(X)$ se distribuye normalmente. Entonces se sabe que

$$E(X) = \exp \{E[\log(X)] + \text{Var}[\log(X)]/2\} \quad (3.5)$$

$$\text{Var}(X) = [\exp \{\text{Var}[\log(X)]\} - 1] \exp \{2E[\log(X)] + \text{Var}[\log(X)]\}.$$

Si se comparan estas expresiones con (3.2) y (3.3) se puede concluir que esta última

es una aproximación razonable cuando $\lambda=0$ (aunque $\log(X)$ no tenga exactamente una distribución normal) siempre que

$$\exp \{ \text{Var} [\log(X)] \} - 1 \simeq \text{Var} [\log(X)]$$

lo cual sucede cuando $\text{Var} [\log(X)]$ es pequeña.

Ahora, si $\lambda_0 \neq 0$, no es posible utilizar el argumento anterior debido a que $X^{(\lambda_0)}$ no puede tener, estrictamente, una distribución normal y por tanto no se pueden obtener las expresiones equivalentes a (3.5). Sin embargo, para el caso $\lambda=1$, se puede observar fácilmente que (3.2) y (3.3) proporcionan igualdades estrictas.

IV. INTERVALOS DE CONFIANZA BASADOS EN $(\lambda)_{\bar{X}}$

Otra ventaja de utilizar el procedimiento aquí propuesto consiste en la obtención de intervalos de confianza para $E(X)$ de manera relativamente simple. Para esto, nótese que $(^{(\lambda_0)} \bar{X})^{(\lambda_0)}$ es un estimador del parámetro μ_0 perteneciente a una distribución normal con varianza σ_0^2 . Es decir, en el supuesto de que $X_i^{(\lambda_0)} \sim N(\mu_0, \sigma_0^2)$, $i=1, \dots, n$ con $X_1, \dots, X_n$, mutuamente independientes, se tiene

$$P_r \{ (^{(\lambda_0)} \bar{X})^{(\lambda_0)} - t_{n-1}(\alpha/2) \frac{S}{\sqrt{n}} \leq \mu_0 \leq (^{(\lambda_0)} \bar{X})^{(\lambda_0)} + t_{n-1}(\alpha/2) \frac{S}{\sqrt{n}} \} = 1 - \alpha \quad (4.1)$$

$$\text{con } (^{(\lambda_0)} \bar{X})^{(\lambda_0)} = \frac{1}{n} \sum_{i=1}^n X_i^{(\lambda_0)}, \quad S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n [X_i^{(\lambda_0)} - ((\lambda_0)_{\bar{X}})^{(\lambda_0)}]^2}$$

y $P_r [t_{n-1} > t_{n-1}(\alpha/2)] = \alpha/2$ con $t_{n-1} \sim t$ de Student con $n-1$ grados de libertad.

De (4.1), se obtiene de inmediato

$$\begin{aligned} P_r \{ [(^{(\lambda_0)} \bar{X})^{(\lambda_0)} - t_{n-1}(\alpha/2) \frac{S}{\sqrt{n}}]^{-(\lambda_0)} \leq \mu_0^{-(\lambda_0)} \\ \leq [(^{(\lambda_0)} \bar{X})^{(\lambda_0)} + t_{n-1}(\alpha/2) \frac{S}{\sqrt{n}}]^{-(\lambda_0)} \} = 1 - \alpha \end{aligned}$$

por lo que

$$\left(\left[\frac{1}{n} \sum_{i=1}^n X_i^{(\lambda_0)} - t_{n-1}(\alpha/2) \frac{S}{\sqrt{n}} \right]^{(\lambda_0)}, \left[\frac{1}{n} \sum_{i=1}^n X_i^{(\lambda_0)} + t_{n-1}(\alpha/2) \frac{S}{\sqrt{n}} \right]^{(\lambda_0)} \right) \quad (4.2)$$

constituye un intervalo aproximado del 100 (1- α)% de confianza para la mediana poblacional que $^{(\lambda_0)} \bar{X}$ pretende estimar. Se dice que es un intervalo aproximado debido a la suposición de normalidad, que no es estrictamente válida.

En la práctica se desconoce el valor λ_0 , por lo cual se deberá utilizar $\hat{\lambda}$. Asimismo, el en términos de los estimadores máximo-verosímiles $\hat{\mu}$ y $\hat{\sigma}$, viene a ser

$$\left(\left[\hat{\mu} - t_{n-1}(\alpha/2) \hat{\sigma} / \sqrt{n-1} \right]^{\hat{\lambda}}, \left[\hat{\mu} + t_{n-1}(\alpha/2) \hat{\sigma} / \sqrt{n-1} \right]^{\hat{\lambda}} \right) \quad (4.3)$$

Ahora bien, una manera sencilla de construir intervalos de confianza aproximados para $E(X)$ consiste en aplicar el factor de corrección (3.4) a ambos extremos del intervalo (4.3); de esta manera se obtiene el siguiente intervalo de aproximadamente 100(1- α)% de confianza para $E(X)$

$$c(\lambda) \left\{ \hat{\mu} \pm t_{n-1}(\alpha/2) \hat{\sigma} / \sqrt{n-1} \right\}^{\hat{\lambda}} \quad (4.4)$$

en donde

$$c(\hat{\lambda}) = \begin{cases} \left\{ \left[1 + \sqrt{1 - 2\hat{\lambda}(\hat{\lambda} - 1)(1 + \hat{\lambda}\hat{\mu})^2 [n/(n-1)]\hat{\sigma}^2} \right] / 2 \right\}^{1/\hat{\lambda}} & \text{si } \hat{\lambda} \neq 0 \\ \exp \{ [n/(n-1)]\hat{\sigma}^2 / 2 \} & \text{si } \hat{\lambda} = 0 \end{cases} \quad (4.5)$$

Se presenta a continuación otra manera de obtener intervalos de confianza, la cual puede usarse cuando n es grande. La varianza de $^{(\lambda)}\bar{X}$ para cualquier valor de λ puede aproximarse de manera similar a como se obtuvieron las aproximaciones (3.1) y (3.3), o sea

$$\begin{aligned} \text{Var}(^{(\lambda)}\bar{X}) &\simeq [E(^{(\lambda)}\bar{X})]^{2-2\lambda} & \text{Var}[(^{(\lambda)}\bar{X})^{(\lambda)}] \\ &= [E(^{(\lambda)}\bar{X})]^{2-2\lambda} & \text{Var}(X^{(\lambda)})/n \end{aligned} \quad (4.6)$$

a partir de esta expresión, el error estándar de $^{(\lambda)}\bar{X}$ viene a ser

$$\text{d.e. } (^{(\lambda)}\bar{X}) \simeq [E(^{(\lambda)}\bar{X})]^{1-\lambda} \frac{[\text{d.e.}(X^{(\lambda)})]}{\sqrt{n}} \quad (4.7)$$

el cual puede estimarse mediante

$$\text{d.e. } (^{(\lambda)}\bar{X}) \simeq (\hat{\mu}^{-(\lambda)})^{1-\lambda} \frac{\hat{\sigma}}{\sqrt{n-1}} = (^{(\lambda)}\bar{X})^{1-\lambda} \frac{\hat{\sigma}}{\sqrt{n-1}} \quad (4.8)$$

Esta última expresión comprende, como casos particulares, las desviaciones estándar de las medias geométrica ($\lambda=0$) y armónica ($\lambda=1$) que fueron obtenidas previamente por Norris (1940) mediante la aplicación del Teorema de Laplace-Liapounoff.

Una vez obtenida la relación (4.8), se procede a utilizar resultados asintóticos que, aun cuando no son estrictamente válidos porque no consideran que la distribución de las variables originales afecta a la matriz de varianza-covarianza asintótica (véase Guerrero y Johnson, 1979), sí proporcionan aproximaciones razonables. Por lo tanto, como una aproximación a los extremos de un intervalo del $100(1-\alpha)\%$ de confianza para la mediana de X , se sugiere usar

$$\hat{\mu}^{-(\lambda)} \pm Z(\alpha/2) (\hat{\mu}^{-(\lambda)})^{1-\lambda} \frac{\hat{\sigma}}{\sqrt{n-1}} \quad (4.9)$$

con $P_r[Z > Z(\alpha/2)] = \alpha/2$ y $Z \sim N(0,1)$.

Así se construyó el intervalo (4.4) a partir de (4.3), también aquí se propone modificar el intervalo (4.9) multiplicándolo por el factor $\hat{c}(\hat{\lambda})$ definido en (4.5); de esta manera el intervalo (4.9) se convierte en

$$\hat{c}(\hat{\lambda}) \{ \hat{\mu}^{-(\hat{\lambda})} \pm Z(\alpha/2) (\hat{\mu}^{-(\hat{\lambda})})^{1-\hat{\lambda}} \hat{\sigma} / \sqrt{n-1} \}. \quad (4.10)$$

Una extensión importante de (1.3) consiste en introducir ponderaciones para los valores de la variable. La definición correspondiente a un promedio paramétrico ponderado, será entonces

$${}^{(\lambda)} \bar{X}' = [(\sum w_i)^{-1} \sum_{i=1}^n w_i X_i^{(\lambda)}]^{-(\lambda)}, X_i > 0, i=1, \dots, n \quad (4.11)$$

donde w_i es la ponderación asociada con la observación X_i .

Para el caso que se utilicen promedios ponderados, los intervalos obtenidos previamente no son adecuados y deben modificarse. El cambio sugerido consiste básicamente en sustituir

$$\hat{\mu} \text{ por } \hat{\mu}' = \sum_{i=1}^n w_i X_i^{(\lambda)} / \sum_{i=1}^n w_i \text{ y } \hat{\sigma}^2/n \text{ por } \sum_{i=1}^n w_i^2 \hat{\sigma}^2 / (\sum_{i=1}^n w_i)^2.$$

Esta sugerencia está basada en que si $X_i^{(\lambda)} \sim N(\mu_0, \sigma_0^2)$ para toda i , entonces

$${}^{(\lambda)} \bar{X}' \sim N(\mu_0, \sum_{i=1}^n w_i^2 \sigma_0^2 / (\sum_{i=1}^n w_i)^2)$$

mientras que ${}^{(\lambda)} \bar{X}' \sim N(\mu_0, \sigma_0^2/n)$. Por lo tanto, los intervalos (4.3) y (4.9) vienen a ser, respectivamente,

$$\{ \hat{\mu}' \pm t_{n-1}(\alpha/2) \sqrt{n \sum_{i=1}^n w_i^2 / \left[(n-1) \left(\sum_{i=1}^n w_i \right)^2 \right]} \hat{\sigma} \}^{-(\hat{\lambda})} \quad (4.12)$$

y

$$\hat{\mu}^{r(\hat{\lambda})} \pm Z(\alpha/2) \{ \mu^{r(\lambda)} \}^{1-\lambda} \sqrt{n \sum_{i=1}^n w_i^2 / \left[(n-1) \left(\sum_{i=1}^n w_i \right)^2 \right] \hat{\sigma}^2} \quad (4.13)$$

En consecuencia, los intervalos para E(X) equivalentes a (4.4) y (4.10) se obtienen multiplicando (4.12) y (4.13) por el factor

$$\hat{K}(\hat{\lambda}) = \begin{cases} \left\{ \left(1 + \sqrt{1 - 2\hat{\lambda}(\hat{\lambda}-1)(1+\hat{\lambda}\hat{\mu})^{-2} n \sum_{i=1}^n w_{i\sigma}^2 / \left[(n-1) \left(\sum_{i=1}^n w_i \right)^2 \right]} \right) / 2 \right\}^{1/\hat{\lambda}} & \text{si } \hat{\lambda} \neq 0 \\ \exp \left\{ n \sum_{i=1}^n w_{i\sigma}^2 / \left[2(n-1) \left(\sum_{i=1}^n w_i \right)^2 \right] \right\} & \text{si } \hat{\lambda} = 0 \end{cases} \quad (4.14)$$

Land (1974) presenta otras formas alternativas para obtener intervalos de confianza, que quizá proporcionen mejores aproximaciones que los intervalos aquí desarrollados, pero que son también más complicados de calcular.

V. EJEMPLOS

1. Indices de precios

Cuando se construyen índices de precios, las fórmulas que se utilizan involucran, por lo general, promedios del tipo (1.3) o (4.11); es decir, frecuentemente se utiliza

$$^{(\lambda)}I_{m,0} = \left[\frac{1}{n} \sum_{i=1}^n (P_{mi} / P_{io})^{(\lambda)} \right]^{-1/\lambda} \quad (5.1)$$

o

$$^{(\lambda)}I'_{m,0} = \left[\left(\sum_{i=1}^n w_i \right)^{-1} \sum_{i=1}^n w_i (P_{mi}/P_{oi})^{(\lambda)} \right]^{-1/\lambda} \quad (5.2)$$

como índice de precios de n artículos, el i -ésimo de los cuales tiene el precio P_{oi} en el período base y P_{mi} en el período m . Por ejemplo, si se utiliza $w_i = P_{oi}Q_{oi}$ en la fórmula (5.2), en donde Q_{oi} representa la cantidad del artículo i consumida en el período base, se tiene:

$$(1)_{I'_{m,0}} = \sum_{i=1}^n P_{mi}Q_{oi} / \sum_{i=1}^n P_{oi}Q_{oi}, \quad (0)_{I'_{m,0}} = \left[\prod_{i=1}^n (P_{mi} / P_{oi})^{P_{oi}Q_{oi}} \right]^{1 / \sum_{i=1}^n P_{oi}Q_{oi}}$$

y

$$^{(-1)}I'_{m,0} = \sum_{i=1}^n P_{oi}Q_{oi} / \left[\sum_{i=1}^n P_{oi}Q_{oi}(P_{oi}/P_{mi}) \right]$$

que son los índices ponderados definidos por Fisher (1927, pp. 466-468). En particular el índice $^{(1)}I'_{m,0}$ anterior es de uso bastante común y se le conoce con el nombre de Índice de Laspeyres; por consiguiente, al usar la ponderación $w_i = P_{oi}Q_{oi}$, se puede pensar que $^{(\lambda)}I'_{m,0}$ constituye una generalización paramétrica del Índice de Laspeyres. Similarmente, si en la ponderación interviene la cantidad consumida en el período m , en lugar de la del período base, o sea $w_i = P_{oi}Q_{mi}$, los índices que se obtengan serán generalizaciones paramétricas del llamado Índice de Paasche que viene dado por la fórmula

$$^{(1)}I'_{m,0} = \sum_{i=1}^n P_{mi}Q_{mi} / \sum_{i=1}^n P_{oi}Q_{mi}$$

Como un ejemplo numérico del procedimiento sugerido en este trabajo para determinar un índice de precios, se utilizarán datos sobre los precios de 151 artículos específicos dentro de la categoría de "Frutas Frescas" pertenecientes al concepto genérico de "Alimentos y bebidas no alcohólicas" que maneja el Banco de México, S. A. El período base será el año de 1978 y se analizarán los meses de febrero y noviembre de 1979. Cabe hacer notar que las ponderaciones por emplear en la fórmula (5.2) son los gastos familiares mensuales determinados para el año de 1963.

El primer paso en la determinación de los índices consistió en estimar el valor de λ para los dos meses considerados; los valores estimados y sus respectivos intervalos del 95% de confianza, como se aprecia en las gráficas 1 y 2, son

febrero : $\lambda = 0.75, (1.08, -0.42)$

noviembre : $\hat{\lambda} = 0.55, (-0.03, 1.08)$

donde se sugiere utilizar la media ya sea aritmética o geométrica (corregida por sesgo) para calcular el índice del mes de noviembre y la media armónica (corregida) para el índice de febrero. Los resultados obtenidos se resumen en la tabla 1, en donde los intervalos de confianza se calcularon con las fórmulas para las muestras grandes. La importancia de estos intervalos de confianza se reflejan en el siguiente comentario de Allen (1975, p. 7): "La variación de los relativos de precios alrededor del promedio, considerado como el índice, es tan importante como el mismo índice. Si no existiera variación, el índice sería trivial; lo que se necesita conocer es qué tanta variación existe".

De la tabla 1 puede colegirse lo siguiente:

- a) Para el mes de febrero, si se utiliza un promedio simple del tipo (5.1), la mediana difiere significativamente de la media, cosa que no ocurre si se utiliza un promedio ponderado del tipo (5.2) y que muestra el efecto que en el presente caso ejerce la ponderación, esto es, volver aproximadamente simétrica la distribución original de relativos de precios.
- b) Para noviembre, la mediana y la media no son significativamente diferentes. Esto queda corroborado por el hecho de que los valores 0 y 1 se encuentran ambos dentro del intervalo de confianza para λ , y se sabe que para $\lambda=1$ la mediana y la media deben ser iguales.

Gráfica 1
 FUNCION DE LOG-VEROSIMILITUD MAXIMIZADA: FEBRERO

Gráfica 2
 FUNCION DE LOG-VEROSIMILITUD MAXIMIZADA: NOVIEMBRE

Tabla 1

VALORES ESTIMADOS E INTERVALOS DEL 95% DE CONFIANZA

Parámetro	Promedio simple (5.1)		Promedio ponderado (5.2)	
$\hat{\lambda}$ Febrero: $\lambda = -0.75$				
Mediana	0.9844,	(0.9273,1.0415)	1.0236,	(0.9424,1.1058)
Valor medio	1.1262,	(1.0609,1.1915)	1.0251,	(0.9438,1.1074)
Varianza	0.2053	-----	0.1342	-----
$\hat{\lambda}$ Noviembre: $\lambda = 0.55$				
Mediana	1.0906,	(1.0425,1.1387)	1.1608,	(1.0923,1.2304)
Valor medio	1.1093,	(1.0603,1.1582)	1.1610,	(1.0925,1.2306)
Varianza	0.0924	-----	0.0991	-----

Por otro lado, el valor de la media armónica -corregida por sesgo- que se sugiere utilizar para el mes de febrero es 1.1368 sin ponderar y 1.0174 ponderada, valores que, como puede observarse en la tabla 1, caen dentro de los respectivos intervalos de la confianza para el valor medio. Similarmente, la sugerencia de utilizar ya sea la media aritmética (1. 1093 sin ponderar y 1.1745 ponderada) o la media geométrica corregida por sesgo (1.1110 sin ponderar y 1.1447 ponderada) se justifica no sólo por el hecho de que los valores 0 y 1 pertenecen al intervalo de confianza para λ , sino porque los valores calculados de estas medias pertenecen a los respectivos intervalos de confianza para el valor medio real.

Dentro de las implicaciones que estos resultados pueden tener desde el punto de vista de la práctica, sobresale el siguiente: si el Banco de México, S. A. hubiese tenido especial interés en calcular el índice de los precios con base en los 151 artículos con los que aquí se trabajó, entonces la oficina de Precios del Banco hubiera estimado dicho índice mediante la fórmula

⁽¹⁾ $\Gamma_{m,0}$, tanto para febrero como para noviembre.

En el caso de noviembre se hubiera obtenido entonces el índice 1.1745, pero no se conocería la variación existente alrededor de este valor (por ejemplo, no se podría decir si el índice buscado difiere significativamente de los valores 1.1 ó 1.25), en cambio con el método propuesto en este trabajo, se puede utilizar el mismo valor 1.1745 como estimación puntual, pero con la ventaja adicional de poder hacer aseveraciones con cierto rigor estadístico, de tal forma que es posible afirmar que el índice real bien podría ser igual a 1.1 (ya que este valor pertenece al intervalo de confianza para el valor medio de los relativos precios) y se dudaría que llegase a ser 1.25 (puesto que 1.25 no está dentro del intervalo de confianza) a menos que un suceso extraño e inesperado tuviera lugar.

Para el mes febrero la situación es más grave, puesto el índice estimado con $^{(1)}I'_{m,0}$ sería 1.1221, el cual si se compara con el valor medio estimado $^{(-0.75)}I'_{m,0} = 1.0251$ que aparece en la tabla 1, indica una sobres estimación de 9.46%. Lo que es más importante aún es que el valor 1.1221 no queda cubierto por el intervalo de confianza para el valor medio real de los relativos de precios, lo cual se traduce en que dicho valor debería ser rechazado como candidato para índice de precio de mes de febrero.

La conjetura que surge del resultado anterior es que existirá sobrestimación del índice de precios si se usa $^{(1)}I'_{m,0}$, cuando $\hat{\lambda} < 1$, lo cual ocurre si algunos de los precios observados en el período m son extremadamente grandes comparados con los precios del período base. Por el contrario, la fórmula $^{(1)}I'_{m,0}$ subestimaré el valor medio de los relativos de precios si $\hat{\lambda} > 1$; para que el valor estimado de λ sea mayor que 1 se requiere que algunos de los precios registrados para el periodo m sean extremadamente pequeños en relación con los precios de los mismos artículos en el período base. Debido a que el primer caso es el más factible de ocurrir en la realidad, es de esperar entonces que el promedio $^{(1)}I'_{m,0}$ tienda a sobrestimar el índice de precios, pero no es posible indicar la magnitud de tal sobrestimación a menos que se utilice el método aquí sugerido y se compare el resultado con $^{(1)}I'_{m,0}$.

Un aspecto que debe tomarse en consideración al utilizar el procedimiento en la práctica, es que las comparaciones que se efectúen entre meses deberán ser en términos de valores medios (o medianas) de relativos de precios referidos al mismo período base, puesto que un

cambio de período base implicaría un cambio en la forma de la distribución de relativos de precios y, por consiguiente, el tipo de promedio paramétrico adecuado también cambiaría. Además, debe notarse que no existe razón para pretender emplear el mismo promedio paramétrico en el cálculo de índices para diversos períodos, a menos que se establezca un compromiso entre los promedios óptimos para cada uno de los períodos, a fin de utilizar el mismo promedio para todos y cada uno de dichos períodos.

Quizá la más importante característica del método de selección de índices de precios aquí propuesto es que se pueden obtener intervalos de confianza de una manera sencilla y "natural". Desde luego que la validez de los mismos depende de que se satisfagan los supuestos del método.

2. Simulación con distribuciones asimétricas

Para tener una idea de la bondad del método aquí propuesto, se generaron muestras aleatorias de tamaño $n=20, 50$ y 100 de las siguientes distribuciones asimétricas:

$$\text{exponencial} - f_X(x) = 2e^{-2x}; x > 0$$

$$\text{beta} - f_X(x) = [\beta(3,2)]^{-1} x^2(1-x); 0 < x < 1.$$

Los resultados de la aplicación del procedimiento se muestran en la tabla 2, a continuación.

Una de las conclusiones inmediatas que se pueden obtener de este pequeño estudio de simulación es que las fórmulas desarrolladas en este trabajo producen resultados razonablemente buenos. Esto se deduce al comparar las estimaciones con los valores (teóricos) reales. En particular, se puede advertir que el factor de corrección (4.5) permite obtener aproximaciones adecuadas.

Tabla 2.

RESULTADOS DE LA SIMULACION

Parámetro	Valor Teórico	Valor estimado e intervalos ^{1/} del 95% de confianza		
		n=20	n=50	n=200
Distribución exponencial				
λ	-	0.26 (0.00,0.50)	0.30 (0.12,0.50)	0.29 (0.18,0.40)
Med(X)	0.3466	0.2202 (0.1088,0.3995)	0.3818 (0.2676,0.5566)	0.3806 (0.3287,0.4380)
E(X)	0.5000	0.3865 (0.1909,0.7013)	0.5730 (0.4015,0.7901)	0.5292 (0.4570,0.6091)
Var(X)	0.2500	0.2124	0.3602	0.2492
Distribución beta				
λ	-	1.39 (0.24,2.75)	1.09 (0.42,1.80)	1.27 (0.91,1.64)
Med(X)	0.6143	0.6142 (0.5176,0.7051)	0.5923 (0.5272,0.6565)	0.6179 (0.5898,0.6457)
E(X)	0.6000	0.6011 (0.5066,0.6901)	0.5884 (0.5296,0.6550)	0.6088 (0.5899,0.6459)
Var(X)	0.0400	0.0406	0.0512	0.0410

^{1/} Cuando aparecen dos intervalos de confianza, el primero de ellos se calculó de acuerdo con (4.3) para Med(X) o (4.4) para E(X), y el segundo de acuerdo con (4.9) o (4.10).

VI . CONCLUSIONES

La definición del promedio paramétrico en términos de la transformación de Box y Cox, permite hacer uso del método de máxima verosimilitud para determinar, de manera objetiva, el promedio y su medida de dispersión asociada adecuados para describir el conjunto de datos con que se cuente.

El utilizar promedios paramétricos, es equivalente a transformar los datos y hacer estimaciones en la escala transformada; como una aproximación al sesgo introducido por la transformación, se presenta un factor multiplicativo que puede estimarse con base a los datos; mediante la aplicación de este factor es posible corregir por sesgo tanto al promedio como a su medida de dispersión.

Como una aplicación en la práctica del método propuesto, se presenta la determinación de índices de precios con sus correspondientes intervalos de confianza. Los resultados obtenidos con base en una muestra de 151 artículos específicos, para dos meses arbitrarios, permiten observar la utilidad que el método puede tener en la práctica.

Así mismo, para corroborar los resultados teóricos, se realizó un pequeño estudio de simulación con dos distribuciones asimétricas. Los resultados logrados muestran una bondad bastante razonable del método.

REFERENCIAS

1. Allen, R.G.D. (1975) Index Numbers in Theory and Practice. The MacMillan Press, Ltd.
2. Box, G.E.P. y Cox, D.R. (1964) "An Analysis of Transformations" Journal of the Royal Statistical Society B-26; 211-252.
3. Draper, N.R. y Cox, D.R. (1969) "On Distributions and their Transformation to Normality". Journal of the Royal Statistical Society B-31; 472-476.
4. Fisher, I. (1927) The Making of Index Numbers. Publications of the Pollack Foundation for Economic Research No. 1 Third Edition, Sir Isaac Pitman and Sons, Ltd.
5. Guerrero, G.V.M. y Johnson, R.A. (1979) "Transformation of Grouped or Censored Data to Near Normality". Technical Report # 542 University of Wisconsin-Madison.
6. Hoyle, M.H. (1973) "Transformations -An Introduction and a Bibliography". International Statistical Review 41; 203-223
7. Land, C.E. (1974). "Confidence Interval Estimation for Means after Data Transformations to Normality". Journal of the American Statistical Association 69; 795-802.
8. Norris, N. (1940) The Standard Errors of the Geometric and Harmonic Means and Their Application to Index Numbers". Annals of Mathematical Statistics 11; 445-448.
9. Norris, N. (1976) "General Means and Statistical Theory". The American Statistician 30; 8-12.
10. Patterson, R.L. (1966) "Difficulties Involved in the Estimation of a Population Mean Using Transformed Sample Data". Thechnometrics 8; 535-537.

BANCO DE MEXICO, S.A.

SUBDIRECCION DE INVESTIGACION ECONOMICA

SERIE DE DOCUMENTOS DE INVESTIGACION

1. ESTRUCTURA FINANCIERA Y EXPERIENCIA CAMBIARIA: MÉXICO 1954-1977. Guillermo Ortiz y Leopoldo Solís
2. EL FINANCIAMIENTO DEL GASTO PÚBLICO EN UNA ECONOMÍA EN CRECIMIENTO: EL CASO DE MEXICO. Alain Ize. Noviembre, 1978.
3. ALGUNOS ASPECTOS DEL ENDEUDAMIENTO PUBLICO EXTERNO EN MEXICO. Ernesto Zedillo. Diciembre, 1978.
4. UNA APLICACION DEL MODELO BAYESIANO DE DECISION EN EL ANALISIS DE FUNCIONES DE PRODUCCION AGRICOLA. Héctor E. González Méndez. Diciembre, 1978.
5. POLITICA MACROECONOMICA EN EL CORTO PLAZO: UNA RESEÑA. Alain Ize. Marzo, 1979.
6. ESTUDIOS DE MONEDA Y BANCA Y POLITICA MONETARIA SOBRE MEXICO: SELECCION BIBLIOGRAFICA DE 1943 A 1978. Abril, 1979.
7. COMERCIO EXTERIOR MEXICO-ESTADOS UNIDOS: PROBLEMAS DE COMPARABILIDAD ESTADISTICA. Jorge Carriles Rubio. Mayo, 1979.
8. EXPLOTACION OPTIMA DE RESERVAS PETROLERAS EN UN CONTEXTO MACROECONOMICO. José Córdoba. Mayo, 1979.
9. ASPECTOS DEFLACIONARIOS DE LA DEVALUACION DEL PESO MEXICANO DE 1976. José Córdoba y Guillermo Ortiz. Mayo, 1979.
10. EXTRACCION OPTIMA DE PETROLEO Y ENDEUDAMIENTO EXTERNO: EL CASO DE MEXICO. Ernesto Zedillo. Junio, 1979.
11. IMPUESTOS DIRECTOS: PROGRESIVIDAD OPTIMA. Jesús Seade. Septiembre, 1979.
12. OPCIONES DE POLITICA ECONOMICA 1979-1982. Sócrates Rizzo y Leopoldo Solís. Septiembre, 1979.
13. INTERMEDIARIOS FINANCIEROS Y MERCADOS IMPERFECTOS DE CAPITAL. Guillermo Ortiz. Septiembre, 1979.

14. ESTIMACIONES DE EQUILIBRIO GENERAL DE LOS EFECTOS DE LAS DISTORSIONES EN LOS MERCADOS DE FACTORES: EL CASO DE MEXICO. José J. Sidaoui y Richard H. Sines. Octubre, 1979.
15. UN ANALISIS DE LA INFLACION EN MEXICO. Alain Ize. Octubre, 1979.
16. ANALISIS DE LOS COMPONENTES DEL CAMBIO ESTRUCTURAL CON UN MODELO DE EQUILIBRIO GENERAL, 1970-75. José J. Sidaoui y Richard H. Sines. Enero, 1980.
17. TIPOS DE CAMBIO FLOTANTES Y DESLIZ CAMBIARIO: LAS EXPERIENCIAS DE ALGUNOS PAISES EN DESARROLLO. Guillermo Ortiz y Leopoldo Solís. Enero, 1980.
18. UN MODELO DE INFLACION Y CRECIMIENTO EN UNA ECONOMIA CAPITALISTA EN DESARROLLO. Alain Ize. Enero, 1980.
19. CRECIMIENTO E INFLACION: ALTERNATIVAS CAMBIARIAS PARA MEXICO. Guillermo Ortiz y Leopoldo Solís. Febrero, 1980.
20. COMPORTAMIENTO DE LA CAPTACION BANCARIA EN MEXICO. Héctor E. González Méndez. mayo, 1980.
21. LA ENCUESTA DE TURISMO RECEPTIVO, REPORTE METODOLOGICO. Alberto Vargas Aguayo. Junio, 1980.
22. AJUSTE ESTACIONAL DE UNA SERIE DE TIEMPO MEDIANTE EL USO COMPLEMENTARIO DE METODOS TRADICIONALES Y LA TECNICA DE BOX-JENKINS. Gabriel Vera Ferrer y Víctor M. Guerrero. Julio, 1980.
23. DISTRIBUCION DEL FINANCIAMIENTO OTORGADO POR EL SISTEMA BANCARIO MEXICANO A LA BANCA PRIVADA Y MIXTA. Víctor M. Guerrero y Gabriel Vera Ferrer. Julio, 1980.
24. LA MIGRACION INDOCUMENTADA A ESTADOS UNIDOS: UN NUEVO ENFOQUE. Juan Díez Cañedo. Julio, 1980.
25. UN MODELO FINANCIERO DE DESEQUILIBRIO A CORTO PLAZO PARA LA ECONOMIA MEXICANA. Alain Ize. Julio, 1980.
26. ESTIMACION DE LA FUNCION DE IMPORTACIONES PARA MEXICO. Javier Salas. Agosto, 1980.
27. UNA ALTERNATIVA PARA LA MEDIA ARITMETICA EN EL CALCULO DE PROMEDIOS SIMPLES DE RELATIVOS DE PRECIOS: LA MEDIA GEOMETRICA. Gabriel Vera Ferrer y Víctor M. Guerrero. Agosto, 1980.

28. LA DEMANDA DE DINERO EN MEXICO: PRIMERAS ESTIMACIONES. Guillermo Ortiz. Septiembre, 1980.
29. ECONOMIAS DE ESCALA Y CONCENTRACION BANCARIA: EL CASO DE MEXICO. Héctor E. González Méndez. Octubre, 1980.
30. LA ESTABILIDAD DE LA DEMANDA DE DINERO EN MEXICO. Guillermo Ortiz. Noviembre, 1980.
31. EL TAMAÑO DE LA FAMILIA Y LA DISTRIBUCION DEL INGRESO EN MEXICO: UN ENSAYO EXPLORATORIO. Gabriel Vera Ferrer, Diciembre, 1980.
32. PROMEDIOS PARAMETRICOS: SU SELECCION Y EMPLEO EN LA DETERMINACION DE INDICES DE PRECIOS. Víctor M. Guerrero, Enero, 1981.