

# Graduación no-paramétrica con suavidad y estructura impuestas por el analista: aplicaciones demográficas para México\*

**Víctor M. Guerrero\*\***

**Eliud Silva\*\*\***

*Se presenta un método estadístico de carácter no-paramétrico para graduar datos demográficos de manera que se obtenga no sólo suavidad, sino que los datos graduados sigan cierta estructura impuesta por el analista. El principal objetivo es que éste sea capaz de controlar tres aspectos fundamentales de la graduación: la fidelidad de los datos graduados a los datos originales, la suavidad de dichos datos graduados, y la cercanía de los mismos a determinada estructura. Las ilustraciones empíricas utilizan datos referidos a la realidad demográfica de México y hacen uso de diversos indicadores al alcance de los analistas interesados.*

Palabras clave: comparabilidad, Filtro de Kalman, graduación, índice de estructura, índice de suavidad, Mínimos Cuadrados Generalizados, Modelo de Componentes No-Observables, Modelo de Espacio de Estados, Modelo de Señal Más Ruido.

Fecha de recepción: 5 de enero de 2011.

Fecha de aceptación: 26 de junio de 2012.

## Nonparametric Graduation with Smoothness and Structure Imposed by the Analyst: Demographic Applications for Mexico

*A non-parametric statistical method is presented for graduating demographic data so as to obtain not only smoothness, but also to ensure that the graduating data follow a certain structure imposed by the analyst. The main aim is for it to be able to control three fundamental aspects of graduation: the faithfulness of the graduated data to the original data, the smoothness of these graduated data, and their proximity to a particular struc-*

\* Trabajo ganador del tercer lugar del Premio Gustavo Cabrera 2010, VII Convocatoria, en la categoría de Mejor Investigación en el campo de Demografía o Población. Los autores reconocen la valiosa aportación de los dictaminadores anónimos, así como la del editor, para la mejora del presente documento. Víctor M. Guerrero agradece a la Asociación Mexicana de Cultura, A.C. el apoyo que le brindó para la realización de esta investigación.

\*\* Departamento de Estadística, Instituto Tecnológico Autónomo de México. Correo electrónico: <guerrero@itam.mx>.

\*\*\* Escuela de Actuaría, Universidad Anáhuac del Norte. Correo electrónico: <jsilvaurrutia@hotmail.com>.

*ture. The empirical illustrations use data regarding the demographic reality of Mexico and make use of various indicators available to interested analysts.*

Key words: comparability, Kalman Filter, graduation, structure index, smoothness index, Generalized Least Squares, Unobservable Components Model, State Space Model, Signal Plus Noise Model.

## **Introducción**

Los datos demográficos, ya sea que provengan de censos, de encuestas o de estadísticas vitales, comúnmente sufren diversas anomalías. Si bien algunas de ellas son atribuibles a errores humanos (de los informantes, de los encuestadores o del personal encargado de transcribir y capturar los datos), ésta no es la única fuente de errores ni necesariamente la más importante, pues ciertas cuestiones climáticas (como los huracanes, las sequías, etcétera) o fuera del control de los analistas (terremotos, huelgas, etcétera) provocan también la generación de errores en las bases de datos. Desde luego, estos errores distorsionan el patrón verdadero del fenómeno que los datos pretenden cuantificar y su presencia hace necesaria la aplicación de herramientas que corrijan, o al menos mitiguen, el efecto perverso de tales distorsiones en lo que respecta a los resultados de posibles análisis que se efectúen con los datos. Ello ocurre de esta manera sin importar lo elaborados que puedan ser tales análisis, los cuales pueden ir desde una mera descripción superficial de los patrones más relevantes en los datos, hasta un análisis confirmatorio de alguna teoría que explique el comportamiento de la población bajo estudio.

Una de tales “herramientas correctoras”, que se utiliza con mucha frecuencia en la práctica, es la graduación de datos. Puede aplicarse tanto en las instituciones gubernamentales, con fines muy generales para percibir de manera lo más clara posible los patrones subyacentes en los datos, o en organizaciones privadas con fines muy particulares, para tomar decisiones específicas de las mismas. Esto se debe en buena medida a que es muy fácil efectuar la graduación de datos debido a que se basa en ideas muy simples y su instrumentación computacional es relativamente fácil; asimismo porque el costo asociado con su aplicación es muy bajo (un programa computacional sencillo permite hacer la aplicación).

Tanto la sencillez de representación como la facilidad y el bajo costo de implementación computacional se mantienen en la propuesta que aquí se presenta. Adicionalmente, el nuevo método que se sugiere produce ganancias en cuanto a la operatividad de conceptos como la fidelidad a los datos originales, la suavidad de los datos graduados y la cercanía de éstos a una estructura que se considera como meta. Para ello primero se formalizan estas ideas mediante una representación de los datos disponibles a través de un modelo estadístico de componentes no-observables, para el cual se asocia en forma natural el método de estimación de parámetros denominado Mínimos Cuadrados Generalizados (MCG). Posteriormente se muestra que es posible cuantificar la suavidad y la cercanía a la estructura meta por medio de unos índices expresables en forma de porcentajes, con lo cual se sigue que es factible controlar estas características de la graduación propuesta simplemente al fijar valores para tales índices, de acuerdo con el criterio del analista.

La importancia de efectuar el control mencionado radica en que con ello se pueden realizar comparaciones válidas entre datos con suavidades y cercanías a estructuras límites similares, algo que de otra manera no tendría mucho sentido intentar siquiera comparar. Finalmente, los cálculos de la graduación aquí propuesta se realizan de manera muy sencilla al emplear una herramienta de cálculo muy poderosa conocida como Filtro de Kalman, que es aplicable cuando el Modelo de Componentes No-Observables se interpreta en la forma de un Modelo de Espacio de Estados.

La estructura del documento es la siguiente: en la próxima sección se presentan brevemente algunos modelos y técnicas no-paramétricas que suelen usarse en la graduación de datos; en la tercera se mencionan algunas técnicas de carácter demográfico que sirven para analizar y proyectar datos. Posteriormente, en la cuarta sección se presenta el método propuesto para llevar a cabo la graduación; ahí se muestran las ideas fundamentales, que se expresan por medio de ecuaciones y dan origen al Modelo de Componentes No-Observables por utilizar. En la quinta sección se presentan los índices de estructura y suavidad que se utilizan para cuantificar los respectivos conceptos y que permiten fijar los porcentajes deseados para los mismos. Una vez que se han fijado los valores para los índices, se demuestra que se pueden deducir los valores para las constantes respectivas, las cuales forman parte de la especificación del modelo. En la sexta sección se pretende mostrar, mediante diversas aplicaciones a datos demográficos de la realidad

mexicana, la utilidad de la metodología propuesta y el tipo de resultados que se pueden obtener de la misma. En la última sección se exponen algunas conclusiones para destacar los pros y contras del uso de la metodología propuesta.

### **Métodos no-paramétricos**

Haberman y Renshaw (1996) definen la graduación como un grupo integral de técnicas y principios que permiten realizar el ajuste de probabilidades o datos en general para suavizarlos de manera que se puedan realizar inferencias y análisis del tipo que sea necesario. La graduación de datos puede efectuarse mediante procedimientos que típicamente se clasifican en paramétricos o no-paramétricos. La segunda clase de técnicas no requiere del supuesto de que una cierta función represente el comportamiento de los datos globalmente y en forma estricta; por el contrario, se pretende suavizar las fluctuaciones que básicamente oscurecen la tendencia subyacente en los datos observados y obtener representaciones localmente aceptables. Los métodos no-paramétricos son más flexibles y robustos, en tanto que son menos los supuestos que respaldan su uso. Además, en muchas ocasiones resulta más fácil emplear los métodos no-paramétricos que su contraparte paramétrica, debido al menor rigor implícito en su aplicación.

La propuesta de este trabajo se encuentra dentro del ámbito de los métodos no-paramétricos, ya que se pretende simplificar el análisis de los datos observados después de que se les ha eliminado una parte de la variabilidad que no es intrínseca de ellos. De hecho, los datos originales se convierten en estimaciones una vez que se les han cancelado las fluctuaciones que oscurecen su tendencia. Dentro de las técnicas no-paramétricas que se utilizan más frecuentemente se encuentran los métodos gráficos, los promedios móviles ponderados, el método del núcleo y la graduación en general, en especial la que se aplica a los datos demográficos (Copas y Haberman, 1983; Papaioannou y Sachlas, 2004). Por ejemplo, algunas aplicaciones de los modelos no-paramétricos en demografía permiten obtener estimaciones de mortalidad en edades avanzadas (Fledelius *et al.*, 2004). Asimismo, en el trabajo de Debón *et al.* (2006) se muestran diversas comparaciones entre métodos no-paramétricos y de suavizamiento realizado con Modelos Aditivos Generalizados, en particular con *splines*, también dentro de un contexto demográfico.

La técnica de graduación que se usa más frecuentemente en la práctica es la de Whittaker y Henderson. Esta forma de graduación se aplica a los datos originales para obtener datos suavizados de manera tal que se satisfagan dos criterios: bondad de ajuste (es decir, fidelidad a los datos originales) y suavidad de los datos resultantes. Al lector interesado en detalles acerca de este tipo de graduación y otros métodos similares se le recomienda consultar el libro de London (1985). Para lograr que los dos criterios se cumplan, se pondera en forma relativa su importancia mediante una constante o parámetro de suavizamiento. De hecho, el método surge al resolver un problema de minimización de la función siguiente:

$$(\mathbf{v}-\mathbf{u})' W(\mathbf{v}-\mathbf{u}) + \lambda \mathbf{v}' K'_d K_d \mathbf{v} \tag{1}$$

en donde sobresale la constante de suavizamiento  $\lambda > 0$  recién mencionada, la cual se supone conocida al efectuar la minimización. El vector  $\mathbf{u} = (u_1, \dots, u_n)'$  contiene los valores observados de la variable en cuestión, y el vector de valores graduados  $\mathbf{v} = (v_1, \dots, v_n)'$  es el que se desea obtener. Se usa también la matriz diagonal  $W = \text{diag}(w_1, \dots, w_n)$  que sirve para asignar ponderaciones a las diferencias entre valores observados y graduados. Finalmente, la matriz  $K_d$ , que permite aplicar diferencias a datos contiguos, es de dimensión  $(n - d) \times n$  y está definida de forma tal que su  $ij$ -ésimo elemento está dado por la expresión

$$K_d(i,j) = (-1)^{d+i-j} d! / [(j-i)!(d-j+i)!] \tag{2}$$

la cual es válida para  $i = 1, \dots, n-d$  y  $j = 1, \dots, n$ , junto con  $K_d(i, j) = 0$  para  $j < i$  ó  $j > d+i$ .

El caso particular  $d = 2$  del método de Whittaker y Henderson fue redescubierto por Hodrick y Prescott en un trabajo que elaboraron dichos autores en 1980 dentro de un contexto de análisis económico, aunque el artículo correspondiente se publicó apenas en 1997 (véase Hodrick y Prescott, 1997). Por este motivo dicho método se conoce como Filtro de Hodrick y Prescott (HP) en el área económica, donde se utiliza ampliamente para estimar tendencias y realizar análisis de ciclos económicos (véase Guerrero, 2008, para mayores detalles sobre el citado Filtro HP). El problema de minimización correspondiente al valor  $d = 2$  puede plantearse como

$$\min_{Y_t^s} \sum \frac{1}{\sigma_0^2} (Y_t - Y_t^s)^2 + \frac{1}{\sigma_1^2} (\nabla^2 Y_t^s)^2 \tag{3}$$

donde ahora se define a  $Y_t$  como la variable observada en el tiempo de observación  $t$ , mientras que  $Y_t^S$  es su correspondiente valor suavizado de tendencia, que no es observable y se pretende estimar,  $\sigma_0^2$  es la varianza del componente cíclico, definido como la desviación del dato observado respecto a la tendencia  $\{Y_t - Y_t^S\}$ , mientras que  $\sigma_1^2$  es la varianza del crecimiento de la tendencia,  $\nabla^2 Y_t^S$ . Por su lado, el parámetro de suavizamiento es de la forma  $\lambda = \sigma_0^2 / \sigma_1^2$ , lo cual permite interpretarlo como una constante con la cual se puede establecer un balance entre la fidelidad de la serie suavizada a los datos originales y la suavidad de la tendencia que se obtiene. Al resolver el problema de minimización, en términos del vector de datos observados  $\mathbf{Y} = (Y_1, \dots, Y_n)'$  se produce el siguiente estimador de la tendencia de los datos. Para detalles acerca de este resultado véase Hodrick y Prescott (1997) o Guerrero (2008),

$$\hat{\mathbf{Y}}^S = (\mathbf{I}_n + \lambda \mathbf{K}_1' \mathbf{K}_2)^{-1} \mathbf{Y} \tag{4}$$

con  $\hat{\mathbf{Y}}^S = (\hat{Y}_1^S, \dots, \hat{Y}_n^S)'$ ,  $\mathbf{I}_n$  la matriz identidad de dimensión  $n \times n$  y  $\mathbf{K}_2$  la matriz definida mediante la expresión [2], con el valor  $d = 2$ , es decir, es de la forma

$$\mathbf{K}_2 = \begin{pmatrix} 1 & -2 & 1 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & & 0 & 0 & 0 & 0 \\ & & & & & \dots & & & & \\ 0 & 0 & 0 & 0 & 0 & \dots & 0 & 1 & -2 & 1 \end{pmatrix} \tag{5}$$

La expresión [4] es precisamente el estimador producido por el Filtro HP cuyo uso requiere, además del vector de datos  $\mathbf{Y}$ , conocer el valor del parámetro de suavizamiento  $\lambda$ , puesto que los demás elementos de esa expresión son conocidos. La solución que dieron Hodrick y Prescott para elegir dicha constante de suavizamiento está basada en argumentos del dominio de las frecuencias, aplicables al análisis de ciclos económicos, con los cuales se obtuvo como resultado el valor  $\lambda = 1600$ , para series económicas trimestrales de Estados Unidos en la época de la posguerra, de longitud aproximada a los 100 datos. Para otros casos el analista debe ser responsable de elegir el valor más apropiado, teniendo en mente que un valor pequeño conducirá a una tendencia poco suave (muy semejante a los datos que se observaron originalmente). En cambio, al elegir un valor elevado para dicha cons-

tante la tendencia será muy suave (*i. e.*, se aproximará a la línea recta estimada para los datos observados mediante el método de Mínimos Cuadrados Ordinarios).

Dado el éxito del Filtro HP para realizar análisis de ciclos económicos, Laxton y Tetlow (1992) propusieron una extensión de dicho resultado y desarrollaron el filtro conocido como Filtro HP Multivariado (HPMV), con el cual se extiende la aplicabilidad de la herramienta original, en el sentido de que además de las ideas de fidelidad a los datos y suavidad se puede incorporar la cercanía a una estructura dictada por algún tipo de teoría del fenómeno en estudio. De esta forma se obtiene el filtro correspondiente al minimizar nuevamente una función de los valores estimados y que ahora considerará también la existencia de un error aleatorio asociado con la discrepancia entre lo que indica la teoría y lo que muestran los datos observados. En consecuencia, el Filtro HPMV sirve para estimar la tendencia  $Y_t^S$  (que no es observable) de una variable  $Y_t$  (observable) mediante la solución del siguiente problema de minimización:

$$\min_{Y_t^S} \sum (Y_t - Y_t^S)^2 + \lambda_1 (\nabla^2 Y_t^S)^2 + \lambda_2 \xi_t \quad [6]$$

donde aparecen ahora dos parámetros de suavizamiento,  $\lambda_1$  y  $\lambda_2$ . Desde luego debe notarse la similitud entre las expresiones [3] y [6], ya que esta última es una extensión de la anterior, en tanto que ahora incluye también el error aleatorio  $\xi_t$  asociado con alguna teoría del fenómeno que involucra a  $Y_t^S$ . Una presentación ampliada y con detalle del desarrollo del Filtro HPMV se puede consultar en Boone (2000).

### **Algunas técnicas demográficas de modelación**

La metodología que se propone en este trabajo para corregir anomalías en los datos demográficos parte de la idea de que se puede representar el comportamiento de tales datos mediante un modelo estadístico muy general. Conviene, por lo tanto, conocer algunas técnicas que también se basan en representaciones o modelos generales para tratar de encontrar similitudes en los distintos enfoques. El Método de Componentes es, sin lugar a dudas, el que más se utiliza en la práctica para efectuar proyecciones demográficas. El método como tal no ha variado en su esencia, y en términos generales se ha usado para estudiar el comportamiento futuro de los diversos componentes demográficos en

forma separada, esto es, la fecundidad, la migración y la mortalidad para horizontes previamente determinados (George *et al.*, 2004).

El Método de Componentes presenta variantes que permiten hacer supuestos acerca del patrón que siguen, por ejemplo, las tasas de mortalidad. A partir de los supuestos, las técnicas pueden agruparse de la siguiente manera: *a)* técnicas extrapolativas; *b)* métodos en que se piensa que la mortalidad de cierta área geográfica es válida para otras áreas; y *c)* modelos estructurales en que los cambios de las tasas de mortalidad se asocian con cambios en ciertas variables socioeconómicas. Para los grupos de técnicas *a)* y *b)*, las posibilidades existentes incluyen el uso de modelos del tipo Auto-Regresivo Integrado y de Promedios Móviles (ARIMA), como sucede con el método que propusieron Lee y Carter (1992), o bien modelos paramétricos como los de Makeham, Gompertz o Helligman y Pollard, entre otros. De igual manera, las tablas de vida de diferentes lugares del mundo se pueden utilizar como tablas base de referencia, dentro de las cuales están las *tablas modelo* que presentan diferentes niveles de mortalidad y estructuras, así como la función *logit* y otras más.

Otros métodos que también se encuentran dentro de las categorías *a)* y *b)* tienen su fundamento en *tablas límite de mortalidad*, de manera que hacen uso de los niveles más bajos alcanzados para interpolar las tablas intermedias. La primera propuesta de tablas límite de mortalidad fue presentada por Bourgeois-Pichat (1952) con el supuesto principal de que los niveles límite se alcanzarían en el largo plazo. La hipótesis que subyace en este tipo de métodos surge de la idea de que la mortalidad cambia como función del nivel y la estructura de la mortalidad, dependiendo de la región del mundo a la que corresponda. Por otra parte, en lo que toca a los límites de la sobrevivencia humana los trabajos de Olshansky *et al.* (1990, 2001) y de Oeppen y Vaupel (2002) ofrecen aportaciones importantes a la literatura porque estudian la reducción en mortalidad que se necesita para alcanzar una esperanza de vida al nacer que aumente de 80 a 120 años y cómo puede afectar esto a distintas áreas de la política pública.

En el caso *b)*, por ejemplo, se utiliza la técnica de alcanzar una meta establecida de antemano. Dicha técnica se basa en la idea de que, para una población dada, las tasas de mortalidad convergen con las que se observan en otra población y se consideran como una meta por alcanzar. La población meta en cuestión debe elegirse de manera tal que brinde un conjunto de metas creíbles que puedan ser alcanzadas por la población que se proyecta. La elección de la población meta se



basa comúnmente en similitudes que se refieren a características culturales, socioeconómicas, avances en la medicina y a las causas primarias de mortalidad (Olshansky, 1988). Una alternativa de presentar las metas consiste en referirlas a lo que se conoce como retraso de la causa; con este tipo de enfoque la población meta que se elige es una cohorte más joven de la misma población en estudio, en lugar de la misma cohorte de una población diferente. El objetivo se ubica por lo general en las implicaciones que conllevan el retraso o la eliminación total de la ocurrencia de una o más causas de mortalidad (Manton *et al.*, 1980 y Olshansky, 1987).

### **Metodología propuesta**

Esta metodología surge al reconocer que existe una fuerte conexión entre suavizar una serie de tiempo y estimar su tendencia (Guerrero y Silva, 2010). De hecho, al suavizar una serie se busca un valor central que represente el comportamiento local de la variable, y eso es lo que en esencia se busca también al estimar la tendencia. Debido a ello conviene recordar algunos resultados importantes acerca de la estimación de tendencias para series de tiempo univariadas. En ese contexto la extracción de señal con el Filtro de Wiener y Kolmogorov, el Filtro de Kalman con suavizamiento y Mínimos Cuadrados Penalizados produce resultados que son equivalentes a los que arroja el Filtro HP que emplean los analistas de ciclos económicos. De manera semejante se ha demostrado que el método estadístico de estimación de MCG produce resultados idénticos a los que resultan de los filtros anteriores (Guerrero, 2007). Además, también se puede apreciar que el inverso de la matriz de Error Cuadrático Medio (ECM) es la suma de dos matrices de precisión. A partir de estos hechos Guerrero propuso medir la participación a la precisión de la tendencia estimada, que corresponde al elemento de suavidad del modelo estadístico implícito en el proceso de estimación. Esa medida de participación de la suavidad conduce a un índice de suavidad que depende sólo del parámetro de suavizamiento y del número de datos de la serie en estudio. Por lo tanto, dada una serie de longitud fija, el índice de suavidad brinda la posibilidad de decidir el valor del parámetro de suavidad como función sólo de un porcentaje deseado de suavidad, el cual puede elegirse en forma anticipada.

El enfoque tradicional para suavizar una serie de datos utiliza el parámetro de suavizamiento  $\lambda$  seleccionado con ayuda de algún crite-

rio numérico, por ejemplo el criterio de Akaike, conocido como AIC, que se minimiza en forma automática sin que el analista se entere de los efectos de la elección que surja (a este respecto véase Hastie y Tibshirani, 1990). Sin embargo debe reconocerse que al suavizar los datos con un valor específico de  $\lambda$  se obtiene también un porcentaje de suavidad específico para la tendencia. Así pues, desde un punto de vista meramente descriptivo, un analista de los datos de la serie que aplique el suavizamiento debería, por lo menos, reportar el porcentaje de suavidad alcanzado con el parámetro  $\lambda$  que utilice. En esta línea de pensamiento se considera aun mejor fijar por adelantado un monto deseado de suavidad para la tendencia, en lugar de elegir el valor de  $\lambda$  en forma automática. Esta idea es semejante a la que respalda la costumbre actual de fijar de antemano el nivel de confianza (digamos en 95%) para estimar parámetros en un análisis estadístico, ya que así se pueden establecer comparaciones válidas entre dos o más intervalos de confianza. Este argumento se extiende en el presente trabajo al caso de fijar el porcentaje de suavidad, junto con el de estructura, que se pretende alcanzar con la estimación de la tendencia. Lo cual es necesario, de nuevo, para poder establecer comparaciones válidas. En resumen, lo que aquí se propone es calibrar los parámetros de suavidad y estructura que intervienen en el proceso de estimación de la tendencia de datos demográficos, de manera que se reduzca la subjetividad en el uso del procedimiento.

Es importante reconocer que el solo hecho de poder medir la suavidad con el índice propuesto permite al analista comparar los resultados de dos series suavizadas no sólo mediante inspección visual, sino numéricamente. Es en este sentido que la decisión acerca de cuál de las dos series tiene mayor suavidad se puede tomar objetivamente, o al menos con base en los datos disponibles, no en creencias subjetivas. Por estos motivos, la sugerencia que aquí se hace es usar el Filtro HPMV para estimar tendencias en datos demográficos incorporando las ideas de suavidad y estructura. Para presentar la propuesta formalmente se usará primero un modelo de señal más ruido, o sea,

$$Y_t = Y_t^S + \eta_t \quad [7]$$

donde  $Y_t$  denota la variable de interés (digamos la mortalidad),  $Y_t^S$  es la señal, que en el presente caso representa la tendencia de la mortalidad suavizada, y  $\eta_t$  es el ruido, que básicamente oscurece el comportamiento de la tendencia, ya que en la práctica sólo se observa  $Y_t$ . Al

penalizar por falta de suavidad y alejamiento de estructura respecto a  $Y_t^S$  surge el siguiente problema de minimización:

$$\min_{Y_t^S} \sum_{t=1}^n (Y_t - Y_t^S)^2 + \lambda_1 (\nabla^d Y_t^S)^2 + \lambda_2 \delta_t \tag{8}$$

con  $\delta_t$  el error aleatorio asociado con un modelo demográfico en donde se hace uso de la variable  $Y_t^S$ . Este problema es similar al que plantea Boone (2000), mediante el cual se intenta estimar ahora los valores de  $Y_t^S$  como solución de [8].

Esta perspectiva consiste en definir primero un índice de suavidad que sirva de ayuda para elegir las constantes  $\lambda_1$  y  $\lambda_2$ . La metodología propuesta brinda la posibilidad de interpretar entonces los resultados de acuerdo con una teoría demográfica que permite realizar comparaciones válidas entre tendencias de los datos. Dicha estimación se efectúa mediante un balance de los siguientes tres elementos: los datos observados originalmente, la suavidad de la tendencia que se obtenga como resultado, y la cercanía de dicha tendencia a cierta estructura teórica presupuesta como meta. El modelo estadístico que se emplea surge del planteamiento de las siguientes representaciones: *i*) los datos observados pueden expresarse como una tendencia oscurecida por un error aleatorio; *ii*) el patrón de suavidad subyacente de la tendencia es de carácter polinomial de orden uno; y *iii*) la estructura teórica que se supone como meta proviene de una fuente de información externa a los datos originales y sirve para incorporar una meta en los datos suavizados. Estas tres representaciones dan origen a las expresiones que siguen y que, en conjunto, forman el modelo

$$Y = Y^S + \eta, \quad \eta \sim (\mathbf{0}, \sigma_\eta^2 I_n) \tag{9}$$

$$K_2 Y^S = \varepsilon, \quad \varepsilon \sim (\mathbf{0}, \sigma_\varepsilon^2 I_{n-2}), \quad E(\varepsilon \eta') = 0 \tag{10}$$

y

$$U = Y^S + \delta, \quad \delta \sim (\mathbf{0}, \sigma_\delta^2 I_n), \quad E(\delta \eta') = 0, \quad E(\delta \varepsilon') = 0 \tag{11}$$

donde el símbolo  $\sim$  significa “distribuido como” (vector de medias, matriz de varianza-covarianza). De esta forma la ecuación [8] expresa el vector de datos originales como un vector de valores de tendencia

$\mathbf{Y}^S$  más un vector de ruido aleatorio  $\eta$ , con varianza de cada elemento dada por  $\sigma_\eta^2$ . En [9] se tiene una ecuación que induce suavidad en el comportamiento  $\mathbf{Y}^S$  al suponer un patrón polinomial de grado uno, esto es,  $Y_t^S = 2Y_{t-1}^S + Y_{t-2}^S + \varepsilon_t$  para  $t = 3, \dots, n$  donde  $\varepsilon_t$  es un error aleatorio con varianza  $\sigma_\varepsilon^2$ . Por último, en [10] se postula una experiencia demográfica con estructura límite o, dicho de otra manera, se hace uso de otra fuente de información para combinar sus datos con los que se observaron originalmente. A partir de las ecuaciones [9] a [11] se obtiene el sistema:

$$\begin{pmatrix} \mathbf{Y} \\ \mathbf{0} \\ \mathbf{U} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_n \\ \mathbf{K}_2 \\ \mathbf{I}_n \end{pmatrix} \mathbf{Y}^S + \begin{pmatrix} \boldsymbol{\eta} \\ -\boldsymbol{\varepsilon} \\ \boldsymbol{\delta} \end{pmatrix}, \text{ con } \begin{pmatrix} \boldsymbol{\eta} \\ -\boldsymbol{\varepsilon} \\ \boldsymbol{\delta} \end{pmatrix} \sim \left( \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \boldsymbol{\Sigma} \right)$$

$$\text{donde } \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_\eta^2 \mathbf{I}_n & 0 & 0 \\ 0 & \sigma_\varepsilon^2 \mathbf{I}_{n-2} & 0 \\ 0 & 0 & \sigma_\delta^2 \mathbf{I}_n \end{pmatrix} \quad [12]$$

Por ello, se puede emplear MCG para estimar  $\mathbf{Y}^S$  y así se obtiene lo siguiente:

$$\begin{aligned} \hat{\mathbf{Y}}^S &= \left[ \begin{pmatrix} \mathbf{I}_n \\ \mathbf{K}_2 \\ \mathbf{I}_n \end{pmatrix}' \boldsymbol{\Sigma}^{-1} \begin{pmatrix} \mathbf{I}_n \\ \mathbf{K}_2 \\ \mathbf{I}_n \end{pmatrix} \right]^{-1} \begin{pmatrix} \mathbf{I}_n \\ \mathbf{K}_2 \\ \mathbf{I}_n \end{pmatrix}' \boldsymbol{\Sigma}^{-1} \begin{pmatrix} \mathbf{Y} \\ \mathbf{0} \\ \mathbf{U} \end{pmatrix} \\ &= (\sigma_\eta^{-2} \mathbf{I}_n + \sigma_\varepsilon^{-2} \mathbf{K}_2' \mathbf{K}_2 + \sigma_\delta^{-2} \mathbf{I}_n)^{-1} (\sigma_\eta^{-2} \mathbf{Y} + \sigma_\delta^{-2} \mathbf{I}_n \mathbf{U}) \end{aligned} \quad [13]$$

Entonces, si se hace  $\lambda_1 = \sigma_\eta^2 / \sigma_\varepsilon^2$  y  $\lambda_2 = \sigma_\eta^2 / \sigma_\delta^2$ , se llega a

$$\hat{\mathbf{Y}}^S = (\mathbf{I}_n + \lambda_1 \mathbf{K}_2' \mathbf{K}_2 + \lambda_2 \mathbf{I}_n)^{-1} (\mathbf{Y} + \lambda_2 \mathbf{U}) \quad [14]$$

que es el estimador buscado, cuya matriz de varianza-covarianza está dada por

$$\Gamma = \text{Var}(\hat{\mathbf{Y}}^S) = (\mathbf{I}_n + \lambda_1 \mathbf{K}'_2 \mathbf{K}_2 + \lambda_2 \mathbf{I}_n)^{-1} \sigma_n^2 \quad [15]$$

Una manera alternativa de expresar los resultados anteriores surge al reconocer que  $\hat{\mathbf{Y}}^S = \mathbf{M}(\mathbf{Y} + \lambda_2 \mathbf{U})$  y que  $\Gamma = \mathbf{M} \sigma_n^2$  con  $\mathbf{M} = (\mathbf{I}_n + \lambda_1 \mathbf{K}'_2 \mathbf{K}_2 + \lambda_2 \mathbf{I}_n)^{-1}$ . Por lo tanto, si ahora se escribe

$$\mathbf{M} = (\mathbf{I}_n + \frac{\lambda_1}{1 + \lambda_2} \mathbf{K}'_2 \mathbf{K}_2)^{-1} (1 + \lambda_2)^{-1} \quad [16]$$

la ecuación [13] puede reescribirse como

$$\hat{\mathbf{Y}}^S = (\mathbf{I}_n + \alpha \lambda_1 \mathbf{K}'_2 \mathbf{K}_2)^{-1} (\alpha \mathbf{Y} + (1 - \alpha) \mathbf{U}), \text{ con } \alpha = (1 + \lambda_2)^{-1} \quad [17]$$

Esta última expresión permite apreciar que  $\hat{\mathbf{Y}}^S \rightarrow \mathbf{U}$  si  $\alpha \rightarrow 0$ , de forma que la suavidad inducida por [10] desaparece y la tendencia converge a la estructura indicada en [11]. Por otro lado, si  $\alpha \rightarrow 1$ ,  $\hat{\mathbf{Y}}^S \rightarrow (\mathbf{I}_n + \lambda_1 \mathbf{K}'_2 \mathbf{K}_2)^{-1} \mathbf{Y}$  y se obtiene entonces el resultado usual del Filtro HP. Debe notarse que el valor de  $\alpha$  tiene que conocerse de antemano para calcular  $\hat{\mathbf{Y}}^S$ . Además, esta tendencia puede interpretarse como la combinación de dos fuentes de información cuyos pesos los decide implícitamente el analista al elegir el valor de la constante  $\alpha$ . Existen dos posibles perspectivas para elegir los valores de las constantes de suavizamiento. La primera, llamada (A), indica elegir los valores de  $\lambda_1$  y  $\lambda_2$ , de manera que se establezca un balance entre la suavidad y la estructura. A partir del conocimiento de  $\lambda_2$  se obtiene el correspondiente valor de  $\alpha$ . La perspectiva (B), en cambio, indica elegir los valores de  $\lambda_1$  y  $\alpha$ , de manera que se tiene que decidir en principio el porcentaje de suavidad y después cuál de las dos fuentes de información (la observada y la propuesta como meta) tiene mayor credibilidad. Desde el punto de vista del cálculo de la estimación numérica de los valores de tendencia, el vector suavizado  $\hat{\mathbf{Y}}^S$  se puede obtener directamente al aplicar el Filtro de Kalman con suavidad. Para aplicar este filtro se debe notar que los modelos [9] y [11] producen las expresiones siguientes, válidas para todo valor de  $t$ ,

$$\mathbf{Y}_t = \mathbf{Y}_t^S + \eta_t \text{ y } \mathbf{U}_t = \mathbf{Y}_t^S + \delta_t, \eta_t \sim (0, \sigma_\eta^2), \delta_t \sim (0, \sigma_\delta^2), \mathbf{E}(\eta_t \varepsilon_t) = 0 \quad [18]$$

por lo cual se sigue que

$$\alpha Y_t + (1-\alpha)U_t = \alpha Y_t^S + \alpha \eta_t + (1-\alpha)Y_t^S + (1-\alpha)\delta_t = Y_t^S + \gamma_t \quad [19]$$

con

$$\gamma_t = \alpha \eta_t + (1-\alpha)\delta_t \sim (0, \sigma_\gamma^2) \quad [20]$$

y

$$\sigma_\gamma^2 = \alpha^2 \sigma_\eta^2 + (1-\alpha)^2 \sigma_\delta^2 \quad [21]$$

A partir de estas ecuaciones puede escribirse un Modelo de Espacio de Estados con las siguientes ecuaciones de medición y transición, respectivamente:

$$\alpha Y_t + (1-\alpha)U_t = \mathbf{c}_t' \mathbf{X}_t + \gamma_t \quad [22]$$

y

$$\mathbf{X}_t = A_t \mathbf{X}_{t-1} + \mathbf{w}_t, \quad [23]$$

donde los vectores y matrices involucrados son de la forma

$$\mathbf{X}_t = \begin{pmatrix} Y_t^S \\ Y_{t-1}^S \end{pmatrix}, A_t = \begin{pmatrix} \alpha & -1 \\ 1 & 0 \end{pmatrix}, \mathbf{c}_t = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ y } \mathbf{w}_t = \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix} \quad [24]$$

Una vez expresado el modelo en forma de espacio de estados es factible utilizar el Filtro de Kalman como en Guerrero, 2008, pero en lugar de usar los datos originales  $Y_t$  como en aquel artículo, ahora se hace uso de los datos combinados  $\alpha Y_t + (1-\alpha)U_t$ , para lo cual se requiere que el valor de  $\alpha$  sea conocido.

### Los índices de suavidad y su empleo para elegir las constantes de suavizamiento

Para medir la proporción de suavidad relativa  $\sigma_\eta^{-2} \mathbf{I}_n$  respecto a la precisión total que se logra con el proceso de estimación, que está dada por  $\sigma_\eta^{-2} \mathbf{I}_n + \sigma_\delta^{-2} \mathbf{I}_n + \sigma_\varepsilon^{-2} \mathbf{K}_2' \mathbf{K}_2$ , se propone utilizar el índice

$$\Lambda(\sigma_\eta^{-2} \mathbf{I}_n; \sigma_\eta^{-2} \mathbf{I}_n + \sigma_\delta^{-2} \mathbf{I}_n + \sigma_\varepsilon^{-2} \mathbf{K}_2' \mathbf{K}_2) = \text{tr}[\sigma_\eta^{-2} \mathbf{I}_n (\sigma_\eta^{-2} \mathbf{I}_n + \sigma_\delta^{-2} \mathbf{I}_n + \sigma_\varepsilon^{-2} \mathbf{K}_2' \mathbf{K}_2)^{-1}] / n \quad [25]$$

en donde  $\text{tr}(\cdot)$  denote la traza de una matriz, mientras que  $\sigma_\eta^{-2} \mathbf{I}_n$ ,  $\sigma_\delta^{-2} \mathbf{I}_n$  y  $\sigma_\varepsilon^{-2} \mathbf{K}_2' \mathbf{K}_2$  son matrices positivas definidas de dimensión  $n \times n$ .

Este índice es una medida que cuantifica la precisión relativa y que tiene las siguientes cuatro propiedades: 1) toma valores dentro del intervalo (0, 1), de manera que puede interpretarse como una proporción propia; 2) es invariante bajo transformaciones lineales de la variable  $Y$  involucrada, por lo cual los resultados son válidos aun si se transforma la variable linealmente; 3) se comporta en forma lineal, de manera que pueden aplicarse directamente herramientas de álgebra lineal para hacer los cálculos necesarios; 4) la suma de las precisiones relativas es la unidad, lo cual significa que

$$\Lambda(\sigma_{\eta}^{-2}I_n; \sigma_{\eta}^{-2}I_n + \sigma_{\delta}^{-2}I_n + \sigma_{\epsilon}^{-2}K_2'K_2) + \Lambda(\sigma_{\delta}^{-2}I_n; \sigma_{\eta}^{-2}I_n + \sigma_{\delta}^{-2}I_n + \sigma_{\epsilon}^{-2}K_2'K_2) + \Lambda(\sigma_{\epsilon}^{-2}K_2'K_2; \sigma_{\eta}^{-2}I_n + \sigma_{\delta}^{-2}I_n + \sigma_{\epsilon}^{-2}K_2'K_2) = 1 \quad [26]$$

La demostración de que  $\Lambda$  es la única medida escalar que cumple con las cuatro propiedades enunciadas previamente, se sigue de la prueba que aparece en Theil, 1963, para el caso de dos matrices positivas definidas  $A$  y  $B$ , donde el índice está dado por  $\Lambda(A; A+B)$ . Lo que se requiere para adaptar la prueba a la situación presente es reconocer que, por ejemplo, ahora  $\sigma_{\eta}^{-2}I_n$  juega el papel de  $A$  y la matriz  $\sigma_{\delta}^{-2}I_n + \sigma_{\epsilon}^{-2}K_2'K_2$  juega el de  $B$ . Este índice es útil para cuantificar la precisión relativa atribuible a la suavidad y a la estructura inducida en el modelo, que forman parte de la matriz de precisión  $\Gamma^{-1}$  dada por el inverso de [15]. En consecuencia, se define el índice de suavidad

$$S(\lambda_1, \lambda_2; n) = \text{tr}[\sigma_{\epsilon}^{-2}K_2'K_2(\sigma_{\eta}^{-2}I_n + \sigma_{\epsilon}^{-2}K_2'K_2 + \sigma_{\delta}^{-2}I_n)^{-1}] / n \quad [27]$$

$$= 1 - \text{tr}[(I_n + \lambda K_2'K_2)^{-1}] / n$$

con

$$\lambda = (\sigma_{\eta}^{-2} + \sigma_{\delta}^{-2})^{-1} \sigma_{\epsilon}^{-2} = \lambda_1(1 + \lambda_2)^{-1} = \alpha\lambda_1. \quad [28]$$

Ahora bien, como el parámetro  $\lambda$  está asociado con la suavidad de  $\alpha Y + (1 - \alpha)U$ , su valor se puede elegir con la ayuda del índice de suavidad  $S(\lambda_1, \lambda_2; n)$ . Posteriormente, ya que el parámetro  $\lambda_1$  se asocia con la suavidad de los datos originales  $Y$ , puede obtenerse a partir de los valores de  $\lambda$  y  $\alpha$ ; esto se debe a que  $\lambda_1 = \lambda/\alpha$  con  $\alpha > 0$ . Otra forma de seleccionar los parámetros necesarios comienza por elegir a  $\lambda_1$  de manera que se fije la suavidad deseada para  $Y$  y después se elige el

valor de  $\alpha \in (0,1)$  entonces se deduce el valor de  $\lambda = \alpha \lambda_1$  que sirve para determinar la suavidad de la combinación convexa  $\alpha \mathbf{Y} + (1 - \alpha) \mathbf{U}$ .

Debe notarse que el porcentaje de suavidad para  $\mathbf{Y}$  debe ser mayor o igual que el de la combinación debido a que: *i)*  $\lambda = \alpha \lambda_1 \leq \lambda_1$ , pues  $0 < \alpha \leq 1$ , y *ii)* el índice de suavidad es una función monótona creciente de  $\lambda$ . Además, el parámetro  $\lambda_1$  se elige a partir del índice de suavidad  $S(\lambda_1; n) = 1 - \text{tr}[(\mathbf{I}_n + \lambda_1 \mathbf{K}'_2 \mathbf{K}_2)^{-1}] / n$ , el cual se asocia con la suavidad de  $\mathbf{Y}$ , puesto que equivale a usar el índice de suavidad aplicable a  $\alpha \mathbf{Y} + (1 - \alpha) \mathbf{U}$  cuando  $\alpha = 1$ , en cuyo caso  $\lambda_2 = 0$  y el estimador de la tendencia resulta ser  $\hat{\mathbf{Y}}^S = (\mathbf{I}_n + \lambda_1 \mathbf{K}'_2 \mathbf{K}_2)^{-1} \mathbf{Y}$ . Desde luego, para incluir tanto la suavidad como la estructura demográfica en la solución, se debe elegir  $\alpha \in (0,1)$ , de manera que se obtenga  $\lambda_2 > 0$  y se utilice el estimador dado por  $\hat{\mathbf{Y}}^S = (\mathbf{I}_n + \alpha \lambda_1 \mathbf{K}'_2 \mathbf{K}_2)^{-1} (\alpha \mathbf{Y} + (1 - \alpha) \mathbf{U})$ . Por último, es importante hacer la siguiente aclaración acerca del máximo porcentaje de suavidad que puede lograrse con un conjunto de datos determinado. Debido a que  $\text{tr}[(\mathbf{I}_n + \lambda \mathbf{K}'_d \mathbf{K}_d)^{-1}] \rightarrow d$  conforme  $\lambda \rightarrow \infty$  con  $d$  el orden de las diferencias en la matriz  $\mathbf{K}^d$  (Eilers y Marx, 1996: 94), se sigue que la máxima suavidad que se puede alcanzar con  $n$  observaciones cumple con  $S(\lambda; n) \rightarrow 1 - d/n$  conforme  $\lambda \rightarrow \infty$ . Este resultado es útil para conocer de manera anticipada el máximo porcentaje de suavidad que es factible alcanzar en aplicaciones prácticas y con base en ello decidir la suavidad deseada para la tendencia.

A continuación se muestra la estrategia para suavizar un conjunto de datos  $\{Y_1, \dots, Y_N\}$  con el Filtro HPMV, de manera que se incorpore la información de los datos de estructura demográfica  $\{U_1, \dots, U_N\}$ . Se aplican las siguientes etapas, las cuales difieren sólo en la segunda para las perspectivas (A) y (B): 1. Suavizar los datos  $\{Y_1, \dots, Y_N\}$  sin considerar la existencia de  $\{U_1, \dots, U_N\}$ . Para esto se fija primero un porcentaje de suavidad deseado para la tendencia y se aplica el procedimiento de Guerrero, 2008. Como resultado se deduce el valor de  $\lambda_1$  y se obtiene la tendencia con suavidad de  $100S(\lambda_1; n)\%$  (por ejemplo, de 80%). 2. A. Decidir el porcentaje de suavidad que será intercambiado por estructura de manera que el porcentaje de suavidad de la tendencia se reduzca (digamos de 80% a 75%). Para lograr esto se fija el valor de  $100S(\lambda_1, \lambda_2; n)\%$  y se obtiene el valor que corresponda para  $\alpha \in (0,1)$ . 2. B. Decidir la credibilidad que se le asigna a cada una de las estructuras de datos disponibles (la observada y la considerada como meta) al fijar el valor de  $\alpha \in (0,1)$ , y simplemente medir la suavidad que se haya obtenido como resultado. 3. Aplicar el procedimiento de suavizamiento con estructura mediante la aplicación del Filtro de Kalman



al conjunto de datos  $\{\alpha Y_t + (1 - \alpha)U_t\}$ . El suavizamiento que se logre con esto tendrá suavidad de  $100S(\lambda_1, \lambda_2; n)\%$  y porcentaje de estructura (es decir, cercanía a  $U$ ) de  $100[S(\lambda_1; n) - S(\lambda_1, \lambda_2; n)]\%$ .

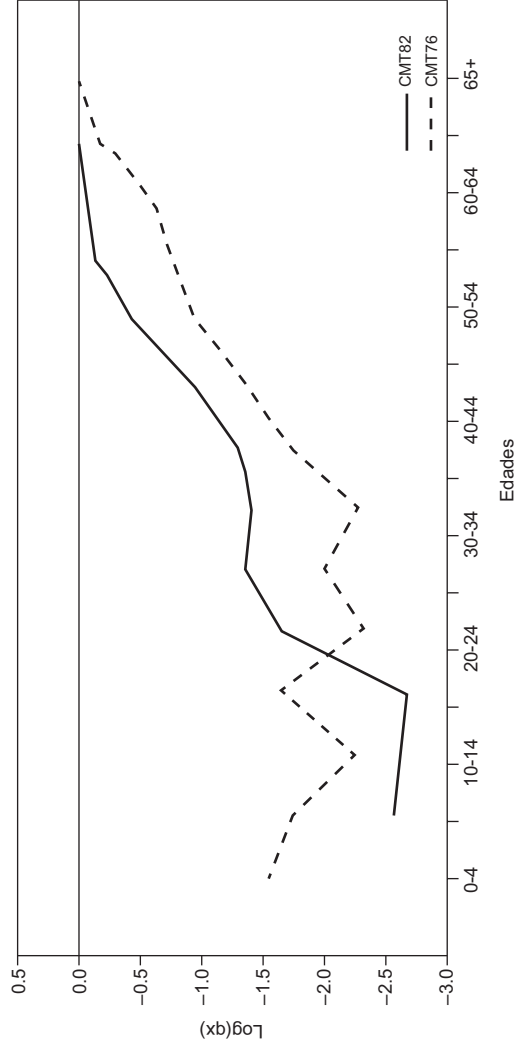
### **Aplicaciones a datos demográficos de México**

En las siguientes aplicaciones se ilustran las dos perspectivas que se sugieren en el marco teórico del trabajo. Por un lado, se presentan ejemplos donde se combinan fuentes de información para obtener tendencias estimadas a partir del nivel de credibilidad que el analista otorgue a cada fuente. Por otra parte, se presentan aplicaciones donde se maneja una estructura por edades de un indicador demográfico para algún país que esté en su última etapa de transición demográfica y se considera a esta estructura como meta futura, a la cual se pretende que aspire la situación actual mexicana para un determinado horizonte de previsión. De acuerdo con estas dos perspectivas, se rebasa la eventual dificultad que pudiese representar la distinta longitud de las series de indicadores demográficos utilizados, gracias a la utilización del Filtro de Kalman. En general se hablará de tendencias en un sentido amplio y no exclusivamente en términos predictivos.

#### *Mortalidad en la Ciudad de México en el siglo XVIII*

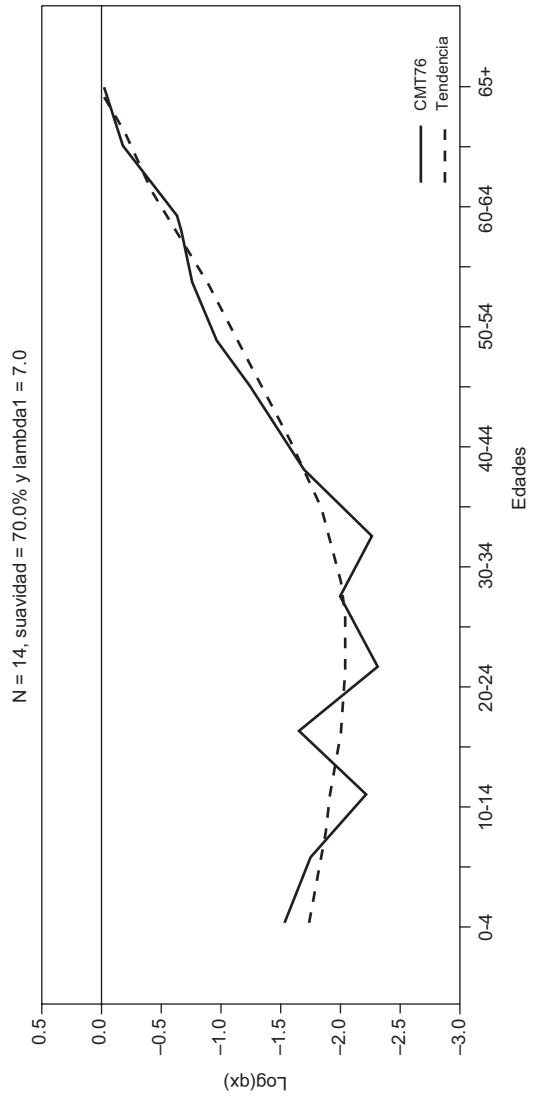
Se muestra que la metodología presentada constituye una herramienta útil en la tarea de aproximar la mortalidad de la población de la Ciudad de México en el siglo XVIII. Se tienen datos de naturaleza paleodemográfica que se obtuvieron en dos momentos diferentes: 1976 y 1982 (Hernández, 1999). Tales datos fueron generados a partir de restos óseos ubicados en la Catedral Metropolitana de la Ciudad de México. En el primer momento de 1976 se observó un total de 1 642 individuos, con una esperanza de vida promedio de 24 años. Estos restos corresponden a una población civil integrada mayoritariamente por criollos y españoles peninsulares, y en menor medida por mestizos e indígenas (Márquez y Civera, 1987). El segundo momento fue en 1982, cuando se tuvo acceso a una serie de datos de restos para otra población civil, principalmente mestiza y probablemente contemporáneos a los estudiados en 1976. Estos últimos con una esperanza de vida de 22.6 años (Hernández, 1991).

GRÁFICA 1  
Mortalidad en la Ciudad de México en el siglo XVIII con CMT76 y CMT82



FUENTE: Logaritmos de tasas de mortalidad de Hernández, P. (1999), "Los estudios paleodemográficos en México", *Revista Argentina de Antropología Biológica*, vol. 2, pp. 335-355.

**GRÁFICA 2**  
**Tendencia estimada para los datos de 1982**



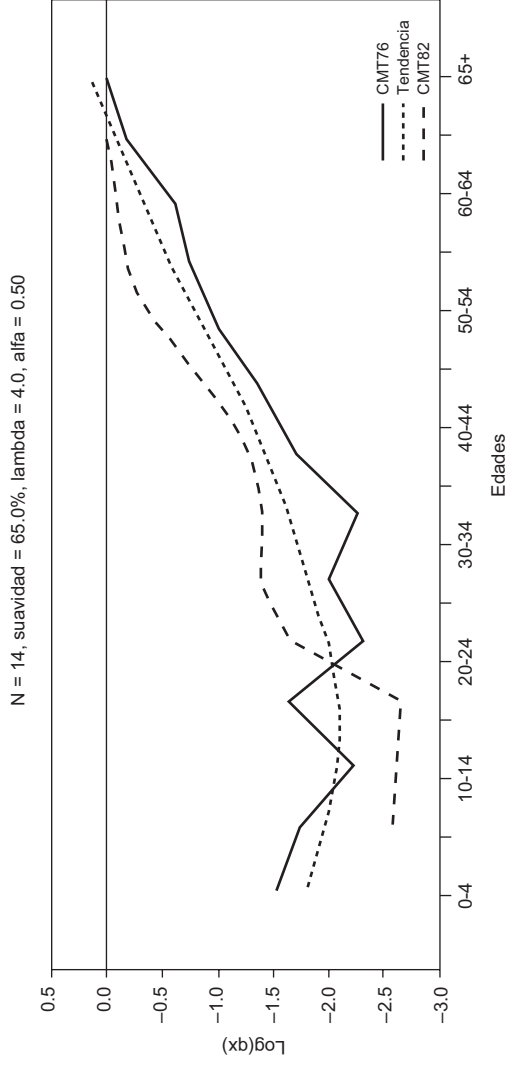
FUENTE: Cálculos propios y logaritmos de tasas de mortalidad de Hernández, 1999.

La idea es estimar una tendencia única que proporcione una visión de la mortalidad en dicho siglo en la Ciudad de México, es decir, se pretende combinar la información generada en ambos momentos. Al hacer esto se puede otorgar mayor o menor credibilidad a las fuentes de acuerdo con las creencias o conocimientos que tenga *a priori* el analista. En la gráfica 1, donde se usan logaritmos de la mortalidad ( $\text{Log}(q_x)$ ), se aprecia que el tamaño de las series es distinto. Sin embargo esto no representa un problema adicional, pues con la metodología propuesta se genera una tendencia tan larga como sea la serie de mayor longitud, en este caso la de 1976. Se puede apreciar que las tasas de mortalidad de 1976 son más bajas desde los 20-24 años de edad que con la base de datos recabados en 1982 y ocurre a la inversa para antes de los 20 años de edad. Se decidió iniciar el procedimiento con los datos de la serie más reciente, es decir, con los correspondientes a 1982 (CMT82). Cabe advertir que se podría haber elegido como datos iniciales los de 1976 (CMT76) y se tendría una gran similitud de los resultados en general. Con 70% de suavidad, se puede apreciar en la gráfica 2 la tendencia de dichos datos y, con base en la tendencia estimada, se podría apreciar cuáles hubiesen sido los niveles de mortalidad para grupos de edades que rebasan incluso los que no se estimaron en dicha fuente.

Si se estima la tendencia considerando la mortalidad de 1976 se observa en la gráfica 3 que el procedimiento trata de conciliar las dos fuentes y que se preservan las características de ambas tendencias. Al intercambiar suavidad por estructura, se eligió 65% de suavidad y, en este caso, se otorgó la misma credibilidad a ambas fuentes, *i. e.*, se usó  $\alpha = 0.50$ .

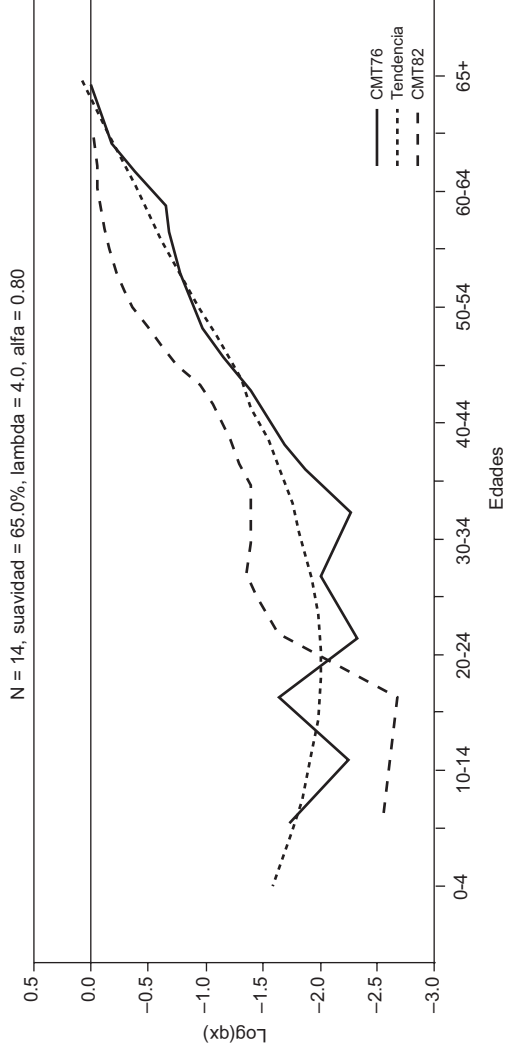
La propuesta metodológica de ninguna manera excluye la posibilidad de que el analista decida asignar distinta credibilidad a las fuentes, con base en sus conocimientos adicionales derivados de su cercanía con el tema de investigación. Así pues, con fines puramente ilustrativos se supondrá que se desea otorgar distinta credibilidad a las fuentes y con ello se puede apreciar que la tendencia estimada se apega más a la fuente que el analista considera más confiable. En primer lugar, al otorgar mayor credibilidad a los datos de 1976 con  $\alpha = 0.20$  se obtiene la gráfica 4. Posteriormente se asigna mayor credibilidad a los de 1982 al elegir  $\alpha = 0.80$  y se obtiene la gráfica 5. Así pues, la credibilidad de las fuentes se otorga mediante el valor del parámetro  $\alpha$  de la expresión [17], de manera que una situación neutra o de misma credibilidad para ambas fuentes conduce a utilizar  $\alpha = 0.50$ .

GRÁFICA 3  
Tendencia estimada con ambas fuentes de información: 1976 y 1982



FUENTE: Cálculos propios y logaritmos de tasas de mortalidad de Hernández, 1999.

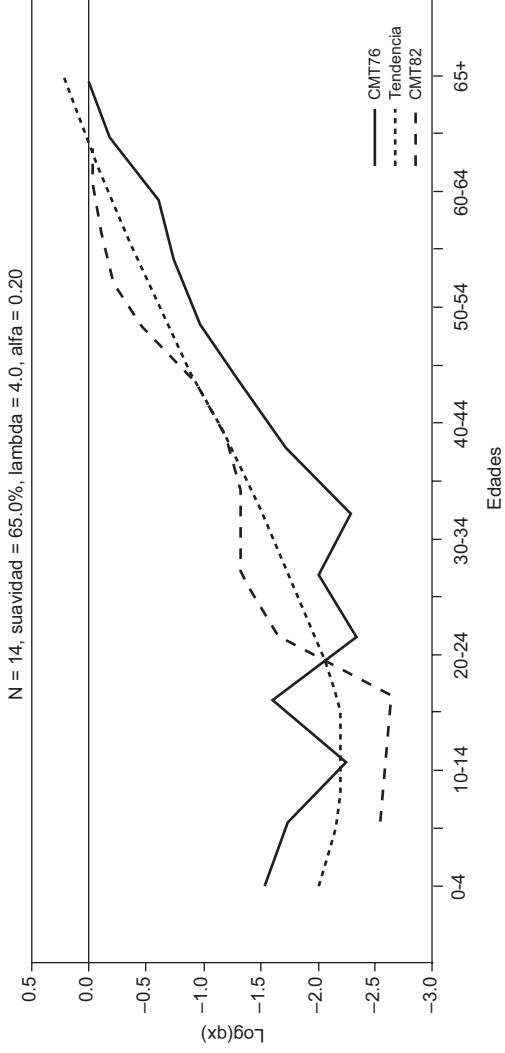
GRÁFICA 4  
Tendencia estimada con mayor credibilidad asignada a los datos de 1976



FUENTE: Cálculos propios y logaritmos de tasas de mortalidad de Hernández, 1999.

GRÁFICA 5

Tendencia estimada con mayor credibilidad en los datos de 1982



FUENTE: Cálculos propios y logaritmos de tasas de mortalidad de Hernández, 1999.

*Hacia una fecundidad típica de la cuarta etapa  
de la transición demográfica*

Por medio de la presente propuesta metodológica es posible hacer aplicaciones tocantes a la fecundidad, por ejemplo considerar la fecundidad masculina mexicana en relación con la de otro país más avanzado en su transición demográfica, o bien estimar la fecundidad “mixta” a partir de los reportes de la fecundidad femenina y masculina existentes en México (Quilodrán y Sosa, 2001). En particular, en la aplicación de esta sección se emplean las tasas específicas de fecundidad de Suecia de 2006 <<http://www.humanfertility.org>> y se les visualiza como una meta de estructura demográfica de fecundidad que se aspiraría fuera alcanzada por todas las edades de la población femenina de México en un horizonte de  $n$  décadas. Para el caso mexicano se parte de estimaciones de Conapo, 2006.

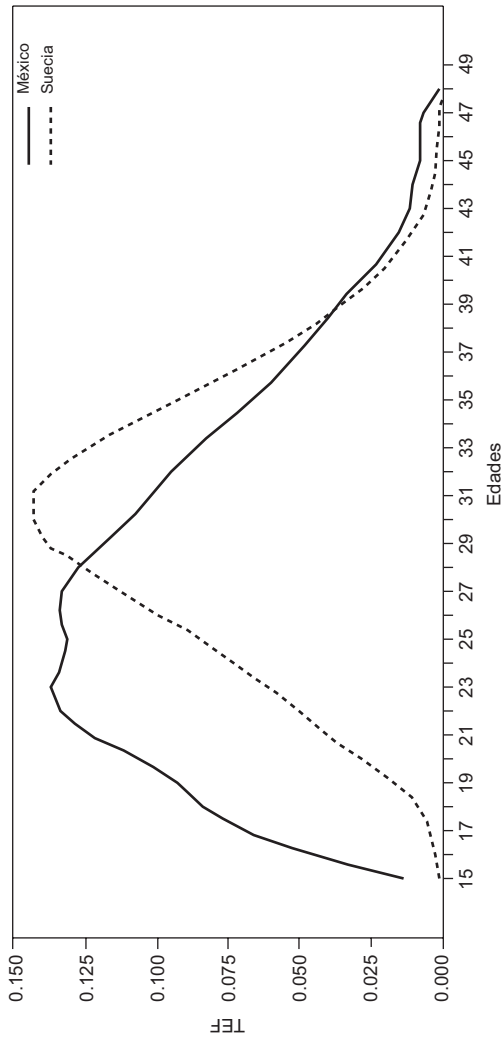
En la gráfica 6 se aprecia la diferencia estructural entre ambas experiencias de fecundidad. La mexicana es asimétrica a la derecha y parece bimodal, mientras que la sueca es sencillamente simétrica. También se observa que las colas de la distribución sueca son más consistentes con lo que plantea la teoría al inicio y al final de las edades fecundas. Dentro de este mismo rubro, en el caso mexicano se podría pensar que el registro de fecundidad no obedece necesariamente a los mejores estándares del mundo o, al menos, es claro que difiere bastante del de Suecia.

En este ejemplo se eligen como datos iniciales a los de México y se estima su tendencia con suavidad de 75%. Como se advierte en la gráfica 7, la tendencia estimada queda ligeramente por encima para las edades iniciales, en la parte superior la situación de aparente bimodalidad se atenúa, y al final de las edades la estimación se vuelve suave, pero la condición de asimetría se preserva. Como se expone en el algoritmo del procedimiento, ahora se reemplaza cierto porcentaje de suavidad por estructura. En este caso, al intercambiar suavidad por estructura, la gráfica 8 muestra que se alcanza una suavidad de 70.9 por ciento.

De esta manera se percibe la estructura a la que podría aspirar la fecundidad mexicana en cierto horizonte de décadas por definir, en relación con la estructura sueca. Es importante notar que la tendencia estimada busca apegarse a la estructura sueca sin alejarse por completo de la estructura mexicana. Debe ser claro que el horizonte asociado con la transición de una estructura a otra depende en gran medida de

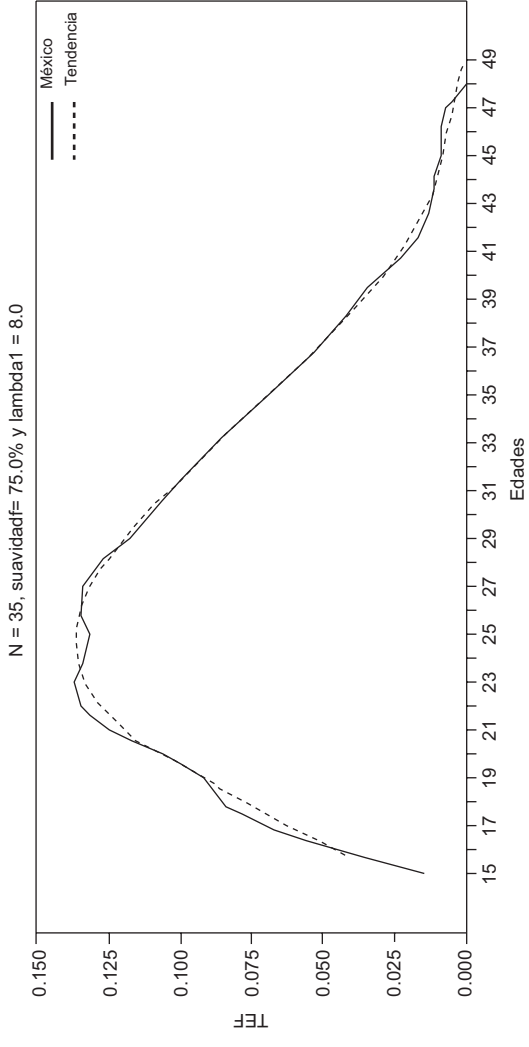


GRÁFICA 6  
Tasas específicas de fecundidad



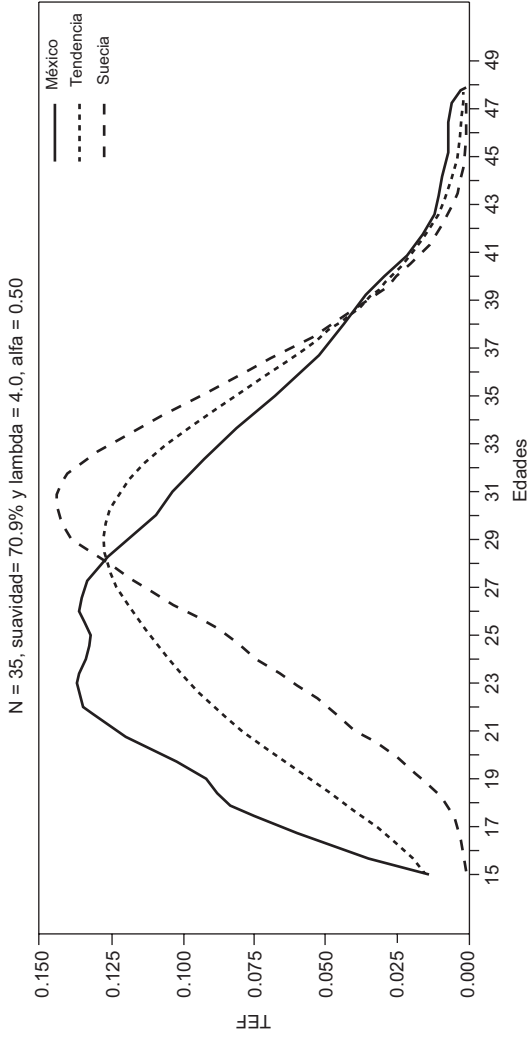
FUENTE: Tasas específicas de fecundidad de Suecia de 2006, disponible en <<http://www.humanfertility.org>> y Conapo, 2006, *Indicadores demográficos básicos*, México <<http://www.conapo.gob.mx>> (11 de marzo de 2006).

**GRÁFICA 7**  
**Tendencia inicial con datos de México, 2006**



FUENTE: Cálculos propios y Conapo, 2006.

GRÁFICA 8  
Tendencia estimada con las estructuras de fecundidad de México y Suecia



FUENTE: Cálculos propios con datos del Conapo, 2006 y <<http://www.humanfertility.org>>.

la agudeza y del conocimiento que el analista tenga sobre el tema. Nótese que la tasa específica de fecundidad en edad 15 es la misma entre la tendencia y la mexicana, y también se podría decir que la *masa* de la función de distribución de las probabilidades de fecundidad se redistribuye en un rango más amplio de edades. Desde luego esta estructura conllevaría nuevos retos en cuanto a las políticas de salud pública de los mexicanos, entre otras muchas implicaciones.

### *Estimaciones de mortalidad infantil en México (1990-2020)*

La tasa de mortalidad infantil tiene un gran impacto en el cálculo de la esperanza de vida al nacer y constituye un indicador del desarrollo social y económico de un país. Por ello las discrepancias que pudieran existir entre distintas fuentes sobre este indicador pueden conducir a diferentes perspectivas sobre la situación real en determinada nación. Así pues, se considera oportuno hacer una aplicación que combine dos fuentes de información respecto de la mortalidad infantil para el caso mexicano. En primer término se emplean las estimaciones de Aguirre (2009), y en segundo las que elaboró Conapo (2010). La longitud de las series es distinta en cada caso, por lo que la estimación de la tendencia con la metodología propuesta será tan larga como la serie de Conapo.

Una vez realizada la estimación de la tendencia con determinado porcentaje de suavidad, se contrastan los resultados con las cifras que reporta la Secretaría de Salud (SS). De esta manera, en principio se genera una estimación de la tendencia de la mortalidad infantil en México para el periodo 1990 a 2020. Como se ilustra en la gráfica 9, desde 1990 hasta 2005 las estimaciones que elaboró Conapo superan a las de Aguirre: la diferencia es más elevada al inicio y se reduce hacia 2005. Por otro lado, se observa que los datos de la Secretaría de Salud de 2000 a 2008 no presentan claramente una tendencia similar a la de las otras dos fuentes; de hecho parece que el valor de este indicador es relativamente constante durante el periodo de referencia.

Para esta aplicación se eligieron como datos iniciales los de la serie de estimaciones de Aguirre, que incluye una cantidad de datos menor que la de Conapo. Se propuso entonces una suavidad de 75% y, como se observa en la gráfica 10, se genera una tendencia cuya curvatura es acorde con las previsiones que elaboró dicho especialista. Como valor agregado se obtienen estimaciones subsecuentes, o vistos de otra forma,

pronósticos de la tasa de mortalidad infantil desde 2006 hasta 2020. Nótese que esto ocurre como consecuencia de la extensión de la serie de Conapo y del uso del Filtro de Kalman.

Al intercambiar una parte del porcentaje de suavidad inicial de 75% por estructura, se obtiene una suavidad final de 70.6%, como se presenta en la gráfica 11. Por otra parte, se decidió brindar la misma credibilidad a ambas fuentes de información, es decir, se eligió  $\alpha = 0.50$ . Con estos parámetros se genera una tendencia que cubre el periodo de 1990 a 2020, donde se aprecia un cruce entre la tendencia estimada y las proyecciones de Conapo de 2014 y 2015. Asimismo, las cifras que reporta la Secretaría de Salud muestran una dinámica distinta de la tendencia estimada a partir de las fuentes iniciales, aunque también se produce un cruce en el año 2001.

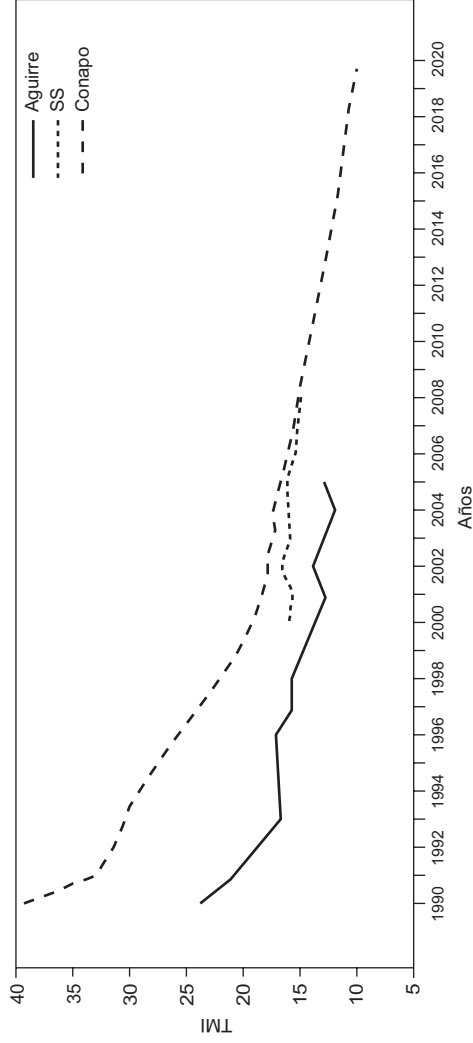
#### *Hacia una mortalidad típica de la cuarta etapa de la transición demográfica*

Para elaborar proyecciones de población desde una perspectiva demográfica el método más empleado es el de Componentes, según se mencionó en la tercera sección. Con este método, de manera resumida, se proyectan las variables demográficas de mortalidad, fecundidad y migración, y se generan escenarios sobre lo que se espera en el futuro con base en el conocimiento y la experiencia del analista. Dentro de la tarea predictiva de la variable de mortalidad existen varias alternativas, como son: el uso de *tablas límite de mortalidad*, la suposición de determinado comportamiento de mortalidad dentro de un contexto ubicado en la última parte de la transición demográfica, etcétera. Es en este espacio donde también se tiene oportunidad de aplicar la propuesta metodológica de manera natural.

En este ejemplo se presenta una estimación de estructura con suavidad de mortalidad mediante una tendencia generada para la experiencia mexicana de 2010 (MXH10) y la japonesa de 2008 (JPH08), en el supuesto de que en cierto horizonte de “*H*” décadas se presentará tal comportamiento. Para hacer la aplicación se usan las tasas específicas de mortalidad por sexo que elaboró Conapo (2010) y para Japón las de la base de datos de <<http://www.mortality.org>>. En otras palabras, se estima la tendencia de manera que el escenario actual mexicano aspire al escenario de Japón. Cabe advertir que se tienen resultados análogos tanto de estructura como de suavidad al hacer la

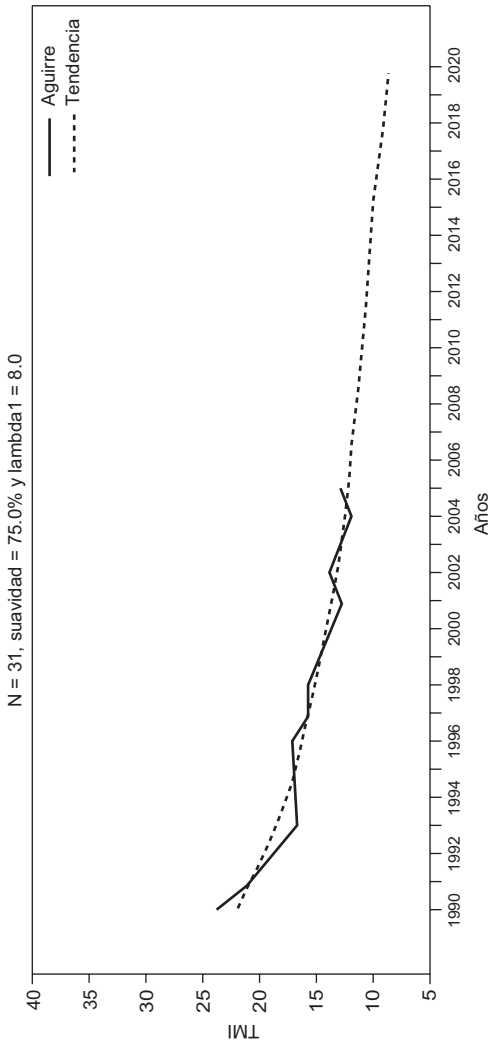
## GRÁFICA 9

### Tasas de mortalidad infantil en México con tres fuentes distintas



FUENTE: Aguirre, A. (2009), "La mortalidad infantil y la mortalidad materna en el siglo XXI", *Papeles de Población*, núm. 15, pp. 75-99; Conapo (2010), *Indicadores demográficos básicos*, México, Consejo Nacional de Población, disponible en <<http://www.conapo.gob.mx>> (11 de agosto de 2010) y datos de la Secretaría de Salud, 2000-2008, disponible en <<http://www.sinais.salud.gob.mx/mortalidad>> (11 de agosto de 2010).

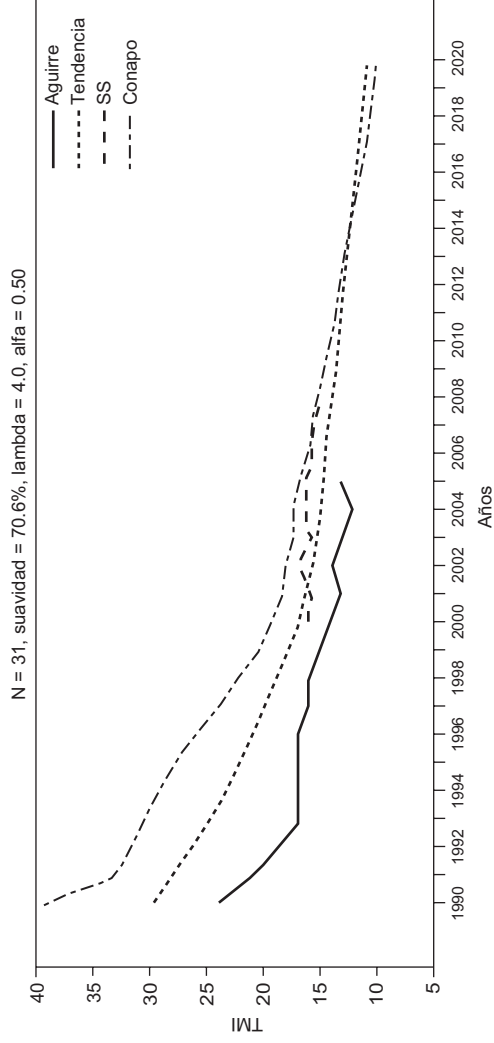
GRÁFICA 10  
Tendencia inicial con datos de Aguirre



FUENTE: Cálculos propios con datos de Aguirre, 2009.

## GRÁFICA 11

### Tendencia estimada con las fuentes de información de Aguirre y Conapo combinadas



FUENTE: Cálculos propios con datos de Aguirre, 2009 y Conapo, 2010.



aplicación con datos del sexo femenino o bien para el total de la población.

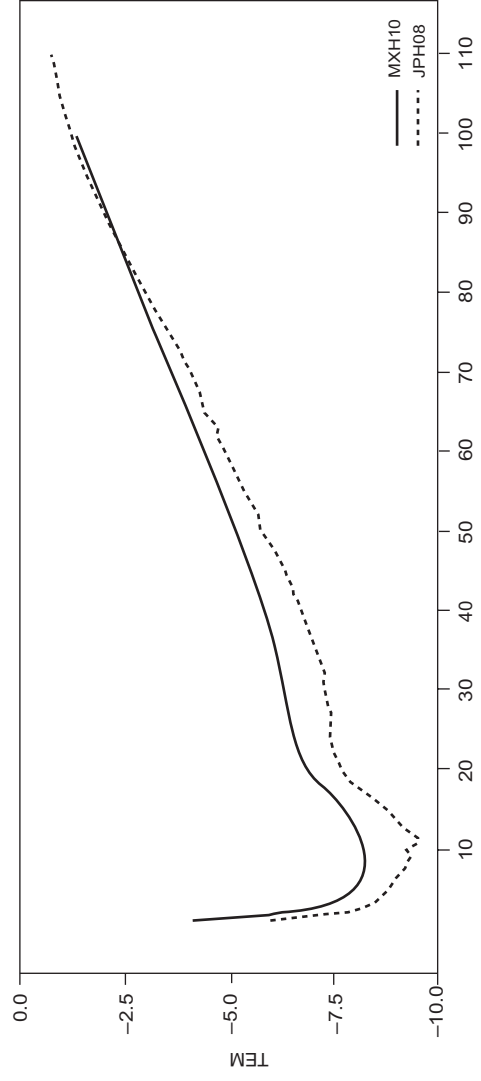
Conviene destacar algunos detalles de la gráfica 12, por ejemplo que la serie de Japón llega hasta los 110 años, mientras que la mexicana se queda en 100 (pese a ello la metodología es aplicable); además, en ambas experiencias de mortalidad se presenta el comportamiento típico de las distintas etapas de la vida, como se ha propuesto en la literatura (Thiele, 1871; Heligman y Pollard, 1980), y también se observa que en casi todo el rango de las series la mortalidad japonesa es inferior hasta el entorno de los 85 años de edad, desde donde la mexicana se vuelve más baja. En este sentido se debe tomar en cuenta que lo que se considera como experiencia mexicana son estimaciones, no datos observados, por lo cual también se explica la suavidad de la serie mexicana frente a la japonesa (sin que se haya aplicado el método aquí propuesto a los datos nacionales). Por último, la serie japonesa a partir de los 85 años presenta una ligera curvatura, semejante a las respectivas experiencias de mortalidad que se observan en algunos países altamente desarrollados y que se han documentado ampliamente en la literatura demográfica (véase Ham, 2005).

No se requiere una suavidad alta para estimar la tendencia de la serie mexicana, y con 65% de suavidad se obtienen resultados satisfactorios (véase la gráfica 13). Por medio de la aplicación de la metodología, la tendencia llega hasta los 110 años de edad, límite de edad máximo de la serie japonesa, y la estimación para niños menores de un año queda ligeramente baja.

Posteriormente, al asignar la misma credibilidad a ambas fuentes de información, es decir, al usar  $\alpha = 0.50$ , se intercambia la suavidad por la estructura y la suavidad se reduce aproximadamente a 60%, con lo cual se obtiene la tendencia que muestra la gráfica 14.

Una vez estimada la tendencia se puede advertir cómo podría ser la mortalidad mexicana en el horizonte futuro prefijado por el analista. En esencia, la estimación resultante es una mezcla entre ambas estructuras y con curvatura en edades jóvenes. Sin embargo el efecto de mezcla no va más allá de los 85 años, y si se quisiera lograr ese efecto se tendría que asignar otro valor al parámetro  $\alpha$  (pero se considera que México dista mucho de alcanzar una estructura de tal naturaleza).

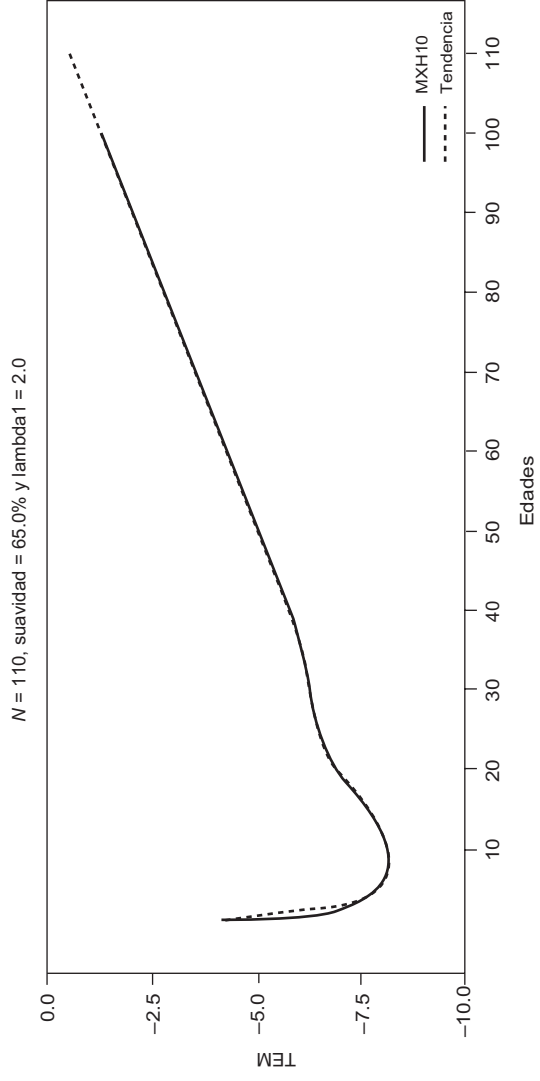
GRÁFICA 12  
Tasas específicas de mortalidad masculina para México en 2010 y Japón en 2008



FUENTE: Conapo (2010), *Indicadores demográficos básicos*, México, Consejo Nacional de Población, disponible en <<http://www.conapo.gob.mx>> (11 de agosto de 2010), y para Japón disponible en la base de datos <<http://www.mortality.org>> (11 de agosto de 2010).

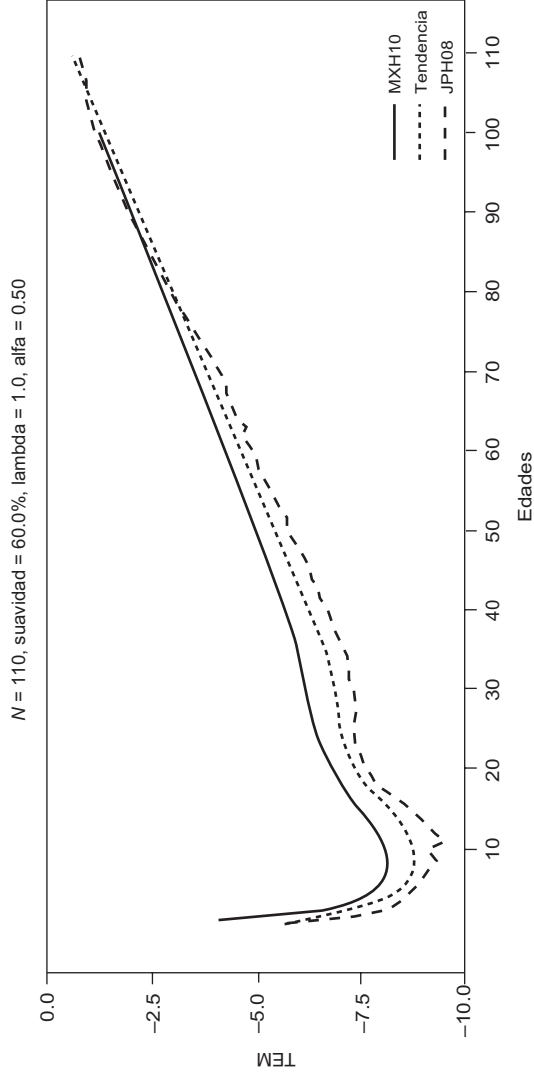
GRÁFICA 13

Tendencia inicial para las tasas específicas de mortalidad masculina en México



FUENTE: Cálculos propios y Conapo, 2010.

**GRÁFICA 14**  
**Tendencia estimada con fuentes de información mexicana y japonesa**



FUENTE: Cálculos propios con datos de Conapo, 2010 y <<http://www.mortality.org>>.

## Conclusiones

La metodología que se describe en este trabajo puede utilizarse para estimar tendencias de datos demográficos que tengan en cuenta la fidelidad a los datos, así como la suavidad que se desea y la estructura proveniente de alguna teoría, o bien, que se le pueda considerar como una meta que se aspire alcanzar en algún horizonte futuro. El control que el analista puede tener sobre estas características de la tendencia, de acuerdo con sus intereses y su conocimiento del fenómeno, hacen que los resultados sean comparables en distintos análisis. Debido a su facilidad de implementación en paquetes del tipo Matlab o R o RATS (en particular se utilizó RATS y los programas están disponibles a solicitud del interesado), es factible que el método sea utilizado por analistas con poca experiencia o con deficiente formación demográfica, por lo que ése es un punto a cuidar en la práctica, porque los resultados que se obtengan dependerán fuertemente de los insumos que se usen, es decir, de los datos originales y de la teoría o meta propuesta. Además debe tenerse en mente que es el analista quien decide cuál de las dos perspectivas de la metodología debe utilizarse en cada caso.

En las aplicaciones aquí mostradas, así como en otras realizadas con la metodología propuesta, se ha observado que el método produce mejores resultados en tanto mayor sea la desagregación de los indicadores demográficos. Es de resaltar el hecho de que la metodología puede aplicarse sin dificultad alguna, aunque la longitud de las series que se combinen no sea la misma o incluso si se presenta la situación de datos faltantes; esto se debe al uso del Filtro de Kalman con suavizamiento, que en los casos mencionados (longitud distinta de las series o datos faltantes) se aplica automáticamente por el procedimiento sin realizar en realidad el suavizamiento, sino sólo el filtrado. Es posible extender la propuesta metodológica para incorporar más fuentes de información, sin embargo ello representa una nueva línea de investigación en la cual los autores están trabajando actualmente. De igual manera, una línea futura de investigación la constituye una extensión del método a casos en que sea necesario aplicar diferentes suavidades o estructuras por rangos de edades de la población. Finalmente se evidencia, de acuerdo con el criterio de los autores, que la herramienta metodológica propuesta representa una herramienta alternativa más para desarrollar diferentes estudios cuantitativos relacionados con distintos fenómenos demográficos, en el sentido de complementar o potenciar lo hasta ahora existente en la literatura especializada. En

concreto, un problema demográfico que se podría resolver con el empleo de la presente propuesta es el de la llamada “conciliación demográfica”. Adicionalmente, y si así se deseara, se podría calcular la varianza de la estimación de acuerdo con la expresión [15].

## Bibliografía

- Aguirre, A. (2009), “La mortalidad infantil y la mortalidad materna en el siglo XXI”, *Papeles de Población*, núm. 15, pp. 75-99.
- Bell, F. y M. Miller (2005), “Life Tables for the United States Social Security Area 1900-2100”, *Actuarial Study*, núm. 120, US Social Security Administration, <[www.ssa.gov/OACT/NOTES/as120/LifeTables\\_Tbl\\_7.html](http://www.ssa.gov/OACT/NOTES/as120/LifeTables_Tbl_7.html)> (11 de febrero de 2010).
- Boone, L. (2000), “Comparing Semi-Structural Methods to Estimate Unobserved Variables: The HPMV and Kalman Filters Approaches”, *Economics Department Working Papers*, núm. 240, OCDE.
- Bourgeois-Pichat, J. (1952), “Essai sur la mortalité ‘biologique’ de l’homme”, *Population*, vol. 7, núm. 3, pp. 381-394.
- Conapo (2006), *Indicadores demográficos básicos*, México, Consejo Nacional de Población <<http://www.conapo.gob.mx>> (11 de marzo de 2010).
- Conapo (2010), *Indicadores demográficos básicos*, México, Consejo Nacional de Población <<http://www.conapo.gob.mx>> (11 de agosto de 2010).
- Copas, J. y S. Haberman (1983), “Non Parametric Graduation Using Kernel Methods”, *Journal of the Institute of Actuaries*, núm. 110, pp. 135-156.
- Debón, A., F. Montes y R. Sala (2006), “A Comparison of Nonparametric Methods in the Graduation of Mortality: Application to Data from the Valencia Region (Spain)”, *International Statistical Review*, núm. 74, pp. 215-233.
- Eilers, P. y B. Marx (1996), “Flexible Smoothing with B-Splines and Penalties”, *Statistical Science*, núm. 11, pp. 89-121.
- Fledelius, P., M. Guillen, J. Nielsen y K. Petersen (2004), “A Comparative Study of Parametric and Nonparametric Estimators of Old-Age Mortality in Sweden”, *Journal of Actuarial Practice*, núm. 11, pp. 101-126.
- Galindo, C. y M. Ordorica (2007), “Estimación de nacimientos ocurridos y registrados, México 1950-2000”, *Papeles de Población*, núm. 54, pp. 39-86.
- George, V., S. Smith, D. Swason y J. Tayman (2004), “The Methods and Materials of Demography”, en J. Siegel y D. Swanson (coords.), *Population Projections*, San Diego, Elsevier Academic Press.
- Guerrero, V.M. (2007), “Time Series Smoothing by Penalized Least Squares”, *Statistics and Probability Letters*, núm. 77, pp. 1225-1234.
- Guerrero, V.M. (2008), “Estimating Trends with Percentage of Smoothness Chosen by the User”, *International Statistical Review*, núm. 76, pp. 187-202.

- Guerrero, V.M. y E. Silva (2010), "Non-parametric and Structured Graduation of Mortality Rates", *Population Review*, núm. 49, pp. 13-26.
- Guerrero, V.M., R. Juárez y P. Poncela (2001), "Data Graduation Based on Statistical Time Series Methods", *Statistics and Probability Letters*, núm. 52, pp. 169-175.
- Haberman, S. y A. Renshaw (1996), "Generalized Linear Models and Actuarial Science", *The Statistician*, núm. 45, pp. 407-436.
- Ham, R. (2005), "La supervivencia más allá de 100 años y más", *Estudios Demográficos y Urbanos*, vol. 20, núm. 1 (58), pp. 103-124. Disponible en <<http://biblioteca.colmex.mx/revistas>>.
- Hastie, T. y R. Tibshirani (1990), *Generalized Additive Models*, Londres, Chapman and Hall.
- Heligman, L. y H. Pollard (1980), "The Age Pattern of Mortality", *Journal of the Institute of Actuaries*, núm. 107, pp. 49-80.
- Hernández, P. (1991), "Los restos óseos del atrio de la Catedral metropolitana. Temporada 1982. México", tesis, México, Escuela Nacional de Antropología e Historia.
- Hernández, P. (1999), "Los estudios paleodemográficos en México", *Revista Argentina de Antropología Biológica*, núm. 2, pp. 335-355.
- Hodrick, R. y E. Prescott (1997), "Post-War U.S. Business Cycles: An Empirical Investigation", *Journal of Money, Credit and Banking*, núm. 29, pp. 1-16.
- Laxton, D. y R. Tetlow (1992), "A Simple Multivariate Filter for the Measurement of Potential Output", *Technical Report*, núm. 59, Ottawa, Bank of Canada.
- Lee, R. y L. Carter (1992), "Modeling and Forecasting U.S. Mortality", *Journal of the American Statistical Association*, núm. 87, pp. 659-675.
- London, D. (1985), *Graduation: The Revision of Estimates*, Reno, ACTEX Publications.
- Manton, K., C. Patrick y E. Stallard (1980), "Mortality Model Based on Delays in Progression of Chronic Diseases: Alternative to Cause Elimination Model", *Public Health Reports*, núm. 95, pp. 580-588.
- Márquez, L. y M. Civera (1987), "Paleodemografía de una muestra de población del periodo colonial mexicano. Estudios de Antropología Biológica", III Coloquio de Antropología Física "Juan Comas", México, UNAM, pp. 405-417.
- Oeppen, J. y J. Vaupel (2002), "Broken Limits to Life Expectancy", *Science*, núm. 296, pp. 1029-1031.
- Olshansky, S.J. (1987), "Simultaneous/multiple Cause-delay (SIMCAD): An Epidemiological Approach to Projecting Mortality", *Journal of Gerontology*, núm. 42, pp. 358-365.
- Olshansky, S.J. (1988), "On Forecasting Mortality", *Milbank Quarterly*, núm. 66, pp. 482-530.
- Olshansky, S.J., B. Carnes y C. Cassel (1990), "In Search of Methuselah: Estimating the Upper Limits to Human Longevity", *Science*, núm. 250, pp. 634-640.
- Olshansky, S., B. Carnes y A. Désesquelles (2001), "Prospects for Human Longevity in an Aging World", *Science*, núm. 291, pp. 1491-1492.

- Ordorica, M. (2001), "Hoy. Un momento importante para revisar las estimaciones demográficas", *Papeles de Población*, núm. 28, pp. 155-163.
- Ortega, A. (2003), "La paleodemografía: ¿un instrumento para simular el comportamiento demográfico del pasado? Análisis comparativo con la demografía histórica en la Ciudad de México del siglo XIX", *Estudios Demográficos y Urbanos*, vol. 19, núm. 1 (55), pp. 181-214. Disponible en <<http://biblioteca.colmex.mx/revistas>>.
- Papaioannou, T. y A. Sachlas (2004), "Graduation of Mortality Rates Revisited. University of Piraeus" <<http://www.stat.ucl.ac.be/Samos2004/abstracts/2004/Papaioannou.pdf>> (26 de julio de 2010).
- Quilodrán, J. y V. Sosa (2001) "Un primer acercamiento a la estimación de niveles de fecundidad masculina en México", *Revista de Información y Análisis*, núm. 15, Aguascalientes, INEGI, pp. 58-67.
- Theil, H. (1963), "On the Use of Incomplete Prior Information in Regression Analysis", *Journal of the American Statistical Association*, núm. 58, pp. 401-414.
- Thiele, P. (1871), "On a Mathematical Formula to Express the Rate of Mortality throughout the Whole of Life", *Journal of the Institute of Actuaries*, núm. 16, pp. 313-329.
- University of California y Max Planck Institute for Demographic Research (2000), "Human Mortality Database: Chile, Japan, United Kingdom and United States" <<http://www.mortality.org>> (11 de febrero de 2010).

### Acerca de los autores

**Víctor M. Guerrero** es actuario por la UNAM y maestro y doctor en Estadística por la Universidad de Wisconsin-Madison. Está adscrito al Departamento de Estadística del ITAM como profesor de tiempo completo. Ha publicado varios libros y una gran cantidad de artículos en revistas especializadas de estadística, principalmente sobre análisis de series de tiempo y pronósticos.

**Eliud Silva** es actuario egresado de la UNAM, diplomado en Modelos Econométricos Dinámicos por el ITAM, maestro en Demografía por El Colegio de México y doctor en Ingeniería Matemática con especialidad en Estadística por la Universidad Carlos III de Madrid. Ha impartido cursos en Naciones Unidas (sede en México), en la UNAM, El Colegio de México, el ITAM, el ITESM y en la Universidad Carlos III de Madrid. Actualmente es profesor de tiempo completo de la Universidad Anáhuac. Cuenta con artículos publicados en revistas nacionales y extranjeras relacionados con series de tiempo y tópicos demográficos. Fue ganador del tercer lugar del Concurso de Investigación Demográfica Gustavo Cabrera de El Colegio de México, en su edición 2010.