

# Proyecciones probabilísticas de la población argentina

## Probabilistic projections of the Argentine population

Lucía Andreozzi

*Universidad Nacional de Rosario CONICET, Argentina*

### *Resumen*

En este trabajo se aplican modelos para datos funcionales (MDF) a las cifras de mortalidad y fecundidad de Argentina para obtener sus correspondientes pronósticos probabilísticos y a partir de ellos, proyecciones de población. Este ejercicio resulta interesante frente a las dificultades presentadas en la ronda censal 2020, ya que podría aportar cifras provisorias ante la espera de su realización definitiva. La aplicación de estos modelos a datos de Argentina en base a datos del periodo 1980-2010, año del último censo disponible, demuestra que cuando se cuenta con información para períodos no muy extensos de tiempo y una calidad aceptable de las fuentes de datos, se pueden emplear con éxito para realizar pronósticos probabilísticos. En segundo lugar, se aplica a datos de mortalidad la metodología alternativa de pronósticos coherentes con el fin de evitar una divergencia entre ambos géneros. Finalmente, a partir de los pronósticos de fecundidad (basados en dos hipótesis de fecundidad) y de mortalidad (independientes y coherentes) se elaboran pronósticos estocásticos de la población para hombres y mujeres. En todos los casos evaluados las cifras estimadas se encuentran por encima de las oficiales. Los resultados aquí obtenidos presentan valores coherentes con la teoría demográfica y principalmente garantizan la consistencia probabilística.

*Palabras clave:* Mortalidad, fecundidad, datos funcionales, pronósticos coherentes.

### *Abstract*

In this study, models for functional data (MDF) are applied to the mortality and fertility data from Argentina to obtain their corresponding probabilistic forecasts and, from them, population projections. This exercise is interesting given the difficulties presented in the 2020 census round, as it could provide provisional figures pending its final completion. The application of these models to data from Argentina based on data from the period 1980-2010, the year of the last available census, shows that when there is information for not very long periods of time and an acceptable quality of the data sources, can be used successfully to make probabilistic forecasts. Second, the alternative consistent forecasting methodology is applied to mortality data in order to avoid a divergence between the two genders. Finally, from the fertility forecasts (based on two fertility hypotheses) and mortality (independent and coherent), stochastic forecasts of the population for men and women are elaborated. In all the cases evaluated, the estimated quantities are above the official ones. The results here are consistent with demographic theory and mainly correspond to probabilistic consistency.

*Keywords:* Mortality, fertility, functional data, consistent forecasts.

## INTRODUCCIÓN

**E**s inevitable que las proyecciones contengan incertidumbre y si bien la situación demográfica actual es conocida, las tendencias en los nacimientos, muertes y migraciones están sujetas a cambios impredecibles. Los errores cometidos en la elaboración de proyecciones varían de acuerdo con el horizonte de pronóstico, el nivel de desarrollo del país y su tamaño. Mientras que las tendencias generales en la mortalidad, la fecundidad y la migración pueden discernirse y proyectarse para el futuro con una “confianza” (en el uso no estadístico de la palabra) razonable, una sustancial incertidumbre se haya asociada a la tendencia particular de un país o región. Cuantificar esta incertidumbre es de gran utilidad para los usuarios de las proyecciones y dado que los escenarios alternativos que se elaboran como parte de una proyección demográfica tienen diferentes implicancias en el futuro, conocer una medida de la incertidumbre asociada a cada resultado permite decidir en base a más información. Actualmente la posible variación en los resultados se expresa proveyendo escenarios alternativos con variantes para la fecundidad (y con menos frecuencia en la mortalidad y la migración). De este modo los escenarios alto y bajo se utilizan para indicar un rango de posibles futuros, no obstante, ninguna probabilidad específica se asocia a estos resultados.

En 1998 el Consejo Nacional de Investigaciones de la Academia de Ciencias de los Estados Unidos conformó un panel de expertos para analizar el tema de las proyecciones en profundidad (Bongaarts y Bulatao, 2000). El panel concluyó que las proyecciones elaboradas para el año 2050 se basaban en supuestos razonables y proveían pronósticos plausibles para las tendencias demográficas en las próximas décadas. Esta conclusión no abarcaba las proyecciones realizadas para países específicos, dado que el panel solamente examinó la metodología general de las proyecciones a nivel mundial. Además, entre las conclusiones se destacaba que las proyecciones elaboradas no otorgan la importancia necesaria a un aspecto relevante: la incertidumbre asociada a los pronósticos, de modo que el panel indicó que sería adecuado que las cifras de proyecciones contemplaran de forma explícita esta característica y remarcaba la necesidad de desarrollar métodos que la cuantifiquen.

La irrupción más clara y concisa de los métodos probabilísticos de pronóstico en el área demográfica la concretó el método propuesto por Lee y Carter (Lee y Carter, 1992). El mismo posee actualmente numerosas variantes y extensiones. Originalmente los autores proponen una metodolo-

gía que permite modelar y extraer las tendencias observadas en las tasas de mortalidad a largo plazo e implementan dicha metodología para pronosticar la mortalidad de los Estados Unidos hasta el año 2065. Lee y Miller (2001) y Booth *et al.* (2002) han propuesto modificaciones al método de Lee-Carter, como la elección del periodo de ajuste, el método para el ajuste del parámetro de nivel y la elección de las tasas base para el pronóstico. La propuesta de Lee y Miller (2001) es ampliamente utilizada y la variante propuesta por Booth *et al.* (2002) ha demostrado ser al menos tan precisa como la de Lee y Miller en el corto plazo, ver Booth *et al.* (2005) y Booth *et al.* (2006a). Varios desarrollos han incorporado una estructura de error heterocedástica Poisson como Brouhns *et al.* (2002) y Wilmoth (1993)). Otros autores han extendido la aplicabilidad del modelo de Lee-Carter a los factores de reducción de la mortalidad, ver Renshaw y Haberman (2003a) o examinado el uso de más de un término en el modelo, como en el caso de Booth *et al.* (2002) y Renshaw y Haberman (2003b). También existen dos extensiones recientes que incluyen suavizados semi-paramétricos en el modelo; Jong y Tickle (2006) combinan un suavizado por *splines* y una estimación por medio del filtro de Kalman para ajustar una versión generalizada del modelo de Lee y Carter. Hyndman y Ullah (2007) proponen utilizar el paradigma de los datos funcionales para modelar las componentes demográficas; el mismo considera que los logaritmos de las tasas de cada año son una función continua de las edades. Hyndman y Ullah (2007) suavizan la mortalidad a través de regresiones *spline* penalizadas para luego ajustar un modelo mediante una descomposición en componentes principales. Estos métodos son comparados por Booth *et al.* (2006b)

Los mayores avances se presentaron en relación a la mortalidad, en cambio, la fecundidad y las migraciones han recibido menor atención por parte de los investigadores. En cuanto al pronóstico de la fecundidad se destaca, en primer lugar, que los métodos elaborados para dicha componente se encuentran menos desarrollados que los destinados al análisis de la mortalidad. La fecundidad presenta dificultades a la hora de ser pronosticada debido a los cambios estructurales en su comportamiento altamente asociados a cambios en las pautas culturales. Lee (1993) encontró necesario preespecificar el valor medio a largo plazo de la fecundidad total e imponer valores límites para reducir la amplitud del intervalo de predicción e implementó un método paralelo al del Lee y Carter. Por otro lado, un enfoque de componentes principales fue empleado por Bozik y Bell (1987), quienes utilizan en su análisis los primeros cuatro componentes principales y modelos ARIMA. Un método similar emplea Hyndman y Ullah (2007)

como parte de un análisis funcional, dicho método es el que se utiliza en el presente trabajo. Los autores emplean un mismo modelo base para analizar la mortalidad y la fecundidad. La metodología propuesta por Hyndman y Ullah (2007) difiere en varios aspectos importantes con respecto al modelo de Lee-Carter; en primer lugar, está enmarcado en el paradigma de datos funcionales (Ramsay y Silverman, 2005), por otro lado, emplea suavizados semi-paramétricos con el fin de reducir la aleatoriedad inherente a los datos observados y en la etapa de descomposición de las tasas suavizadas permite utilizar componentes principales clásicos o robustos. Estos últimos tienen en cuenta dentro del análisis, los posibles *outliers*. En relación con el modelado de *outliers*, Li y Chan (2005) proponen una versión robusta del modelo de Lee-Carter, pero esta no se ubica dentro del paradigma de los datos funcionales.

Los métodos para pronosticar la migración están aún menos desarrollados y son habitualmente mucho más simples (George y Perreault, 1992). La razón principal de esta falta de desarrollo radica en la carencia de series temporales confiables, detalladas y que representen adecuadamente los flujos migratorios internacionales. Además, las tendencias observadas en esta componente son más volátiles, debido a los acontecimientos políticos y económicos y los cambios en la legislación. Un método que puede solucionar la falta de datos es estimar la migración neta como la diferencia entre el incremento del tamaño de la población y el crecimiento natural utilizando la ecuación compensadora, también llamada ecuación de crecimiento. Más aún, para pronósticos de población subnacionales que involucran efectos de la migración interna, este método es el único disponible ya que es frecuente que no se registren datos sobre migración a ese nivel de desagregación, ver Miller (2003); Miller y Lee (2004). Además, el uso de la migración neta no contempla un aspecto importante: la bondad de pronóstico depende de la desagregación de acuerdo con diferentes grupos o tipos de inmigrantes y emigrantes (Hilderink *et. al.*, 2002 y Rogers, 1990). Por otro lado, Beer (1997) encontró consistencia entre pronósticos de series de tiempo de inmigración total, emigración y migración neta. También es posible obtener los pronósticos de migración neta desagregados por edad (Keilman y Pham, 2004), utilizando una versión reducida del modelo multi-exponencial desarrollado en Rogers y Castro (1981) y Rogers y Little (1994).

Durante los últimos años se han desarrollado múltiples enfoques para pronósticos probabilísticos (Booth *et al.*, 2006a) y dentro de esta categoría han cobrado una particular relevancia los métodos para datos funcionales

(Ramsay y Silverman, 2005). Estos métodos, de reciente aparición, constituyen un nuevo marco para el análisis de series de tiempo, que ha sido adoptado, entre otras finalidades para realizar pronósticos para todas las componentes demográficas (Hyndman y Ullah, 2007). Un aporte interesante a este enfoque lo hacen Hyndman *et al.* (2013) al introducir la idea de pronósticos coherentes en el paradigma de datos funcionales. La idea principal de esta propuesta radica en que la diferencia entre los pronósticos de grupos de interés debe permanecer constante a través del tiempo, reproduciendo la relación presente en los datos observados. Los grupos de interés pueden ser subregiones geográficas o géneros, por mencionar algunas posibilidades.

Las proyecciones de población son el producto demográfico más utilizado por los no demógrafos y sin embargo no es el principal objeto de estudio por parte de los científicos de población. Los usuarios de estos resultados podrían sorprenderse por la falta de una teoría rigurosa, o una base histórica para los escenarios que subyacen en las proyecciones más usadas. Si bien las tendencias generales de la fecundidad, la mortalidad y la migración pueden discernirse y proyectarse hacia el futuro con resultados razonables, existe una incertidumbre considerable vinculada a cada tendencia específica y a un país o región particular. La cuantificación de esta incertidumbre es útil para los usuarios de las proyecciones porque permite tener en cuenta las diferentes implicancias de un pronóstico. En los pronósticos actuales, la incertidumbre se expresa típicamente proporcionando escenarios alternativos basados en la variación de la trayectoria para la fecundidad, pero rara vez, para la mortalidad y la migración. Escenarios altos y bajos se utilizan para abarcar una gama de posibles futuros. Sin embargo, ninguna medida de probabilidad específica se adjunta a esta gama, y su significado es, por lo tanto, ambiguo.

Es importante destacar que día a día se suman propuestas novedosas para proyectar la población, como por ejemplo la propuesta de emplear la teoría de conjuntos de difusos (Argote Cusi, 2018), sin embargo, las más difundidas y empleadas son pocas. Las técnicas probabilísticas para proyectar la población podrían clasificarse de acuerdo a tres tipos de aproximaciones para el cálculo de proyecciones probabilísticas: i) la bayesiana (Raftery *et al.*, 2014), ii) utilizando métodos de simulación estocástica (García-Guerrero, 2014), y iii) utilizando las técnicas de suavización exponencial y datos funcionales (Hyndman y Booth, 2008; Mendoza Ponce *et al.*, 2019).

## LOS DATOS Y EL CENSO 2020

Los modelos para datos funcionales se aplican a tasas de mortalidad y fecundidad, para ello los datos requeridos son las defunciones por edad simple y género, los nacimientos según edad de la madre y la población por edad simple y género. Es importante destacar que durante la estimación es necesario contar con las cantidades netas tanto de eventos como de población y por ello no es la tasa en sí el dato base, sino que el mismo se construye a partir de los registros de nacimientos, defunciones y las cifras de población por edad. En relación con los datos provenientes del registro de hechos vitales, los mismos son suministrados por la Dirección de Estadísticas e Información de Salud (DEIS), a través de bases de datos que contienen el registro de los hechos individuales que permiten generar la matriz por edades para un periodo de tiempo de aproximadamente 30 años. Al momento de corte se disponen de datos de mortalidad para el periodo (1980-2012) y de fecundidad para el periodo (1980-2011). Por otro lado, no existe una estimación oficial de la población en base a los cuatro últimos Censos, menos aún por edades simples. La idea del presente trabajo, en relación con las cifras de población, es utilizar una única fuente oficial que no requiera ajustes y sea coherente (preferentemente elaborada para todos los años mediante el mismo procedimiento o metodología), por ello, se decide utilizar la población por edades simples presentada por CELADE y estimada en forma conjunta con el INDEC para el periodo 1980-2012. Se decide recortar el periodo base de aplicación del modelo al mencionado periodo, para evaluar la construcción de una proyección de población para el año 2020 mientras los datos de entrada no se alejen de la última ronda censal disponible.

## MODELO PARA DATOS FUNCIONALES

Se definen a continuación los insumos básicos del modelo para datos funcionales (MDF) que se desarrolla en este trabajo:

$B_t(x)$  = Nacimientos en mujeres de edad  $x$  ocurridos durante el año calendario  $t$ .

$D_t(x)$  = Muertes de personas de edad  $x$  ocurridas en el año calendario  $t$ .

$P_t(x)$  = Población de edad  $x$  al primero de enero del año  $t$ .

$E_t(x)$  = Población de edad  $x$  expuesta al riesgo al 30 de junio del año  $t$ .

donde  $x = 0, 1, 2, \dots, p-1, p^+$  y  $t = 1, \dots, n$ . Con  $p^+$  se indica el grupo final abierto (generalmente “80 o más”, “85 o más”, etc.).

La tasa de mortalidad específica por edad en el año calendario  $t$  es  $m_t(x) = D_t(x)/E_t(x)$  y  $f_t(x) = B_t(x)/E_t^M(x)$  es la tasa de fecundidad específica por edad para el año  $t$  (el superíndice  $M$  se utiliza para indicar que se trata de la población de mujeres). Los nacimientos se dividen por sexo usando  $\rho$ , la razón de sexo al nacer.

La migración neta denotada con  $G_t$  se estima utilizando la ecuación de crecimiento:

$$\begin{aligned} G_t(x, x+1) &= P_{t+1}(x+1) - P_t(x) + D_t(x, x+1) \text{ con } x = 0, 1, 2, \dots, p-2 \\ G_t(p-1^+, p^+) &= P_{t+1}(p^+) - P_t(p-1) + D_t(p-1, p^+) \\ G_t(B, 0) &= P_{t+1}(0) - B_t + D_t(B, 0) \end{aligned}$$

Donde  $D_t(x, x+1)$  se refiere a las muertes en el año calendario  $t$  de personas de edad  $x$  al principio del año  $t$ ;  $D_t(p-1, p^+)$  indica las muertes en el año calendario  $t$  de personas de edad  $p-1$  y mayores al comienzo del año  $t$  y  $D_t(B, 0)$  son las muertes en el año calendario  $t$  de los nacidos durante el año  $t$ , de modo similar se definen para la migración neta,  $G_t(x, x+1)$ ,  $G_t(p-1, p^+)$  y  $G_t(B, 0)$ . Las muertes  $D_t(x, x+1)$  se estiman utilizando un enfoque de tabla de vida estándar. Ver Preston *et al.* (2001) para más detalles.

Con el fin de obtener pronósticos demográficos individualmente para cada componente demográfica, y luego para la población, primero se desarrollan modelos funcionales de series de tiempo sobre las cinco componentes demográficas específicas:  $m_t^M(x)$ ,  $m_t^V(x)$ , (tasas de mortalidad por edad para mujeres y varones),  $f_t(x)$  (tasas de fecundidad por edad de la madre),  $G_t^M(x, x+1)$  y  $G_t^V(x, x+1)$  (migración neta por edad de mujeres y varones). Para modelar cada cantidad se sigue el enfoque de Hyndma y Ullah (2007) y los resultados obtenidos a partir de los cinco modelos se emplean luego en la simulación de la población futura.

Se denota con  $y_t^*(x)$  la cantidad a ser modelada, tasas de mortalidad, fecundidad o números de migración neta para la edad  $x$  en el año  $t$ . Si bien es posible plantear una transformación general de Box y Cox (1964) sobre  $y_t^*$ , que permite modelar una tasa cuya variación aumenta con el valor de  $x$ , es decir, cuando la variabilidad de las tasas es mayor a medida que las tasas son mayores, en la mayoría de las aplicaciones se implementa directamente la transformación logaritmo.

Se supone el siguiente modelo para las observaciones transformadas  $y_t(x)$ :

$$y_t(x) = s_t(x) + \sigma_t(x)\varepsilon_{t,x} \quad (1)$$

$$s_t(x) = \mu(x) + \sum_{k=1}^K \beta_{t,k}\phi_k(x) + e_t(x), \quad (2)$$

donde  $s_t(x)$  es una función suave subyacente de  $x$ ,  $\varepsilon_{t,x}$  son variables aleatorias, independientes e idénticamente distribuidas y la definición de  $\sigma_t(x)$  permite a la variancia cambiar con la edad y con el tiempo. Esto significa que las observaciones transformadas son la suma de la cantidad a modelar,  $s_t(x)$ , una función suave de la edad y un error (Ecuación 1). La ecuación sobre  $s_t(x)$  describe la dinámica a través del tiempo, en esta ecuación  $\mu(x)$ , es la media de  $s_t(x)$  a través de los años,  $\{\varphi_k(x)\}$  es un conjunto de  $K$  funciones base ortogonales (Ecuación 2) calculadas utilizando una descomposición en componentes principales funcionales de la matriz  $[\hat{s}_t(x) - \hat{\mu}(x)]$  y  $e_t(x)$  es el error del modelo (el cual se supone no correlacionado serialmente). La dinámica del proceso está controlada por los coeficientes  $\{\beta_{t,k}\}$ , los cuales tienen un comportamiento independiente uno de otro (garantizado por la utilización del método de componentes principales). Existen tres fuentes de variación en el modelo: representa la variación aleatoria con respecto a la distribución relevante, para los nacimientos, muertes y migrantes (Poisson o Normal);  $e_t(x)$  representa el residuo que surge al modelar  $s_t(x)$  utilizando un conjunto de funciones bases y además existe una aleatoriedad inherente al modelo de series de tiempo para cada  $\{\beta_{t,k}\}$  que ejerce los cambios en la dinámica de la curva suave  $\{s_t(x)\}$ . Es importante destacar que es posible implementar este enfoque para edades simples y también para grupos quinqueniales.

El modelo general, que se aplica a cada una de las componentes demográficas, se estima a través de las etapas que se enumeran a continuación:

1. Se estiman las funciones suaves  $s_t(x)$  a través de regresión no paramétrica sobre  $y_t(x)$  para cada año  $t$ , (ecuación 1).
2. A  $\hat{\mu}(x)$  se la define como la media de  $\hat{s}_t(x)$  a través de los años;
3. Los coeficientes  $\beta_{t,k}$  y las bases  $\varphi_k(x)$ , con  $k = 1, \dots, K$  se estiman aplicando análisis de componentes principales funcionales sobre la matriz  $[\hat{s}_t(x) - \hat{\mu}(x)]$ ; (ecuación 2).
4. Se ajusta un modelo de series de tiempo a  $\hat{\beta}_{t,k}$  donde  $k = 1, \dots, K$ . Para ello es posible utilizar un modelo ARIMA

(Box y Jenkins, 1976) o modelos de espacio de estado de innovaciones (Hyndman *et al.*, 2008).

Aunque el valor de  $K$  debe ser especificado, Hyndman y Ullah (2007) sostienen que el método es insensible al valor elegido siempre y cuando sea lo suficientemente grande. Esto significa, que el costo al elegir  $K$  grande es pequeño (más allá del tiempo computacional), mientras que seleccionar un  $K$  pequeño puede producir menor exactitud en los pronósticos. Hyndman y Booth (2008) utilizan  $K = 6$  para todos los componentes demográficos; esta cantidad parece ser mayor a la que cualquier componente demográfica requiere.

## PRONÓSTICOS COHERENTES

En demografía se denomina pronósticos “coherentes” a aquellos que para hombres y mujeres (u otros subgrupos) difieren siempre en la misma proporción través del tiempo, esencialmente, lo que la metodología impone es que la diferencia entre los pronósticos de ambos grupos sea estacionaria a través del tiempo. Hyndman *et al.* (2013) proponen un método para el pronóstico coherente para las tasas de mortalidad de dos o más subpoblaciones, llamado también método de pronóstico funcional cociente-producto, el mismo modela y pronostica la media geométrica de las tasas de las subpoblaciones y el cociente de las tasas de las subpoblaciones para producir pronósticos de las tasas que no divergen en el tiempo. Además, Hyndman *et al.* (2013) proponen una definición precisa de los pronósticos coherentes, sostienen que los mismos se obtienen cuando los pronósticos de la razón de tasas específicas por edad para dos subpoblaciones cualesquiera convergen a un conjunto de constantes apropiadas. Inicialmente se plantea el problema de pronosticar las tasas específicas de mortalidad por edad para varones y mujeres, de todos modos, el método puede aplicarse a cualquier número de subpoblaciones e incluso a otras componentes demográficas, como, por ejemplo, la migración. Para aplicar el método del cociente-producto para varones y mujeres se definen las raíces cuadradas de los productos y cocientes entre las tasas suavizadas de cada sexo:

$$p_t(x) = \sqrt{s_{t,M}(x)s_{t,V}(x)},$$

$$c_t(x) = \sqrt{s_{t,M}(x)/s_{t,V}(x)}.$$

Las cantidades se modelan en lugar de las tasas de mortalidad específicas por edad observadas. La ventaja de este enfoque es que el cociente y el producto se comportaran de forma aproximadamente independiente uno de otro siempre que tengan variancias similares. En la escala logarítmica estas cantidades son sumas y diferencias aproximadamente no correlacionadas (si  $X$  e  $Y$  son dos variables aleatorias, luego  $\text{Corr}(X - Y, X + Y) = \text{Var}(X) - \text{Var}(Y)$ . Luego si  $X$  e  $Y$  tiene igual variancia,  $X - Y$ ,  $X + Y$  y son no correlacionadas). Si existen diferencias sustanciales en las variancias de las subpoblaciones luego el producto y el cociente no serán no correlacionados. Este hecho no produce sesgo en los pronósticos, pero los hace menos eficientes. En cualquier caso, el método, continuaría siendo mejor que tratar las poblaciones de forma independiente ya que aún de este modo se estarían imponiendo restricciones de coherencia. Es importante tener en cuenta que, el logaritmo de la raíz del producto representa el promedio entre el logaritmo de la mortalidad femenina y masculina, mientras que el logaritmo de la raíz del cociente es la mitad de la diferencia entre los logaritmos de hombres y mujeres. Para modelar  $p_t(x)$  y  $c_t(x)$  Hyndman y Ullah (2007) utilizan modelos funcionales de series de tiempo:

$$\log[p_t(x)] = \mu_p(x) + \sum_{k=1}^K \beta_{t,k} \phi_k(x) + e_t(x),$$

$$\log[c_t(x)] = \mu_r(x) + \sum_{l=1}^L \gamma_{t,l} \psi_l(x) + w_t(x),$$

donde las funciones  $\{\phi_k(x)\}$  y  $\{\psi_l(x)\}$  son las componentes principales obtenidas de la descomposición de  $\{\log(p_t(x))\}$  y  $\{\log(c_t(x))\}$  respectivamente,  $\beta_{t,k}$  y  $\gamma_{t,l}$  son los correspondientes *scores* de las componentes principales. La función  $\mu_p(x)$  es la media del conjunto de curvas  $\{\log(p_t(x))\}$  y  $\mu_r(x)$  es la media de  $\{\log(c_t(x))\}$ . Los términos de error, dados por  $e_t(x)$  y  $w_t(x)$ , tienen media cero y son no correlacionados.

Los pronósticos de cada componente demográfica se obtienen a partir de los pronósticos de los coeficientes  $\beta_{t,1}, \dots, \beta_{t,K}$  y  $\gamma_{t,1}, \dots, \gamma_{t,K}$ , realizados en forma independiente unos de otros. No hay necesidad de considerar modelos vectoriales dado que los coeficientes son no correlacionados por construcción (Hyndman y Ullah, 2007). Lo mismo sucede con los, que son aproximadamente no correlacionados entre si

debido al uso de productos y cocientes. Los coeficientes del modelo para la raíz del producto,  $\{\beta_{t,1}, \dots, \beta_{t,K}\}$  se pronostican utilizando modelos ARIMA. Al ajustar el modelo ARIMA, se utiliza el algoritmo automático de selección desarrollado por Hyndman *et al.* (2008) para seleccionar el orden apropiado tanto para la parte promedio móvil como para la parte autorregresiva. Los coeficientes del modelo para la raíz del cociente  $\{\gamma_{t,1}, \dots, \gamma_{t,L}\}$  se pronostican utilizando modelos estacionarios ARMA(p, q) (Box *et al.*, 2008) o procesos ARFIMA(p,d,q) (Granger y Joyeux (1980), Hosking (1981)), donde el orden de integración es fraccionario, también conocidos como procesos de memoria larga.

## SIMULACIONES DE POBLACIÓN

Se simulan poblaciones a través de la ecuación compensadora adaptada para permitir un error aleatorio observacional. La población inicial es la observada al primero de enero en el año  $n + 1$ . Para elaborar los pronósticos de población es necesario simular un gran número de trayectorias de cada componente demográfica;  $m_t^M(x)$   $m_t^V(x)$   $f_t(x)$ .  $G_t^M(x, x + 1)$  y  $G_t^V(x, x + 1)$ . En forma secuencial y por simulación de Monte Carlo se obtienen  $N$  conjuntos de valores futuros para las tasas de mortalidad, fecundidad y en caso de que sea posible para los saldos migratorios. A partir de las componentes generadas, empleando el algoritmo que se detalla a continuación es posible generar  $N$  poblaciones simuladas.

## RESULTADOS

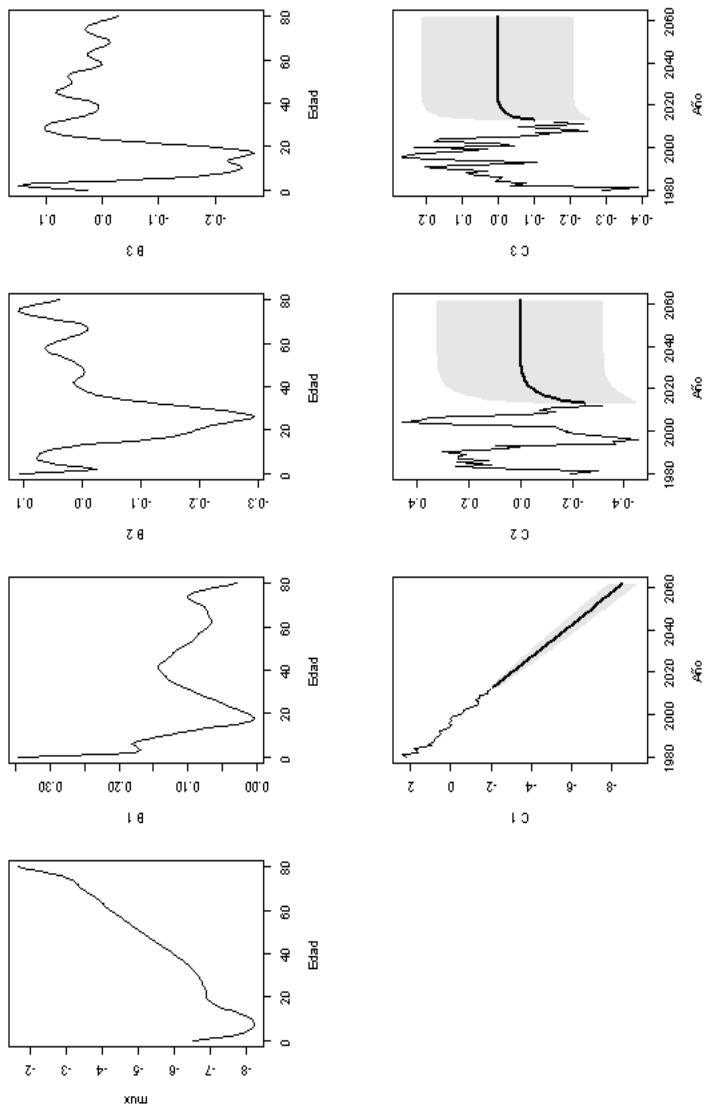
Los datos suavizados se descomponen mediante el uso de análisis de componentes principales funcionales. En relación con el número de bases funcionales se establece  $K = 6$ , por lo que se obtienen seis bases y sus correspondientes coeficientes asociados (es importante destacar que gráficamente se presentan siempre los primeros tres pares base-coeficiente por motivos de practicidad e interpretabilidad). Hyndman sostiene que es difícil hallar una explicación para los coeficientes y bases funcionales más allá del segundo o tercer par base-coeficiente. Las Figuras 1, 2 y 3 representan los resultados para el total de la población, los varones y las mujeres, respectivamente. La Figura 1 presenta la media, las bases, los coeficientes funcionales estimados y el pronóstico de los coeficientes con sus intervalos de pronóstico de 80 por ciento de confianza para el total, mientras que las Figuras 2 y 3 contienen los mismos resultados para varones y mujeres res-

pectivamente. Estas Figuras deben interpretarse de la siguiente manera: el primer recuadro de la fila superior representa el comportamiento promedio de la mortalidad a través de las edades. A partir de la segunda columna, la fila superior debe interpretarse en forma conjunta con el correspondiente recuadro de la fila inferior, es decir, cada base debe interpretarse en correspondencia con su coeficiente asociado. Por ejemplo, en el caso de la mortalidad total, el coeficiente muestra un decrecimiento de la mortalidad a través del tiempo, pero para interpretarlo correctamente este comportamiento debe ponerse en correspondencia con su base asociada (recuadro de la fila superior), ya que la misma indica en qué edades se manifiesta este descenso y con qué intensidad. En este caso la base funcional indica que el decrecimiento se da especialmente en los primeros años de vida (mortalidad infantil y primeros años) y en menor medida para los mayores de 40 años. Sin embargo, ese decaimiento no es tan notorio para personas de alrededor de 20 años, este hecho se refleja en la base ya que la misma adquiere un valor cercano a cero para la mencionada edad. Como se dijo antes, generalmente, no es posible hallar una explicación lógica con un correlato coherente con los hechos demográficos, más allá del segundo o tercer par de coeficientes y bases. Por otro lado, la media estimada ( $\hat{\mu}$ ) representa el perfil promedio de la mortalidad a lo largo de la vida (obtenido como el promedio de las funciones a través de los años).

En primer lugar, se analizan los resultados obtenidos para ambos sexos, principalmente se destaca que en el comportamiento de la mortalidad total se solapan las tendencias de ambos sexos. Las seis primeras bases explican 91.6, 3.8, 1.6, 0.7, 0.6 y 0.4 por ciento de la variación en los datos, dejando 1.3 por ciento sin explicar. La primera base presenta una estructura similar a la de los varones, pero más suave, mientras que la segunda base presenta un comportamiento similar a la misma base en mujeres. Por último, la tercera base tiene un comportamiento similar a la tercera base estimada para los varones. En otras palabras, el comportamiento de la mortalidad total está dominado en la primer y tercer base por el comportamiento que presenta la mortalidad en los varones, mientras que la segunda base estaría dominada por la pauta que sigue la mortalidad de las mujeres.

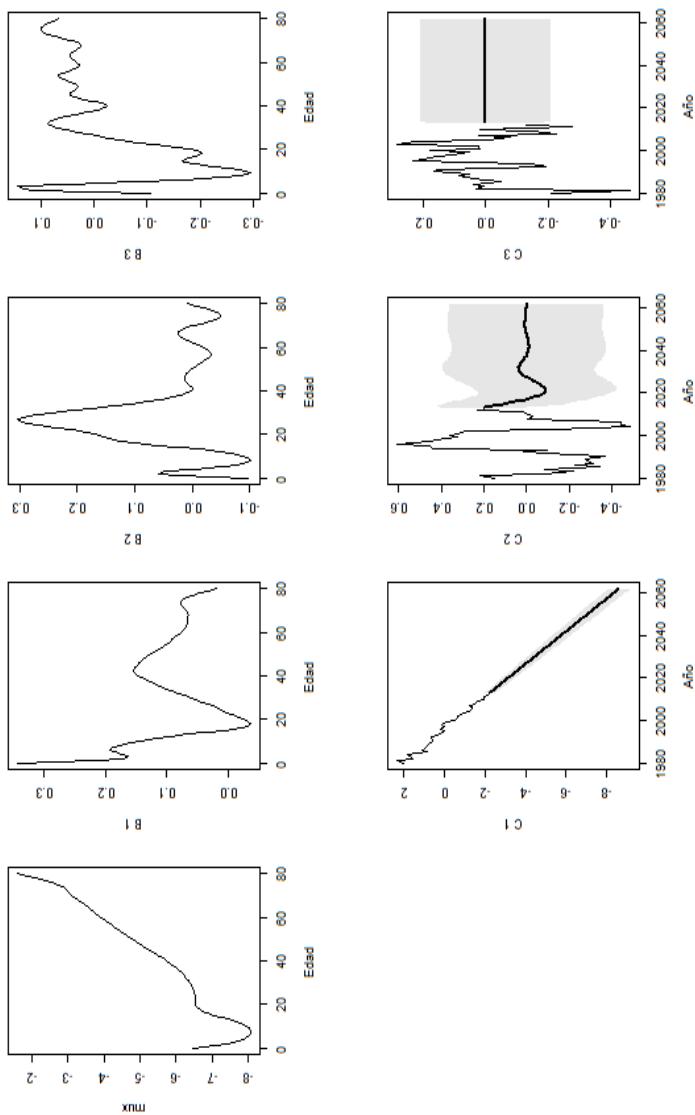
Resulta más interesante analizar los resultados para cada sexo, en el caso de los varones, las seis primeras bases explican 90.3, 5, 1.4, 0.8, 0.7 y 0.5 por ciento de la variación presente en los datos, dejando 1.3 por ciento sin explicar.

Figura 1: Funciones base, coeficientes asociados, pronósticos e intervalos de pronóstico del 80 por ciento de confianza. Modelo de datos funcionales para la mortalidad de ambos sexos



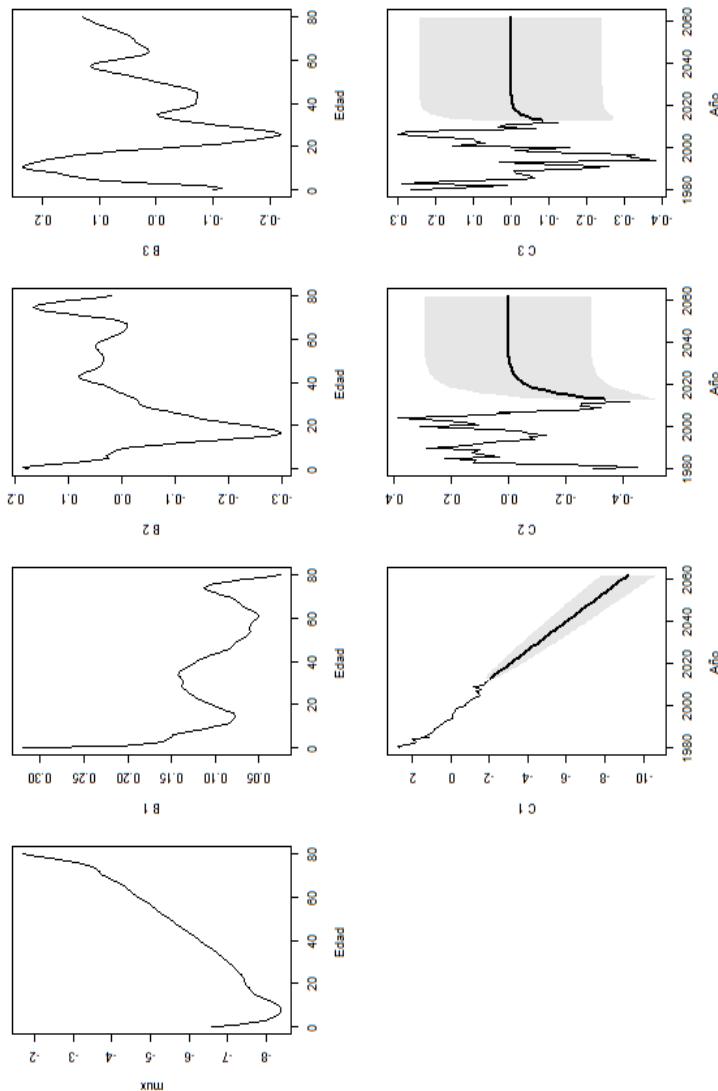
Fuente: elaboración propia.

Figura 2: Funciones base, coeficientes asociados, pronósticos e intervalos de pronóstico del 80 por ciento de confianza. Modelo de datos funcionales para la mortalidad de Varones



Fuente: elaboración propia.

Figura 3: Funciones base, coeficientes asociados, pronósticos e intervalos de confianza. Modelo de datos funcionales para la mortalidad de Mujeres



Fuente: elaboración propia.

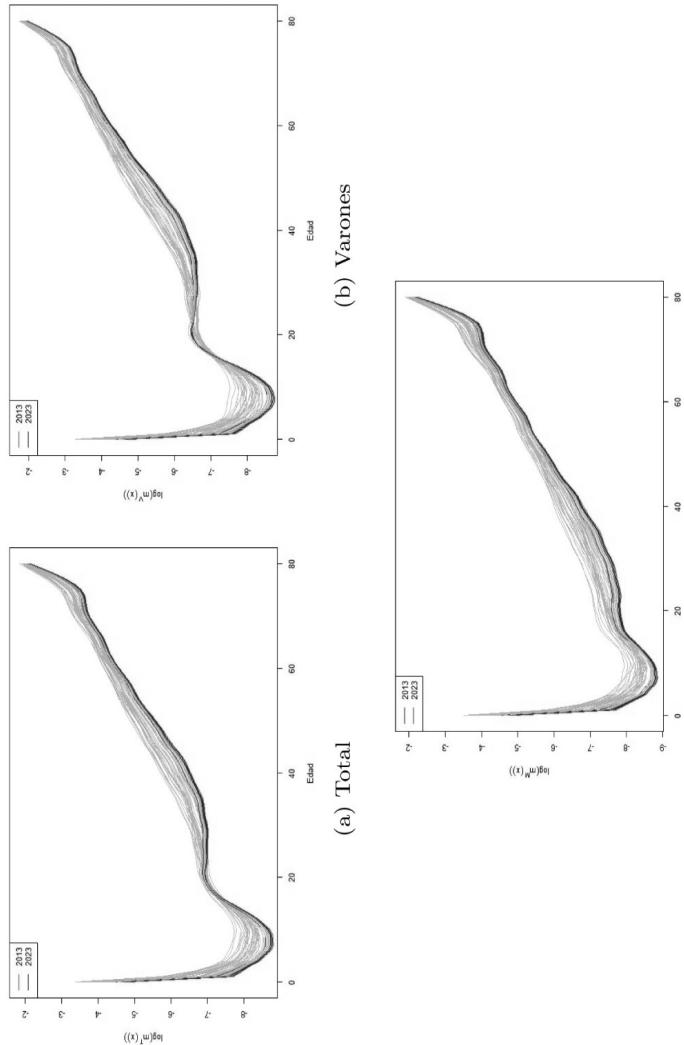
El coeficiente correspondiente a la primera base representa una mortalidad que desciende a través del tiempo, con un pico en al año 1982 que podría atribuirse al conflicto bélico por la soberanía de las Islas Malvinas en abril de ese año. La primera base funcional estimada indica de qué modo este descenso presente en el coeficiente se manifiesta para las distintas edades; en otras palabras, indica que grupos etarios sufren en mayor o en menor medida el descenso a través del tiempo, es decir, el pico en las edades iniciales representa el descenso en la mortalidad infantil, seguido de valores altos para la niñez y las edades entre 40 y 50 años, que presentan también descensos, pero con menor magnitud. El descenso para la franja etaria 40-50 está vinculado a la caída en la mortalidad a nivel mundial atribuible en su mayoría a la merma en las muertes por afecciones cardíacas u otras patologías debido al avance de la medicina. El coeficiente asociado a la segunda base presenta períodos de subas moderadas alrededor de 1980 y 2010 y una suba marcada a mediados de la década de los 90, bajas menores se presentan en los restantes períodos de tiempo. La base que corresponde a este coeficiente indica que este comportamiento se manifiesta principalmente en la franja de entre 18 y 30 años aproximadamente. Por último, el tercer par base-coeficiente podría estar asociado a un diferencial en la mortalidad de varones de 10 a 20 años con respecto al resto de las edades y, su base asociada indicaría que la mortalidad para dicho grupo con relación a las demás edades fue baja en 1980, luego presentó una suba con fluctuaciones hasta alcanzar su valor máximo alrededor del año 1990, para luego comenzar a descender y alcanzar valores mínimos (relativos o locales) entre 2005 y 2010. En síntesis, para los hombres, las edades que dominan el cambio en la mortalidad resultan, la franja de edades cercanas a cero, el grupo de 40 a 50 años, el grupo de 18 a 30 años y el grupo formado por las edades en torno a los 15 años. La forma en la que ejercen el cambio está dada por las bases, un descenso para los primeros tres grupos y fluctuaciones a lo largo del periodo para los restantes.

Si se analizan los resultados relativos a las mujeres, las bases explican 92.2, 2.2, 1.8, 0.9, 0.6 y 0.5 por ciento de la variación en los datos respectivamente, dejando 1.8 por ciento sin explicar. Se destaca, que la franja etaria de cero a cinco años presenta el valor más alto observado para la primera base estimada, esto indica que se trata de las edades que determinan en mayor medida el decrecimiento que presenta el coeficiente asociado a la primera base. Interpretado de otro modo, es el grupo que presentó el mayor descenso en la mortalidad durante el pe-

riodo 1980-2010. En segundo lugar, se encuentran los grupos de mujeres de alrededor de 30 y 75 años, que también presentan un descenso en el periodo en estudio, pero, con una menor magnitud o intensidad. Luego, el coeficiente asociado a la primera base claramente representa el descenso en la mortalidad de mujeres sucedido entre 1980 y 2010. Además, se destaca un leve aumento en la fluctuación de dicho coeficiente en periodo 2005-2010 y si bien el fenómeno podría ser producto del azar, es posible plantear algún tipo de hipótesis vinculadas a sus causas. El segundo par base-coeficiente se referiría a un comportamiento diferencial entre las mujeres de 15 a 20 años en relación con el resto de las edades. Para este par, el coeficiente presenta una gran fluctuación a través de los años. Evaluando base y coeficiente en forma conjunta se puede interpretar que la mortalidad de mujeres de 15 a 20 años es la que menos presenta fluctuación. Por último, la tercera base, que es generalmente a partir de la cual resulta más difícil hallar una interpretación, podría estar representando la diferencia en la mortalidad para edades previas versus edades posteriores a los veinte años y su base asociada indicaría un comportamiento fluctuante durante el periodo en estudio, con valores altos al inicio del periodo y sobre el 2005, y valores bajos a principios de los 90 y a finales del periodo. Sin embargo, es importante tener mayor cautela al tratar de definir el comportamiento que modela el tercer par coeficiente-base. En síntesis, los grupos que dominan los cambios en la mortalidad de mujeres desde 1980 a 2010 son las niñas de 0 a 5 años, mujeres de alrededor de 30 años y las adultas mayores de alrededor de 75 años, en segundo lugar, las mujeres de alrededor de 20 años y por ultimo las levemente mayores y menores a los 20 años. La forma en que estos grupos dominan los cambios está dada por su base asociada, descenso en mayor o menor medida para los tres primeros grupos mencionados y fluctuaciones a lo largo del tiempo para los restantes.

Una vez estimado el modelo se pronostican los coeficientes utilizando modelos  $\hat{\beta}$  utilizando modelos ARIMA, los pronósticos de los coeficientes pueden verse en las mismas Figuras que presentan la descomposición, como una línea que se extiende más allá del año 2012, rodeada del sombreado color gris que representa el intervalo de pronóstico de 80 por ciento de confianza. Los coeficientes  $\hat{\beta}$  pronosticados permiten construir las tasas y obtener las variancias de pronóstico a fin de contar con intervalos de confianza para las tasas. En la Figura 4 se presentan las tasas observadas, en gris claro y los pronósticos, en gris oscuro, para el periodo 2013-2023. Se destaca un comportamiento diferente entre sexos principalmente alrededor de los 20 años.

Figura 4: Pronósticos de la mortalidad a 10 años. Total, varones y mujeres. Argentina (2013-2023)



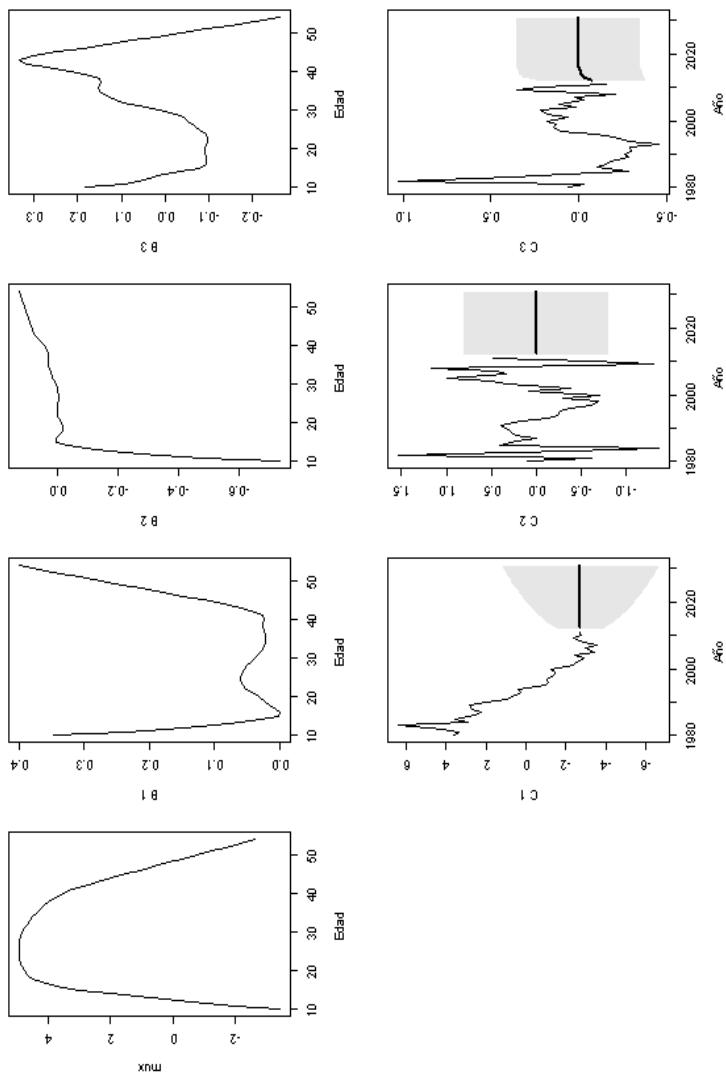
Fuente: elaboración propia.

En los varones no se detecta para la mencionada edad el mismo descenso que se observa para la mayoría de las edades, es decir, que para los varones de entre 15 y 22 años la tasa de mortalidad pronosticada se mantiene en el mismo nivel aun cuando por lo general se presenta un descenso de la mortalidad. En el caso de las mujeres el descenso más marcado de presenta entre los 15 y 45 años y el menor descenso entre los 10 y 15 años y alrededor de los 60 años.

## FECUNDIDAD

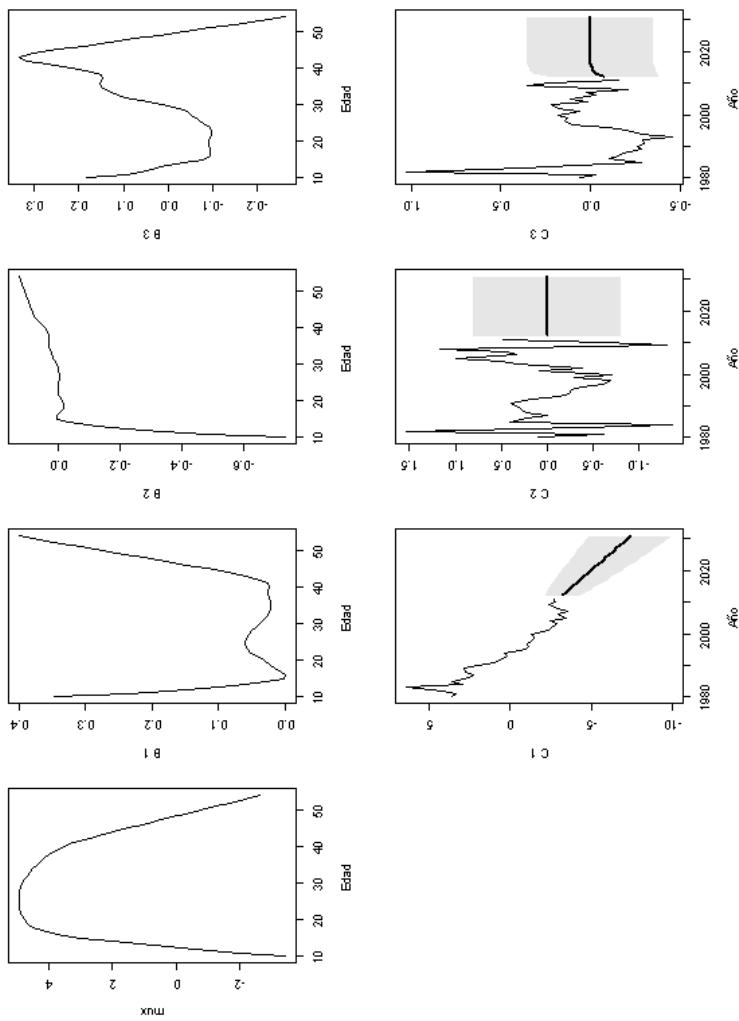
Se aplica el enfoque de datos funcionales para el análisis de las tasas de fecundidad según edad de la madre en Argentina durante el periodo 1980 a 2011. Estos últimos datos se convierten en datos funcionales estimando una curva suave para las tasas a través del tiempo. Las curvas suavizadas se estiman utilizando *B-splines* con restricción de concavidad mientras que  $\sigma_t^2$  se estima a partir de  $\{f_t(x)\}^{-1} E_t(x)^{-1}$ . Como se mencionó Hyndman y Ullah (2007) imputan, previo al suavizado, los datos para las edades más extremas (le asigna cinco nacimientos por mil mujeres a edades menores de 15 y mayores de 49), sin embargo, en la presente aplicación se decide respetar los datos oficiales. De este modo los resultados posiblemente reflejen no sólo tendencias demográficas sino también la tendencia en errores o subregistro, permitiendo un análisis de este aspecto. El ajuste del modelo para datos funcionales aplicado a la fecundidad se muestra en las Figuras 5 y 6. Si bien bases y coeficientes de ambas Figuras son idénticos, las mismas se diferencian en los pronósticos de los coeficientes beta, es decir la etapa cuatro del modelo, dado que, en este punto, se plantean dos hipótesis relativas a la fecundidad y su comportamiento en el futuro (se incluyen dos supuestos: una primera hipótesis que sostiene que la fecundidad continuará descendiendo y una segunda hipótesis que plantea que la fecundidad ya ha alcanzado su máximo descenso, por ello se seleccionan dos modelos: uno con pendiente y otro sin pendiente). Las seis primeras bases estimadas explican 92, 5.9, 0.9, 0.5, 0.3 por ciento y 0.1 por ciento de la variación en los datos respectivamente, dejando 0.3 por ciento sin explicar. Sin embargo, resulta complejo encontrar un comportamiento lógico más allá de la segunda componente principal.

Figura 5: Funciones base, coeficientes asociados, pronósticos e intervalos del pronóstico del 80 por ciento de confianza. Modelo para datos funcionales de fecundidad - sin pendiente



Fuente: elaboración propia.

Figura 6: Funciones base, coeficientes asociados, pronósticos e intervalos de confianza. Modelo para datos funcionales de fecundidad - con pendiente



Fuente: elaboración propia.

Cada base debe interpretarse en correspondencia con su coeficiente asociado. Por ejemplo, el primer coeficiente presenta un decrecimiento en la fecundidad a través del tiempo que, al ponerlo en correspondencia con su base asociada (fila superior), indica que el comportamiento de ese decrecimiento se da particularmente en mujeres de aproximadamente diez años, para las mayores de 40 años y en forma más leve en las mujeres de alrededor de 25 años. Como el coeficiente marca un descenso sostenido es posible pensar que lo que principalmente se observa es un descenso en los registros con error (reflejados en la base mediante valores altos en los extremos) mientras que el pico que se destaca en el centro es realmente el descenso en las tasas de fecundidad que se da en las edades intermedias, el grupo de mujeres de entre 20 y 30 años.

Además, en los coeficientes de la primera base se destaca un valor alto en el año 1983, este valor posiblemente no se deba a errores de registro sino a un cambio social asociado con el retorno de un régimen de gobierno democrático en la República Argentina luego de varios años de dictadura militar.

En relación con el segundo par coeficiente-base, la misma puede pensarse como la diferencia entre la fecundidad adolescente versus edades avanzadas. La base presenta valores bajos para edades próximas a los diez años, valores relativamente altos para mujeres mayores de 40 años y valores cercanos a cero para las edades restantes, es por ello, que se la puede asociar a un comportamiento diferente entre las edades que presentan los valores bajos y altos. Para dicha base el comportamiento a través del tiempo está determinado por el coeficiente asociado, el cuál presenta valores altos para los años 1982 y 1983, un descenso hacia el año 2000 seguido de un aumento hasta 2010, año en el que se observa una caída abrupta. Sin embargo, es importante remarcar que poner en correspondencia las bases y coeficientes con comportamientos demográficos derivados de la teoría resulta particularmente complejo en el caso de la fecundidad, más allá de la primera o segunda base. Finalmente, el tercer par base-coeficiente podría vincularse con el traslado de la edad de la maternidad hacia edades más avanzadas, de acuerdo con lo que indica la base la cuál presenta valores altos alrededor de los 40 años y el coeficiente que presenta una alta fluctuación con valores más bajos al inicio del periodo y más altos hacia el final.

En resumen, la descomposición en bases y coeficientes permite identificar el patrón de descenso y las edades que lo determinan, el diferencial en la fecundidad adolescente versus edades avanzadas y el

corrimiento de la maternidad a edades alrededor de los 40 años y su estabilización en torno a la misma.

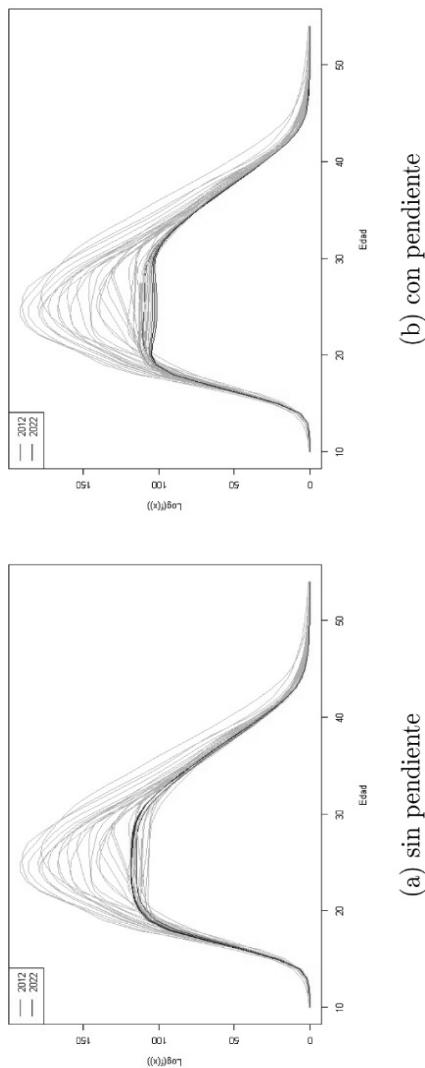
Una vez que se estima el modelo para la fecundidad y se evalúa la bondad de su ajuste, se pronostican los coeficientes  $\hat{\beta}_k$  mediante modelos ARIMA. La selección automática de modelos se realiza utilizando el algoritmo de Hyndman y Khandakar (2008) y en esta etapa se incluyen dos supuestos: una primera hipótesis que sostiene que la fecundidad continuará descendiendo y una segunda hipótesis que plantea que la fecundidad ya ha alcanzado su máximo descenso, por ello se seleccionan dos modelos: uno con pendiente y otro sin pendiente, en correspondencia con los supuestos planteados. A partir de la combinación entre los coeficientes pronosticados y las bases estimadas por el modelo, se generan los pronósticos de las curvas de fecundidad. Una vez estimado el modelo para datos funcionales, los pronósticos de los coeficientes  $\hat{\beta}$  obtenidos bajo las dos hipótesis permiten construir los pronósticos de las tasas y sus intervalos (Figura 7).

## PRONÓSTICOS COHERENTES

Hyndman *et al.* (2013) proponen una metodología alternativa para realizar los pronósticos por género: los pronósticos coherentes. Los mismos tienen como finalidad evitar una divergencia entre los resultados para ambos subgrupos, de este modo los pronósticos para los varones y mujeres difieren siempre en la misma proporción través del tiempo y más aún en los pronósticos se mantiene la misma diferencia que se presenta en los datos observados. Se construyen las funciones  $p_t(x)$  y  $c_t(x)$  de las tasas de mortalidad de varones y mujeres. La función  $p_t(x)$  es la media geométrica de las tasas suavizadas de hombres y mujeres y para la mortalidad esta medida declina a través del tiempo, aunque se detectan leves ascensos en edades menores de cinco años y entre los 15 y 30 años, ver Figura 9 (primer panel). La función  $c_t(x)$  es la raíz cuadrada del cociente entre las tasas suavizadas de cada sexo, en este caso se observa una alta variabilidad en las edades entre los 20 y 40 años, ver Figura 8 (primer panel). Además, entre los 40 y 70 años, aproximadamente, dicha función presenta un descenso a través del tiempo.

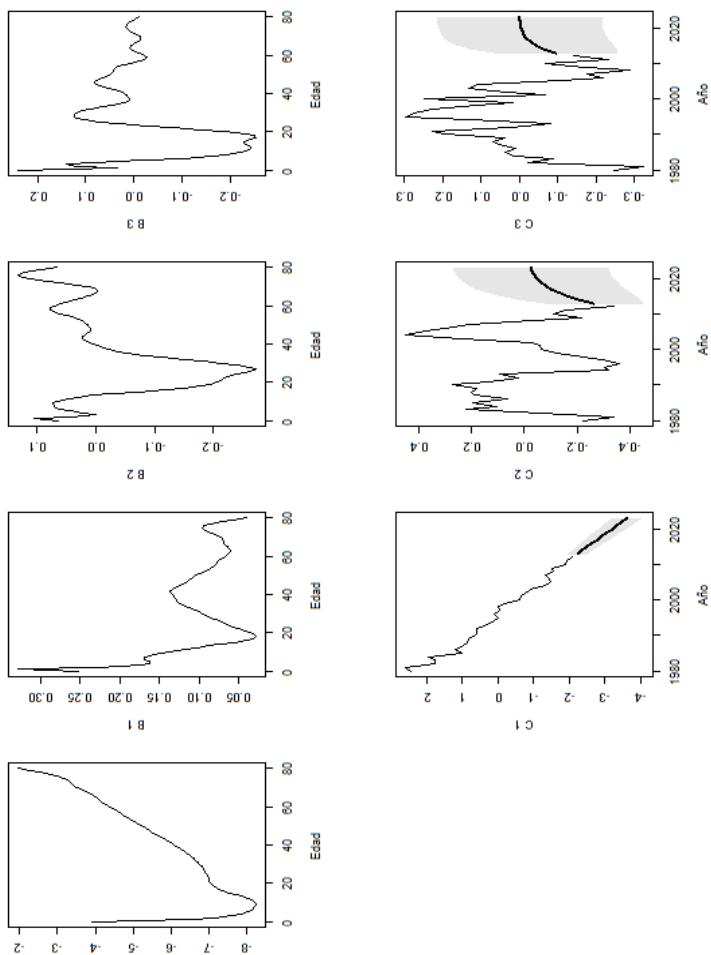
En las Figuras 7 y 8 se presentan las estimaciones obtenidas a partir del modelo; funciones base, con los coeficientes correspondientes y sus pronósticos.

Figura 7: Pronósticos de la fecundidad a 10 años. Hipótesis sin y con pendiente. Argentina (2012-2022)



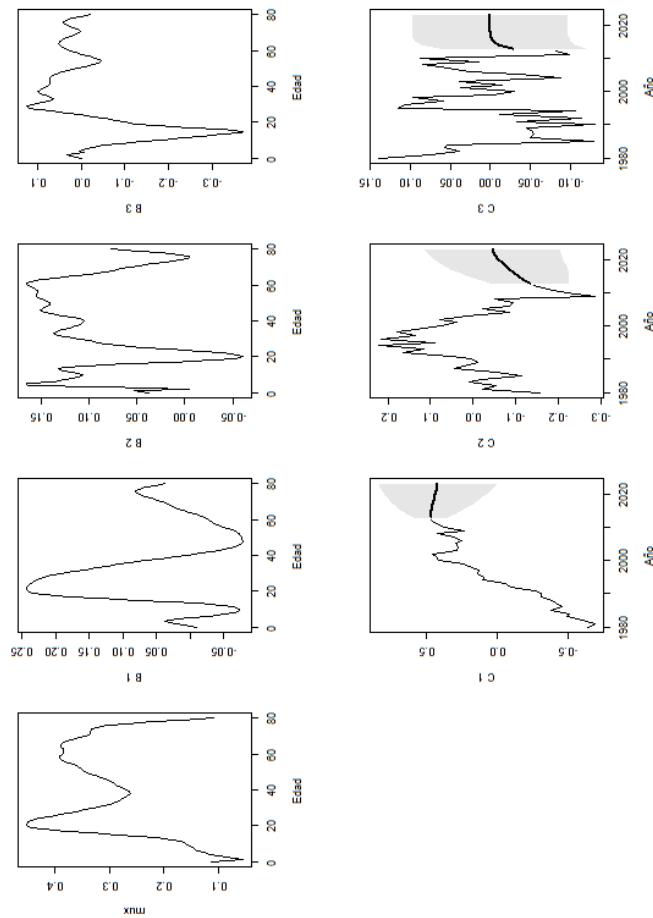
Fuente: elaboración propia.

Figura 8: Funciones base, coeficientes asociados, pronósticos e intervalos de confianza de 80 por ciento de confianza. Modelo para datos funcionales de la Mortalidad -  $p_t(x)$  (Argentina, 1980-2012)



Fuente: elaboración propia.

Figura 9: Funciones base, coeficientes asociados, pronósticos e intervalos de promóstico de 80 por ciento de confianza. Modelo para datos funcionales de la Mortalidad -  $c_t(x)$  (Argentina, 1980-2012)



Fuente: elaboración propia.

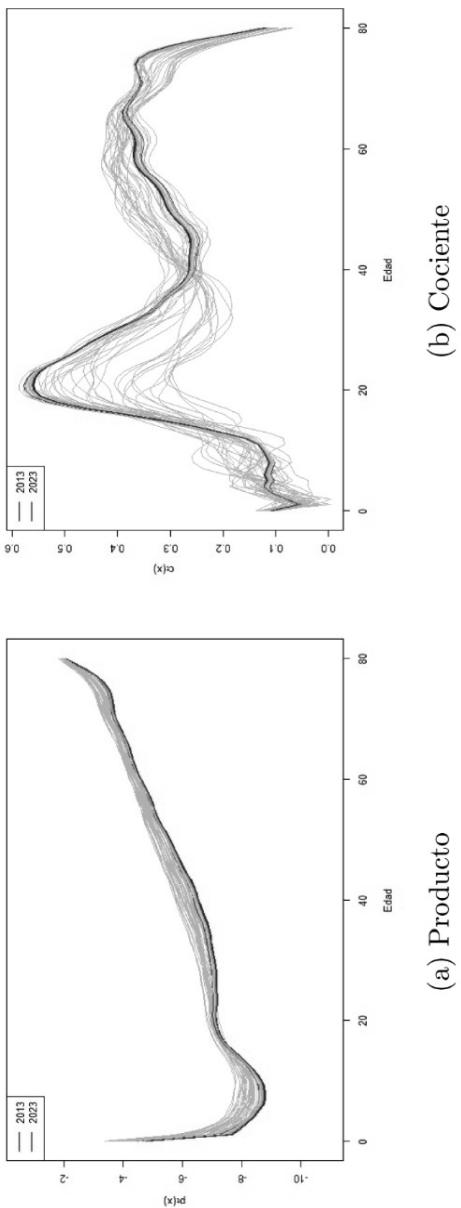
Es importante destacar que el logaritmo de la raíz cuadrada del producto representa el promedio entre el logaritmo de la mortalidad femenina y masculina, mientras que el logaritmo de la raíz cuadrada del cociente es la mitad de la diferencia entre el logaritmo de la mortalidad de hombres y mujeres, es decir, que los resultados de bases y coeficientes se refieren a un comportamiento promedio de la mortalidad para el producto, y a una diferencia entre la mortalidad de ambos géneros para el cociente. En el modelo de  $p_t(x)$ , la primera componente principal muestra un descenso a través del tiempo, que, según el comportamiento de la base correspondiente, resulta más rápido en la infancia y más lenta en las edades cercanas a los 20 años y edades avanzadas. Este término retiene 92.9 por ciento de la variabilidad.

El porcentaje de variación debida a las seis primeras bases es 92.9, 2.8, 1.5, 0.8, 0.6 y 0.4 por ciento, respectivamente. En relación con el modelo estimado para  $c_t(x)$  se observa una tendencia creciente para la primera componente, que se estabiliza hacia el año 2000, y, de acuerdo con la base correspondiente se refiere a edades entre 20 y 40 años. Este par base-coeficiente está asociado a la diferencia en la mortalidad entre géneros y estaría indicando que, en edades entre 20 y 40 años y mayores de 70 años, se incrementa la diferencia entre la mortalidad de hombres y mujeres. Esta componente representa 78.8 por ciento de la variabilidad, y de la segunda a la sexta base explican 8.4, 3.2, 2.6, 1.5 y 1.2 por ciento, respectivamente. En ambos modelos resulta complejo hallar un correlato con comportamientos demográficos teóricos. La última etapa del modelo consiste en el pronóstico de los coeficientes  $\beta_{n,k}$  y  $\gamma_{n,k}$ , los mismos se realizan por modelos ARIMA en el caso de  $p_t(x)$ , y por modelos ARFIMA en el caso de  $p_t^*(x)$ . Luego, a partir de los pronósticos de los coeficientes (Figura 10) es posible calcular los pronósticos de las tasas para ambos sexos.

## PROYECCIONES DE POBLACIÓN

A partir de los pronósticos de fecundidad (basados en las dos hipótesis de fecundidad) y de mortalidad (independientes y coherentes) se elaboran pronósticos estocásticos de la población para hombres y mujeres. En relación con la componente migratoria, dado que el análisis de esta componente arrojó resultados atípicos, se supone que las cantidades de migrantes y emigrantes son iguales generando así un saldo migratorio nulo.

Figura 10: Pronósticos del producto y cociente de Mortalidad (Argentina, 2013-2023)



Fuente: elaboración propia.

Teniendo en cuenta que los datos de fecundidad y mortalidad se modelan hasta 2011 y 2012, respectivamente, la simulación se basa en datos de las dos componentes demográficas para el periodo 1980-2011 y se toma como población base la del año 2012.

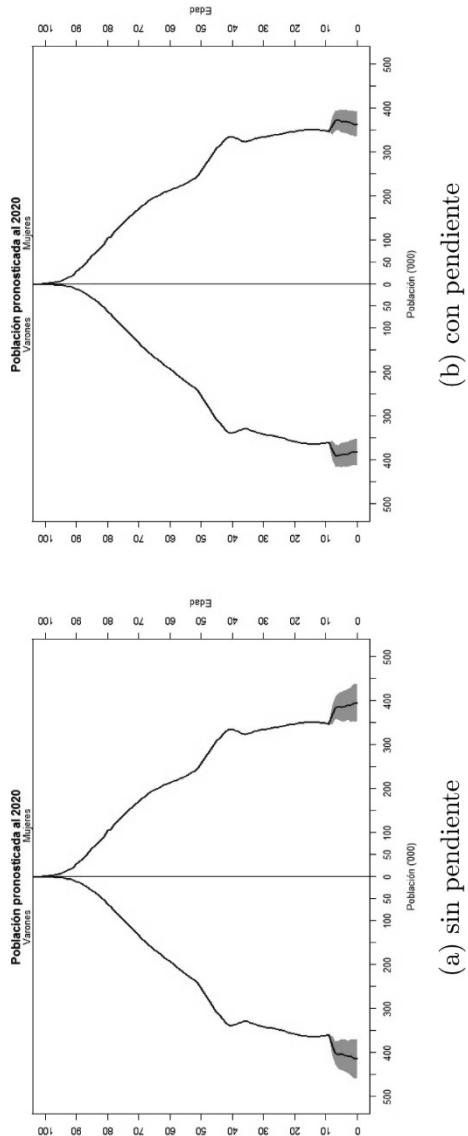
Con base en  $N = 1000$  trayectorias simuladas para cada componente se generan los pronósticos de población y sus intervalos de confianza. Los pronósticos para hombres y mujeres se obtienen para un periodo de diez años. En la Figura 11 se muestran las pirámides de población del año 2020 realizados a partir de los pronósticos coherentes de mortalidad y de acuerdo con los dos supuestos de fecundidad, sin y con pendiente respectivamente. En la Figura 12 se muestran las pirámides de población del año 2020 realizadas a partir de los pronósticos independientes de mortalidad y de acuerdo con los dos supuestos de fecundidad, al igual que en el caso anterior. Los intervalos de confianza de 80 por ciento se obtienen en base a 1000 simulaciones. Se elige el año 2020 dado que el mismo está establecido oficialmente para realizar el próximo Censo Nacional de Población y Vivienda en la Argentina. En general, los intervalos amplios en las edades inferiores a los diez años reflejan la mayor incertidumbre asociada al pronóstico de la fecundidad. Por otro lado, la amplitud de dichos intervalos puede estar asociada a las mismas causas que generan problemas en la estimación de la migración, ya que en ambos casos las cifras obtenidas se obtienen a través de fórmulas recursivas basadas en datos de fecundidad y mortalidad. Es importante destacar que en (Hyndman *et al.*, 2008), donde se presenta el pronóstico de población de Australia para el año 2020, también se observa una amplitud marcada en los intervalos de pronóstico correspondientes a las edades iniciales y que comienza a descender alrededor de los diez años.

Los resultados oficiales brindados por el INDEC y los que se obtienen en presente trabajo a través de los modelos para datos funcionales se resumen en las siguientes cifras:

La población total para el año 2020 proyectada por el INDEC basada en datos del Censo Nacional de población y Vivienda 2010, es de 45,376,763 de habitantes, 22,273,132 varones y 23,10,631 mujeres.

La población pronosticada (proyectada probabilísticamente) imponiendo coherencia entre géneros es de 46,479,523 habitantes (46,016,405-46,983,910), 22,876,092 varones (22,641,598-23,135,384) y 23,603,430 mujeres (23,376,855-23,850,370), para el supuesto de fecundidad estable y 46,128,313 habitantes (45,813,878-46,468,696), 22,696,071 varones (22,533,200-22,866,642) y 23,432,242 mujeres (23,279,489-23,596,665) para el supuesto de fecundidad descendente.

Figura 11: Pirámides de población año 2020 (Argentina) – Pronósticos coherentes

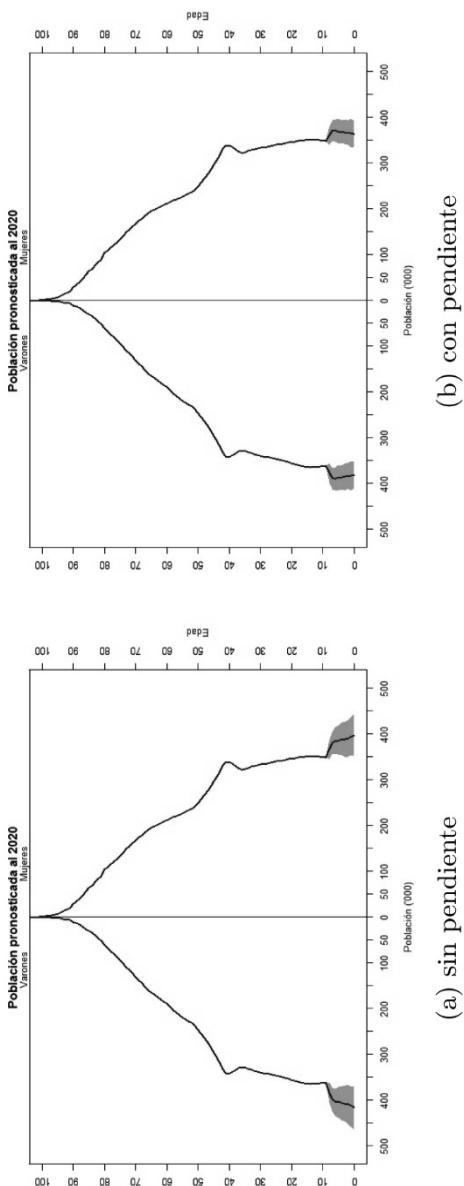


(a) sin pendiente

(b) con pendiente

Fuente: elaboración propia.

Figura 12: Pirámides de población año 2020 (Argentina) – Pronósticos Independientes



Fuente: elaboración propia.

(a) sin pendiente

(b) con pendiente

La población pronosticada con pronósticos independientes entre géneros es de 46,113,514 habitantes (45,633,580-46,619,389), 22,690,469 varones (22,436,905-22,951,933) y 23,423,045 mujeres (23,188,880, 23,668,318) para el supuesto de fecundidad estable y 46,104,694 habitantes (46,099,118-46,586,127), 22,685,361 varones (22,445,377-22,932,227) y 23,419,333 mujeres (23,192,940, 23,658,456) para el supuesto de fecundidad descendente.

A fin de hacer un resumen de las cifras que facilite la comparación se calculan las diferencias, para el total (ambos sexos), varones y mujeres de las cuatro cifras estimadas versus las oficiales, Tabla 1. En la comparación, en todos los casos las cifras estimadas se encuentran por encima de las oficiales con una diferencia promedio de 2.6 por ciento. Se destaca que la mayor diferencia con la proyección oficial se observa para varones, consistentemente, a lo largo de las cuatro estimaciones.

Tabla 1: Diferencias relativas porcentuales. Las estimaciones oficiales del INDEC *versus* de las cuatro estimaciones de elaboración propia – Argentina, 2020

Fecundidad Género/Tipo	Estable		Descendente	
	Coherente	Independiente	Coherente	Independiente
Total	2.43	1.59	1.66	1.58
Varones	2.71	1.82	1.90	1.82
Mujeres	2.16	1.35	1.42	1.35

Fuente: elaboración propia.

Las cifras que se obtienen a través de pronósticos coherentes con hipótesis de fecundidad estable presentan los porcentajes más altos de diferencia tanto para el total como para ambos géneros; 2.43, 2.71 y -2.16 por ciento. Mientras que las cifras calculadas en forma independiente para el total y cada género, correspondientes a la hipótesis de fecundidad descendente, son las que presentan la menor diferencia con la proyección oficial, 1.58, 1.82 y 1.35 por ciento.

Es importante remarcar que esta comparación no es un indicador de bondad de ajuste, ya que todas las cifras que se comparan son pronósticos, en cambio da cuenta de la divergencia en las metodologías y supuestos que subyacen a los números obtenidos. La similitud entre las cifras oficiales y las obtenidas por medio de pronósticos independientes e hipótesis de fecundidad descendente radica, posiblemente, en que en ambos procesos no se impone coherencia entre las cifras de varones y mujeres, además en ambos casos se supone una fecundidad descendente. El INDEC en su

proyección oficial, por medio de una función logística y en base a la tendencia histórica que presenta la Tasa Global de Fecundidad (TGF) desde al año 1981, genera tasas proyectadas hasta el año 2040 (Tabla 2) (INDEC-CELADE, 2013). Estas proyecciones conforman el supuesto de la fecundidad que subyace en la población obtenida. Si bien el documento técnico que detalla la metodología empleada para la elaboración de las proyecciones destaca que esta componente presenta un descenso mucho más atenuado que el observado en décadas pasadas y en base a ello recomienda tener cautela.

Tabla 2: Tasa global de fecundidad. Argentina. Periodo 2010-2040

Año	TGF
2010	2.41
2015	2.28
2020	2.18
2025	2.10
2030	2.05
2035	2.01
2040	1.98

Fuente: elaboración INDEC.

Un ejercicio más exhaustivo de comparación con las cifras oficiales excede el objetivo del presente trabajo, sin embargo, podría avanzarse a un análisis detallado de los resultados como proponen Narro Robles, J. *et al.* (2012). Dicho ejercicio resultaría de gran utilidad frente a las crecientes críticas a los operativos censales, tanto por su costo, como por sus problemas de omisión.

## CONCLUSIONES

En síntesis, al realizar un pronóstico de población es importante tener en cuenta múltiples fuentes de variación. Por otro lado, la cuantificación de esta incertidumbre es útil para los usuarios de las proyecciones porque permite tener en cuenta las diferentes implicancias de un pronóstico. El método que aquí se utiliza comprende las fuentes más importantes de variabilidad: la variabilidad observada a través del modelo Poisson que genera las muertes y nacimientos y la variación dinámica de las tasas a través de los modelos de series de tiempo. Este aspecto garantiza una representación más completa que la mayoría de los enfoques propues-

tos para este tipo de datos. En los pronósticos de población, las correlaciones entre los errores de pronóstico de las componentes demográficas afectan la amplitud de los intervalos de predicción, en este enfoque las correlaciones a través de las edades se tienen en cuenta mediante el uso de suavizados y la correlación temporal mediante el uso de modelos para series de tiempo. Estas correlaciones se incorporan naturalmente en las simulaciones de cada componente demográfica. De todos modos, dado que los métodos de series de tiempo que se utilizan son útiles para pronosticar a corto y a mediano plazo, la amplitud de los intervalos muestra que los métodos se vuelven menos informativos a lo largo del tiempo y recomiendan no utilizarlos con un horizonte mayor a los 20 años (Hyndman y Ullah, 2007).

Con relación a los resultados estimados para la componente mortalidad, para los hombres, las edades que dominan el cambio en dicha componente resultan; la franja de edades cercanas a cero, el grupo de 40 a 50 años, el grupo de 18 a 30 años y el grupo formado por las edades en torno a los 15 años. La forma en la que ejercen el cambio está dada por las bases, un descenso para los primeros tres grupos y fluctuaciones a lo largo del periodo para los restantes. En cambio, para las mujeres, los grupos que dominan los cambios en la mortalidad para el periodo 1980 a 2010 son las niñas de 0 a 5 años, mujeres de alrededor de 30 años y las adultas mayores de alrededor de 75 años, en segundo lugar, las mujeres de alrededor de 20 años y por ultimo las levemente mayores y menores a los 20 años. En segundo lugar, se desarrolla la variante para pronósticos coherentes de mortalidad, los mismos tienen como finalidad evitar una divergencia entre los resultados para ambos subgrupos, de este modo los pronósticos para los varones y mujeres difieren siempre en la misma proporción a lo largo del tiempo y más aún en los pronósticos se mantiene la misma diferencia que se presenta en los datos observados.

Luego, en el análisis de la fecundidad, la descomposición en bases y coeficientes permite identificar el patrón de descenso y las edades que lo determinan, el diferencial en la fecundidad adolescente versus edades avanzadas y el corrimiento de la maternidad a edades alrededor de los 40 años y su estabilización en torno a la misma.

Finalmente, en base a  $N = 1000$  trayectorias simuladas para cada componente se generan los pronósticos de población y sus intervalos de confianza. Los pronósticos para hombres y mujeres se obtienen para un periodo de 10 años y se comparan con las cifras oficiales de INDEC, luego se

evidencian, los distintos supuestos y las diferencias que constituyen a cada técnica.

Los resultados aquí obtenidos presentan medidas adecuadas de bondad de ajuste, cifras pronosticadas coherentes con la teoría demográfica y principalmente garantizan la consistencia probabilística. Las tasas y cifras de población pronosticadas están acompañadas de su intervalo de pronóstico, anexando de este modo una cuantificación de la incertidumbre asociada al resultado informado.

## REFERENCIAS BIBLIOGRÁFICAS

- Argote-Cusi, M., 2018, “El uso de lógica difusa en proyecciones de población: el caso de México”, en *Papeles de Población*, 24(95). Consultado en <https://rppoblacion.uaemex.mx/article/view/10119>
- Beer, J. D., 1997, “The effect of uncertainty of migration on national population forecasts: the case of the Netherlands”, in *Journal of Official Statistics*, 13:227–243.
- Bongaarts, J. y Bulatao, R. A., 2000, *Beyond Six Billion: Forecasting the World's Population. Panel on Population Projections, Committee on Population, National Research Council*. National Academy Press.
- Booth, H., Hyndman, R., Tickle, L., y de Jong, P., 2006a, “Lee-Carter mortality forecasting: a multi-country comparison of variants and extensions”, in *Demographic Research*, 15(9):289–310.
- Booth, H., Hyndman, R. J., Tickle, L., y de Jong, P., 2006b, *Lee-Carter mortality forecasting: a multi-country comparison of variants and extensions*.
- Booth, H., Maindonald, J., y Smith, L., 2002, “Applying Lee-Carter under conditions of variable mortality decline”, in *Population Studies*, 56(3):325–336.
- Booth, H., Tickle, L., y Smith, L., 2005, “Evaluation of the variants of the Lee-Carter method of forecasting mortality: A multi-country comparison”, in *New Zealand Population Review*, 31(1):13–34.
- Box, G. y Cox, D., 1964, “An analysis of transformations”, in *Journal of the Royal Statistical Society. Series B (Methodological)*, 26(2):211–252.
- Box, G. y Jenkins, G., 1976, *Time Series Analysis: Forecasting and Control*. Holden-Day.
- Box, G., Jenkins, G., y Reinsel, G., 2008, *Time Series Analysis: Forecasting and Control*. Wiley, Hoboken, New Jersey.
- Bozik, J. y Bell, W., 1987, *Forecasting age-specific fertility using principal components*.

- Brouhns, N., Denuit, M., y Vermunt, J., 2002, “A Poisson log-bilinear regression approach to the construction of projected lifetables”, in *Insurance: Mathematics and Economics*, 31(3):373–393.
- García-Guerrero, V., 2014, *Proyecciones y políticas de población en México*. COLMEX, México.
- George, M. y Perreault, J., 1992, *Methods of external migration projections and forecasts*. National population forecasting in industrialized countries, p. 87103.
- Granger, C. W. J. y Joyeux, R., 1980, “An introduction to long memory time series models and fractional differencing”, in *Journal of Time Series Analysis*, 1(1):15–29.
- Hilderink, H., der Gaag, N. V., Wissen, L. V., Jennissen, R., Román, A., Salt, J., Clarke, J., y Pinkerton, C., 2002, *Analysis and forecasting of international migration by major groups*, part iii. Working papers and studies, 3.
- Hosking, J., 1981, “Fractional differencing”, in *Biometrika*, 68(1):165–176.
- Hyndman, R., Booth, H., 2008, “Stochastic population forecasts using functional data models for mortality, fertility and migration”, in *International Journal of Forecasting* 24, 323-342.
- Hyndman, R., Booth, H., y Yasmeen, F., 2013, “Coherent Mortality Forecasting: The Product-Ratio Method with Functional Time Series Models”, in *Demography*, 50(1):261–283.
- Hyndman, R. y Khandakar, Y., 2008, “Automatic time series forecasting: the forecast package for r”, in *Journal of Statistical Software*, 3(27).
- Hyndman, R., Koehler, A., Ord, J., y Snyder, R., 2008, *Forecasting with Exponential Smoothing The State Space Approach*. Springerlink.
- Hyndman, R. y Shang, H., 2009, “Forecasting functional time series (with discussion)”, in *Journal of the Korean Statistical Society*, 38(3):199–221.
- Hyndman, R. y Shang, H., 2010, “Rainbow plots, bagplots and boxplots for functional data”, in *Journal of Computational and Graphical Statistics*, 19:29–45.
- Hyndman, R. y Ullah, M., 2007, “Robust forecasting of mortality and fertility rates: A functional data approach”, in *Computational Statistics and Data Analysis*, 51:4942– 4956.
- Jong, P. D. y Tickle, L., 2006, “Extending Lee-Carter mortality forecasting”, in *Mathematical Population Studies*, 13(1):1–18.
- Keilman, N. y Pham, D., 2004, *Empirical errors and predicted errors in fertility, mortality and migration forecasts in the european economic area*, Discussion Papers, Research Department of Statistics Norway, 386.

- Lee, R., 1993, “Modeling and forecasting the time series of us fertility: age distribution, range, and ultimate level”, in *International Journal of forecasting*, 9:187–202.
- Lee, R., 2000, “The Lee-Carter method for forecasting mortality, with various extensions and applications”, in *North American actuarial journal*, 4(1):80–91.
- Lee, R., 2006, *Mortality Forecasts and Linear Life Expectancy Trends*, pp. 19–40. National Social Insurance Board, Stockholm.
- Lee, R. y Carter, L., 1992, “Modeling and Forecasting U. S. Mortality”, in *Journal of the American Statistical Association*, 87:659–671.
- Lee, R. y Miller, T., 2001, “Evaluating the performance of the Lee-Carter Method for Forecasting Mortality”, in *Demography*, 38:537–549.
- Lee, R. y Tuljapurkar, S., 1994, “Stochastic population forecasts for the United States: Beyond high, medium, and low”, in *Journal of the American Statistical Association*, 89:1175–1189.
- Li, S. y Chan, W., 2005, “Outlier analysis and mortality forecasting: The United Kingdom and scandinavian countries”, in *Scandinavian Actuarial Journal*, (3):187–211.
- Mendoza Ponce, A.V., Corona Núñez, R. O., Galicia Sarmiento, L., y García Guerrero, V. M., 2019, “Proyecciones poblacionales y económicas bajo diferentes escenarios para México”, en *Papeles de Población*, 25(99), 9-43. Epub 25 de febrero de 2020.<https://doi.org/10.22185/24487147.2019.99.02>
- Miller, T., 2003, *California s uncertain population future. technical appendix. Special Report: the growth and aging of California s population: demographic and fiscal projections, characteristics and service needs. center of Economics and Demography of aging*. CEDA.
- Miller, T. y Lee, R., 2001, “Evaluating the performance of the Lee-Carter method for forecasting mortality”, in *Demography*, 38(4):537–549.
- Miller, T. y Lee, R., 2004, *A probabilistic forecast of net migration to the United States*. Report I in Stochastic infinite horizon forecasts for social security and related studies, 10917.
- Narro Robles, J., Hernández Bringas, H., y Flores Arenales, R., 2012, “El censo de población de 2010: cuatro millones más de mexicanos de lo previsto, ¿el final de una política de Estado?”, en *Papeles de Población*, 18(74). Consultado en <https://rppoblacion.uaemex.mx/article/view/8410>
- Preston, S., Heuveline, P., y Guillot, M., 2001, *Demography, Measuring and Modelling Population Processes*. Blackwell, Oxford.
- Raftery, A.E., Alkema, L., Gerland, P., 2014, “Bayesian population projections for the United Nations. Statistic”, in *Sci.* 29, 58-68.
- Ramsay, J., Silverman, B., Liu, A., Belliveau, J., y Dale, J., 1997, *Functional Data Analysis*.

- Ramsay, J. O. y Silverman, B. W., 2005, *Functional data analysis 2nd ed.* Springer, New York.
- Renshaw, A. y Haberman, S., 2003a, “Lee-Carter mortality forecasting: A parallel generalized linear modelling approach for england and wales mortality projections”, in *Applied Statistics*, 52(1):119–137.
- Renshaw, A. y Haberman, S., 2003b, “Lee-Carter mortality forecasting with age-specific enhancement”, in *Insurance: Mathematics and Economics*, 33(2):255–272.
- Renshaw, A. y Haberman, S., 2003c, “On the forecasting of mortality reduction factors”, in *Insurance: Mathematics and Economics*, 32(3):379–401.
- Rogers, A., 1990, “Requiem for the net migrant”, in *Geographical Analysis*, 22:283–300.
- Rogers, A. y Castro, L., 1981, *Model migration schedules*. Research Report, International Institute for Applied Systems Analysis, pp. 81–30.
- Rogers, A. y Little, J. S., 1994, “Parameterizing age patterns of demographic rates with the multiexponential model schedule”, in *Mathematical Population Studies*, 4:175–195.
- Serfaty, E., Foglia, L., Masaútis, A., y Negri, G., 2007, “Mortalidad por causas violentas en adolescentes y jóvenes de 10- 24 años”, in *Rev Vértex.*, 40:25–30.
- Wilmoth, J. , 1993, *Computational methods for fitting and extrapolating the Lee–Carter model of mortality change*. Technical report. department of demography. University of California, Berkeley.

## RESUMEN CURRICULAR DE LA AUTORA

### *Lucía Andreeozzi*

Es Doctora en Demografía de la Universidad Nacional de Córdoba, Magíster en Estadística Aplicada, Investigadora Asistente de CONICET y docente de la Facultad de Ciencias Económicas y Estadística de la Universidad Nacional de Rosario (UNR) en la Licenciatura en Estadística y en la Licenciatura en Economía. Dicta seminarios de posgrado en la Especialización en Estadística Actuarial (UNR) y la Maestría en Salud Pública (Instituto Lazarte). Coordinadora Estadística de la Usina de Datos de la UNR. Integrante del Grupo de Estudio de Economía y Género de la FCEyE.

Dirección electrónica: andreeozzi.lu@gmail.com

Registro ORCID: <https://orcid.org/0000-0002-1723-5725>

Artículo recibido el 3 de septiembre de 2021 y aprobado el 25 de octubre de 2021