

Pronósticos multivariados de poblaciones con series de tiempo: el caso de la ZMCM contrastado con datos del Censo 2010*

Eliud Silva**

Manuel Ordorica***

En este artículo se utilizan series de tiempo multivariadas para mostrar que es factible realizar pronósticos de población con esta perspectiva y se evidencia cómo éstos son satisfactorios de acuerdo con datos derivados del Censo de Población y Vivienda 2010 para la Zona Metropolitana de la Ciudad de México (ZMCM). Se enuncia de manera general la metodología empleada para realizar los pronósticos partiendo de técnicas de desagregación, ajustes y estimación de pronósticos multivariados. Se considera que las herramientas metodológicas capaces de proyectar de manera eficiente resultan indispensables para elaborar gran diversidad de programas públicos cuyo principal objetivo sea un desarrollo sustentable.

Palabras clave: series de tiempo, compatibilidad, pronósticos multivariados, modelo VAR(p), horizonte de pronóstico.

Fecha de recepción: 21 de octubre de 2011.

Fecha de aceptación: 10 de abril de 2012.

Multivariate Population Forecasts with Time Series: The Case of the MCMZ Contrasted with the 2010 Census Data

This article uses multivariate time series to show that it is feasible to undertake population forecasts using this perspective and shows how they are satisfactory according to data from the Population and Housing Census 2010 for the Mexico City Metropolitan Zone (MCMZ). It explains the methodology used to make predictions based on disaggregating techniques, adjustments and multivariate forecast estimates. The methodological tools capable of making effective forecasts are regarded as essential to developing a wide range of public programs whose primary goal is sustainable development.

Key words: time series, compatibility, multivariate forecasts, var (p) model, forecast horizon.

* Los autores agradecen profundamente los comentarios de los dictaminadores anónimos para mejorar este trabajo.

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

*** El Colegio de México. Correo electrónico: <mordori@colmex.mx>.

Introducción

Con base en la literatura especializada resulta evidente que las herramientas estadísticas de series de tiempo se pueden emplear para la descripción, análisis y pronósticos de diversos problemas demográficos. En este sentido, por citar algunos casos, desde la década de los ochenta se han elaborado diversas investigaciones, como por ejemplo: Land y Cantor, 1983; Carter y Lee, 1986; Thompson *et al.*, 1989; McNown y Rogers, 1992; Lee y Tuljapurkar, 1994; Bell, 1997; Keilman *et al.*, 2002; McNown y Rajbhandary, 2003; Laporte y Fergusonb, 2003; Brücker *et al.*, 2003; McNown y Ridao-Cano, 2005; Jeon y Shields, 2008; Girosi y King, 2008; Tuljapurkar *et al.*, 2004; Hyndman y Booth, 2008; Alonso *et al.*, 2008; Okita *et al.*, 2009; Goldstein *et al.*, 2009; Cornwell, 2009. Con esta perspectiva, es decir, al emplear series de tiempo en el contexto de proyecciones de población se advierte que se pueden asociar a las estimaciones puntuales futuras, intervalos de predicción con determinado nivel de confianza, donde se prioriza un apego a la objetividad derivada de los mismos datos, y, por otra parte, se considera que ésta resulta ser una visión complementaria con ciertas propuestas recientes, como la elaborada por King y Samir (2011).

Una de las mayores preocupaciones que sistemáticamente afectan a las instituciones generadoras de información, como el INEGI (Instituto Nacional de Estadística y Geografía) y el Conapo (Consejo Nacional de Población) en México y a los ministerios y consejos de población de otros países, así como a las instancias públicas y privadas que la utilizan, es lo tocante a las proyecciones o pronósticos de población. Universalmente se aspira a que éstos sean lo más confiables posible, para que con base en ellos se puedan sostener programas o planes cuyo diseño e implementación parten del supuesto de que se satisfagan determinados pronósticos oficiales y sean lo más precisos posible. Un pronóstico sin una apropiada precisión, por ejemplo en el sector público, es información que puede poner en jaque la viabilidad de algún programa social, ya sea en términos económicos o sociales, por decir lo menos.

Ante la eventual problemática que puede acarrear la baja capacidad predictiva de una proyección demográfica, en este trabajo se hace referencia a una alternativa cuyos elementos se han sugerido originalmente en el ámbito económico, pero que resultan eficientes en el ámbito demográfico. Así pues, el objetivo del presente trabajo es exponer el potencial que tiene la propuesta metodológica elaborada por

Silva *et al.* (2011), donde se emplean series de tiempo multivariadas para pronósticos de población. Se particulariza en el caso de los resultados del Censo 2010 para la Zona Metropolitana de la Ciudad de México (ZMCM). Asimismo, y como hipótesis, se tiene que la herramienta metodológica representa una alternativa para, en concreto, enriquecer los mecanismos de proyecciones demográficas, y en general para realizar una mejor planeación en distintos órdenes, así como para propiciar un ambiente idóneo a considerar por parte de los tomadores de decisiones.

Metodología

Se sigue la propuesta metodológica de Silva *et al.* (2011) que se sustenta en el manejo de modelos de series de tiempo multivariados denominados VAR(p) (VAR, del inglés Vector Autoregressive). Para la implementación de la metodología se emplea el software estadístico EViews versión 5.0 en lo referente a la estimación y diagnóstico del modelo VAR(p), en tanto que se utiliza Matlab versión 7.0 para las desagregaciones y pronósticos multivariados. Cabe mencionar que en general, para este fin, se podrían utilizar algunos otros softwares, como SAS, STATA, R o RATS.

La lógica de utilizar series de tiempo multivariadas emerge de la concepción de cómo está compuesta la ZMCM. Esta zona se concibe como un conjunto de municipios y delegaciones que a su vez constituyen anillos; a saber, es una ciudad central y tres anillos (en el cuadro 1 se explica cómo fueron consideradas dichas unidades geográficas). Aún cuando se podrían estimar cuatro modelos independientes, con o sin la perspectiva de series de tiempo, para las unidades geográficas referidas, al tener evidencia de que existe dependencia temporal entre las series (lo cual se puede verificar de manera formal mediante las denominadas pruebas de *cointegración*), resultaría un *desperdicio* el no aprovechar toda la información disponible. Con esto se considera justificable el uso de la perspectiva multivariada.

Debe notarse que para la desagregación no se cuenta con datos específicos censales de 1940 a 2000 relativos a la ciudad central y los anillos, sino que sus series se deben ir reconstruyendo a partir de los respectivos datos de municipios y delegaciones (a pesar de que en algunos casos no existen como tales en las primeras décadas que se consideran en el análisis) (véase el cuadro 2). Lo mismo ocurre en

CUADRO 1
Composición de la ZMCM

	Delegaciones	Municipios
<i>Ciudad central</i>	Benito Juárez, Cuauhtémoc, Miguel Hidalgo y Venustiano Carranza	
<i>Primer anillo</i>	Azcapotzalco, Coyoacán, Cuajimalpa, Gustavo A. Madero, Iztacalco, Iztapalapa y Álvaro Obregón	Naucalpan y Nezahualcóyotl
<i>Segundo anillo</i>	La Magdalena Contreras, Tláhuac, Tlalpan y Xochimilco	Atenco, Atizapán, Coacalco, Chimalhuacán, Ecatepec, Huixquilucan, La Paz, Tlahuepantla, Tultitlán y Cuautitlán Izcalli
<i>Tercer anillo</i>	Milpa Alta	Acolman, Cuautitlán, Chalco, Chiautla, Chicoloapan, Chiconcuac, Ixtapaluca, Jaltenco, Melchor Ocampo, Nextlalpan, Nicolás Romero, Papalotla, Tecámac, Teoloyucán, Teotihuacán, Tepetlaoxtoc, Tepotzotlán, Texcoco, Tezoyuca, Tultepec y Zumpango

FUENTE: Definición de la ZMCM adoptada con base en la propuesta del INEGI.

CUADRO 2

Censos de población (1940-2010) y conteos (1995-2005) por anillos concéntricos

<i>Delegación o municipio</i>	1940	1950	1960	1970	1980	1990	1995	2000	2005	2010
Ciudad de México	1 448 422	2 234 795	2 832 133	2 902 969	2 595 823	1 930 267	1 760 359	1 688 401	1 677 358	1 721 137
Benito Juárez	n.d.	n.d.	n.d.	n.d.	n.d.	407 811	369 956	359 334	355 017	385 439
Cuauhtémoc	n.d.	n.d.	n.d.	n.d.	814 983	595 960	540 382	515 132	521 348	531 831
Miguel Hidalgo	n.d.	n.d.	n.d.	n.d.	543 062	406 868	364 398	351 846	353 534	372 889
Venustiano Carranza	n.d.	n.d.	n.d.	n.d.	692 896	519 628	485 623	462 089	447 459	430 978
Total ciudad central	1 448 422	2 234 795	2 832 133	2 902 969	5 595 823	1 930 267	1 760 359	1 688 401	1 677 358	1 721 137
Acapotzalco	63 000	187 864	370 724	534 554	601 524	474 688	455 131	440 558	425 298	414 711
Coyoacán	35 248	70 005	169 811	339 446	597 129	640 066	653 489	639 021	628 063	620 416
Cuajimalpa de Morelos	6 025	9 676	19 199	36 200	91 200	119 669	136 873	151 127	173 625	186 391
Gustavo A. Madero	41 567	204 833	579 180	1 186 107	1 513 360	1 268 068	1 256 913	1 233 922	1 193 161	1 185 772
Iztacalco	11 212	33 943	198 904	477 331	570 377	448 322	418 982	410 717	395 025	384 326
Iztapalapa	25 393	76 621	254 355	522 095	1 262 354	1 490 499	1 696 609	1 771 673	1 820 888	1 815 786
Álvaro Obregón	32 313	93 176	220 011	456 709	639 213	642 753	676 930	685 327	706 567	727 034
Naucalpan de Juárez	11 051	12 737	16 021	387 184	730 170	786 551	839 723	857 511	821 442	833 779
Nezahualcóyotl	n.d.	n.d.	n.d.	580 436	1 341 230	1 256 115	1 233 868	1 224 924	1 140 528	1 110 565
Total primer anillo	225 809	688 855	1 828 205	4 520 062	7 346 557	7 126 731	7 368 518	7 414 780	7 304 597	7 278 780

(continúa)

CUADRO 2
(concluye)

<i>Delegación o municipio</i>	1940	1950	1960	1970	1980	1990	1995	2000	2005	2010
La Magdalena Contreras	13 159	21 955	40 724	75 429	173 105	195 041	211 898	221 762	228 927	239 086
Tláhuac	13 843	19 511	29 880	62 419	146 923	206 700	255 891	302 483	344 106	360 265
Tlalpan	19 249	32 767	61 195	130 719	368 974	484 866	552 516	580 776	607 545	650 567
Xochimilco	33 313	47 082	70 381	116 493	217 481	271 151	332 314	368 798	404 458	415 007
Atenco	5 023	5 424	7 341	10 616	16 418	21 219	27 988	34 393	42 739	56 243
Atizapán de Zaragoza	n.d.	n.d.	n.d.	44 322	202 248	315 192	427 444	467 262	472 526	489 937
Coacalco de Berriozábal	1 736	2 315	3 984	13 197	97 353	152 082	204 674	252 270	285 943	278 064
Chimalhuacán	7 399	13 004	76 740	19 946	61 816	242 317	412 014	490 245	525 389	614 453
Ecatepec de Morelos	10 501	15 226	40 815	216 408	784 507	1 218 135	1 457 124	1 620 303	1 688 258	1 656 107
Huixquilucan	12 034	13 491	16 229	33 527	78 149	131 926	168 221	193 156	224 042	242 167
La Paz	3 052	4 194	7 880	32 258	99 436	134 782	178 538	213 045	232 546	253 845
Tlahuepanlta de Baz	14 626	29 005	105 447	366 915	778 173	702 807	713 143	720 755	683 808	664 225
Tultitlán	6 638	9 237	15 479	52 317	136 829	246 464	361 434	432 411	472 867	524 074
Cuautitlán Izcalli	n.d.	n.d.	n.d.	n.d.	173 754	326 750	417 647	452 976	498 021	511 675
Total segundo anillo	140 573	213 211	476 095	1 174 566	3 335 166	4 649 432	5 720 846	6 350 635	6 711 175	6 955 715
Milpa Alta	14 786	18 212	24 379	33 694	53 616	63 654	81 102	96 744	115 895	130 582
Acolman	7 234	9 422	12 230	20 964	32 316	43 276	54 468	61 181	77 035	136 558
Cuautitlán	10 701	13 622	20 509	41 156	39 527	48 858	57 373	75 831	110 345	140 059

Chalco	17 994	22 056	29 725	41 450	78 393	106 450	175 521	222 201	257 403	310 130
Chiautla	3 479	3 867	4 862	7 266	10 618	14 764	16 602	19 559	22 664	26 191
Chicoloapan	2 417	3 229	4 719	8 750	27 354	57 306	71 351	77 506	170 035	175 053
Chiconcuac	3 587	4 520	6 031	8 399	11 371	14 179	15 448	17 977	19 656	22 819
Ixtapaluca	7 765	10 787	20 472	36 722	77 862	137 357	187 690	293 160	429 033	467 361
Jaltenco	2 236	2 767	3 322	4 738	7 847	22 803	26 238	31 608	26 359	26 328
Melchor Ocampo	4 156	4 928	6 537	10 834	17 990	26 154	33 455	37 724	37 706	50 240
Nextlalpan	2 349	2 611	3 602	4 360	7 380	10 840	15 053	19 755	22 507	34 374
Nicolás Romero	17 283	23 346	29 617	47 504	112 645	184 134	237 064	269 393	306 516	366 602
Papalotla	815	706	735	1 088	1 769	2 387	2 998	3 469	3 766	4 147
Tecánac	7 718	9 104	11 971	20 882	84 129	123 218	148 432	172 410	270 574	364 579
Tecoloyucan	5 656	7 446	9 939	15 477	28 836	41 964	54 454	66 486	73 696	63 115
Teotihuacán	6 667	8 348	10 477	16 283	30 140	30 486	39 183	44 536	46 779	53 010
Tepetlaoxtoc	5 379	5 373	5 471	7 068	10 019	16 120	19 380	22 687	25 507	27 944
Tepotzotlán	7 879	10 073	12 682	21 902	27 079	39 647	54 419	62 247	67 724	88 559
Texcoco	24 812	32 265	42 525	65 628	105 851	140 368	173 106	203 681	209 308	235 151
Tezoyuca	2 161	2 532	3 424	4 770	7 567	12 416	16 338	18 734	25 372	35 199
Tultepec	3 638	5 517	7 744	11 480	22 910	47 323	75 996	93 364	110 145	91 808
Zumpango	14 073	17 498	22 677	36 105	51 393	71 413	91 642	99 781	127 988	159 647
Total tercer anillo	172 785	218 859	293 650	466 520	846 612	1 255 117	1 647 313	2 010 054	2 556 013	3 009 456

NOTA: Los datos de 2005 corresponden a datos con estimación elaborados por el INEGI.

n.d. = Dato no disponible.

FUENTE: Recopilación propia en los censos respectivos (INEGI, 1940-2010) y en los conteos (INEGI, 1995 y 2005).

cuanto a las estadísticas vitales, las cuales se aproximan por medio de porcentajes. A continuación se citan, de manera general, los cuatro pasos principales de la metodología.

Paso 1. Previo

Se realiza la recolección de datos provenientes de los censos de población desde 1940 hasta 2000, de los conteos de 1995 y 2005, e información anual de estadísticas vitales (nacimientos y muertes) de 1941 a 2000. Siguiendo un algoritmo de naturaleza demográfica (véase el anexo) se generan series anuales de tasas acumulativas de crecimiento demográfico PGR^1 (*Population Growth Rate*), tomando como momento inicial el año de 1940 y estimando el crecimiento en referencia con dicho año y cada uno de los años restantes para cada unidad geográfica. Éstas son las llamadas series preliminares.

Paso 2. Desagregación

Se define un vector de variables no observables de tamaño 4 (k) (número de unidades geográficas consideradas) $\mathbf{Z}_{Dt} = (Z_{ccpt}, Z_{frpt}, Z_{srpt}, Z_{trpt})'$, donde los subíndices se refieren a la unidad geográfica respectiva: ciudad central (*ccpt*), primer anillo (*frpt*), segundo anillo (*srpt*), tercer anillo (*trpt*). Esto se realiza para todo el tiempo deseado de desagregación anual, que comienza en 1941 y termina en 2000; es decir, para $t = 1, \dots, mn$ donde $n = 6$, número de periodos censales completos y m es la frecuencia anual ($m = 10$ años en una década). Luego se define otro vector apilado $\mathbf{Z}_D = (\mathbf{Z}'_{D1}, \dots, \mathbf{Z}'_{Dmn})'$ en el cual, si se observa, se disponen en forma alterna las series correspondientes a los años señalados. Como se trata de variables no observables, se requiere estimarlas a partir de series observables más alguna información adicional, lo cual se obtiene definiendo de manera análoga los vectores apilados para las series preliminares $\mathbf{W}_{Dt}$ y $\mathbf{W}_D$. Con base tanto en $\mathbf{W}_D$ como en $\mathbf{Y}_D = C_D \mathbf{Z}_D$ (combinación lineal que contiene la información censal), donde $\mathbf{Y}_D$ es un vector que contiene los datos agregados $\mathbf{Z}_D$ de (datos censales) y C_D es una matriz de constantes conocidas; $\mathbf{Z}_D$ se estima con

¹ Para regresar a la escala original, es decir, para obtener la población total en términos absolutos para el año t , se multiplica la población de 1940 por $(1+PGR_t)$ del año respectivo.

$$\hat{\mathbf{Z}}_D = \mathbf{W}_D + \mathbf{A}_D(\mathbf{Y}_D - \mathbf{C}_D \mathbf{W}_D) \quad [1]$$

Una forma sencilla de interpretar la expresión [1] es afirmar que las series desagregadas finales (estimadas) son iguales a las series preliminares más un factor de ajuste (o corrección), el cual está vinculado con las restricciones por satisfacer, es decir, con los datos censales. Se remite al lector al artículo de Guerrero y Nieto (1999) para mayores detalles sobre el método y las definiciones empleadas.

Con la desagregación se transita de un número pequeño de observaciones a otro mucho más elevado. O sea, de siete observaciones (datos censales de 1940, 1950, ..., 2000) a un total de 60 observaciones (1941, 1942, ..., 1999, 2000), y además esta cantidad de datos se genera para cada una de las unidades geográficas definidas. El tipo de desagregación que se realiza se puede entender como una interpolación, con la propiedad de que se garantizan optimalidad y propiedades deseables en términos de estimación. También es importante destacar que las series preliminares no deben tener como meta el alcanzar los niveles de las observaciones censales, sino que en todo caso son ellas las que proporcionan la *estructura* a las series desagregadas finales. Estas últimas sí que deben satisfacer las restricciones expuestas en el segundo sumando de la ecuación [1]. A manera de validación de la metodología para la desagregación se verifica que los datos del Censo 1995 caigan respectivamente dentro de los intervalos de estimación a un nivel de 95% de confianza.

Paso 3. Pronósticos multivariados

Para los pronósticos multivariados se define el vector $\mathbf{Z}_t = (Z_{cpt}, Z_{jpt}, Z_{spt}, Z_{npt})'$ de $k = 4$ variables observadas en el tiempo t , es decir, para las series desagregadas correspondientes a las unidades geográficas (para más detalles véase Lütkepohl, 2005). Se asume que el conjunto de series desagregadas admite una representación de un modelo VAR(p)

$$\Pi(B)\mathbf{Z}_t = \mathbf{D}_t + \mathbf{a}_t \quad [2]$$

donde $\Pi(B)$ es una matriz polinomial de orden p finita en el operador retraso B tal que $BX_t = X_{t-1} \forall X$ y t , $\mathbf{D}_t$, $\mathbf{a}_t$, son elementos determinísticos y $\{\mathbf{a}_t\}$ es un vector de ruido blanco. Se define además el vector

$\mathbf{Z}_F = (\mathbf{Z}'_{N+1}, \dots, \mathbf{Z}'_{N+H})'$ de valores futuros con horizonte de pronóstico de 20 años. Se ha demostrado en la literatura que el pronóstico lineal de $\mathbf{Z}_{N+h}$ es óptimo en el sentido de error cuadrático medio mínimo, para $h = 1, \dots, H$, además de ser condicionalmente insesgado.

Por otra parte, se hacen dos modificaciones sobre los intervalos de confianza de los pronósticos multivariados: los ajustes por ser generados a partir de procesos estimados y la adherencia de los errores heredados de la desagregación. Para este último punto se consideran matrices de covarianza de los residuales correspondientes así como los datos respectivos de las dos últimas décadas, es decir, de 1980 a 2000, para incluirlas dentro del horizonte de pronóstico de 2001 a 2020. La eficiencia de los pronósticos multivariados se respalda tomando en cuenta los datos del Censo 2005 y verificando que estas observaciones caigan respectivamente dentro de los intervalos de predicción a un nivel de 95% de confianza.

Paso 4. Verificación del potencial de la técnica

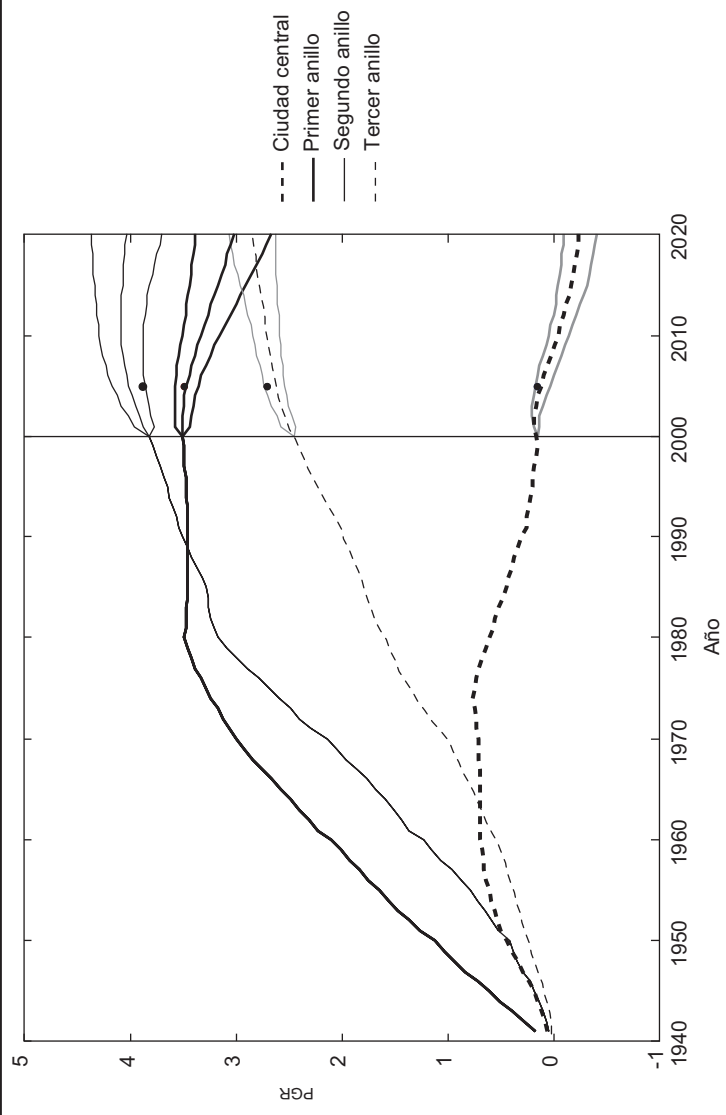
Un elemento teórico adicional considerado por Silva *et al.* (2011) lo constituyen las llamadas pruebas de compatibilidad, cuyo objetivo esencial es señalar qué tan acordes o no son los valores futuros propuestos con base en el comportamiento de las series poblacionales observadas. La propuesta de este estadístico fue planteada por Guerrero y Peña (2000, 2003), y de manera general viene dado por

$$K = \mathbf{d}' \boldsymbol{\Omega}^{-1} \mathbf{d} \sim \chi_M^2 \quad [4]$$

donde $\mathbf{d} = \mathbf{Y} - \mathbf{CE}(\mathbf{Z}_F | \mathbf{Z})$. Entonces se dice que $\mathbf{Y}$ es incompatible con $\mathbf{CE}(\mathbf{Z}_F | \mathbf{Z})$ al nivel de significancia $100 \alpha\%$ si $K_{calc} > \chi_M^2(\alpha)$ con M grados de libertad (M número de metas futuras). Esta prueba se puede realizar de manera marginal para validar las metas con base en alguna única serie objeto de estudio o bien de manera conjunta, es decir para todas las unidades geográficas de manera simultánea. Se habla entonces de compatibilidad parcial y compatibilidad total. Con base en estas pruebas estadísticas, así como con los intervalos de predicción, se verifica el potencial de la propuesta metodológica.

En el cuadro 3 se presenta una síntesis de los pasos comprendidos para la aplicación de la metodología con la finalidad de que el lector los pueda seguir y dimensionar globalmente.

GRÁFICA 1
Pronósticos multivariados para la ZMCM, 2001-2020



NOTA: Intervalos de predicción de 95%. Los puntos respectivos son los datos del Censo 2005.
FUENTE: Cálculos propios.

CUADRO 3

Pasos de la metodología

<i>Paso 1. Previo</i> 1.1. Recolección de datos censales (1940-2000) y de conteos (1995 y 2005) para el D.F. y el Estado de México. 1.2. Recolección de estadísticas vitales anuales (1941-2000) para el D.F. y el Estado de México. 1.3. Construcción de series preliminares con algoritmo demográfico (véase el anexo). 1.4. Generación de las series <i>PCR</i> para cada unidad geográfica. 1.5. Cálculo de las series <i>PCR</i> equivalentes de los datos censales y conteos.	<i>Paso 2. Desagregación</i> 2.1. Estimación de modelo VAR(p) con series preliminares. 2.2. Elección del parámetro p del modelo, con base en pruebas estadísticas. 2.3. Validación de supuestos del modelo VAR(p). 2.4. Estimación de la desagregación puntual y por intervalo. 2.5. Valoración de la desagregación con base en datos registrados del conteo 1995. 2.6. Resguardo de matrices de residuales derivadas de la desagregación de 1980-2000.
<i>Paso 3. Pronósticos multivariados</i> 3.1. Definición de horizonte de pronóstico: 2001-2020. 3.2. Estimación de pronósticos puntuales. 3.3. Estimación de intervalos de predicción al 90, 95 y 99% de confianza. 3.4. Adherencia de matrices de residuales derivadas de la desagregación 1980-2000. 3.5. Corrección de pronósticos por ser un proceso estimado. 3.6. Valoración de los pronósticos con base en datos registrados del conteo 2005.	<i>Paso 4. Verificación de potencial de la técnica</i> 4.1. Recolección de los datos poblaciones del Censo 2010 para cada una de las unidades geográficas. 4.2. Cálculo de las <i>PCR</i> equivalentes de 1940 a 2010 por cada unidad geográfica. 4.3. Elaboración de pruebas de compatibilidad parcial y total. 4.4. Valoración de pronósticos vs. datos de población. 4.5. Obtención de elementos para evidenciar el potencial de la técnica.

Resultados para pronósticos de 2010

Una vez expuesta brevemente la metodología de Silva *et al.* (2011) se valora qué tan eficiente resulta ésta en cuanto a su capacidad predictiva con base en los datos censales del año 2010, es decir, se procede a conocer qué potencial tiene la propuesta metodológica. El insumo para esta tarea es el conjunto de resultados de volúmenes de población que se han alcanzado en la ZMCM y que provienen de los resultados del levantamiento del Censo de Población y Vivienda 2010 a cargo del INEGI, algo relativamente análogo a lo que se realizó con datos de los conteos 1995 y 2005 respectivamente.

Con base en los resultados del Censo se calcularon las *PGR* correspondientes para los crecimientos (o eventuales decrementos) de la ciudad central y los anillos concéntricos (se utiliza la fórmula VII del anexo). Durante la elaboración de Silva *et al.* (2011), desde la perspectiva estadística se tenía la certeza de que los intervalos de predicción iban a contener los niveles de población para 2010 y que para los conteos subsecuentes (si es que los hubiera) sería necesario actualizar los pronósticos dotándolos de mayor información. Sin embargo, por tratarse de una manera distinta de hacer las proyecciones poblacionales, con la mirada demográfica existía un sano grado de escepticismo ante la verificación del potencial de la propuesta de los pronósticos multivariados.

En la práctica resulta inverosímil que los pronósticos puntuales de cualquier fenómeno donde se susciten de manera intrínseca efectos aleatorios generados por diversas causas, cualquiera que sea el método con el que se elaboren, tengan un error nulo o una precisión absoluta. Por esta razón, resulta recomendable acompañarlos con sus respectivos intervalos de predicción. Más aún, si todo intervalo de predicción es capaz de contener los datos que se desea pronosticar, entonces se puede evidenciar que se tiene capacidad predictiva, y en eso radica justamente el potencial de la técnica aquí propuesta.

Para un juicio objetivo y orientador acerca de la viabilidad del cumplimiento de metas futuras en diversos contextos, adicionalmente a los intervalos de predicción, resultan útiles las pruebas de compatibilidad. Una pregunta natural que se puede formular es en torno a la relación que pudiera existir entre los intervalos de predicción y las pruebas de compatibilidad. De manera empírica, por medio de diversas simulaciones elaboradas por los autores, se ha observado que las regiones donde se satisface la compatibilidad contienen a los interva-

los de predicción. Por lo tanto, en el marco de los denominados pronósticos restringidos multivariados, una buena manera de fijar metas futuras por alcanzar es emplear todo el conjunto de puntos contenido en los intervalos de predicción.

Conforme a las ideas básicas anteriores se ha optado por buscar la pertenencia de las *PGR* observadas del Censo 2010 en los intervalos de predicción con los distintos niveles de confianza que generalmente se emplean: 90, 95 y 99% (los resultados se muestran en el cuadro 4). En resumen, considerando tanto los intervalos de predicción como los resultados de las pruebas de compatibilidad se lograron los siguientes hallazgos para cada una de las unidades geográficas:

- a) Ciudad central. En ningún caso el valor observado de la *PGR* cae dentro de los intervalos de predicción. Se piensa que esto se explica en gran medida debido a los resultados de la política del Bando 2, que se impuso en 2000 con la idea de incentivar el repoblamiento y así aprovechar la infraestructura urbana ahí existente (véase Delgadillo, 2009; Paquete y Yescas, 2009). De hecho la *PGR* observada es incompatible con la dinámica de la serie porque el valor del estadístico es de 20.57 y es significativo al 1%; se rompe con la dinámica de la tendencia que presentaba esta serie. Otra perspectiva de este mismo resultado sería que la política *agresiva* de repoblamiento en el Distrito Federal tuvo éxito, sin embargo haría falta precisar si eso era lo que se esperaba para 2010, tanto para los planificadores como para los tomadores de decisiones, o más bien fue algo circunstancial cuyas consecuencias no se han previsto en lo tocante a su sustentabilidad, ya sea en el corto, mediano o largo plazos.
- b) Primer anillo. El valor de la *PGR* es 3.47, que cae casi dentro del intervalo de predicción de 90%; asimismo resulta estar contenida en los dos últimos intervalos de predicción. En este caso el estadístico de compatibilidad fue de 2.68 y no es significativo al 1%; también se aprecia que con base en el dato censal hubo un incremento poblacional más alto que la resultante a partir de la estimación puntual obtenida.
- c) Segundo anillo. La *PGR* observada fue de 3.90 y está contenida en todos los intervalos de predicción; de hecho, es la única de todas las unidades geográficas que posee esta característica. Su estadístico de compatibilidad fue de 2.83 y es no significativo

al 1%. A diferencia de los otros dos anillos, el dato censal es menor que la estimación puntual generada.

- d) Tercer anillo. Su *PGR* es 2.86 y se alcanza a cubrir hasta el intervalo de predicción de 99%, en tanto que su estadístico de compatibilidad parcial fue de 5.22 y es no significativo también al 1%. Al igual que en el caso del primer anillo, el dato censal es más alto que la estimación puntual obtenida.

Con base en la gráfica 2 y con los estadísticos de compatibilidad parcial y total, los resultados del Censo 2010 serían compatibles si no hubiera tenido efecto alguno la política de repoblamiento impuesta en la ciudad central; es decir, los datos censales de 2010 son compatibles con la dinámica de crecimiento poblacional de la ZMCM exclusivamente en los anillos. En términos cualitativos se puede afirmar que el ritmo de crecimiento es menos pronunciado para el segundo anillo y más pronunciado de lo previsto para las demás unidades geográficas. Para subsecuentes pronósticos convendría tomar los datos observados de 2010 y así hacer los ajustes respectivos.

De acuerdo con la gráfica 2, se pueden distinguir aproximadamente las denominadas etapas del metropolitano (véase Sobrino, 2007): *i*) concentración (la tasa de crecimiento demográfico de la ciudad central supera a la de la periferia); *ii*) desconcentración (la periferia alcanza un mayor ritmo de crecimiento poblacional respecto a la ciudad central); *iii*) despoblamiento (la ciudad central inicia un saldo neto migratorio negativo y prosigue hasta el decrecimiento absoluto de su población), y *iv*) repoblamiento (la ciudad central retoma su crecimiento demográfico). En un trabajo de investigación futuro convendría entonces, a partir de los datos censales de 2010, hacer un estudio acerca de la movilidad de la población de la ZMCM en relación con el lugar de residencia en retrospectiva.

Conclusiones

Se utilizan métodos estadísticos de series de tiempo multivariadas con el propósito de explicar que esta perspectiva resulta apropiada para los pronósticos de población. Así pues, el enfoque estadístico expuesto puede ser útil y adecuado para propiciar un panorama idóneo para la toma de decisiones, más allá de que el conjunto de herramientas estadísticas que lo conforman haya sido originalmente propuesto en el ámbito económico.

CUADRO 4

PCR observadas vs. intervalos de predicción

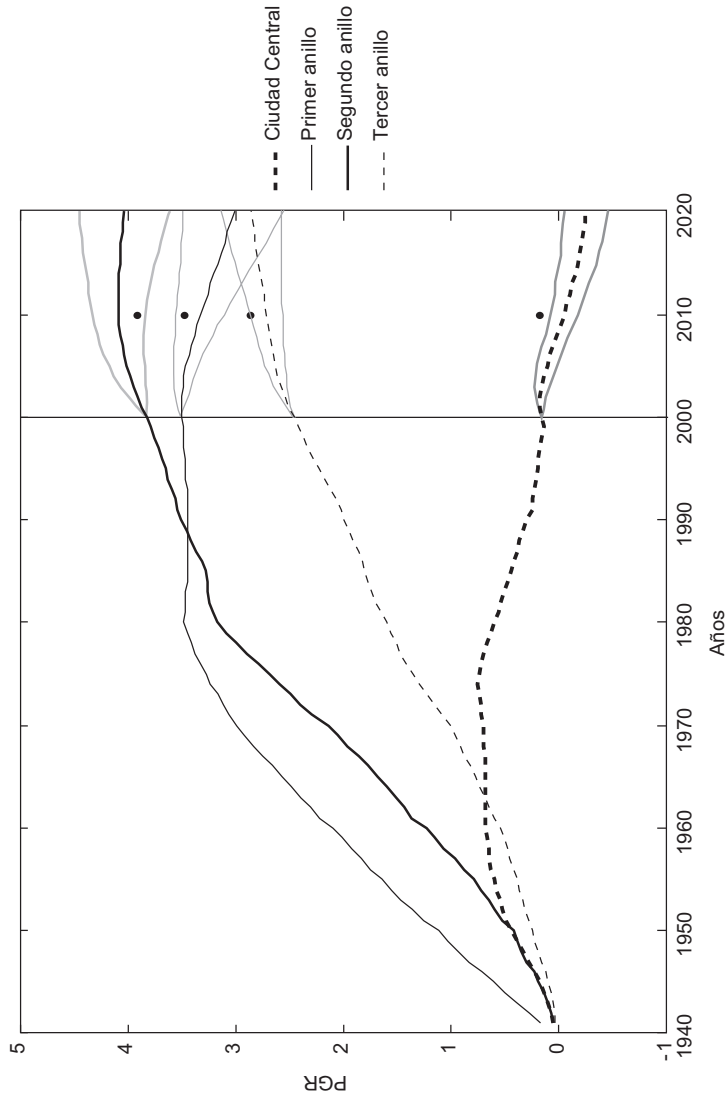
Unidad geográfica	PCR 2010	Valores de límites de intervalos de predicción					
		90%		95%		99%	
		LI	LS	LI	LS	LI	LS
Ciudad central	0.17	-0.14	0.03	-0.15	0.04	-0.18	0.07
Primer anillo	3.47	3.18	3.46	3.16*	3.49*	3.11*	3.54*
Segundo anillo	3.90	3.92	4.23	3.89*	4.26*	3.83*	4.32*
Tercer anillo	2.86	2.61	2.80	2.60	2.82	2.56*	2.86*

LI: límite inferior del intervalo de predicción; LS: límite superior del intervalo de predicción. Se marca con * el intervalo donde el valor observado pertenece al intervalo de predicción.

FUENTE: Cálculos propios.

GRÁFICA 2

Pronósticos irrestrictos para la ZMCM, 2001-2020



NOTA: Intervalos de predicción de 95%. Los puntos de colores respectivos son los datos del Censo 2010.

FUENTE: Cálculos propios.

Para subsecuentes pronósticos y conforme a la presente propuesta, sería oportuna la actualización de los pronósticos sobre las mismas series demográficas con datos censales de 2010. Asimismo se considera que se podría utilizar la herramienta para hacer propuestas de planes de crecimiento demográfico donde se utilicen nuevamente las llamadas pruebas de compatibilidad. Por otra parte, se está ante la disyuntiva de deducir qué es más eficiente: *a)* aplicar la desagregación hasta 2010, o *b)* usar pronósticos restringidos para satisfacer la meta (datos observados) de 2010, para pronosticar posteriormente en ambos casos.

Una implementación de la propuesta de Silva *et al.* (2011) para todos los estados del país podría resultar extremadamente costosa en términos de tiempo, presupuesto y esfuerzo. Sin embargo sería más factible y provechoso en términos estratégicos replicar la aplicación de la metodología en las zonas metropolitanas del país. La idea sería en principio establecer las unidades geográficas que las definen y luego seguir los pasos subsecuentes, dado que se tiene una alta concentración demográfica en ellas y que existe una alta generación de riqueza (PIB) en las mismas.

Finalmente es digno de mencionar que con horizontes de pronósticos más pequeños, por ejemplo cinco años, los resultados suelen ser más satisfactorios en el caso univariado; sin embargo en el caso multivariado, como el ejemplo de la ZMCM, los resultados sugieren que se puede tener eficiencia predictiva en al menos el doble de tiempo. Esto a manera de orientación y con base en diversas pruebas que realizamos los autores, más allá de un resultado teórico formalmente demostrado.

Anexo

El algoritmo para obtener las series preliminares es el siguiente:

- I. Calcular los crecimientos naturales para el D.F. y el Estado de México, siendo denotadas respectivamente por $ngdf_t$ y $ngsm_t$, para $t = 1941, 1942, \dots, 2000$.
- II. Calcular las poblaciones parciales para ambas unidades dfp_t y smp_t , por medio de $dfp_t = dfp_{2000-n} + \sum_{t=2000-n+1}^{2000} ngdf_t$ y $smp_t = smp_{2000-n} + \sum_{t=2000-n+1}^{2000} ngsm_t$ con $n = 60$.

- III. Considerar la proporción de poblaciones de las unidades geográficas respecto a $dfp_{t^*} + sm p_{t^*}$ para los años censales $t^* = 1940, 1950, \dots, 2000$, representados por $rccp_{t^*}$, $rfrp_{t^*}$, $rsrp_{t^*}$ y $rtrp_{t^*}$. Posteriormente calcular las proporciones para años intercensales t , dados por $rccp_t$, $rfrp_t$, $rsrp_t$ y $rtrp_t$, asumiendo un comportamiento lineal. Nótese que las series de las proporciones estimadas deben satisfacer la relación

$$rmzmc p_t = rccp_t + rfrp_t + rsrp_t + rtrp_t$$

para $t = 1941, 1942, \dots, 2000$.

- IV. Calcular las series de las proporciones de población para las unidades geográficas con

$$ccp_t^+ = rccp_t(df p_t + sm p_t), \quad frp_t^+ = rfrp_t(df p_t + sm p_t),$$

$$srp_t^+ = rsrp_t(df p_t + sm p_t) \text{ y } trp_t^+ = rtrp_t(df p_t + sm p_t)$$

para $t = 1941, 1942, \dots, 2000$.

- V. Calcular las diferencias asumidas como atribuibles a la migración para las unidades geográficas:

$$migccp_{t^*} = ccp_{t^*} - ccp_{t^*}^+, \quad migfrp_{t^*} = frp_{t^*} - frp_{t^*}^+,$$

$$migsrp_{t^*} = srp_{t^*} - srp_{t^*}^+ \text{ y } migtrp_{t^*} = trp_{t^*} - trp_{t^*}^+$$

para $t^* = 1940, 1950, \dots, 2000$, donde ccp_{t^*} , frp_{t^*} , srp_{t^*} y trp_{t^*} son datos censales en los años t^* . Luego asumir que la migración presenta un comportamiento uniforme en el tiempo y agregar un décimo de esas diferencias a las estimaciones anuales obtenidas en el paso IV.

- VI. Calcular las series para $t = 1941, 1942, \dots, 2000$ como sigue:

$$ccp_t = ccp_t^+ + 0.1 * migccp_{t^*}, \quad frp_t = frp_t^+ + 0.1 * migfrp_{t^*},$$

$$srp_t = srp_t^+ + 0.1 * migsrp_{t^*} \quad \text{y} \quad trp_t = trp_t^+ + 0.1 * migtrp_{t^*}.$$

- VII. Finalmente, estimar las *PGR* entre 1940 y cada año desde 1941 hasta 2000, para cada unidad geográfica, con la expresión

$$PGR_t = \ln \left(\frac{p_t}{p_{1940}} \right)$$

Bibliografía

- Alonso, A.M., D. Peña y J. Rodríguez (2008), "A Methodology for Population Projections: An Application to Spain", Madrid, Universidad Carlos III de Madrid (Working Papers, Statistics and Econometrics, 8.12).
- Bell, W. (1997), "Comparing and Assessing Time Series Methods for Forecasting Age-Specific Fertility and Mortality Rates", *Journal of Official Statistics*, núm. 13, pp. 279-303.
- Brücker, H., B. Siliverstovs y P. Trübswetter (2003), "International Migration to Germany: Estimation of a Time-Series Model and Inference in Panel Cointegration", documento de trabajo, Berlín, German Institute for Economic Research, DIW Berlin.
- Carter, L. y R. Lee (1986), "Joint Forecasts of US Marital Fertility, Nuptiality, Births and Marriages Using Time Series Models", *Journal of the American Statistical Association*, vol. 81, núm. 396, pp. 902-911.
- Cornwell, D. (2009), "An Exploratory Time Series Analysis of Apprehensions and Linewatch Hours on the Southwest Border", documento de trabajo, Office of Immigration Statistics.
- Delgadillo, V. (2009), "Reseña de 'Los desafíos del Bando 2. Evaluación multidimensional de las políticas habitacionales en el D.F. 2000-2006', de Sergio Tamayo", *Economía, Sociedad y Desarrollo*, vol. 9, núm. 29, pp. 209-220.
- Girosi, F. y G. King (2008), *Demographic Forecasting*, Princeton-Woodstock, Princeton University Press.
- Goldstein, E., J. Dushoff, J. Mad, J. Plotkine, D. Earn y M. Lipsitch (2009), *Reconstructing Influenza Incidence by Deconvolution of Daily Mortality Time Series*, Princeton, Burton H. Singer, Princeton University.
- Guerrero, V.M. y F.H. Nieto (1999), "Temporal and Contemporaneous Disaggregation of Multiple Economic Time Series", *Sociedad de Estadística e Investigación Operativa (TEST)*, vol. 8, núm. 2, pp. 459-469.
- Guerrero, V.M. y D. Peña (2000), "Linear Combination of Restrictions and Forecasts in Time Series Analysis", *Journal of Forecasting*, núm. 19, pp. 103-122.
- Guerrero, V.M. y D. Peña (2003), "Combining Multiple Time Series Predictors: a Useful Inferential Procedure", *Journal of Statistical Planning and Inference*, núm. 116, pp. 249-276.
- Hyndman, R.J. y H. Booth (2008), "Stochastic Population Forecasts Using Function Data Models for Mortality, Fertility, and Migration", *International Journal of Forecasting*, núm. 24, pp. 323-342.
- INEGI (1940-2010), *Censos Generales de Población y Vivienda: Distrito Federal y Estado de México*, México, Dirección General de Estadística, Instituto Nacional de Estadística, Geografía e Informática.
- INEGI (1995), *I Censo de Población y Vivienda 1995: Distrito Federal y Estado de México*, México, Dirección General de Estadística, Instituto Nacional de Estadística, Geografía e Informática.
- INEGI (2005), *II Censo de Población y Vivienda 2005: Distrito Federal y Estado de*

- México, México, Dirección General de Estadística, Instituto Nacional de Estadística, Geografía e Informática.
- INEGI (2010), *Censos Generales de Población y Vivienda 2010: Distrito Federal y Estado de México*, México, Dirección General de Estadística, Instituto Nacional de Estadística y Geografía (Bases de Datos).
- Jeon, Y. y M. Shields (2008), "The Impact of Relative Cohort Size on U.S. Fertility, 1913-2001", *IZA Discussion Papers*, 3587, Bonn, Institute for the Study of Labor (IZA).
- Keilman, N., D. Pham y A. Hetland (2002), "Why Population Forecasts Should be Probabilistic – Illustrated by the Case of Norway", *Demographic Research*, vol. 6, pp. 409-454.
- King, G. y S. Samir (2011), "The Future of Death in America", *Demographic Research*, vol. 25, pp. 1-38.
- Land, K. y D. Cantor (1983), "ARIMA Models of Seasonal Variation in U.S. Birth and Death Rates", *Demography*, núm. 20, pp. 541-568.
- Laporte, A. y B. Ferguson (2003), "Income Inequality and Mortality: Time Series Evidence from Canada", *Journal of Infection and Public Health*, vol. 66, núm. 1, pp. 107-117.
- Lee, R. y S. Tuljapurkar (1994), "Stochastic Population Projections for the United States: Beyond High, Medium and Low", *Journal of the American Statistical Association*, vol. 89, núm. 428, pp. 1175-1189.
- Lütkepohl, H. (2005), *New Introduction to Multiple Time Series Analysis*, Nueva York, Springer.
- McNown, R. y S. Rajbhandary (2003), "Time Series Analysis of Fertility and Female Labor Market Behavior", *Journal of Population Economics*, Springer, vol. 16, núm. 3, pp. 501-523.
- McNown, R. y C. Ridao-Cano (2005), "A Time Series Model of Fertility and Female Labour Supply in the UK", *Journal Applied Economics*, núm. 37, pp. 521-532.
- McNown, R. y A. Rogers (1992), "Forecasting Cause-Specific Mortality Using Time Series Methods", *International Journal of Forecasting*, núm. 8, pp. 413-432.
- Okita, Y., W. Pfau y G. Thanh (2009), "A Stochastic Forecast Model for Japan's Population", documento de trabajo, Tokio, National Graduate Institute for Policy Studies (GRIPS Discussion Paper, 9.6).
- Paquette, C. y M. Yescas (2009), "Producción masiva de vivienda en la Ciudad de México: dos políticas en debate", *Revista de la Organización Latinoamericana y del Caribe de Centros Históricos*, núm. 3, pp. 15-26.
- Silva, E., V. Guerrero y D. Peña (2011), "Temporal Disaggregation and Restricted Forecasting of Multiple Population Time Series", *Journal of Applied Statistics*, vol. 38, núm. 4, pp. 799-815.
- Sobrinho, J. (2007), "Patrones de dispersión intrametropolitana en México", *Estudios Demográficos y Urbanos*, vol. 22, núm. 3 (66), pp. 583-617.
- Thompson, P., W. Bell, J. Long y R. Miller (1989), "Multivariate Time Series

Projections of Parametrized Age-Specific Fertility Rates”, *Journal of the American Statistical Association*, núm. 84, pp. 689-699.

Tuljapurkar, Shripad, Lee Ronald y Li Qi (2004), “Random Scenario Forecasts versus Stochastic Forecasts”, *International Statistical Review*, vol. 72, núm. 2, pp. 185-199.

Acerca de los autores

Eliud Silva es actuario egresado de la UNAM, diplomado en Modelos Econométricos Dinámicos por el ITAM, maestro en Demografía por El Colegio de México y doctor en Ingeniería Matemática con especialidad en Estadística por la Universidad Carlos III de Madrid. Se ha desarrollado profesionalmente en compañías de seguros, como asesor de proyectos de investigación en Consejos de Población y en la Secretaría de Salud; fue estadístico del Ceneval y funcionario de Conaculta. Ha impartido cursos en Naciones Unidas (sede en México); también ha sido docente de licenciatura y posgrado en la UNAM, El Colegio de México, la Universidad Carlos III de Madrid, el ITAM y el Instituto Tecnológico de Monterrey. Fue ganador del tercer lugar del Concurso de Investigación Demográfica Gustavo Cabrera de El Colegio de México en su edición 2010. Ha publicado artículos en revistas nacionales y extranjeras relacionados con series de tiempo y tópicos demográficos. Actualmente es profesor de tiempo completo de la Universidad Anáhuac.

Manuel Ordorica cursó la carrera de actuario en la Facultad de Ciencias de la UNAM; es maestro en Demografía por El Colegio de México, y doctor en Ingeniería con especialidad en Investigación de Operaciones por la UNAM. Su tesis de doctorado “El filtro de Kalman en la planeación demográfica” obtuvo mención honorífica. Fue jefe del Departamento de Evaluación y Análisis Demográfico en la Dirección General de Estadística; director de Estudios de Población en el Consejo Nacional de Población; consultor en Educación en Población de la UNESCO; coordinador de la maestría en Demografía y del doctorado en Estudios de Población en El Colegio de México, y fungió como director del Centro de Estudios Demográficos y de Desarrollo Urbano de dicha institución. Forma parte del consejo editorial de la revista *Population* (INED, París). Pertenece al Sistema Nacional de Investigadores y recibió el Premio Nacional de Demografía 1998. Trabaja en el campo de la demografía matemática. Actualmente es secretario general de El Colegio de México.