

# Pronóstico estocástico de la población mexicana utilizando modelos de datos funcionales

Víctor Manuel García Guerrero

La dinámica demográfica mexicana está en una etapa avanzada del proceso de transición demográfica; por la reducción de la fecundidad, la población actualmente se concentra en las edades productivas. Desde hace poco más de medio siglo y hasta inicios del presente, la mortalidad y fecundidad mantuvieron una tendencia descendente, mientras que la emigración hacia Estados Unidos de América iba en aumento. Durante el siglo XXI, las tendencias cambiaron: la mortalidad se incrementó, el descenso de la fecundidad y la emigración hacia el vecino del norte se estancaron, y se observó un retorno de migrantes y sus familias. Los métodos de proyección demográfica utilizados hasta el momento no están diseñados para dar cuenta de la incertidumbre asociada al futuro poblacional. El objetivo de este trabajo es aplicar un método alternativo de proyección estocástica de la población y analizar sus resultados a la luz de las últimas proyecciones estocásticas calculadas y de las proyecciones oficiales de población. Los métodos estocásticos superan las proyecciones oficiales en el sentido de que no incrementan el rango de error con supuestos subjetivos; sin embargo, discrepan entre sí en cuanto a la amplitud de los intervalos de predicción calculados.

**Palabras clave:** proyecciones de población, pronóstico estocástico de la población, datos funcionales, suavizamiento exponencial, dinámica demográfica de México.

Mexican population dynamics are in an advanced stage of demographic transition; population is now concentrated in the so-called productive ages due to fertility's reduction. Until the beginning of 21st century, mortality and fertility had a downward trend, while emigration to the United States kept an upward trend. In the second half of 2000s decade, Mexico's demographic trends changed: mortality increased, fertility and emigration became stagnant, and an unexpected return of migrants with their families was observed. The usual current demographic projection methods are not designed to take account for the uncertainty associated with future's population. Thus, the objective of this paper is to apply an alternative method of stochastic population projection and analyze its results in light of both the recent stochastic projections calculated for Mexico, and the official population projections. Stochastic methods are better than official projections in the sense that they do not increase the range of error with subjective assumptions; but there are some differences between the stochastic models, mainly related with the prediction intervals' width.

**Keywords:** Population Projections, Population Stochastic Forecasting, Functional Data Models, Exponential Smoothing, Mexico's Demographic Dynamic.

Recibido: 11 de abril de 2014

Aceptado: 12 de agosto de 2014



## Introducción

Las proyecciones de población son una de las técnicas demográficas más importantes para la toma de decisiones. Con ellas, quienes toman decisiones tienen una referencia cuantitativa para distribuir los recursos del Estado de una forma más eficiente. Desde las primeras proyecciones para Inglaterra y Gales realizadas por Edwin Cannan en 1885, el método de las componentes o de las cohorte-componentes ha sido la técnica, la cual consiste en, primero, extrapolar de manera independiente cada variable demográfica (mortalidad, fecundidad y migración) y, después, con los resultados, se calcula la población total futura por edad y sexo (algunos supuestos se tienen que hacer para extrapolar los niveles y tendencias de cada variable); en el tercer paso, y como intento para tratar de evaluar la probabilidad de dicha proyección, se calculan algunas variantes, simulando escenarios extremos de crecimiento de la población; sin embargo, este tipo de variantes de proyección presentan fuertes problemas conceptuales y estadísticos, el primero de ellos es que éstas no tienen una probabilidad de ocurrencia, lo cual puede llevar a una mala percepción de la probabilidad que tiene la variante media, y el segundo es que las variantes altas y bajas son irreales e inconsistentes desde el punto de vista estadístico, en el sentido de que, por ejemplo, en la de alto crecimiento poblacional la fecundidad siempre es alta en todos los años de la proyección, y la esperanza de vida también lo es durante todo ese periodo; esto último implica que se supone que existe una correlación perfecta entre las variables demográficas a lo largo del tiempo y una correlación serial perfecta de cada una de ellas (Lee, 1999; Alho, 1998; y Keilman, 2002).

En Demografía se articulan muchos factores que, al extrapolar, incrementan el grado de incertidumbre en la evolución futura de la población; por ello, la incertidumbre se exagera si la metodología empleada para datos demográficos no es la apropiada o cuenta con una alta dosis de subjetividad; por ejemplo, un aspecto que ha llamado la atención es que el Censo de Población y Vivienda 2010 contabilizó a poco más de 112

millones de habitantes, pero la última proyección del Consejo Nacional de Población (CONAPO) indicaba que serían poco más de 108 millones, una diferencia de cerca de 4 millones (García Guerrero, 2013). Lo que sorprende no es sólo la magnitud de esta diferencia, sino que, además, el dato proyectado fue menor que la población censada. La gravedad de esto es que cualquier planeación con ese dato no cumple sus objetivos y puede provocar que ciertos sectores de la población no sean alcanzados por los planes y programas de gobierno, por lo tanto, una proyección que mida la incertidumbre asociada permitiría informar, al menos de una forma más honesta, lo que se conoce y, sobre todo, lo que se desconoce, acerca de la dinámica demográfica de un país o región.

En general, las proyecciones de población oficiales no tienen una probabilidad asociada (como ninguna estimación puntual), y para evaluar su viabilidad, por ello, se acostumbra calcular escenarios extremos o alternativos de crecimiento poblacional, en los cuales se realizan supuestos sobre la evolución futura de las variables demográficas básicas; por ejemplo, para construir un escenario de crecimiento demográfico alto se supone que, a lo largo de todo el tiempo y de manera independiente —es decir que no interactúan—, la mortalidad permanece en descenso, la fecundidad se mantiene constante o en crecimiento y la migración neta es positiva o nula; por su parte, para desarrollar uno bajo se da por sentado que la mortalidad permanece en descenso, que la fecundidad se mantiene también en descenso y que el saldo neto migratorio permanece constante (en valores negativos).

Estos supuestos no son producto de un modelo matemático, sino de la experiencia del demógrafo que las diseña. Si la proyección programática se encuentra entre ambos escenarios alternativos, entonces se deduce que es consistente y útil para los fines programáticos que se establecen; sin embargo, esta forma de evaluar la incertidumbre no toma ventaja de las herramientas estadísticas y computacionales con las que se cuenta hoy en día, las cuales permiten determinar proyecciones de población menos imprecisas.

Los retos sociales, económicos, políticos y demográficos, aunados a la gran cantidad de información que se maneja en la actualidad, obligan a modificar la manera en que se calculan las proyecciones cambiando el paradigma del que parten por otro que permita considerar adecuadamente la incertidumbre intrínseca de cada una de las componentes demográficas y de la dinámica demográfica en su conjunto (García Guerrero, 2014a). Para realizar este tipo de proyecciones, es necesario contar con una estimación demográfica que abarque el mayor periodo histórico posible.

En virtud de lo anterior, el objetivo del presente trabajo es aplicar la metodología propuesta por Hyndman y Booth (2008) a la información demográfica de México y compararla en cuanto a sus resultados con otro método estocástico propuesto y aplicado para el país (García Guerrero, 2014a).

## Antecedentes

Metodológicamente, se han hecho algunos intentos por medir la incertidumbre inherente a toda prefiguración del futuro. Entre los principales trabajos sobre proyecciones estocásticas de la población se encuentran los de Sykes (1969), Le Bras (1971), Saboia (1974 y 1977), Heyde y Cohen (1985), Cohen (1986), Alho y Spencer (1985 y 2005), Alho (1990), Lee (1993 y 1974), Lee y Carter (1992), Lee y Tuljapurkar (1994) y, en fecha reciente, Hyndman y Booth (2008).

Debido a la falta de información confiable, de buena calidad y de larga data para México, sólo unos cuantos demógrafos han intentado incursionar en el ámbito de los pronósticos poblacionales estocásticos; por un lado, están los trabajos de Ordorica (2004 y 1995), Kesseli y Galindo (2007), González y Guerrero (2007), García Guerrero y Ordorica (2012) y García Guerrero (2014a, 2014b y 2014c), que consideran de manera integral a la dinámica demográfica en el sentido de que incorporan el impacto que tiene la estructura por edades; no obstante, también existen investigaciones que aplican técnicas de series de tiempo para

proyectar la población total de áreas pequeñas en México en los que sólo consideran el volumen total de la población (Silva, Guerrero y Peña, 2011 y Silva y Ordorica, 2013).

Ordorica (2004) realizó una proyección del número total de defunciones del país por medio del análisis estadístico de series de tiempo. Una limitante de ésta es —como bien señala el autor— que no se considera el efecto de la estructura por edad aunque sigue siendo útil para el pronóstico de corto plazo. En otro trabajo, Ordorica (1995) usó una aplicación del Filtro de Kalman para la estimación de la población mexicana a nivel nacional y en el análisis de políticas de población; su metodología es de gran utilidad para el análisis de series de tiempo con parámetros dependientes del tiempo; un elemento muy importante de esa investigación es que analiza el efecto en las proyecciones de población en el caso de que se adopten diferentes niveles de las varianzas de los errores; cabe mencionar que este trabajo es un hito en el campo de las proyecciones poblacionales en el sentido de que es el primero que se hizo para México en el cual se adopta un enfoque estocástico.

Kesseli y Galindo (2007) elaboraron una proyección demográfica de tipo probabilístico con el *Program on Error Propagation* (PEP) desarrollado por Juha Alho,<sup>1</sup> y realizaron un análisis de la incertidumbre que rodea a la información demográfica de México hacia el 2005.

Con el objetivo de evaluar el impacto en los sistemas de pensiones de la incertidumbre en la mortalidad, González y Guerrero (2007) utilizaron la modificación al modelo Lee-Carter propuesta por Li, Lee y Tuljapurkar (2004) para datos incompletos de la mortalidad y, con ello, pronosticaron la mortalidad mexicana para el periodo 2002-2050; luego, usaron la proyección de la fecundidad y migración realizada por el CONAPO (2002) para proyectar la población total del país por edades y sexo; así, los autores concluyeron que "...al comparar el nivel de mortalidad observado en el año 2000 con

1 <http://joyx.joensuu.fi/~ek/pep/pepstart.htm>

el pronóstico obtenido para el periodo 2002 a 2050 con el método LC, se observa que esta diferencia tendrá un impacto negativo, tanto desde la perspectiva de los afiliados, ya que recibirán una pensión menor al momento del retiro, como desde la perspectiva del gobierno, que tendrá que aumentar el gasto destinado a solventar este sistema de pensiones..." (p. 60).

Por último, García Guerrero (2014a) propuso una metodología de corte estocástico para proyectar la población mexicana a nivel nacional, la cual se basa en la proyección de las variables demográficas utilizando variaciones del método de Lee-Carter (1992) —mismo que se aplicó para la mortalidad mexicana en García Guerrero y Ordorica, 2012— y de Lee (1993); con los resultados, el autor realizó miles de simulaciones del método de las componentes demográficas o de las cohorte-componentes, lo cual le permitió medir y reportar la incertidumbre asociada por medio del cálculo de los percentiles correspondientes, no sólo a cada variable demográfica, sino la de la población total y su estructura por edades; lo reportado, él mismo lo actualizó (García Guerrero, 2014b) para la población total y el resto de las variables demográficas.

## Datos

Para la realización del pronóstico, es necesario contar con datos históricos desagregados por sexo y edad de los nacimientos, defunciones, migrantes netos y población total. En México, esta información no se obtiene de manera directa, ya que la calidad de los datos de las estadísticas vitales y los censos, si bien ha mejorado con el tiempo, aún tiene varias áreas de oportunidad; por ello, y con el fin de tener una serie histórica armonizada de la dinámica demográfica, en México —desde hace aproximadamente 15 años y después de cada censo o conteo— se realiza una estimación demográfica llamada conciliación, que provee de insumos para las proyecciones de población: la más reciente y confiable, y que abarca un periodo lo suficientemente largo como para poder calcular proyecciones estocás-

ticas, fue calculada por la Sociedad Mexicana de Demografía (SOMEDE) bajo encargo del CONAPO (SOMEDE, 2011).<sup>2</sup>

De acuerdo con los requerimientos de información del *análisis de datos funcionales* (FDA, por sus siglas en inglés, Hyndman y Booth, 2008), es necesario contar con los nacimientos ocurridos durante el año  $t$  provenientes de las mujeres a edad  $x$ ,  $B_t(x)$ , las defunciones ocurridas en el año  $t$  de las personas a edad  $x$ ,  $D_t(x)$ , la población de edad  $x$  a inicios de año,  $P_t(x)$ , y la población de edad  $x$  expuesta al riesgo (a mitad de año) en el año  $t$ ,  $E_t(x)$ , para  $x = 0, 1, \dots, \omega +$  y  $t = 1, 2, \dots, T$ . Las tasas específicas de mortalidad y fecundidad y las tablas de mortalidad se calculan de la forma usual en Demografía (ver Preston *et al.*, 2001), y la migración neta de las personas a edad  $x$  denotada por  $G_t(x)$  se estima utilizando la ecuación demográfica básica, es decir,  $G_t(x, x+1) = P_{t+1}(x+1) - P_t(x) + D_t(x, x+1)$  para  $x = 0, 1, \dots, \omega - 2$ ;  $G_t(\omega - 1, \omega +) = P_{t+1}(\omega +) - P_t(\omega +) - P_t(\omega - 1) + D_t(\omega - 1, \omega +)$  para el grupo abierto de edades  $\omega +$ , y  $G_t(B, 0) = P_{t+1}(0) - B_t + D_t(B, 0)$  para el primer grupo etáreo (Hyndman y Booth, 2008, p. 325).

## Modelo

Este trabajo se basa en el de Hyndman y Booth (2008), quienes proponen un método muy innovador para pronosticar las poblaciones sustentado en el FDA. De acuerdo con Ramsay y Silverman (2005), este tipo de análisis es una rama de la Estadística que analiza información empírica de curvas, superficies o cualquier otro objeto que varíe sobre un continuo, que usualmente es el tiempo, pero puede ser la localización geoespacial, longitud de ondas, probabilidades, etcétera. Los datos utilizados en el FDA pueden estar tan suavizados que se podría pensar que no tienen error alguno de medición; sin embargo, es posible que tengan una relación muy compleja con la curva que definen, es decir, se supone que, intrínsecamente, los datos provienen de una función muy suavizada

<sup>2</sup> Aunque después, esa conciliación no fue utilizada oficialmente, pero supera en calidad a la conciliación oficial (García Guerrero, 2013).

aunque no se vea reflejado al graficarlos. Muchas técnicas estadísticas, como series de tiempo o regresión multivariada, tienen sus contrapartes en el FDA.

En este marco, a grandes rasgos, el método de Hyndman y Booth modela las tasas vitales por edad y sexo como datos funcionales con coeficientes estimados tanto con descomposición en valores singulares como con series de tiempo. La migración es estimada como la diferencia entre los datos poblacionales históricos y las pronosticadas para cada año. Para tomar en cuenta la variación creciente con los valores de las tasas originales no suavizadas, utiliza la transformación de Box-Cox; después, con el fin de hacer más robusto el pronóstico, asume que los nacimientos y las defunciones siguen una distribución de *Poisson*; finalmente, emplea el método de simulación de monte-carlo para pronosticar la población total.

Cabe mencionar que el método de Hyndman y Booth está pensado de origen para datos provenientes de registros administrativos en contextos donde la información de nacimientos, defunciones y migraciones cuadra casi de manera perfecta con la censada; no obstante, se sugiere que es posible utilizar información estimada (o conciliada) en contextos en los que los registros administrativos tienen una calidad medianamente aceptable, como es el caso de México; por ello, uno de los objetivos de este artículo es aplicar dicha metodología a los datos provenientes de la conciliación demográfica realizada por la SOMEDE (2011) y compararlos con los resultados obtenidos con un método que no busca suavizar la información y se enfoca en pronosticar la dinámica demográfica (García Guerrero, 2014a).

La modelación de datos funcionales propuesta por los autores parte del trabajo previo de Hyndman y Ullah (2007), y consiste en modelar cada una de las componentes demográficas básicas por edad y sexo, denotadas por  $y_t^*(x)$  (i.e. tasa específica de mortalidad, fecundidad o nivel de la migración neta por edad al año  $t$ ), de tal manera que se les transforma de acuerdo con la propuesta de Box y Cox (esto con el fin de que la variabilidad sea congruente con la magnitud de las tasas estimadas, Hyndman y Booth, 2008, p. 326), es decir:

$$y_t(x) = \begin{cases} \frac{1}{\lambda} ([y_t^*(x)]^\lambda - 1) & \text{si } 0 < \lambda \leq 1; \\ \ln(y_t^*(x)) & \text{si } \lambda = 0. \end{cases}$$

donde el valor de  $\lambda$  determina la fuerza de la transformación; debido a que la migración neta puede contener valores negativos, se supone que  $\lambda = 1$ .

De esta manera, Hyndman y Booth suponen que la variable transformada  $y_t(x) = s_t(x) + \sigma_t(x) \varepsilon_{t,x}$ , donde  $s_t(x) = \mu(x) + \sum_{k=1}^K \beta_{t,k} \phi_k(x) + e_t(x)$  es una función de ajuste subyacente que depende de la edad,  $\varepsilon_{t,x}$ , es una variable aleatoria independiente e idénticamente distribuida normal estándar y  $\sigma_t(x)$  es una función que permite incluir una varianza que cambia con el tiempo y la edad; se supone que el cuadrado de dicha función se distribuye como una *Poisson* para la mortalidad y fecundidad con parámetro igual al número de defunciones y nacimientos, respectivamente, y se estima por medio de regresión no-paramétrica para el caso de la migración. Nótese que  $\mu(x)$  determina la dinámica promedio de la función de ajuste subyacente y  $\{\phi_k(x)\}$  es un conjunto de bases ortogonales calculadas a partir de la descomposición de componentes principales (Hyndman y Booth, 2008, p. 326). El conjunto de parámetros  $\{\beta_{t,k}\}$  controlan la dinámica del proceso y se asumen independientes entre sí. Por último,  $e_t(x)$  representa el error residual del modelo, que se supone serialmente no correlacionada.

El modelo anterior es una generalización del de Lee-Carter (1992), el cual, a su vez, está basado en el de edad-cohorte-periodo propuesto por Wilmoth (1990). La metodología de estimación de los parámetros es la siguiente (Hyndman y Booth, 2008, p. 326):

1. Se estiman las funciones de suavizamiento  $s_t(x)$  por medio de alguna regresión no-paramétrica aplicada a  $y_t(x)$  para cada año.
2. Se estima  $\mu(x)$  como el promedio de  $s_t(x)$  a lo largo del tiempo.
3. Se estiman  $\beta_{t,k}$  y  $\phi_k(x)$  utilizando descomposición en valores singulares (DVS) de la matriz  $y_t(x) - \hat{\mu}(x)$ .
4. Se estiman los modelos de series de tiempo que mejor se ajusten a las series de  $\hat{\beta}_{t,k}$  En

este caso, los autores proponen modelos de suavizamiento exponencial en el espacio de estados.

5. A partir de los valores pronosticados se extrapolan las variables haciendo sustitución inversa del resto de los parámetros y la transformación inversa correspondiente.

Ahora bien, para pronosticar la población total, se utiliza el tradicional método de proyección de las componentes demográficas o de la cohorte-componentes para cada una de las realizaciones de la variable futura estimada  $\hat{y}_t^F(x)$ , es decir, que a partir de las estimaciones de las variables aleatorias relacionadas con la variable demográfica se simulan miles de escenarios demográficos para cada año de la proyección, o sea, la población futura, de acuerdo con el modelo femenino dominante, para cada simulación queda determinada por:

$$P_{t+1}^F(x) = \left[ P_t^F(x-1) + \frac{G_{t,t+1}^F(x-1)}{2} \right] \frac{L_x^F}{L_{x-1}^F} + \frac{G_{t,t+1}^F(x)}{2}$$

para  $x \in [1, \omega - 1]$ ,  $P_{t+1}^F(x)$

$$= \left[ P_t^F(\omega - 1) + P_t^F(\omega +) + \frac{G_{t,t+1}^F(\omega-1)}{2} \right]$$

$$\frac{T_{\omega+}^F}{T_{\omega-1}^F} + \frac{G_{t,t+1}^F(\omega+)}{2} \left( 1 + \frac{T_{\omega+}^F}{T_{\omega-1}^F} \right), P_{t+1}^F(0)$$

$$= B_{t,t+1}^F \frac{L_0^F}{l_0^F} + \frac{G_{t,t+1}^F(0)}{2},$$

donde  $B_{t,t+1}(x) = f_t(x) \left[ \frac{P_t^F(x) + P_{t+1}^F(x) + G_{t,t+1}^F(x)/2}{2} \right]$  para  $x \in [15, 49]$  y  $\beta_{t,t+1} = \sum_{x=15}^{49} B_{t,t+1}(x)$ . La distribución por sexo queda determinada suponiendo una razón de masculinidad al nacimiento de 1.05. Para más detalles de la forma determinista de la proyección (que es un caso particular de la forma estocástica), se puede consultar Preston *et al.* (2001).

## Aplicación al caso de México

El periodo histórico comprendido en la estimación o conciliación utilizada comprende los años entre 1960 y el 2010 (SOMEDE, 2011). Para las estimaciones y proyecciones antes mencionadas, se usaron las li-

brerías *demography* (Hyndman *et al.* 2014a) y *forecast* (Hyndman *et al.* 2014b) programadas para el paquete estadístico *R* (*R Development Core Team*, 2005).

## Mortalidad

Para reportar la incertidumbre asociada a la mortalidad futura de mujeres y hombres, se emplea la esperanza de vida al nacimiento, aunque un análisis más detallado debería reportar la mortalidad materna y la infantil, sin embargo, en virtud de la extensión del presente trabajo, sólo se utilizará la esperanza de vida al nacimiento.

En este marco, los resultados apuntan a que, ante la disminución de la esperanza de vida al nacimiento masculina y el estancamiento en la femenina durante los últimos años atribuible principalmente al incremento en las defunciones ocasionadas por los homicidios (Mier y Terán y García Guerrero, 2013; Canudas-Romo y García-Guerrero, 2013), la incertidumbre en este indicador se incrementa de manera considerable, lo que se hace visible en la amplitud de los intervalos de predicción asociados. En efecto, como puede observarse en la gráfica 1, el intervalo de predicción de la población masculina tiende a incrementarse considerablemente rebasando, incluso, el mínimo correspondiente a 1960. De hecho, en la mediana se pronostica una disminución continua de la esperanza de vida masculina hasta el 2013, alcanzando un valor de 70.8 años para luego recuperarse de forma paulatina y llegar a un valor de 71.12 años hacia el 2050. En cambio, las proyecciones estocásticas de García Guerrero (2014b) y las del CONAPO (2014) presentan en la mediana y en la única variante, respectivamente, una tendencia ascendente;<sup>3</sup> no obstante, las proyecciones de García Guerrero, que de aquí en adelante se denotarán como PE1, llegan a un valor mediano de 77.2 años hacia el 2050 y las del CONAPO, al de 77.3 años. El intervalo de predicción de 95% de las primeras abarca el rango entre los

<sup>3</sup> Cabe mencionar que ambas proyecciones estocásticas, las calculadas con el FDA y las de García Guerrero, provienen de las mismas estimaciones, mientras que las del CONAPO, de estimaciones diferentes.

74.2 y los 79.6 años y el de las proyecciones FDA, uno de entre 51.3 y 80 años de esperanza de vida masculina al nacimiento.

Por otro lado, la esperanza de vida de las mujeres al nacimiento se estancó durante los últimos siete u ocho años. Esto se refleja en que en la mediana el pronóstico FDA muestra que dicho estancamiento se puede mantener en el futuro cerca de los 77.6 años de vida, su valor al 2010, llegando a oscilar entre los 66.4 y 83.4 años al 2050, un rango de 17 años. En cambio, las proyecciones PE1 muestran una tendencia creciente en la mediana pasando de 77.6 años a 82.6 años en el 2050, oscilando en ese último año entre 79.7 y 84.9 años, un rango de 5.2.

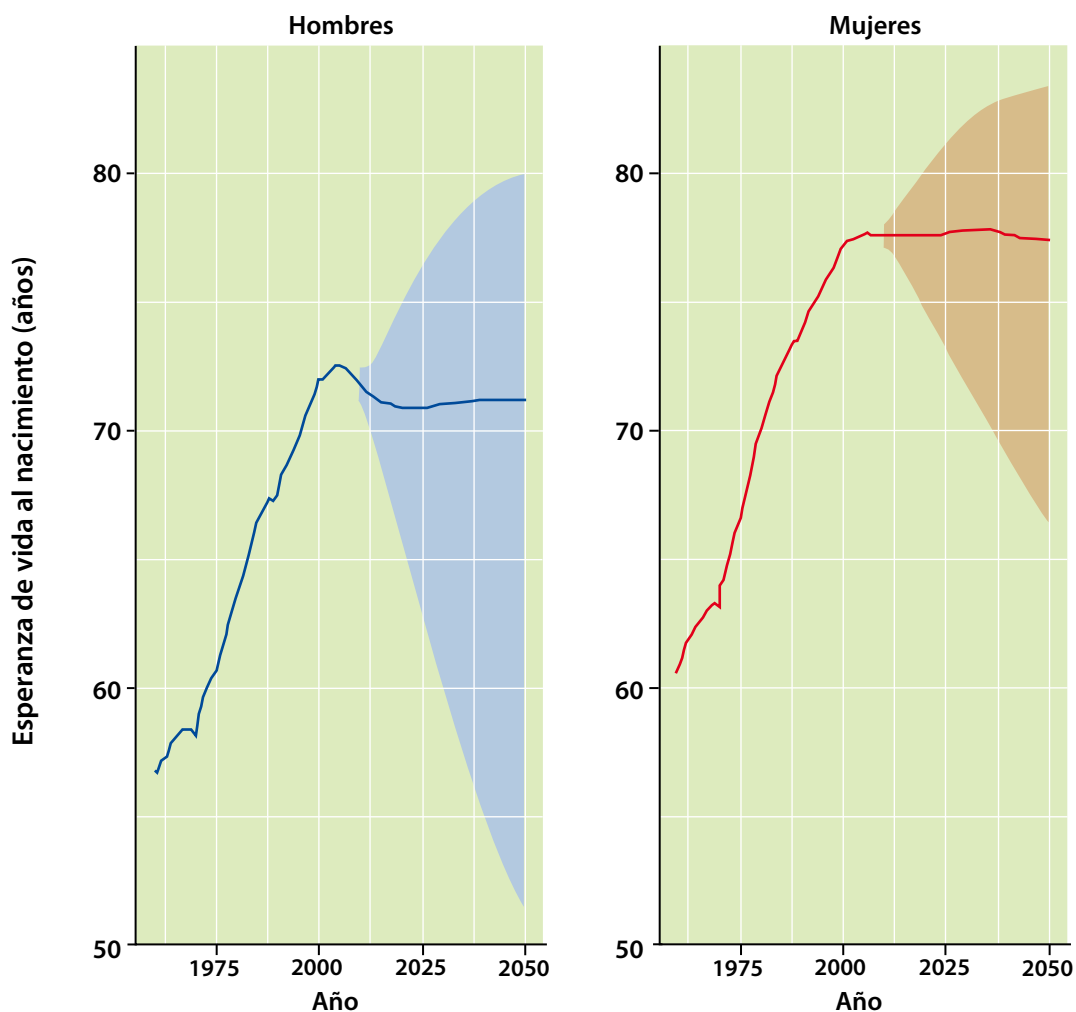
Las proyecciones del CONAPO presentan, asimismo, una tendencia creciente llegando a un valor de 81.6 años al 2050.

## Fecundidad

Respecto a esta variable, las proyecciones FDA no son tan volátiles como en el caso de la mortalidad (ver gráfica 2). De hecho, se mantienen en un descenso casi lineal en la mediana, pasando de 2.3 hijos por mujer en el 2010 a 2.04 en el 2050, pudiendo variar entre 1.6 y 2.7 hijos por mujer en ese último año. Eso señala que, de acuerdo con la evolución histórica de la fecundidad en México, el camino descendente

Gráfica 1

### Esperanza de vida al nacimiento estimada 1960-2009 y pronosticada 2010-2050, por sexo, México





no se encuentra garantizado del todo, sin embargo, tampoco está determinado a mantenerse en 2.08 hijos por mujer a partir del 2030, como supone el CONAPO.

Las proyecciones PE1 señalan una posible evolución futura un poco diferente, mientras que, al igual que las proyecciones FDA, en la mediana indican que la tendencia general seguirá a la baja, al pasar de 2.3 a 1.25 hijos por mujer, esto es cerca de un hijo por mujer menos. En este último año, las proyecciones PE1 señalan que la tasa general de fecundidad (TGF) podría oscilar entre 0.9 y 1.75 hijos por mujer. Estos últimos valores han sido cuestionados por algunos demógrafos, quienes aseguran que son poco plausibles para el contexto mexicano; sin embargo, hasta el momento, no existe evidencia empírica o alguna perspectiva teórica que ayude a refutar estos resultados, pero sí hay experiencias internacionales,

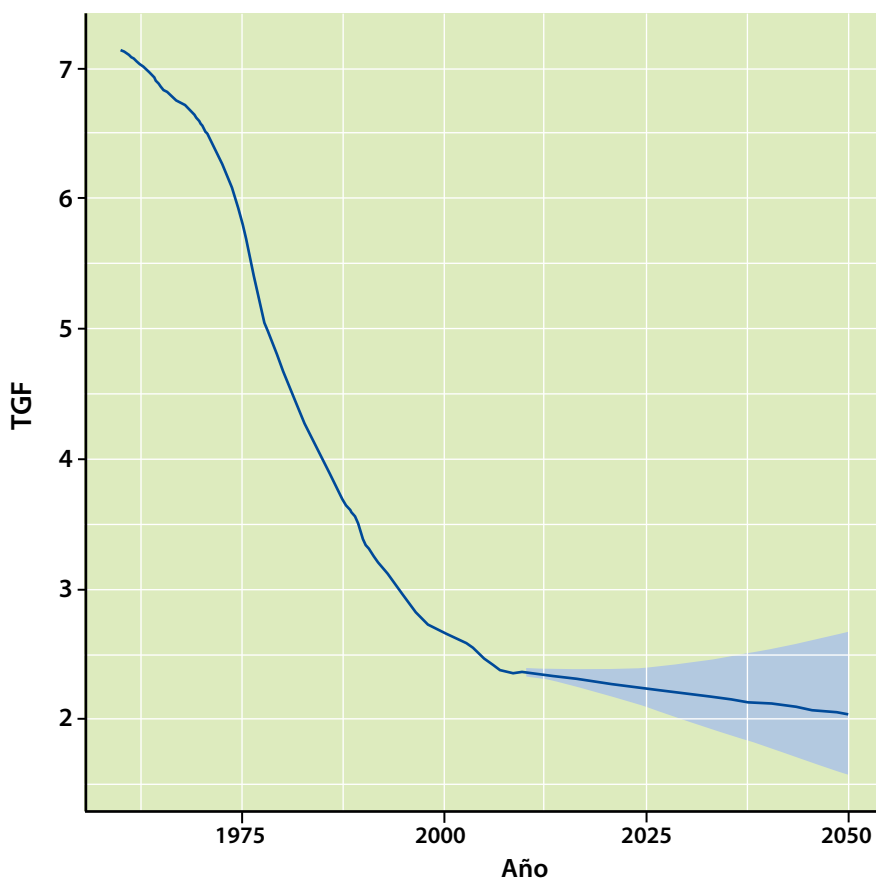
en países de Europa y del este de Asia durante la década de los 90, donde se ha observado una muy baja fecundidad (Goldstein *et al.*, 2009). Al igual que en las proyecciones FDA, las PE1 indican en su límite de predicción superior un estancamiento de la TGF hasta el 2018, pero a diferencia de las primeras, después pronostica un descenso acelerado.

## Migración neta internacional

La proyección de la migración internacional representa un reto metodológico particular. Usualmente, algunos expertos sugieren mantener las tasas de migración neta constantes a lo largo del horizonte de proyección (CONAPO, 2008 y 2003 y DPNU, 2012), lo cual es lo mejor cuando no se sabe nada sobre su dinámica intrínseca; por ejemplo, en las *Perspectivas de la población mundial de la ONU del*

Gráfica 2

### Tasa global de fecundidad estimada 1960-2009 y pronosticada 2010-2050, México



2012 (DPNU, 2012) se formularon dos hipótesis sobre el futuro de la migración neta: en la primera (llamada hipótesis de migración normal), los niveles proyectados de ésta se mantienen constantes hasta el 2015 y, después, suponen que se reducirán gradualmente; en la segunda (denominada hipótesis de migración cero), la migración internacional para cada país se fija en cero a partir del periodo 2010-2015 (DPNU, 2012, p. 12). Por otra parte, la más reciente proyección de población realizada por el CONAPO supone que, en el largo plazo, la migración neta subirá de nuevo a los niveles alcanzados en 1995 (CONAPO, 2012, p. 88). Estos supuestos se contradicen de alguna manera con el carácter programático de tales proyecciones, lo cual implica que el gobierno mexicano esperaría que, en el futuro, su población encontrara mejores oportunidades económicas en el extranjero que en México (García Guerrero, 2013).

Hay demasiados factores que afectan al fenómeno de la migración mexicana: sociales, económicos, políticos y ambientales, entre otros. Éstos son muy inciertos (principalmente los económicos), por lo que la migración también lo es. Hoy en día, el fenómeno migratorio mexicano atraviesa por un momento coyuntural provocado por la reciente crisis económica mundial. De acuerdo con Passel, en los años recientes la migración mexicana a Estados Unidos de América (EE.UU.) ha disminuido de manera inesperada por varias causas, "...la paralización de la migración parece ser el resultado de muchos factores, incluyendo el debilitamiento de los mercados de trabajo y de la construcción de viviendas en E.U., el incremento en la vigilancia fronteriza, aumento en las deportaciones, los crecientes peligros asociados con los cruces de indocumentados en la frontera, la disminución de las tasas de natalidad de México y las más amplias condiciones económicas en México..." (Passel 2012, p. 6).

Estos factores eran imprevisibles al comienzo del siglo XXI, cuando la migración de México a EE.UU. se encontraba en su nivel más alto: en el 2000, la inmigración anual de nuestro país hacia el vecino de norte fue de alrededor de 770 mil personas y, en el 2010, cayó a 140 mil por año, aproximadamente

(Passel 2012, p. 8). Otro efecto colateral de la crisis económica mundial es que muchos migrantes de Centroamérica que intentan cruzar el territorio mexicano rumbo a EE.UU. permanecen en México a la espera de mejores condiciones económicas en éste; no obstante, lo anterior no significa que el estancamiento de la emigración haya sido la única causa del estancamiento de la migración mexicana, sino también lo es el regreso de muchas familias con niños nacidos en Estados Unidos de América.

En este sentido, y como se puede observar en la gráfica 3, el pronóstico de la migración neta internacional considera casi cualquier escenario. Ante la alta volatilidad que ha presentado durante los últimos años, hacia el 2050 considera una migración internacional marcada por una fuerte inmigración o por un éxodo de mexicanos sin precedentes. La diferencia de las proyecciones FDA respecto de las PE1 es que estas últimas, por su construcción metodológica, acotan y conservan bajo control las variables que pueden incrementar la volatilidad. Esto se hizo con el fin de no tener cálculos poblacionales incoherentes, como serían poblaciones totales negativas; sin embargo, con el FDA esto no ocurre a pesar de la alta volatilidad. Por otra parte, respecto a las medianas, en el pronóstico PE1 se mantienen constantes y cercanas a cero, para ambos sexos, a lo largo del horizonte de proyección. En cambio, la proyección FDA muestra ciertas oscilaciones a lo largo del horizonte de proyección; por ejemplo, para el caso de la migración neta internacional masculina, llega a un valor negativo cercano a los 130 mil en el 2019 y a uno positivo de poco más de 30 mil en el 2050; para el caso de la femenina, señala que, en la mediana, desde el primer año del pronóstico, la tendencia es creciente, indicando un cambio de dirección en el flujo predominante de la migración neta hacia el 2030, cuando se observa un cambio de signo negativo a positivo, alcanzando un valor positivo de poco más de 80 migrantes netos hacia el 2050.

## **Población total**

Finalmente, al incorporar las proyecciones de las variables demográficas básicas en la ecuación de-

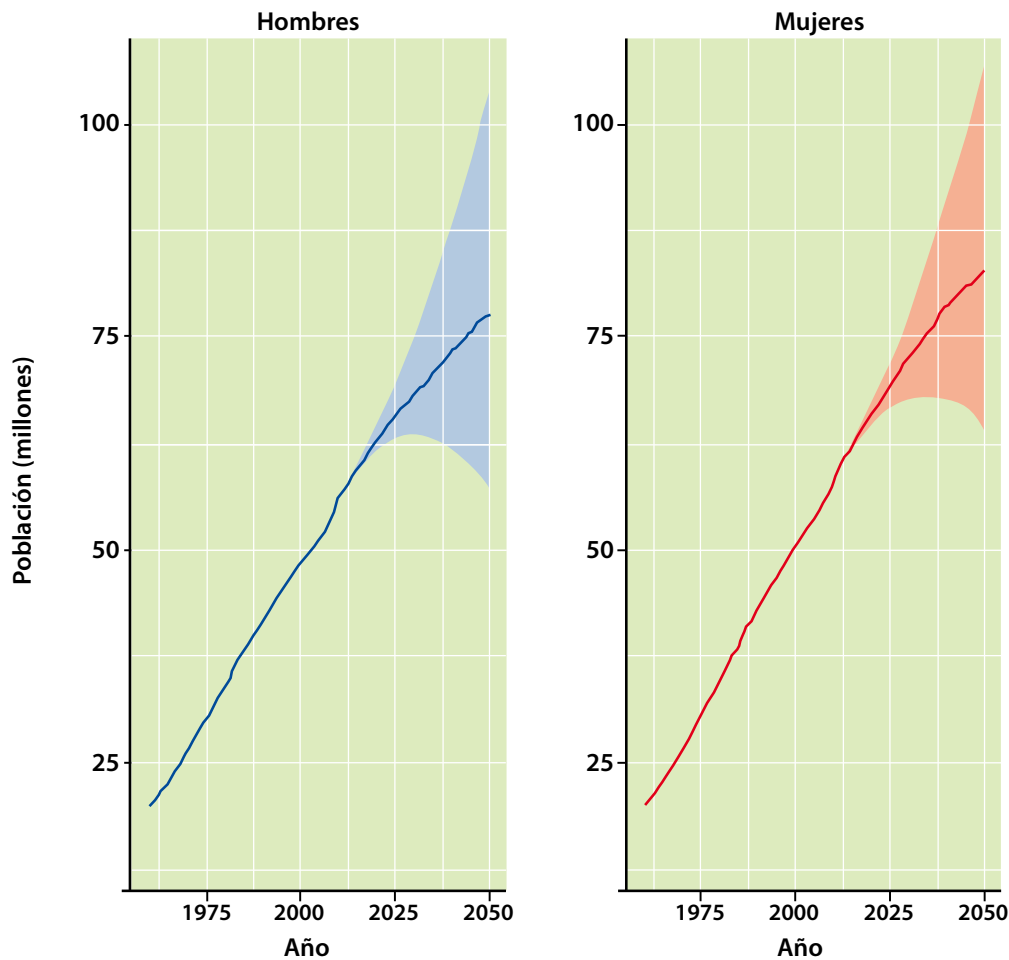
**Migración neta total estimada 1960-2009 y pronosticada 2010-2050, por sexo, México**

mográfica básica por edades y sexo en un modelo de simulación que permite incorporar, a su vez, la incertidumbre asociada a dichas variables, se calcula la población futura probable de la población mexicana. En las gráficas 4 y 5 se muestran, por un lado, los montos totales de población futuros por sexo y las pirámides poblacionales al inicio y al final de horizonte de proyección.

La alta incertidumbre cuantificada para la mortalidad y migración neta internacional se refleja en la forma de trompeta del intervalo de predicción para cada sexo, es decir, es tan probable que continúe un crecimiento ininterrumpido y que, incluso, se llegue a acelerar, como que comience a disminuir a una gran velocidad, pero en la mediana se muestra una clara tendencia a que siga el crecimiento

del país. Como se puede observar en la gráfica 4, el pronóstico FDA señala que la población podría pasar de 114.6 millones de habitantes en el 2010 (distribuidos en 58.6 millones de mujeres y 56 millones de hombres) a 128.7 millones en 2020 (66 millones de mujeres y 62.7 millones de hombres), con una oscilación de entre 126 millones y 131 millones en ese mismo año. Hacia el 2050 se pronostica un valor poblacional mediano de 161 millones (84 millones de mujeres y 77 millones de hombres), pudiendo oscilar entre 127 millones y 196 millones de personas en ese mismo año.

Por otro lado, el pronóstico PE1 parte de una base poblacional de 113 millones de habitantes en el 2010 (58 millones de mujeres y 55 millones de hombres) para llegar a 128 millones en el 2020

**Población total por sexo estimada 1960-2009 y pronosticada 2010-2050, México**

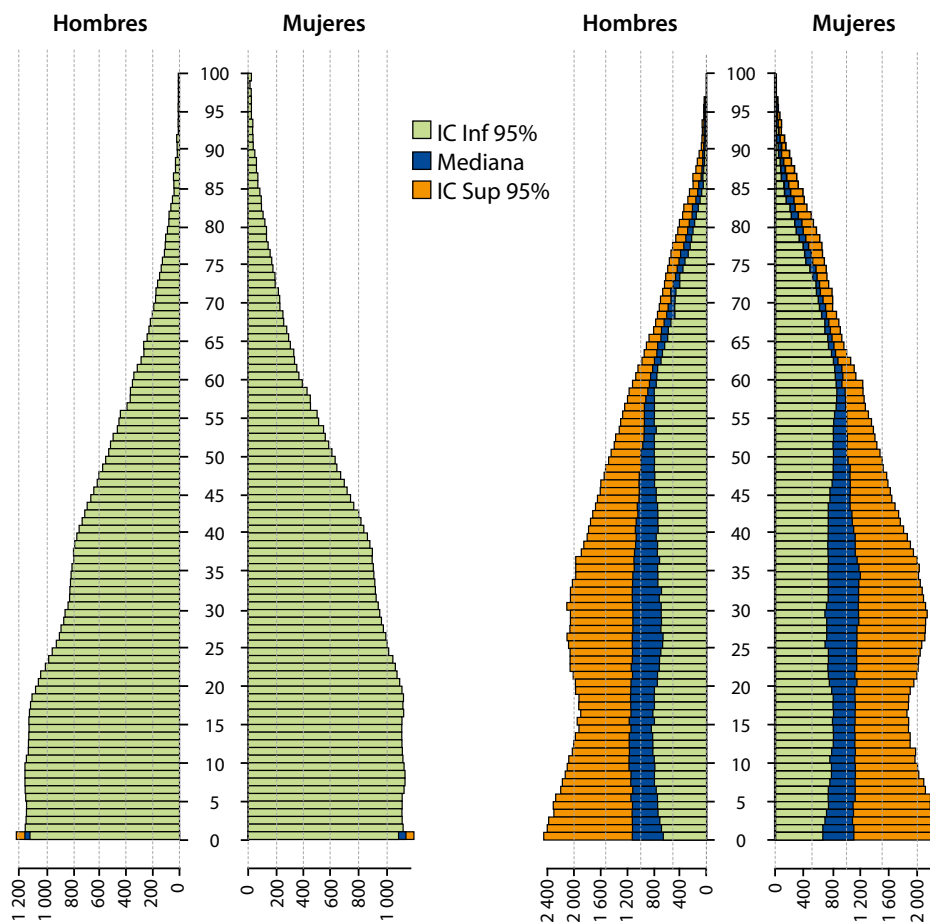
(distribuidos en 65 millones de mujeres y 63 millones de hombres) con una variabilidad de entre 125 millones y 131 millones de personas, cifras muy similares al pronóstico FDA, pero, hacia el 2050, el del PE1 muestra un valor mediano de 145 millones (74 millones de mujeres y 71 millones de hombres), con una oscilación de entre 125 millones y 166 millones. Lo anterior deja al pronóstico FDA 16 millones por arriba del PE1 y, mientras el rango de su intervalo de predicción es de 70 millones, el del PE1 es de 41 millones. Lo anterior se explica porque el pronóstico PE1 no permite que haya mucha volatilidad en las variables demográficas básicas; sin embargo, en el corto plazo, los resultados entre ambos métodos son muy similares.

Por último, la proyección del CONAPO parte, en el 2010, de un valor similar al del FDA: 114.3 millo-

nes (58.5 millones de mujeres y 55.8 millones de hombres). Al 2020, proyecta 127 millones (distribuidos en 66 millones de mujeres y 61 millones de hombres) y al 2050, 151 millones (78 millones de mujeres y 73 millones de hombres). Así, es probable que al 2020 subestime a la población que será censada en ese año y al final del horizonte de proyección queda en la intersección de los intervalos de confianza, por ello, es probable que para ese año, si no hay cambios abruptos en la demografía nacional, se obtenga un monto poblacional muy similar (lo cual no significa que las variables demográficas básicas tengan que operar de la manera en que son proyectadas en el ejercicio del CONAPO).

Lo anterior significa que, a largo plazo, la incertidumbre de las variables demográficas básicas se refleja en una alta incertidumbre de la población

## Estructura por edades y sexo pronosticada al 2010 (izquierda) y al 2050 (derecha), México



total para cada sexo. La planeación demográfica a largo plazo puede ser una tarea difícil y un tanto arriesgada; no obstante, a corto plazo, la incertidumbre se observa moderada y es posible asegurar que la variabilidad de la dinámica demográfica no puede afectar considerablemente la planeación. Las principales diferencias a corto plazo entre las proyecciones calculadas con métodos distintos pueden provenir de la estimación de la población base y de la estimación de la dinámica demográfica histórica (llamada también conciliación).

Respecto a la estructura por edades (ver gráfica 5), al final del horizonte de proyección, el pronóstico FDA indica una estructura etárea muy similar a la de una población estable, pero no en el límite de predicción superior en el que se observa

un rejuvenecimiento abrupto de la población. En cambio, la proyección del CONAPO muestra, para el 2050, una estructura por edades aún concentrada en los grupos jóvenes, llegando a una base todavía predominante, producto de mantener la fecundidad constante. En cambio, en el pronóstico PE1 se presenta una continua disminución de la base de la pirámide, no sólo en la mediana sino en los límites de predicción, lo cual conlleva a que esta proyección es la única que muestra un acelerado proceso de envejecimiento. Éste se debe, en parte, a que no se controla el descenso de la fecundidad y llega a niveles de muy baja fecundidad. Lo anterior implica que, si bien las diferencias entre las distintas proyecciones pueden ser relativamente pocas en el monto poblacional, en la estructura por edades indican escenarios demográficos muy diferentes entre sí.

## Conclusiones

En este trabajo se aplica la perspectiva de datos funcionales propuesta por Hyndman y Booth (2008) para pronosticar la población mexicana. Se comparan los resultados tanto los obtenidos en el pronóstico estocástico desarrollado en García Guerrero (2014a) como los de las proyecciones deterministas del CONAPO (2012). Los tres tipos de ejercicios prospectivos parten de un marco metodológico general muy similar en el que se incorporan las proyecciones o pronósticos de las variables demográficas en la ecuación demográfica básica por edades, también conocido como método de las componentes o de las cohorte-componentes. Éste supone que es por medio de la ecuación demográfica básica que se genera la interacción entre las variables demográficas lo cual, a su vez, resulta una determinada estructura por edades y sexo. Además, supone que es posible pronosticar cada variable demográfica como una serie histórica independiente ya que, al utilizar las tasas específicas de cada fenómeno demográfico, implícitamente se incorpora la información del fenómeno mismo, de su estructura por edades y de la población expuesta al riesgo relacionada y su estructura por edades, misma que, a su vez, contiene la información del fenómeno a modelar y su interacción con las otras dos variables demográficas.

De esta manera, el FDA, en su afán de no dejar fuera ninguna fuente de incertidumbre, calcula intervalos de predicción muy amplios, sobre todo en variables que han sufrido *shocks* inesperados en su dinámica, como es el caso de la mortalidad y migración neta internacional mexicanas. Cuando la dinámica varía dentro de lo esperado y continúa con la tendencia esperada, el FDA estima y pronostica de manera muy eficiente, como es el caso de la fecundidad. Esto es razonable, ya que el FDA ha sido probado sólo para el caso de Australia, país que tiene una calidad en su información demográfica muy superior a la de México hoy en día. El PE1 controla mejor el rango de variación de las componentes demográficas básicas, sobre todo aquellas que están sujetas a muy alta incertidumbre, además de que no utiliza métodos de suaviza-

miento ni fuerza a que exista coherencia entre las variables demográficas en el sentido de Hyndman y Ullah (2007); sin embargo, con el afán de no interferir subjetivamente con el modelo, permite que variables, como la fecundidad, continúen su evolución sin restringirla en virtud de las implicaciones en la dinámica demográfica en su conjunto.

No obstante, ambos métodos son mucho más potentes que las extrapolaciones deterministas que se han venido realizando oficialmente en el sentido que se mencionó con anterioridad, ya que permiten medir la incertidumbre inherente a cualquier fenómeno humano y porque permiten informar de una manera más honesta y completa lo que se sabe y, sobre todo, lo que se desconoce acerca de la dinámica demográfica. Es necesario seguir realizando aplicaciones y propuestas de este tipo para el caso de la información demográfica de México o de países con similitudes en la captación de datos, de tal suerte que, de forma eventual, se llegue a construir un método unificado para la mejor toma de decisiones. Tal vez, las propuestas provenientes de la estadística bayesiana tengan mucho que aportar al respecto.

## Fuentes

- Alho, Juha. "Stochastic methods in population forecasting", en: *International Journal of Forecasting*. Núm. 6, 1990, pp. 521-530.
- \_\_\_\_\_. "Uncertain Population Forecasting", en: *Journal of the American Statistical Association*. Vol. 80, Núm. 390, 1985, pp. 306-314.
- Alho, Juha y B. Spencer. *Statistical demography and forecasting*. Springer Series in Statistics. EE.UU., Springer, 2005.
- Booth, Heather, J. Maindonald y L. Smith. "Age-time interactions in mortality projection: Applying Lee-Carter to Australia", en: *Working Papers in Demography*. No. 85. Demography and Sociology Program, Research School of Social Sciences, The Australian National University, 2002.
- Box, George y G. Jenkins. *Time Series Analysis, Forecasting and Control*. San Francisco, California, EE.UU., Holden-Day, 1976.
- Canudas-Romo, Vladimir y V. M. García-Guerrero. *The Stagnation of the Mexican Life Expectancy in the First Decade of the Twenty First Century: The Impact of Violent Deaths*. Ponencia presentada en la sesión 043: *Crisis mortality: documenting the effects of conflict, famines and natural disasters* en el marco de la XXVIII IUSSP International Population Conference. Busan, Corea del Sur, agosto del 2013.

- Cohen, Joel E. "Population Forecasts and Confidence Intervals for Sweden: A Comparison of Model-Based and Empirical Approaches", en: *Demography*. Vol. 23, Núm. 1, 1986, pp. 105-126.
- CONAPO. Proyecciones de la población de México, de las entidades federativas, de los municipios y de las localidades, 2005-2050. *Documento metodológico*. México, CONAPO, 2008.
- \_\_\_\_\_. *Proyecciones de la población de México, de las entidades federativas, de los municipios y de las localidades, 2000-2050*. México, CONAPO, 2003.
- División de Población de Naciones Unidas (DPNU). Assumptions Underlying the 2010 Revision. *Department of Economic and Social Affairs*, DPNU, 2012. Disponible en <http://esa.un.org/unpd/wpp/Documentation>
- García Guerrero, Víctor Manuel. *Proyecciones y políticas de población en México*. México, Centro de Estudios Demográficos, Urbanos y Ambientales, El Colegio de México, AC, 2014<sup>a</sup>.
- \_\_\_\_\_. "Las proyecciones de la población de México", en: Rabell, C. (coord.). *Los mexicanos: un balance del cambio demográfico*. México, Fondo de Cultura Económica, 2014b, pp. 857-900.
- \_\_\_\_\_. "Las estimaciones y proyecciones demográficas como herramienta para el diseño de políticas de población", en: Valdés, L. M. (coord.). *Hacia una nueva ley general de población en México*. México, Instituto de Investigaciones Jurídicas, UNAM, 2014c, pp. 259-278.
- \_\_\_\_\_. "Las nuevas proyecciones de población 2010-2050", en: *Boletín de la Sociedad Mexicana de Demografía*. México, SOMEDE, AC, 2013.
- \_\_\_\_\_. "Un análisis de las diferencias entre las proyecciones de población 2006-2050 y el censo de población 2010", en: *Coyuntura Demográfica. Revista sobre los Procesos Demográficos en México Hoy*. Vol. I, Núm. 1. México, Sociedad Mexicana de Demografía, AC, 2011. pp. 62-65. Disponible en <http://www.somede.org/coyuntura-demografica/flipping/numero1/#/62>
- García Guerrero, Víctor Manuel y M. Ordorica. "Proyección estocástica de la mortalidad mexicana por medio del método de Lee-Carter", en: *Estudios Demográficos y Urbanos*. Vol. 27, Núm. 2 (80). Centro de Estudios Demográficos, Urbanos y Ambientales, El Colegio de México, AC, 2012, pp. 409-448.
- Goldstein, J., T. Sobotka y A. Jasiloniene. "The End of "Lowest-Low" Fertility?", en: *Population and Development Review*. 35(4), 2009, pp. 663-699.
- Gómez de León, José. *Empirical DEA Models to Fit and Project Time Series of Age-Specific Mortality Rates*. No publicado. Central Bureau of Statistics, Noruega, 1990.
- González Pérez, Carlos y V. Guerrero Guzmán. *Pronósticos estadísticos de mortalidad y su impacto sobre el sistema de pensiones de México*. Trabajo ganador del Premio de Pensiones 2007 otorgado por la Comisión Nacional del Sistema de Ahorro para el Retiro, 2007. Consultado en [http://www.consar.gob.mx/premio\\_pensiones/premio\\_2007.shtml](http://www.consar.gob.mx/premio_pensiones/premio_2007.shtml)
- Heyde, C. C. y J. E. Cohen. "Confidence Intervals for Demographic Projections Based on Products of Random Matrices", en: *Theoretical Population Biology*. Vol. 27, Núm. 2, 1985, pp. 120-153.
- Hyndman, Rob J. y H. Booth. "Stochastic population forecasts using functional data models for mortality, fertility and migration", en: *International Journal of Forecasting*. Vol. 24, Núm. 3, 2008, pp. 323-342.
- Hyndman, R. y S. Ullah. "Robust forecasting of mortality and fertility rates: a functional data approach", en: *Computational Statistics & Data Analysis*. 51, 2007, pp. 4942-4956.
- Hyndman R. (con contribuciones de Heather Booth, Leonie Tickle y John Maindonald). *Demography: Forecasting mortality, fertility, migration and population data. R package version 1.17*. 2014a. Disponible en <http://CRAN.R-project.org/package=demography>
- Hyndman R. (con contribuciones de George Athanasopoulos, Slava Razbash, Drew Schmidt, Zhenyu Zhou, Yousaf Khan, Christoph Bergmeir and Earo Wang). *Forecast: Forecasting functions for time series and linear models. R package version 5.3*. 2014b. Disponible en <http://CRAN.R-project.org/package=forecast>
- Kesseli, Katja y C. Galindo. "The Many Mexicos. Stochastic Forecast 2001-2050", en: *Papeles de Población*. Núm. 51, 2007, pp. 147-185.
- Le Bras, H. *Un modèle de projections aleatoires*. Documento presentado en la conferencia de la IUSSP-Ciudad de México (mimeo), 1971.
- Lee, Ronald. "Probabilistic Approaches to Population Forecasting", en: *Population and Development Review, Supplement: Frontiers of Population Forecasting*. Vol. 24, 1998, pp. 156-190.
- \_\_\_\_\_. "Modeling and forecasting the time series of US fertility: Age distribution, range, and ultimate level", en: *International Journal of Forecasting*. Vol. 9, 1993, pp. 187-202.
- \_\_\_\_\_. "Forecasting Births in Post-Transition Populations: Stochastic Renewal with Serially Correlated Fertility", en: *Journal of the American Statistical Association*. Vol. 69, Núm. 347, 1974, pp. 607-617.
- Lee, Ronald y L. Carter. "Modeling and Forecasting U.S. Mortality". Supplement: *Aging, Health, and Public Policy*, en: *Journal of the American Statistical Association*. Vol. 87, Núm. 419, 1992, pp. 659-671.
- Lee, Ronald y T. Miller. "Evaluating the performance of the Lee-Carter method for forecasting mortality", en: *Demography*. Vol. 38, Núm. 4, 2001, pp. 537-549.
- Lee, Ronald y S. Tuljapurkar. "Stochastic population forecast for the United States: Beyond high, medium and low", en: *Journal of the American Statistical Association*. Vol. 89, Núm. 428, 1994, pp. 1175-1189.
- Li, Nan, R. Lee y S. Tuljapurkar. "Using the Lee-Carter method to forecast mortality for populations with limited data", en: Lutz, Wolfgang y Joshua Goldstein (eds.). *How to deal with uncertainty in population forecasting?*, Reprinted from *International Statistical Review*. 72(1 y 2):1-106, 157-208, 2004, pp.19-36.
- Meyer, Carl. *Matrix Analysis and Applied Linear Algebra*. EE.UU., Society for Industrial and Applied Mathematics Philadelphia, 2000.

- Mier y Terán, Marta y V. M. García Guerrero. "Changements démographiques récents et perspectives futures au Mexique", en: *Problèmes d'Amérique latine*. Núm. 87. Francia, Editions Choiseul, I.S.S.N. 0765-1333, 2012, pp. 49-78.
- Oeppel, Jim y J. Vaupel. "Broken limits to life expectancy", en: *Science*. 296(5570), 2002, pp. 1029-1031.
- Ordorica Mellado, Manuel. "Pronóstico de las defunciones por medio de los modelos autorregresivos integrados de promedios móviles", en: *Papeles de Población*. Núm. 42. México, Universidad Autónoma del Estado de México, 2004, pp.249-264.
- \_\_\_\_\_. *El Filtro de Kalman en la planeación demográfica*. Tesis doctoral. México, División de Estudios de Posgrado de la Facultad de Ingeniería, UNAM, 1995.
- Passel, Jeffrey, D. Cohn y A. González-Barrera (). *Net Migration from Mexico Falls to Zero and Perhaps Less*. EE.UU., Pew Hispanic Center, Pew Research Center, Washington, DC, 2012.
- Preston, Samuel, P. Heuveline y M. Guillot. *Demography: Measuring and Modeling Population Processes*. Londres, Inglaterra, Blackwell Publishers Ltd., 2000.
- R Development Core Team. *R: A language and environment for statistical computing*. Viena, Austria, R Foundation for Statistical Computing, 2005. Disponible en <http://www.R-project.org>
- Saboia, João L. "Autoregressive Integrated Moving Average (ARIMA) Models for Birth Forecasting", en: *Journal of the American Statistical Association*. Vol. 72, Núm. 358, 1977, pp. 264-270.
- \_\_\_\_\_. "Modeling and forecasting populations by time series: The Swedish case", en: *Demography*. 11, 1974, pp. 483-492.
- Silva, E., V. Guerrero y D. Pea. "Temporal disaggregation and restricted forecasting of multiple population time series", en: *Journal of Applied Statistics*. Vol. 38, Issue 4, 2011, pp. 799-815.
- Silva E. y M. Ordorica. "Pronósticos multivariados de poblaciones con series de tiempo: el caso de la ZMCM contrastado con datos del Censo 2010", en: *Estudios Demográficos y Urbanos*. Vol. 28, Núm. 1 (82), 2013.
- Sykes, Z. M. "Some stochastic versions of the matrix model for population dynamics", en: *Journal of the American Statistical Association*. Vol. 44, 1969, pp. 111-130.
- SOMEDE. *Conciliación demográfica de México y entidades federativas 1990-2010*. Inédito, 2011.
- Wilmoth, John R. "Variation in vital rates by age, period, and cohort", en: *Sociological Methodology*. 20, 1990, pp. 295-335.
- \_\_\_\_\_. "Computational methods for fitting and extrapolating the Lee-Carter model of mortality change", en: *Technical Report*. EE.UU. Department of Demography, University of California, Berkeley, 1993.