Proyección de poblaciones carcelarias en Colombia

Sergio David Solano Bejarano
Ingeniero Industrial

Universidad Nacional de Colombia
Facultad de Ciencias
Departamento de Estadística
Bogotá, D.C.
Noviembre de 2018

Proyección de poblaciones carcelarias en Colombia

Sergio David Solano Bejarano
Ingeniero Industrial

Disertación presentada para optar al título de
Master en Ciencias - Estadística

Director
B. Piedad Urdinola Contreras, Ph.D.
Doctor en Demografía

Línea de investigación
Demografía

Universidad Nacional de Colombia
Facultad de Ciencias
Departamento de Estadística
Bogotá, D.C.
Noviembre de 2018

Título en español

Proyección de poblaciones carcelarias en Colombia

Title in English

Prison populations projections for Colombia

Resumen: Se realizaron proyecciones de la población carcelaria en Colombia, usando la información disponible para los años 1991-2016. La información publicada periodicamente no incluye las tasas de transición (ingreso y salida) del sistema, por esta razón se eligieron tres métodos que permiten realizar la proyección a partir de la población observada. Los métodos utilizados son: modelos demográficos para poblaciones pequeñas, modelos ARIMA, y modelos Estado-Espacio. Se comparó el ajuste de cada modelo, sus ventajas y desventajas.

Abstract: The purpose of this thesis is to present projections of the prison population in Colombia, considering available data from 1991-2016. Monthly released data does not include admission or release rates, so we use three methods that work over the total population. The considered methods are: Demographical methods for subnational populations, ARIMA Models, and State-Space Models. We compare the fit of the three models, their advantages and disadvantages.

Palabras clave: Poblaciones carcelarias, series de tiempo, procesos SARIMA, Modelos Estado Espacio, Poblaciones pequeñas

Keywords: Prison populations, time series, SARIMA processes, State Space Models, Subnational Populations

Nota de aceptación

Trabajo de tesis
Aprobado

"Mención Meritoria o Laureada"

Jurado
Jurado uno

Jurado
Jurado dos

Director
B. Piedad Urdinola

Bogotá, D.C., Junio 01 de 2017

Dedicado a

Este documento está dedicado a Hercilia Bejarano, mi madre.

Agradecimientos

Agradezco a Piedad, paciente directora de tesis, por su colaboración; a mis compañeros de estudio: Julian, Jennifer, Juan y Ángela; y a Alanis, por supuesto.

Índice general

Índice general **I**

Índice de tablas **III**

Índice de figuras **IV**

Introducción **VI**

1. Antecedentes teóricos **1**

1.1. Proyecciones de población 1

1.1.1. Proyección de poblaciones pequeñas 2

1.1.2. Aplicaciones nacionales e internacionales 2

1.2. Series de tiempo 3

1.2.1. Modelos SARIMA 5

1.2.2. Modelos de Series de Tiempo Estructurales 8

1.2.3. Modelos Estado-Espacio 9

2. La población carcelaria en Colombia 1991 - 2017 **11**

2.1. Análisis exploratorio 11

2.2. El sistema penitenciario en Colombia 12

2.3. Identificación del modelo 12

3. Modelos SARIMA **14**

3.1. Identificación del modelo 15

3.2. Estimación de parámetros 19

3.3. Tratamiento de outliers 34

3.4. Proyección 37

3.5. Discusión 39

4. Métodos demográficos para poblaciones pequeñas **41**

4.1. Censal ratio 44

4.2. Población privada de la libertad por rango de edad Febrero 2016 a Febrero 2018 46

4.3. Proyecciones Censal Ratio 52

4.4. Discusión 56

5. Modelos Estado-Espacio **58**

5.1. Población privada de la libertad 2013-2016 por delito 59

5.2. Selección de Software 63

5.3. Identificación del modelo 63

5.4. Estimación de parámetros 64

5.5. Proyecciones 2017 - 2020 66

5.6. Discusión 71

6. Comparación **73**

Conclusiones **76**

Trabajo futuro **77**

Bibliografía **78**

Índice de tablas

3.1. Comportamiento de la ACF y PACF en modelos ARMA(p,q) 17
3.2. Comparación de criterios de información 19
3.3. Parámetros del modelo (3,2,1,0,0,2) 20
3.4. Parámetros del modelo (1,2,1,0,0,2) 20
3.5. Comparación de criterios de información 25
3.6. Parámetros del modelo (0,1,1,0,0,1) 25
3.7. Comparación de criterios de información 30
3.8. Parámetros del modelo (1,1,1,0,0,1) 30

4.1. Proyección de la población carcelaria vs real 54

5.1. Parámetros estimados: hurto, hombres 65
5.2. Proyección a treinta y seis meses: hurto, hombres 67
5.3. Proyección de la población carcelaria total, hombres 68
5.4. Proyección de la población carcelaria total, mujeres 69

6.1. Proyecciones y observado enero 2017 a mayo 2018 74
6.2. Error de pronóstico enero 2017 a mayo 2018 74
6.3. Sesgo de pronóstico enero 2017 a mayo 2018 74

Índice de figuras

2.1. Población privada de la libertad 1991 - 2017 11
2.2. Tasa de encarcelamiento según género 1991 - 2017 12
2.3. Población carcelaria por situación jurídica 13

3.1. Variación inter-mensual de población carcelaria 15
3.2. Descomposición de la variación inter-mensual de población carcelaria total . 16
3.3. Descomposición de la variación inter-mensual de población carcelaria total . 16
3.4. Descomposición de la variación inter-mensual de población carcelaria total . 17
3.5. Autocorrelación parcial de la variación inter-mensual 18
3.6. Diagnóstico del SARIMA (1,1,1)(0,0,2), población total 21
3.7. Diagnóstico del SARIMA (1,1,1)(0,0,1), población total 22
3.8. Diagnóstico del SARIMA (3,2,1)(0,0,2), población total 23
3.9. Diagnóstico del SARIMA (1,2,1)(0,0,2), población total 24
3.10. Diagnóstico del SARIMA (1,1,1)(0,0,1), población sindicada 26
3.11. Diagnóstico del SARIMA (2,1,1)(0,0,1), población sindicada 27
3.12. Diagnóstico del SARIMA (0,1,2)(0,0,1), población sindicada 28
3.13. Diagnóstico del SARIMA (0,1,2)(0,0,0), población sindicada 29
3.14. Diagnóstico del SARIMA (1,1,1)(0,0,1), población condenada 31
3.15. Diagnóstico del SARIMA (2,1,1)(0,0,1), población condenada 32
3.16. Diagnóstico del SARIMA (0,1,2)(0,0,1), población condenada 33
3.17. Diagnóstico del SARIMA (0,1,2)(0,0,0), población condenada 34
3.18. Diagnóstico del SARIMA (2,1,1)(2,0,0), población total con variables Dummy 35
3.19. Diagnóstico del SARIMA (0,1,2)(0,0,1), población sindicada con variables Dummy 36
3.20. Diagnóstico del SARIMA (1,1,1)(0,0,1), población condenada con variables Dummy 37

3.21. Proyección de la población carcelaria total 38
3.22. Proyección de la población carcelaria sindicada 38
3.23. Proyección de la población carcelaria condenada 39

4.1. Proyección de la población nacional por rango de edad 43
4.2. Pirámide poblacional, población privada de la libertad, Junio - 2017 47
4.3. Población privada de la libertad. Febrero-2016 a Febrero-2018 47
4.4. Población privada de la libertad. Febrero-2016 a Febrero-2018 48
4.5. Población privada de la libertad. Febrero-2016 a Febrero-2018 49
4.6. Población privada de la libertad. Febrero-2016 a Febrero-2018 50
4.7. Tasa de específica de encarcelamiento. Junio-2017 51
4.8. Tasa específica de encarcelamiento. Junio-2016, Junio -2017 52
4.9. Población privada de la libertad. 2005-2020 53
4.10. Población privada de la libertad. 2005-2020 55
4.11. Población privada de la libertad. 2005-2020 55

5.1. Población carcelaria promedio por delito y género. Julio 2013-Diciembre 2016 59
5.2. Evolución de la población carcelaria masculina, Julio 2013-Diciembre 2016 . 61
5.3. Evolución de la población carcelaria femenina, Julio 2013-Diciembre 2016 . . 62
5.4. Población sindicada y condenada por hurto, hombres 65
5.5. Población carcelaria promedio por delito y género. Octubre 2013-Diciembre 2016 66
5.6. Población carcelaria promedio por delito y género. Octubre 2013-Diciembre 2016 66
5.7. Proyección de población carcelaria o por delito, hombres 2017 - 2020 70
5.8. Proyección de población carcelaria o por delito, mujeres 2017 - 2020 71

6.1. Proyecciones de Población privada de la libertad Enero 2017 - Mayo 2018 . . 73

Introducción

La población carcelaria es el grupo poblacional que se encuentra privado de la libertad en una institución penitenciaria. Dentro de esta población, a aquellos a la espera de juicio se les denomina sindicados, y condenados a quienes están cumpliendo una sentencia.

En Colombia el INPEC (Instituto Nacional Penitenciario y Carcelario), que "es la institución pública administradora del sistema penitenciario y carcelario del país"[18], publica periódicamente la serie histórica de población carcelaria, que da cuenta de la población al cierre de cada mes desde 1991 hasta la fecha, separada por situación judicial (sindicados, condenados) y género. Además publica reportes mensuales desde 2013 con información desagregada por genero, edad, delito, estado judicial y regional.

Con base en estos datos el CONPES 3828 anota que "el promedio anual de crecimiento de la Población Privada de la Libertad (PPL) entre 1993 y 2014 fue de 9,43 % para la población condenada y 4,91 % para la sindicada" [8]. Ambas tasas están muy por encima del 1,35 % anual, crecimiento poblacional nacional estimado por el DANE para 1993 a 2014 [7]. El aumento por encima de la población nacional sugiere dinámicas de crecimiento diferentes al resto de la población.

De otra parte el mismo documento observa que "en el período 1993-2014 la proporción de cupos habilitados creció en 173,59 %, mientras la población privada de la libertad (PPL) creció en un 315,39 %" [8]. El crecimiento acelerado de la población carcelaria, por encima de la oferta de cupos implica un incremento del hacinamiento, lo que a su vez podría incidir en problemas de salud y de orden al interior de las instituciones carcelarias.

En 2016 Colombia ocupó el puesto catorce entre doscientos cincuenta y un países por el tamaño de su población carcelaria (120 914 hbts.) y el cincuenta y uno según la tasa de encarcelamiento (240 por cada 100.000 hbts). Tasa que pasó de 51,5 en el año 2000 a 240 por cada 100.000 hbts en 2016. Con una ocupación del 154 % de las plazas disponibles, resulta relevante contar con proyecciones de la población carcelaria en el corto, mediano y largo plazo.[17]

El CONPES 3828 de 2015 realiza, además, proyecciones de la población carcelaria a través de un modelo de crecimiento exponencial con tasa de crecimiento de 7,23 % anual. Este tipo de modelos, aunque ampliamente usados en proyecciones de población, no resulta conveniente cuando la tasa de crecimiento no es estable a través del tiempo y depende de variables exógenas como: la población nacional, su estructura etaria, las características del sistema penitenciario y del sistema judicial.

En este contexto resulta necesario contar con proyecciones de la población carcelaria en diferentes escenarios, para determinar la necesidad de cupos en el corto y mediano plazo.

Tales proyecciones deben tener en cuenta las dinámicas propias del sistema carcelario y deben permitir estimar el impacto de modificaciones a la política criminal y carcelaria, de forma que permitan una mejor planeación de la cantidad y/o tamaño de los centros carcelarios.

Las poblaciones que se estiman y proyectan a un nivel menor que el nacional se conocen como poblaciones pequeñas (ciudad, departamento, etc...). Dentro de las poblaciones pequeñas se encuentran poblaciones especiales que "se localizan en un área especial por una acción legislativa o administrativa". [32]. La población carcelaria es un tipo de población especial cuyos componentes: nacimientos y defunciones no necesariamente siguen los patrones biológicos que dan vida al modelo exponencial y requieren el uso de técnicas diferentes que permitan un mejor ajuste. Swanson revisa en [32] los métodos de proyección para áreas pequeñas, abordando también las poblaciones especiales y población carcelaria.

Las series de población carcelaria disponibles para Colombia son mediciones de la población al cierre de mes, agrupada por situación judicial y género. Las series utilizadas se tomaron de la página de estadísticas del INPEC [16]. Los proyecciones de población suelen incluir las tasas de nacimiento, migración y mortalidad. En el caso de las poblaciones carcelarias, los equivalentes serían la tasa de ingreso al sistema, la tasa a la que son juzgados, la mortalidad y la duración de las penas, que determina la tasa de salida de los condenados. En el caso de Colombia estos datos no se publican periódicamente, por lo tanto el modelo debe permitir estimar las tasas no observadas. Adicionalmente, estas tasas no son constantes a través del tiempo, pues dependen de la estructura del sistema judicial y del sistema carcelario.

Este trabajo pretende realizar proyecciones de población carcelaria de corto y mediano plazo, comparando los modelos de: series de tiempo, modelos estado espacio y métodos de proyección de poblaciones pequeñas.

El objetivo principal de este documento es generar proyecciones de población privada de la libertad, para Colombia, a partir de los datos históricos del periodo 1991-2016. Los objetivos específicos incluidos son:

- Proyectar la población carcelaria usando los métodos demográficos para áreas pequeñas.
- Proyectar el comportamiento de la población carcelaria a través de series de tiempo SARIMA.
- Aplicar modelos estado espacio para estimar las tasas de transición, cuando: son no observadas, varían en el tiempo y tienen correlación con variables exógenas.
- Comparar el ajuste de los tres métodos.

CAPÍTULO 1

Antecedentes teóricos

En este capítulo se contextualiza de forma teórica el problema de proyección de poblaciones carcelarias, desde una perspectiva demográfica. Posteriormente se presentan, siguiendo a Shumway & Stoffer en [29] y [30], los conceptos asociados al manejo de series de tiempo: modelos SARIMA, Series de tiempo estructurales y modelos estado espacio. El modelo-censal ratio se presenta en el capítulo 4, pues se le realizan ajustes para usarlo con los datos disponibles.

En las secciones dedicadas a los modelos SARIMA y a los modelos estado-espacio se mencionan los métodos de estimación y proyección usados en los paquetes seleccionados.

1.1. Proyecciones de población

"Una estimación poblacional consiste en determinar el tamaño o las características de una población, para el momento actual o para uno anterior, en ausencia de información. Cuando se realizan un conjunto de supuestos sobre el comportamiento de los vitales hacia el futuro, hablamos de proyección, y cuando se escoge un escenario como el más probable, hablamos de pronóstico" [32].

Para incluir la incertidumbre en las proyecciones de población Lee(1994) considera los siguientes métodos [20]:

- El enfoque de escenarios alto, medio y bajo: Asume comportamientos fijos para la fertilidad, la mortalidad y las migraciones durante el período de proyección, basado en algunos supuestos [20].
- Análisis estocásticos
  - Análisis ex-post: consiste en evaluar el error de pronóstico en proyecciones anteriores y aplicarlo a las nuevas proyecciones [20].
  - Simulación estocástica: Permite hacer proyecciones de población, al asignar una distribución de probabilidad a las tasas vitales (mortalidad, natalidad, migraciones) [20].

- Modelos estocásticos de la tasa de crecimiento: Consiste en estimar la tasa de crecimiento del total de la población; aunque permite estimar intervalos de confianza, no permite separar la proyección de las tasas vitales, ni de las franjas etarias [20].
- Matrices de Leslie con modelos estimados para las tasas vitales: Al usar matrices de Leslie se estima la población por rangos etarios para un instante i, y se calcula la población en el instante i + 1 aplicando la natalidad y la mortalidad proyectadas para el período i. Puesto que las series de población carcelaria no se publican separadas por edad, no se abordará esta técnica.

1.1.1. Proyección de poblaciones pequeñas

La proyección de áreas pequeñas es entendida como la proyección a un nivel geográfico menor al nacional. Estas proyecciones pueden incluir, departamentos, ciudades o poblaciones especiales [32]. "Una población especial es un grupo poblacional que se encuentra restringido a un área por una medida administrativa o legislativa. Dentro de los grupos usualmente considerados se encuentran las prisiones, universidades, hospitales e instituciones militares". Este tipo de población puede tener una estructura etaria y de sexo, y unos vitales diferentes al resto de la población; además no suelen envejecer en el mismo lugar, lo que permite mantener una estructura etaria que no varía a través del tiempo [32].

1.1.2. Aplicaciones nacionales e internacionales

Las proyecciones de poblaciones carcelarias oficiales analizadas corresponden, en buena parte, a los métodos expuestos en los capítulos anteriores: Proyecciones por escenarios, proyección de la tasa de crecimiento, modelos ARIMA para las tasas de ingreso y salida.

En Colombia, el CONPES 3828 proyectó la población carcelaria usando la tasa media de crecimiento anual (1993-2014) sin incluir una medida de incertidumbre [8]. Este tipo de proyecciones que de forma subyacente incluyen el supuesto de un sistema estable (que seguirá comportándose de la misma manera a través del tiempo) omiten el efecto de los cambios en las tasas de ingreso y salida del sistema carcelario, asociados a cambios en la política criminal o a cambios en la población.

El Reino Unido hasta 2015 realizaba una proyección por escenarios (alto, medio y bajo), año en el cual cambió a un modelo de proyección de la media y su incertidumbre. La incertidumbre se incluyó a través de un análisis ex-post, de la desviación de la proyección en años anteriores [19].

El departamento de Justicia de los Estados Unidos realizó estimaciones de la población carcelaria por estado para el período 2013-2014. Las estimaciones parten del censo de prisiones 1993-2014. Estas proyecciones se puede enmarcar dentro de las proyecciones de áreas pequeñas [24].

El bureau de estadísticas e investigación del crimen en Australia proyecta las tasas de arresto y sentencia usando modelos ARIMA; a partir de estas tasas proyecta la población carcelaria. Estas proyecciones incluyen un período de validación de tres años. Los resultados mostraban que la serie real se encuentra dentro de los intervalos de confianza de la proyección, cercano a la proyección de la media [35].

Blummstein en [2] desarrolla un método de proyección basado en los componentes demográficos, tasas específicas de arresto por delito y reincidencias, a partir de estos datos proyecta el tamaño y la composición de las poblaciones.

Con los datos libres disponibles en Colombia no se podría utilizar el enfoque presentado por Wan et al. en [35] ni el método de Blummstein presentado en [2], pues las tasas de ingreso al sistema y sentencia no se publican. La tesis propone tres métodos de proyección para situaciones donde no se cuenta con el registro de los vitales o su equivalente en la población analizada. En el escenario de las poblaciones carcelarias en Colombia, con los datos abiertos disponibles, resulta conveniente usar las series SARIMA en los conteos individuales de población sindicada y condenada, y los modelos estado-espacio para la serie bivariada. El método Censal-ratio, para proyectar poblaciones pequeñas, permite proyectar la población carcelaria como proporción de la población total.

1.2. Series de tiempo

Una serie de tiempo es un conjunto de observaciones y_t asociadas a un instante de tiempo t. Es usual referirse como series de tiempo, tanto a las realizaciones y_t como a las variables aleatorias Y_t que las generan [5].

En una regresión lineal clásica, una variable Y es explicada o predicha en función de un conjunto de n variables X. La diferencia entre el valor observado en el instante t y_t y el valor predicho se suponen provenientes de un proceso aleatorio con media cero [6].

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} ... \beta_n x_{tn} + w_t \tag{1.1}$$

También es posible representar las observación y_t en función de las observaciones anteriores como se denota en la ecuación 1.2. El valor presente se denota como t y el inmediatamente anterior como $t-1$.

$$y_t = \beta_0 + \beta_1 y_1 + \beta_2 y_2 + ... + \beta_{t-1} y_{t-1} + w_t \tag{1.2}$$

La variable aleatoria Y en el instante t, es una combinación lineal de los valores observados de la variable aleatoria, más un error aleatorio, donde $\{w_1, w_2, ...w_t\}$ son independientes e idénticamente distribuidas (i.i.d.) [29]. Este supuesto no suele cumplirse, razón por la cual el análisis de series de tiempo se ha desarrollado como un área particular de la estadística [5].

Se define un proceso como ruido blanco a la colección de variables aleatorias z_t, no correlacionadas, con media 0 y varianza σ_w^2 [29].

Bajo un modelo de k parámetros, el estimador máximo verosímil de la varianza es:

$$\hat{\sigma}_k^2 = \frac{SSE(k)}{n} = \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t) \tag{1.3}$$

Para seleccionar el modelo que mejor explique los datos observados se recurre a los criterios AIC (Criterio de Información de Aikake) y BIC (Criterio de Información Bayesiano), que penalizan los modelos con mayor cantidad de parámetros y están definidos como:

$$AIC = ln(\hat{\sigma_k}^2) + \frac{n + 2k}{n} \tag{1.4}$$

$$BIC = ln(\hat{\sigma_k}^2) + \frac{k(ln(n))}{n} \tag{1.5}$$

donde k es la cantidad de parámetros en el modelo y n es la cantidad de observaciones [33]. Tanto en el AIC como en el BIC se selecciona el modelo de k parámetros con menor valor.

Para medir la dependencia entre las observaciones recurrimos a la función de autocovarianza y a la función de autocorrelación (ACF). La función de autocovarianza se define como [29]:

$$\gamma(s,t) = cov(y_s, y_t) = E[(y_s - \mu_s)(y_t - \mu_t)] \tag{1.6}$$

para todo s y t, donde $\mu = EY_t$ y k representa la cantidad de pasos entre dos observaciones. La autocovarianza mide la dependencia lineal entre dos puntos en la misma serie en diferentes instantes. [29].

La función de autocorrelación (ACF) se define como [5]:

$$\rho(s,t) = \frac{\gamma(s,t)}{\sqrt{\gamma(s,s)\gamma(t,t)}} \tag{1.7}$$

Para poder realizar predicciones sobre el estado futuro de una serie de tiempo, es necesario suponer que el comportamiento de la serie es estable a través del tiempo, incluso con un componente aleatorio. La estacionaridad es el concepto que permite articular esta necesidad. Estamos ante un **proceso estacionario** cuando las variables $\{X_1, ..., X_k\}$ tienen la misma distribución conjunta que $\{X_{h+1}, ..., X_{h+k}\}$, para todos los enteros h y k [5].

La estacionaridad débil se presenta cuando $E[Y_t]$ y $E[Y_t, Y_{t+h}]$ son independientes de t, es decir: i) Presenta media μ constante e independiente de t ii) La función de autocovarianza $\gamma(h, h+k)$ depende solamente de la cantidad de pasos que separa las observaciones (k). En adelante, al referirse a estacionaridad, se tratará de estacionaridad débil, a menos que se indique lo contrario [29].

Cuando la serie no tiene un comportamiento estacionario se puede recurrir a diferenciar la serie, este proceso al ser de uso frecuente tiene su propia notación. La diferencia $\bigtriangledown y_t = y_t - y_{t-1}$ se conoce como diferencia de primer orden. La diferencia de segundo orden se define como $\bigtriangledown^2 y_t = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2})$.

El operador de rezago B se define como [29]:

$$By_t = y_{t_1} \tag{1.8}$$

tal que:

$$\bigtriangledown y_t = (1 - B)y_t \tag{1.9}$$

La notación se puede extender de forma que:

$$\bigtriangledown^d y_t = (1-B)^d y_t \tag{1.10}$$

La función de auto-correlación (ACF) de una serie de tiempo estacionaria se define como [29]:

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} \tag{1.11}$$

donde $\gamma(h) = cov(y_{t+h}, y_t) = E[(y_{t-h} - \mu)(y_t - \mu)]$ y $\bar{y}$ denota la media del proceso.

Para estimar la ACF utilizamos la función de autocorrelación muestral definida como [29]:

$$\hat{\rho(h)} = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} = \frac{\sum_{t=1}^{n-h}(y_{t+h} - \bar{y})(y_t - \bar{y})}{\sum_{t=1}^{n}(y_t - \bar{y})^2} \tag{1.12}$$

con $h = 0, 1, ..., n-1$.

Si tenemos las variables aleatorias X, Y, Z y queremos medir la correlación entre las variables Z y X, descontando el efecto de la variable Y, recurrimos a la función de correlación parcial $\rho_{XY|Z}$, que se calcula como [29]:

$$\rho_{XY|Z} = corr\{X - \hat{X}, Y - \hat{Y}\} \tag{1.13}$$

Al adaptar este concepto a series de tiempo tenemos que la función de autocorrelación parcial de un proceso estacionario y_t, α_{hh}, con $h = 1, 2, ..., t$ se define como [29]:

$$\alpha_{11} = corr(y_1, y_0) = \rho(1) \tag{1.14}$$

y

$$\alpha_{hh} = corr(y_h - \hat{y_h}, y_0 - \hat{y_0}), h \geq 2 \tag{1.15}$$

donde $\hat{y_h}$ es la regresión de y_h en $\{y_1, y_2, ..., y_{h-1}\}$ y $\hat{y_0}$ es la regresión de y_0 en $\{y_1, y_2, ..., y_{h-1}\}$.

En un proceso estacionario y_t la PACF es la correlación entre y_t, y_{t+h} una vez controlado por los elementos entre ellos $\{y_{t-h+1}, y_{t-h+2}, ..., y_{t-1}\}$ [29].

1.2.1. Modelos SARIMA

Box & Jenkins proponen una aproximación iterativa de cuatro etapas para la selección de un modelo [3]:

1. Selección de una clase de modelos, con base en la teoría y la práctica.
2. Identificación del modelo, donde se seleccionan un conjunto de parámetros que permitan explicar el sistema con parsimonia.
3. Estimación de parámetros.

4. Chequeo diagnóstico, para detectar fallas en el ajuste. Si se detectan fallas en el ajuste, se regresa al segundo paso.

Shumway en [29] plantea que un proceso es auto-regresivo de orden p o AR(p) si:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} ... + \phi_p y_{t-p} + w_t \tag{1.16}$$

Donde y_t es estacionaria con $\mu = 0$; $\phi_1, \phi_2, ..., \phi_p$ son constantes ($\phi_p \neq 0$) y los w_t son independientes e idénticamente distribuidos. Cuando la $\mu \neq 0$

$$y_t - \mu = \phi_1(y_{t-1} - \mu) + \phi_2(y_{t-2} - \mu) ... + \phi_p(y_{t-p} - \mu) + w_t \tag{1.17}$$

Que se puede representar como:

$$y_t = \mu(1 - \phi_1 ... - \phi_p) + \phi_1(y_{t-1}) + \phi_2(y_{t-2}) ... + \phi_p(y_{t-p}) + w_t \tag{1.18}$$

También es posible utilizar el operador B para representar la serie 1.16 como:

$$y_t - \phi_1(y_{t-1}) - \phi_2(y_{t-2}) ... - \phi_p(y_{t-p}) = w_t \tag{1.19}$$

$$(1 - \phi_1 B - \phi_2 B^2 ... - \phi_p B^p) y_t = w_t \tag{1.20}$$

o en una notación simplificada:

$$\phi(B) y_t = w_t \tag{1.21}$$

Se considera un proceso de promedio móvil de orden q o MA(p) si:

$$y_t = \theta_1 w_{t-1} + \theta_2 w_{t-2} ... \theta_q w_{t-q} + w_t \tag{1.22}$$

Donde $\theta_1, \theta_2, ..., \theta_q$ son parámetros ($\phi_p \neq 0$) y los w_t son independientes e idénticamente distribuidos y y_t tiene media cero. A diferencia de los procesos auto-regresivos, los procesos de promedio movil son estacionarios sin importar el valor de $\theta_1, \theta_2, ..., \theta_q$ [29]. Este proceso también se puede representar usando el operador de rezago al ponerlo en la forma:

$$y_t = (\theta_1 B + \theta_2 B^2 ... \theta_q B^q + 1) w_t \tag{1.23}$$

o en forma simplificada:

$$y_t = \theta(B) w_t \tag{1.24}$$

Un proceso se considera ARMA(p,q) si y_t es estacionaria y [29] :

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \theta_1 w_{t-1} + \theta_2 w_{t-2} ... \theta_q w_{t-q} + w_t \tag{1.25}$$

Con $E(y_t) = \mu = 0$. Cuando $(\mu \neq 0)$ se puede representar como:

$$y_t = \mu(1-\phi_1...-\phi_p)+\phi_1 y_{t-1}+\phi_2 y_{t-2}+...+\phi_p y_{t-p}+\theta_1 w_{t-1}+\theta_2 w_{t-2}...\theta_q w_{t-q}+w_t \quad (1.26)$$

A este modelo se le conoce como ARMA(p,q). Usando las ecuaciones 1.21 y 1.24 se puede representar como:

$$\phi(B)y_t = \theta(B)w_t \quad (1.27)$$

Una técnica común para tratar con series no estacionarias consiste en diferenciarlas hasta que la serie obtenida sea ARMA(p,q) [5]. Si la serie se diferencia d veces el proceso se nota como ARIMA(p,d,q).

Un proceso ARIMA (i,1,j) resulta al considerar una serie de la forma [29]:

$$y_t = \alpha + y_{t-1} \quad (1.28)$$

Tal que el proceso $y_t - y_{t-1}$ es un proceso ARMA(p,q).

Los criterios para identificar el orden de un ARMA son [4]:

- Cuando la función de autocorrelación (ACF) se reduce progresivamente, y la función de autocorrelación parcial (PACF) no tiene picos en rezagos luego de p, es un proceso autoregresivo de orden p AR(p)
- Cuando la función de autocorrelación tiene un pico en el rezago q, y la función de autocorrelación parcial se reduce progresivamente, es un proceso de media movil de orden q MA(q)
- Si ambas funciones se reducen gradualmente, se trata de un ARMA (p,q)

Cuando las series presentan un comportamiento estacional se recurre al proceso de diferenciación estacional de la forma: [33]

$$\bigtriangledown_s y_t = y_t - y_{t-s} = (1 - B^s)y_t \quad (1.29)$$

Un proceso autoregresivo de orden P, promedio móvil de orden Q estacional $ARMA(P,Q)_s$, donde s corresponde a la periodicidad de la serie se define como [29]:

$$\Phi_P(B^s)y_t = \Theta_Q(B^s)w_t \quad (1.30)$$

donde $\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - ... - \Phi_P B^{Ps}$ se denomina operador estacional autoregresivo de orden P y $\Theta_Q(B^s) = 1 + \Theta_1 B^s + \Theta_2 B^{2s} + ... + \Theta_Q B^{Qs}$ se denomina operador estacional promedio movil de orden Q [29].

Un proceso SARIMA(p,d,q)(P,D,Q) se define como:

$$\Phi_P(B^s)\phi_p(B^s) \bigtriangledown^D \bigtriangledown^d y_t = \Theta_Q(B^s)\theta_q(B)w_t \quad (1.31)$$

donde w_t es un proceso gausiano de ruido blanco.

El paquete *astsa* se ha utilizado para estimar los modelos ARIMA en el capitulo 3. La función *sarima* del paquete *astsa* es un manejador de la función *arima* del paquete base. [31]. La función *arima* utiliza la función *optim* para encontrar estimadores de máxima verosimilitud [27].

El paquete *astsa* se ha utilizado para realizar los pronósticos en el capitulo 3. La función $sarima.for$ del paquete *astsa* es un manejador de la función $predict.Arima$ del paquete base [31]. La función $predict.Arima$ utiliza la función $KalmanForecast$. En la siguiente sección se hablará en detalle del filtro de Kalman [27].

Una vez se ha realizado la estimación del modelo, se analiza el comportamiento del error. Para descartar autocorrelación en el error se utiliza la estadística Q de Ljung-Box, definida como:

$$Q = n(n+2)\sum_{n=1}^{H}\frac{\hat{\rho_e}^2(h)}{n-1} \tag{1.32}$$

donde H es el rezago analizado, $\hat{\rho_e}(h)$ es la autocorrelación estimada del error en el rezago h. La hipótesis nula es que los errores son no correlacionados. Bajo la hipótesis nula $Q \sim \chi^2_H$ [21] [29].

1.2.2. Modelos de Series de Tiempo Estructurales

En economía y ciencias sociales hay fenómenos que se pueden descomponer en dos componentes básicos, un comportamiento estable de largo plazo (tendencia) y un comportamiento que se repite cada cierto tiempo, usualmente anual (estacionalidad). Estos fenómenos se pueden representar de la forma: [10]

$$observado = tendencia + estacionalidad + irregularidad. \tag{1.33}$$

o en forma multiplicativa

$$observado = tendencia \times estacionalidad \times irregularidad. \tag{1.34}$$

La ecuación 1.34 puede ser llevada a la forma 1.33 al tomar logaritmo en cada lado de la ecuación.

En los modelos de series de tiempo estructurales, al representar en componentes de tendencia, estacionalidad y ciclo, estos tienen una interpretación directa. Más que representar el proceso que genera los datos, separa la serie en componentes con una interpretación directa de cada uno. En los modelos de series de tiempo estructurales, los componentes no son fijos, sino que se consideran generados por impactos aleatorios [10].

1.2.3. Modelos Estado-Espacio

En un modelo estado espacio una serie de tiempo (multiple) observada $z_1, ..., z_t$ depende de un estado y_t , posiblemente no observado, que se comporta siguiendo un proceso estocástico. La relación entre y_t y z_t está dada por la *ecuación de medida*: [22]

$$z_t = H_t y_t + v_t \tag{1.35}$$

donde H_t es una matriz que puede o no depender del tiempo t y v_t es el error de observación, que se asume usualmente como un proceso de ruido. El vector de estado es generado como:

$$y_t = B_{t-1} y_{t-1} + w_{t-1} \tag{1.36}$$

La matriz B_t es una matriz de coeficientes que puede depender de t y w_t es un proceso de ruido [22].

Una forma de representar los modelos de series de tiempo estructurales es a través de los modelos Estado-Espacio, que además permiten representar otros modelos como Regresión Lineal o Modelos ARIMA [9].

Los modelos Estado-Espacio o Modelos Lineales Dinámicos (DLM), en su forma más simple emplean un vector autoregresivo de la forma [30]:

$$\boldsymbol{y_t} = \Phi \boldsymbol{y_{t-1}} + \boldsymbol{w_t} \tag{1.37}$$

donde $\boldsymbol{y_t}$ es un vector de p componentes. $\boldsymbol{w}_t$ es un vector de $p \times 1$ variables aleatorias independientes e idénticamente distribuidas normalmente con media cero y matriz de covarianzas $\boldsymbol{Q}$. Asumimos que el proceso es normal con un vector $\boldsymbol{y_0}$ con media $\boldsymbol{\mu_0}$ y matriz de covarianzas $\sum$ [30].

$$\boldsymbol{z_t} = A_t \boldsymbol{y_t} + \boldsymbol{v_t} \tag{1.38}$$

donde A_t es de tamaño $q \times p$. $\boldsymbol{v_t}$ se asume ruido blanco, con matriz de covarianzas R. Se suponen $\boldsymbol{w_t}$ y $\boldsymbol{v_t}$ no correlacionados.

Holmes en [12] denomina a los modelos de esta forma modelos Estado Espacio Multivariados Auto Regresivos (MARSS). Sobre esta estructura de modelos se realizan las aplicaciones en el capítulo 5.

Variables exógenas o inputs fijos pueden ingresar en los estados o las observaciones. En este caso tendremos un vector u_t de tamaño $r \times 1$:

$$\boldsymbol{y_t} = \Phi \boldsymbol{y_{t-1}} + \boldsymbol{\Upsilon u_t} + \boldsymbol{w_t} \tag{1.39}$$

$$\boldsymbol{z_t} = A_t \boldsymbol{y_t} + + \boldsymbol{\Gamma u_t} + \boldsymbol{v_t} \tag{1.40}$$

con $\boldsymbol{\Upsilon}$ y $\boldsymbol{\Gamma}$ de dimensión $p \times r$ y $q \times r$, respectivamente.

El objetivo es tener estimaciones de y_t a partir de las observaciones $z_1, z_2, ..., z_s$. Cuando $s < t$ se trata de predicción. Cuando $s = t$ se habla de filtrado y para $s > t$ de suavizado [30].

$$\boldsymbol{y_t^s} = E\{\boldsymbol{y_t}|\boldsymbol{z_s}\} \tag{1.41}$$

$$\boldsymbol{P_t^s} = E\{(\boldsymbol{y_{t1}} - \boldsymbol{y_{t1}^s})|(\boldsymbol{y_{t2}} - \boldsymbol{y_{t2}^s})'\} \tag{1.42}$$

Con valores iniciales $\boldsymbol{x_0^0} = \boldsymbol{\mu_0}$ y $P_0^0 = \Sigma_0$ para $t = 1, 2, ..., n$

$$\boldsymbol{y_t^{t-1}} = \Phi \boldsymbol{y_{t-1}^{t-1}} + \Gamma \boldsymbol{u_t} \tag{1.43}$$

$$\boldsymbol{P_t^{t-1}} = \Phi \boldsymbol{P_{t-1}^{t-1}} \Phi' + \boldsymbol{Q} \tag{1.44}$$

con

$$\boldsymbol{y_t^t} = \boldsymbol{y_t^{t-1}} + K_t(\boldsymbol{z_t} + A_t \boldsymbol{y_t^{t-1}} - \Gamma \boldsymbol{u_t}) \tag{1.45}$$

$$\boldsymbol{P_t^t} = [I - K_t A_t] P_t^{t-1} \tag{1.46}$$

donde

$$\boldsymbol{K_t} = P_t^{t-1} A_t' [A_t P_t^{t-1} A_t' + R]^{-1} \tag{1.47}$$

K_t se conoce como ganancia de Kalman. Predicciones para t >n se realizan con 1.43, con condiciones iniciales y_n^n y P_n^n.

CAPÍTULO 2

La población carcelaria en Colombia 1991 - 2017

2.1. Análisis exploratorio

El INPEC publica mensualmente la serie población carcelaria. Esta serie contiene la población carcelaria desde 1991, separada por situación jurídica (condenados, sindicados) y género.

La población carcelaria total entre 1991 y enero 2017 se ha cuadruplicado, al pasar de 32.036 a 128.125 internos. Ver figura 2.1. Aunque la mayoría de los internos son hombres, la población carcelaria femenina ha crecido a un ritmo aún más acelerado, al quintuplicar su población entre 1991 y enero 2017 (pasa de 1633 personas a 7800).

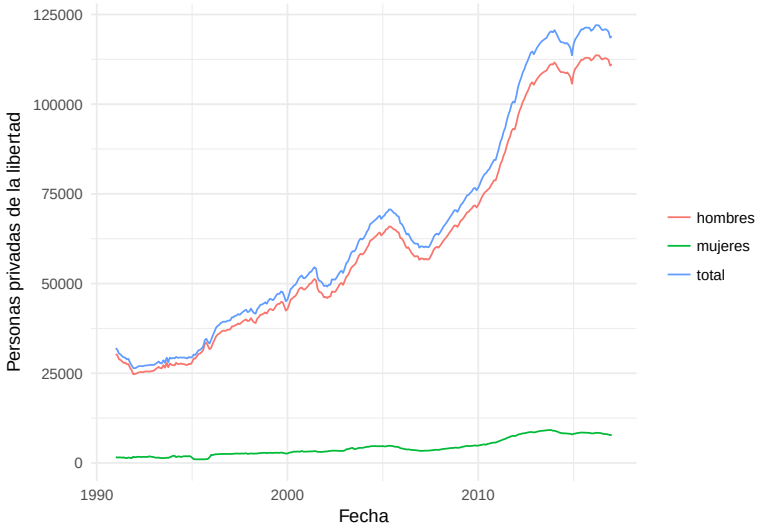

FIGURA 2.1. Población privada de la libertad 1991 - 2017

Fuente: INPEC Elaboración propia

El incremento en la población carcelaria podría tomarse como un efecto del crecimiento de la población colombiana. Para validar este supuesto calculamos la tasa de encarcelamiento, que mide la cantidad de personas encarceladas por cada cien mil habitantes. Este

indicador pasó de 92 personas por cada cien mil habitantes en enero de 1991 a 242 en enero de 2016. Tal incremento se puede ver tanto en hombres como en mujeres. Ver figura 2.2.

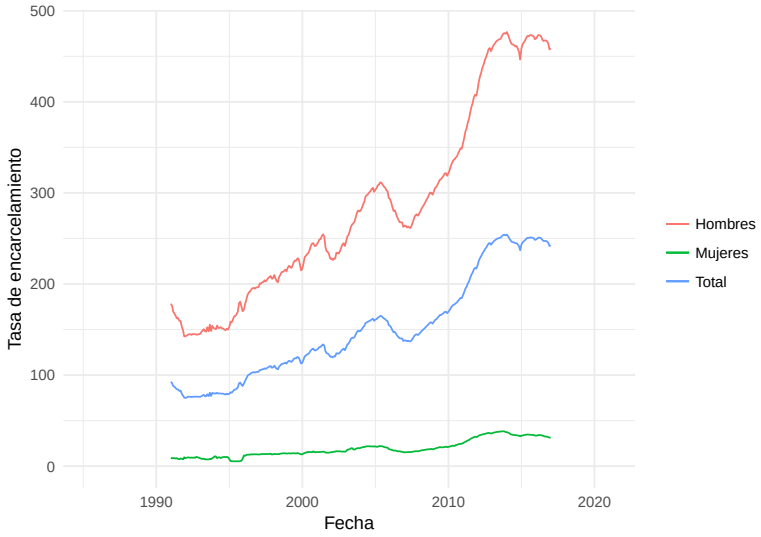

FIGURA 2.2. Tasa de encarcelamiento según género 1991 - 2017

Fuente: INPEC Elaboración propia

La tasa de encarcelamiento es un indicador que varía según la edad y género, siendo más elevado en los hombres que en las mujeres y más algo en los hombres jóvenes, que en los hombres mayores [2]. Otra posible explicación al cambio en la tasa de encarcelamiento es un cambio en la pirámide poblacional en el periodo analizado. No obstante, no podemos confirmar o refutar esta hipótesis pues la serie de tiempo, contenida en los datos de libre acceso no se encuentra desagregada por edad.

2.2. El sistema penitenciario en Colombia

La población carcelaria se ve afectada por dos políticas, la política penitenciaria, que determina las condiciones de privación de la libertad y la política criminal que determina las causas de encarcelamiento y la duración de las penas. [8]

A la población privada de la libertad antes del juicio se le denomina población sindicada, y a aquellos que han sido juzgados y se encuentran cumpliendo la sentencia se les denomina población condenada. La evolución de la población según situación jurídica se puede observar en la figura 2.3

2.3. Identificación del modelo

Podemos modelar el sistema penitenciario de la siguiente manera:

$$S_t = \alpha N_t + (1 - \gamma) S_{t-1} \tag{2.1}$$

$$C_t = C_{t-1} - \omega C_{t-1} + \beta \gamma S_{t-1} \tag{2.2}$$

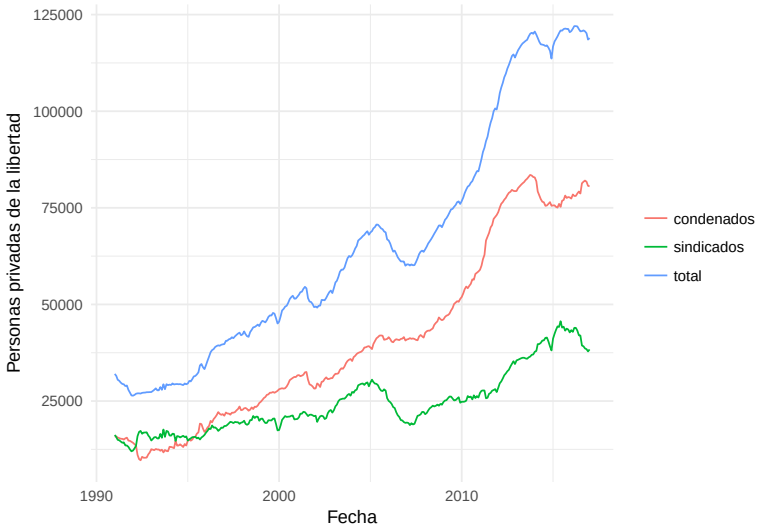

FIGURA 2.3. Población carcelaria por situación jurídica

Fuente: INPEC Elaboración propia

N_t = población nacional en el periodo t
S_t = población de sindicados en el periodo t
C_t = población de condenados en el periodo t
α = proporción de la población libre que ingresa al sistema carcelario
γ = proporción de sindicados que es juzgada cada periodo
β = proporción de sindicados que han sido encontrados culpables
ω = proporción de condenados que terminan su condena cada periodo.

CAPÍTULO 3

Modelos SARIMA

En este capítulo se utilizarán los conteos de población privada de la libertad, separados por situación judicial (sindicados y condenados). Se utilizan 312 observaciones mensuales, desde enero 1991 hasta diciembre 2016.

Se realizan proyecciones de la población carcelaria total, la población sindicada y la condenada. No se considera interacción entre la población sindicada y condenada, y se tratan como series independientes.

Las proyección de la población total será la suma de las proyecciones de sindicados y la de condenados. Realizar la proyección individual de la serie tiene como propósito comparar la tendencia, para ver coherencia entre ambas proyecciones. Un primer ejercicio, antes de retirar outliers arrojó proyecciones con tendencia diferente; mientras la proyección de la serie total arrojaba una tendencia a la baja, la suma de sindicados y condenados indicaba una tendencia al alza, por el contrario las proyecciones finales mantienen una tendencia similar.

En cada una de las series se sigue el siguiente procedimiento: identificación del modelo, donde se busca separar componentes estacionales, la tendencia y la autocorrelación; estimación de parámetros y selección de un modelo; identificación y tratamiento de outliers y proyección. La estimación de los parámetros y la proyección de las series se realiza usando el paquete base del software R [27], y el paquete astsa [31], cuyo uso es discutido en detalle por Shumway [29].

En adelante nos referiremos a la función de autocorrelación muestral como ACF y a la función de autocorrelación parcial muestral como PACF.

Retomando la notación del marcó teórico: Un modelo autoregresivo de orden p, promedio móvil de orden q integrado de orden d; con un componente estacional de rezago anual $S = 12$, integrado de orden D, autoregresivo estacional de orden P, promedio móvil de orden Q se notará: $SARIMA(p, d, q)(P, D, Q)[12]$. En las tablas que contienen los estimados se usa ARp y MAq para el componente autoregresivo y de promedio móvil, y $SMAQ$ y $SARP$ para los componentes estacionales de promedio móvil y autoregresivos. El componente estacional se analiza con periodicidad anual (rezago de 12 periodos), y en los modelos SARIMA donde no se haga explicita, se referirá a 12 periodos.

3.1. Identificación del modelo

Antes de realizar la proyección de una serie de tiempo es necesario identificar el modelo que explique adecuadamente su comportamiento. Aunque se conoce, por la estructura del proceso que genera los datos, que las series de población sindicada y condenada no son independientes, podemos simplificar la proyección, tratándolas como si lo fueran. En este caso, los modelos ARIMA y SARIMA resultan apropiados, pues permiten explicar separadamente cada observación en función del comportamiento histórico de la serie.

El capítulo anterior sugería que la población carcelaria, tanto sindicada como condenada, tiene una marcada tendencia al alza. En este caso una herramienta útil es mostrar gráficamente la variación mes a mes de la población. 3.1. Para separar la serie en una tendencia, un componente estacional y un componente aleatoria, se utiliza la función "decompose"de R base [27].

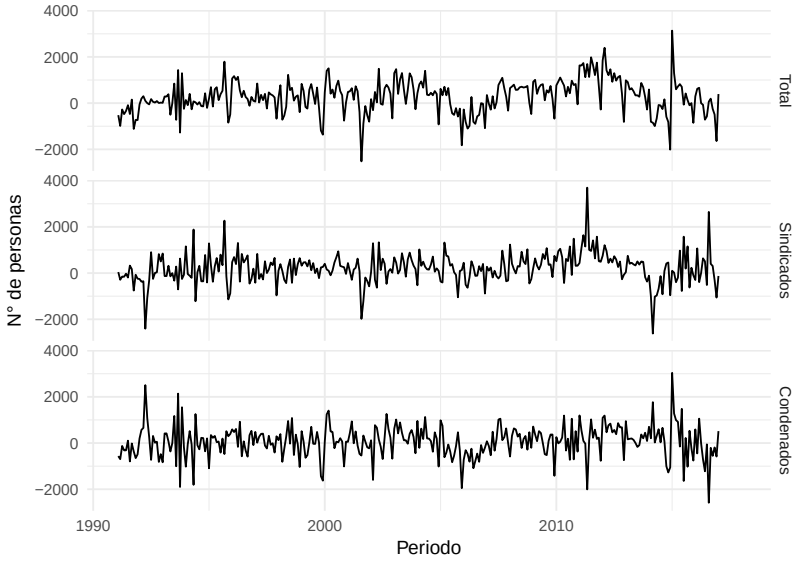

FIGURA 3.1. Variación inter-mensual de población carcelaria, sindicados y condenados

Fuente: INPEC Elaboración propia

La serie tiene un componente estacional marcado, con una reducción de la población carcelaria en diciembre. La variabilidad del componente aleatorio es elevada. La tendencia parece tener cambios estructurales en algunos periodos, por ejemplo reducción de la población carcelaria entre 2005-2007, y 2012 - 2015, e incrementos de la población de magnitud mayor al promedio entre 2008 y 2012.

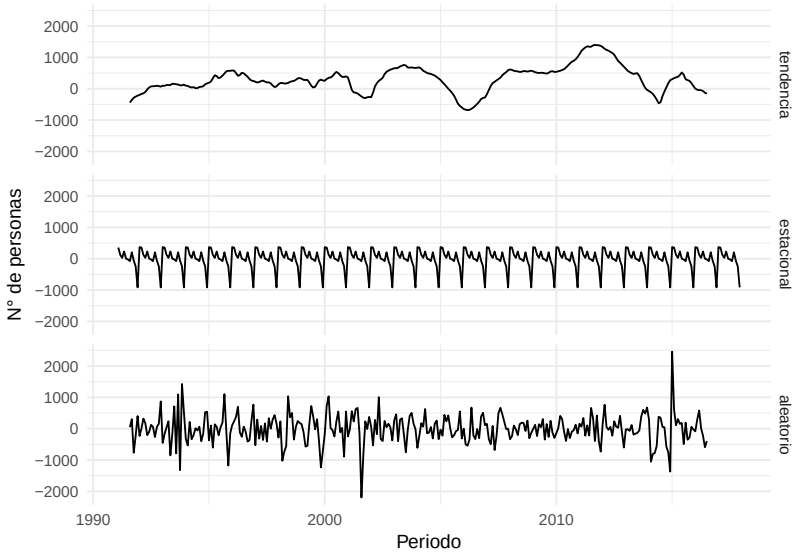

FIGURA 3.2. Variación mensual de población carcelaria descompuesta por tendencia, estacionalidad y componente aleatorio.

Fuente: INPEC Elaboración propia

La mayor parte de la variación de la población total se puede asociar con variaciones en a población sindicada 3.3. La población condenada muestra un comportamiento más estable, con un componente estacional de menor magnitud, en la figura 3.4

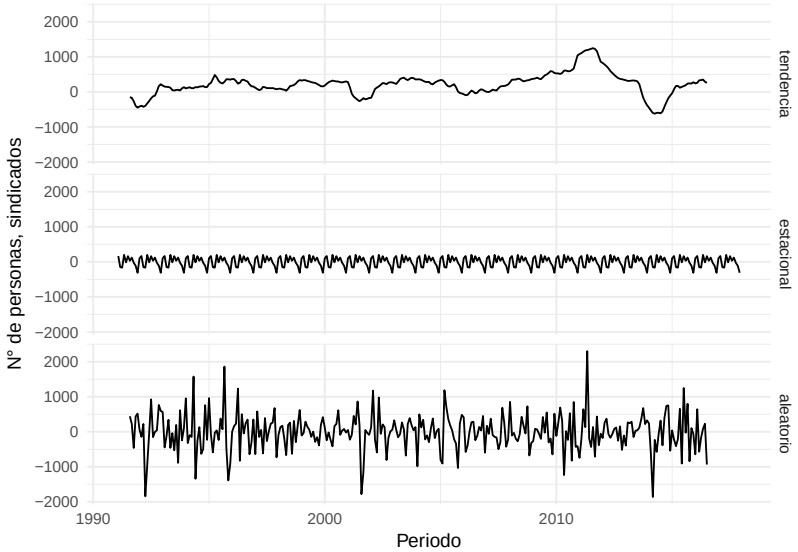

FIGURA 3.3. Variación mensual de población carcelaria sindicada, descompuesta por tendencia, estacionalidad y componente aleatorio.

Fuente: INPEC Elaboración propia

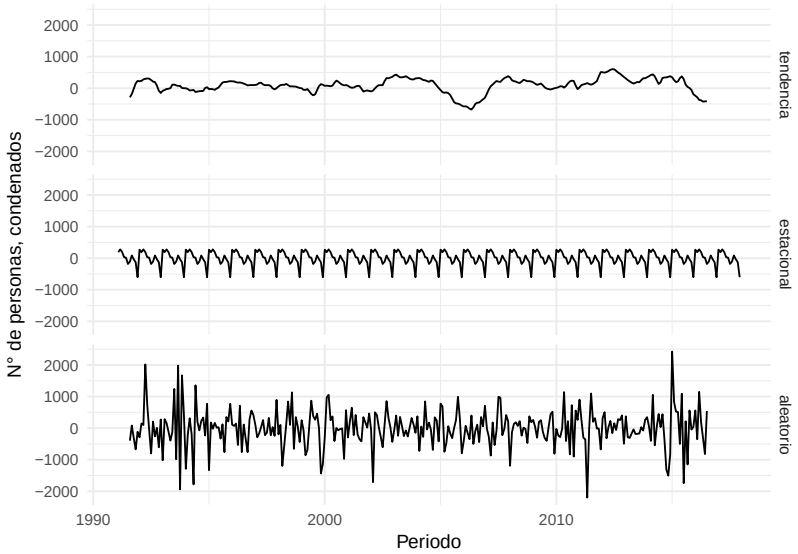

FIGURA 3.4. Variación mensual de población carcelaria condenada, descompuesta por tendencia, estacionalidad y componente aleatorio.

Fuente: INPEC Elaboración propia

El crecimiento sostenido entre 1991 y 20017, sugiere un proceso estocástico integrado. Se trabajo entonces en cada serie diferenciando en los rezagos 1 y 12.

Sobre cada serie diferenciada en rezago 1 (población total, población sindicada y población condenada) se realiza la función de autocorrelación y autocorrelación parcial , que se presentan en la figura 3.5. Estas funciones son usadas como herramienta de diagnóstico, para identificar modelos adecuados en cada serie.

Con base en la tabla 3.1 se realiza una revisión del comportamiento de las series de población carcelaria, población sindicada y población condenada. [29]

TABLA 3.1. Comportamiento de la ACF y PACF en modelos ARMA(p,q)

	AR(p)	**MA(q)**	**ARMA(p,q)**
ACF	reduce progresivamente	se corta en el rezago q	se reduce progresivamente
PACF	se corta luego del rezago p	se reduce progresivamente	se reduce progresivamente

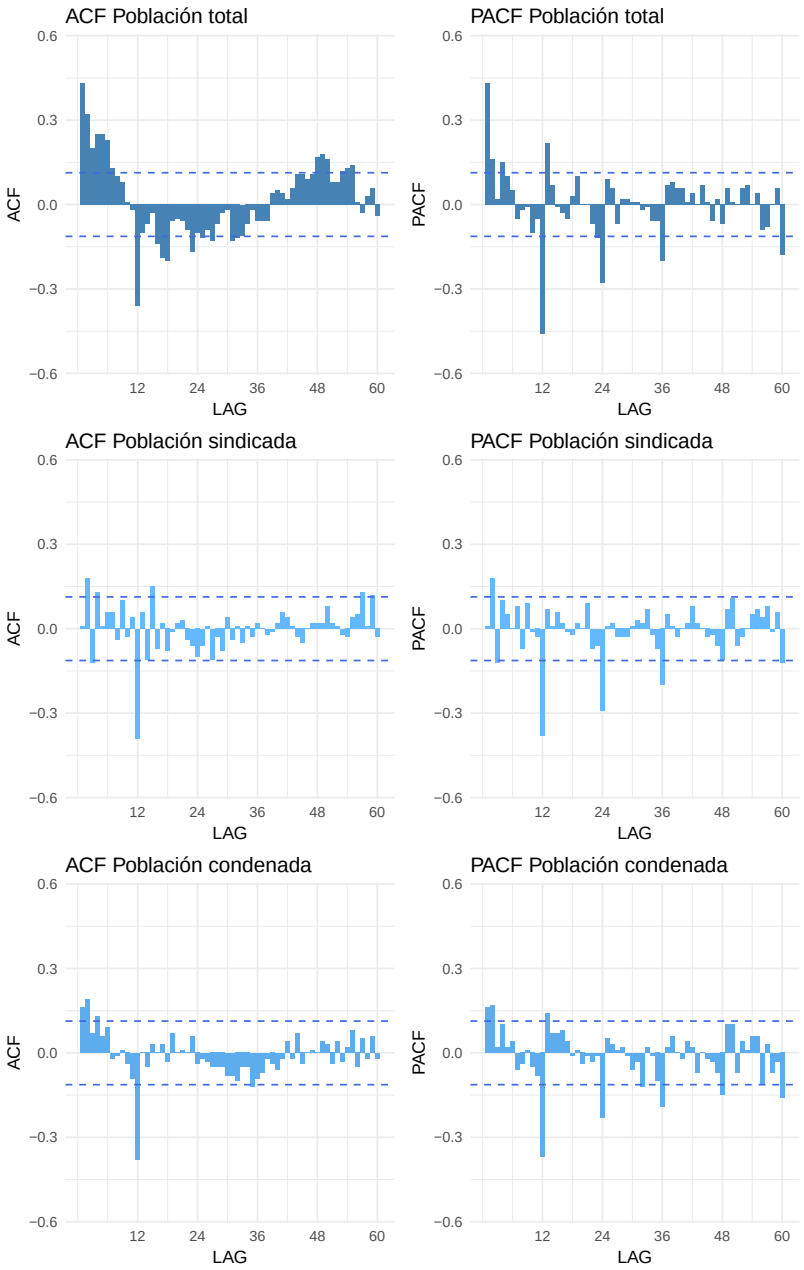

FIGURA 3.5. Autocorrelación parcial de la variación inter-mensual de la población

Fuente: INPEC Elaboración propia

La tabla 3.1 presenta una guía para la interpretación de la ACF y la PACF.

Si se analiza el total de la población carcelaria como una sola serie, es posible realizar una proyección separada, contra la cual contrastar las proyecciones de la población sindicada y de la condenada. En la figura 3.5 podemos observar que tanto la ACF como la PACF decaen lentamente en la población total. La APCF decae progresivamente en los meses 12, 24, 36, mientra que la ACF se corta en el 24, lo que sugiere que el proceso es un promedio móvil de orden 1 en el componente estacional.

En este caso una primera estimación se realiza para la proyección de la población total como SARIMA(1,1,1,0,0,1).

En la figura 3.5 podemos observar que en la población sindicada la ACF decae lentamente, mientras la PACF cae por debajo del error luego del rezago 1. La APCF decae progresivamente en los meses 12, 24, 36, mientra que la ACF se corta en el rezago 12, lo que sugiere que el proceso es un promedio móvil de orden 1 en el componente estacional.

En este caso un primer acercamiento a la proyección de la población sindicada se presenta como SARIMA(1,1,1,0,0,1).

El comportamiento de la ACF y la PACF de la población condenada es similar al de la población sindicada, sugiriendo el mismo modelo: SARIMA(1,1,1,0,0,1).

3.2. Estimación de parámetros

Para cada serie (población total, población sindicada, población condenada) se evalúa el modelo sugerido por el análisis de la ACF y de la PACF; a partir de este modelo se propone un segundo modelo con un parámetro AR adicional, para identificar si se logra un ajuste mejor (medido con el AIC y el BIC); como tercer paso se utiliza la funcion *auto.arima* para identificar el modelo con menor AIC [15]; a partir del paso anterior se propone un modelo con menos parámetros, para buscar una solución más parsimoniosa.

De acuerdo con el diagnóstico se evaluaron los siguientes modelos para proyectar la población total:

1. SARIMA de orden (1,1,1) y estacionalidad (0,0,2)
2. SARIMA de orden (1,1,1) y estacionalidad (0,0,1). Se valida que el BIC y el AIC son menores en el modelo 1, como se esperaría de acuerdo con nuestro diagnóstico y se presenta en la tabla 3.2
3. SARIMA de orden (3,2,1) y estacionalidad (0,0,2). Es el modelo sugerido por la función auto.arima, sin embargo incluye parámetros autoregresivos no significativos (pvalue >0.05). El detalle de los resultados se presenta en la tabla tabla 3.3
4. SARIMA de orden (1,2,1) y estacionalidad (0,0,2). Este modelo tiene un BIC menor al modelo 3, y eliminando dos parámetros no significativos conserva un AIC cercano.

La estadística de Ljung-Box tiene como hipótesis nula H_0 la independencia del error hasta el rezago l. Un p value alto indica que no podemos rechazar la hipótesis nula. Al realizar la prueba sobre los residuales, se espera que en un modelo con buen ajuste, no se rechace la hipótesis nula a ningún rezago (se requieren p values altos).

Para el modelo 1 y el modelo 2, Ljung-Box sugiere que aún hay correlación significativa en los residuos, el modelo 3 y el modelo 4 no tienen correlación significativa en los residuos.

En las proyecciones siguientes para la población total se usará el modelo 4.

TABLA 3.2. Comparación de criterios de información

	Orden	AIC	BIC
1	(1,1,1)(0,0,2)[12]	4885.2	4907.66
2	(1,1,1)(0,0,1)[12]	4897.79	4916.51
3	(3,2,1)(0,0,2)[12]	4874.46	4900.64
4	(1,2,1)(0,0,2)[12]	4874.86	4893.56

TABLA 3.3. Parámetros del modelo (3,2,1,0,0,2)

	Estimate	SE	t.value	p.value
AR1	0.05	0.10	0.50	0.62
AR2	-0.09	0.07	-1.16	0.25
AR3	-0.16	0.07	-2.18	0.03
MA1	-0.68	0.09	-7.25	0.00
SMA1	0.24	0.06	4.10	0.00
SMA2	0.19	0.05	3.88	0.00

TABLA 3.4. Parámetros del modelo (1,2,1,0,0,2)

	Estimate	SE	t.value	p.value
AR1	0.19	0.08	2.39	0.02
MA1	-0.82	0.05	-16.48	0.00
SMA1	0.24	0.06	4.10	0.00
SMA2	0.19	0.05	3.89	0.00

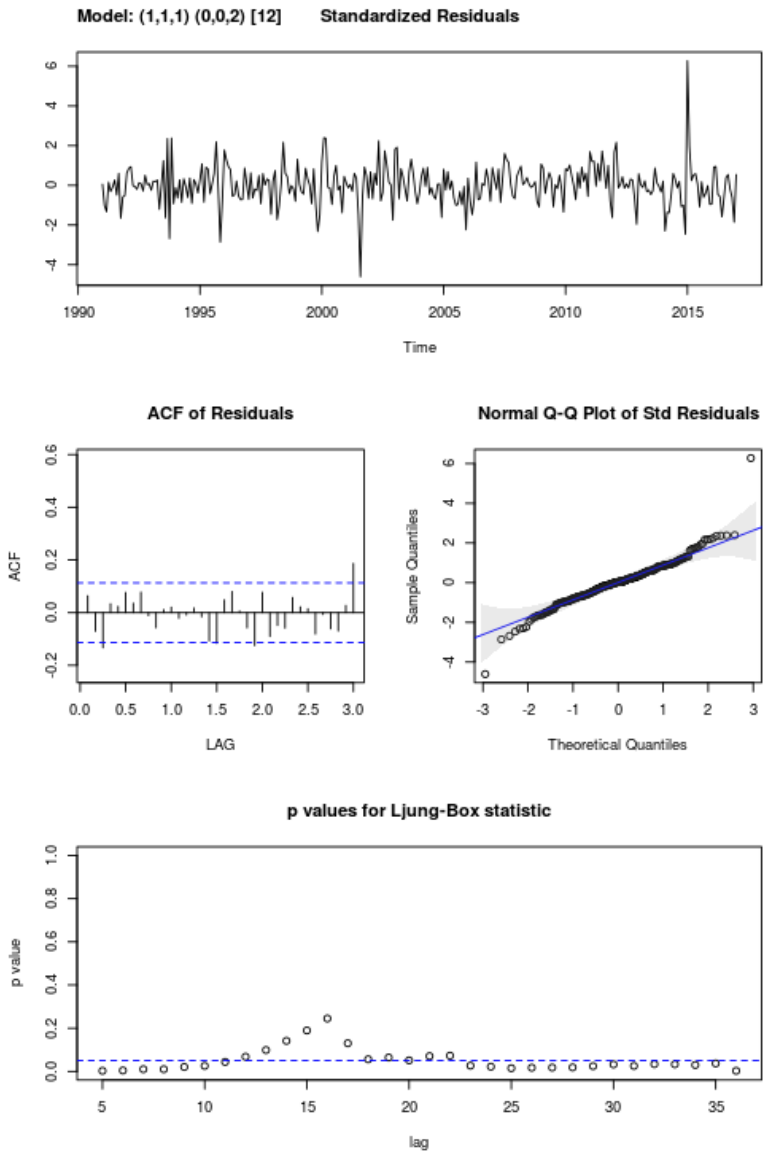

FIGURA 3.6. Diagnóstico del Modelo 1. SARIMA (1,1,1)(0,0,2), para la población carcelaria total.

Fuente: INPEC Elaboración propia

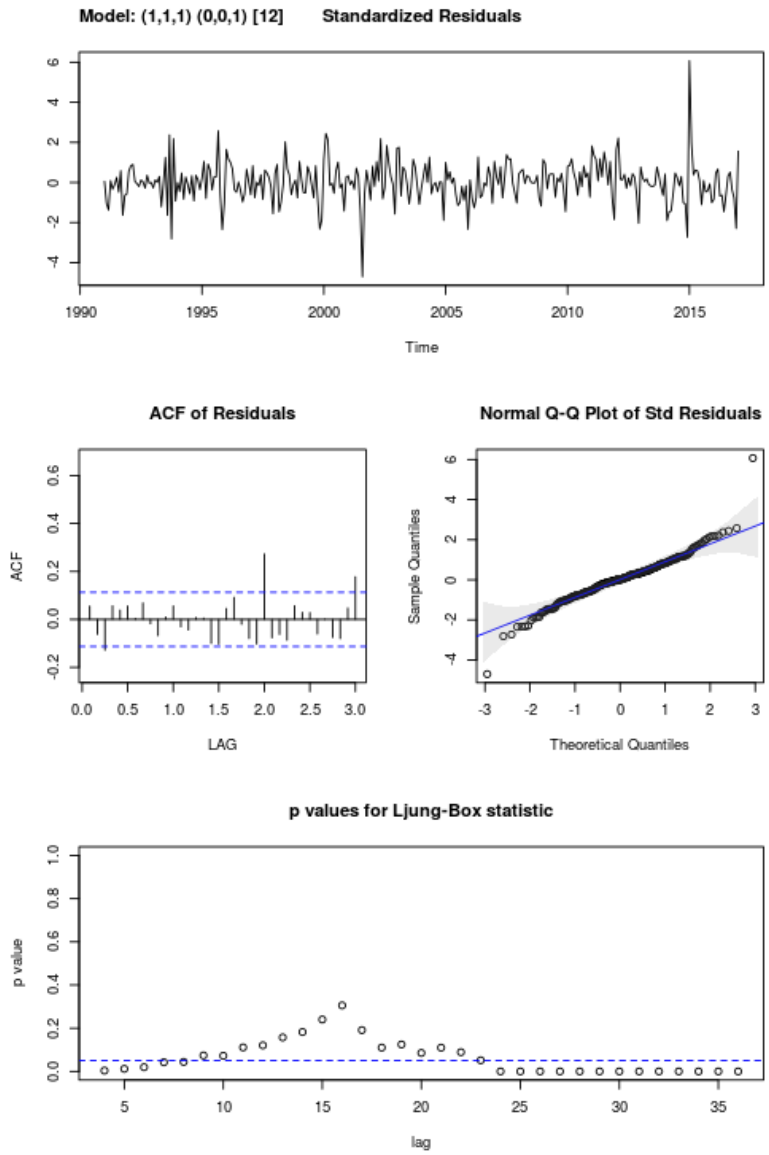

FIGURA 3.7. Diagnóstico del Modelo 2. SARIMA (1,1,1)(0,0,1), para la población carcelaria total.

Fuente: INPEC Elaboración propia

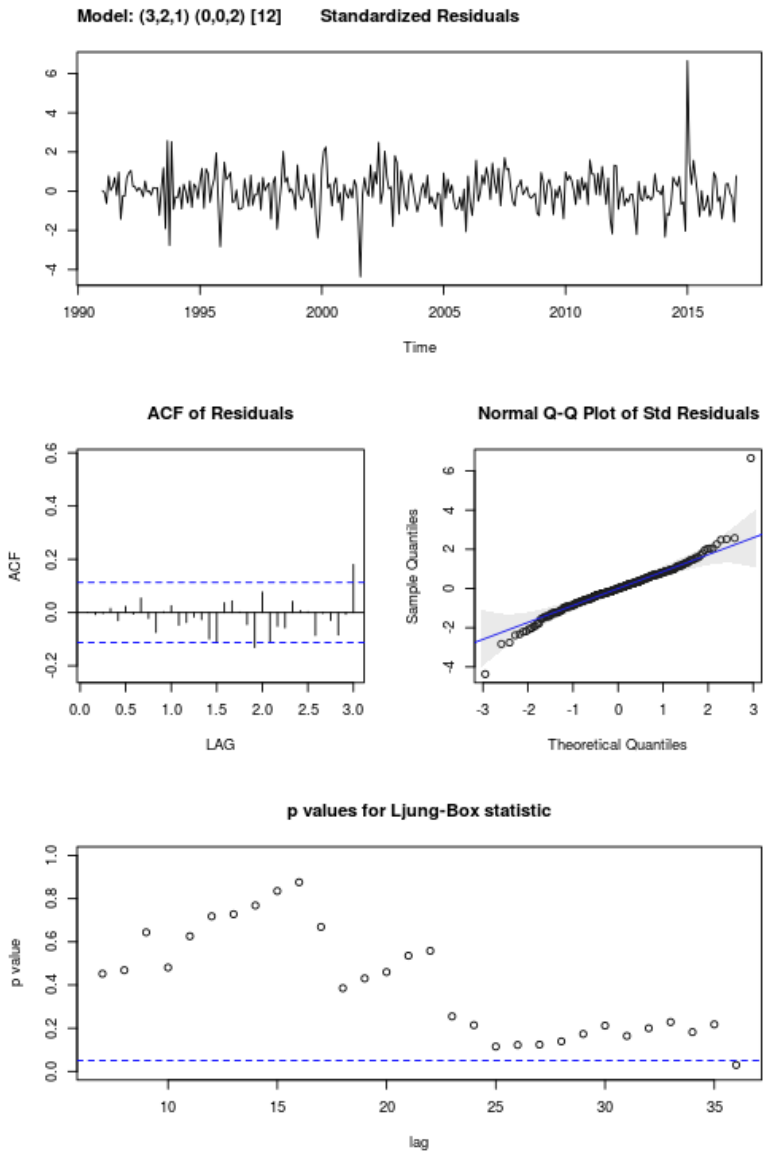

FIGURA 3.8. Diagnóstico del Modelo 3. SARIMA (3,2,1)(0,0,2), para la población carcelaria total.

Fuente: INPEC Elaboración propia

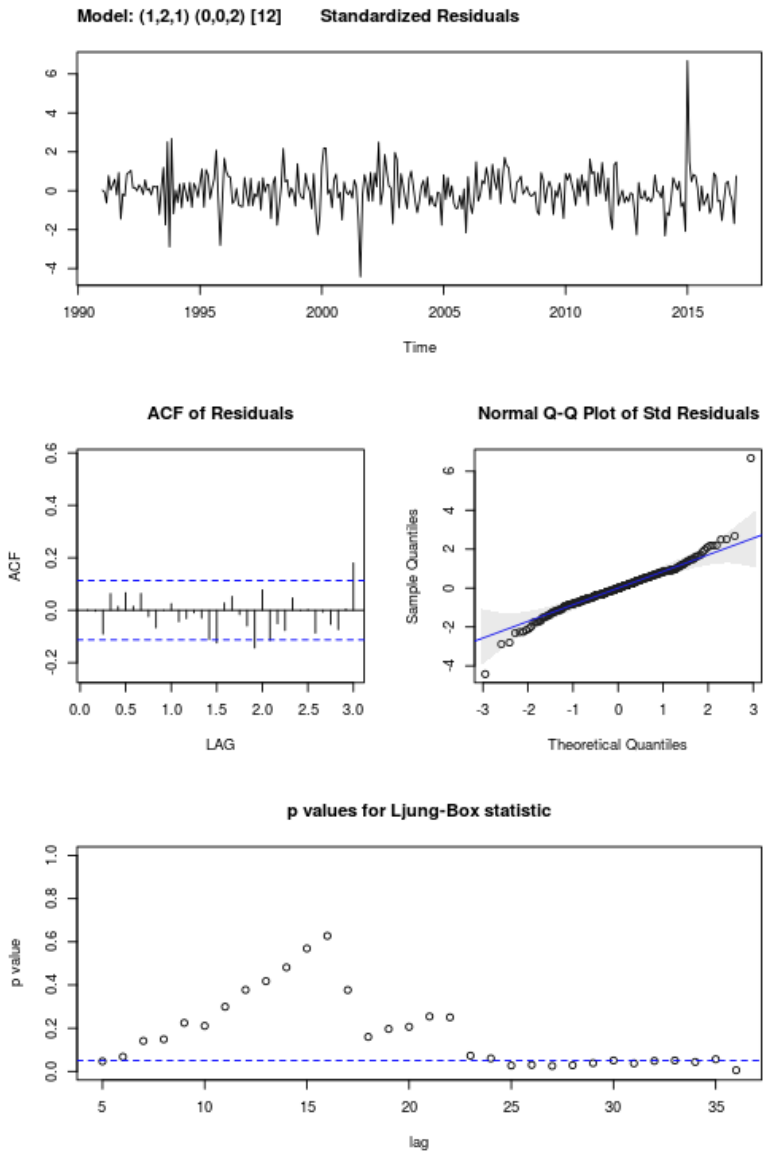

FIGURA 3.9. Diagnóstico del Modelo 4. SARIMA (1,2,1)(0,0,2), para la población carcelaria total.

Fuente: INPEC Elaboración propia

Para la población sindicada se usaron los siguiente modelos, de acuerdo con el diagnóstico realizado de la ACF y la ACPF.

1. SARIMA de orden (1,1,1) y estacionalidad (0,0,1)
2. SARIMA de orden (2,1,1) y estacionalidad (0,0,1). Se valida que el BIC y el AIC son menores que en el modelo 1, tiene el menor AIC de los modelos probados. 3.2
3. SARIMA de orden (0,1,2) y estacionalidad (0,0,1). Es el modelo sugerido por la función auto.arima, tiene el menor BIC y un AIC cercano al modelo 2.
4. SARIMA de orden (0,1,1) y estacionalidad (0,0,0). Este modelo tiene un BIC mayor al modelo 3, y a los dos modelos iniciales.

TABLA 3.5. Comparación de criterios de información

	ORDEN	AIC	BIC
1	(1,1,1)(0,0,1)[12]	4968.8	4987.52
2	(2,1,1)(0,0,1)[12]	4964.1	4986.56
3	(0,1,2)(0,0,1)[12]	4964.26	4982.98
4	(0,1,1)(0,0,0)[12]	4974.16	4985.39

Modelo 3 tiene el menor BIC y tiene pvalues for Ljung-Box que permiten descartar correlación. Los parámetros del modelo se presentan en la tabla 3.6 . El detalle del diganóstico de cada modelo se encuentra en las figuras 3.10 a 3.13

TABLA 3.6. Parámetros del modelo (0,1,1,0,0,1)

	Estimate	SE	t.value	p.value
MA1	0.06	0.06	0.99	0.32
MA2	0.16	0.05	2.91	0.00
SMA1	0.15	0.06	2.57	0.01
Constante	64.94	53.21	1.22	0.22

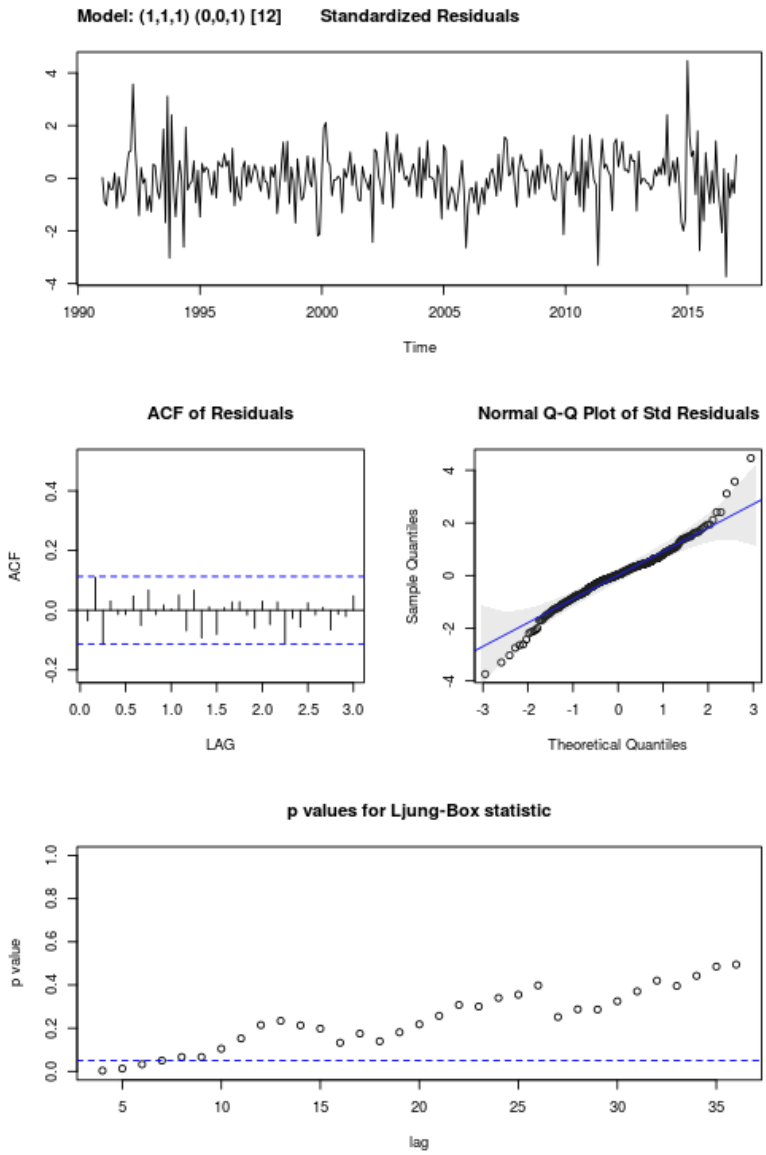

FIGURA 3.10. Diagnóstico del Modelo 1. SARIMA (1,1,1)(0,0,1), población carcelaria sindicada.

Fuente: INPEC Elaboración propia

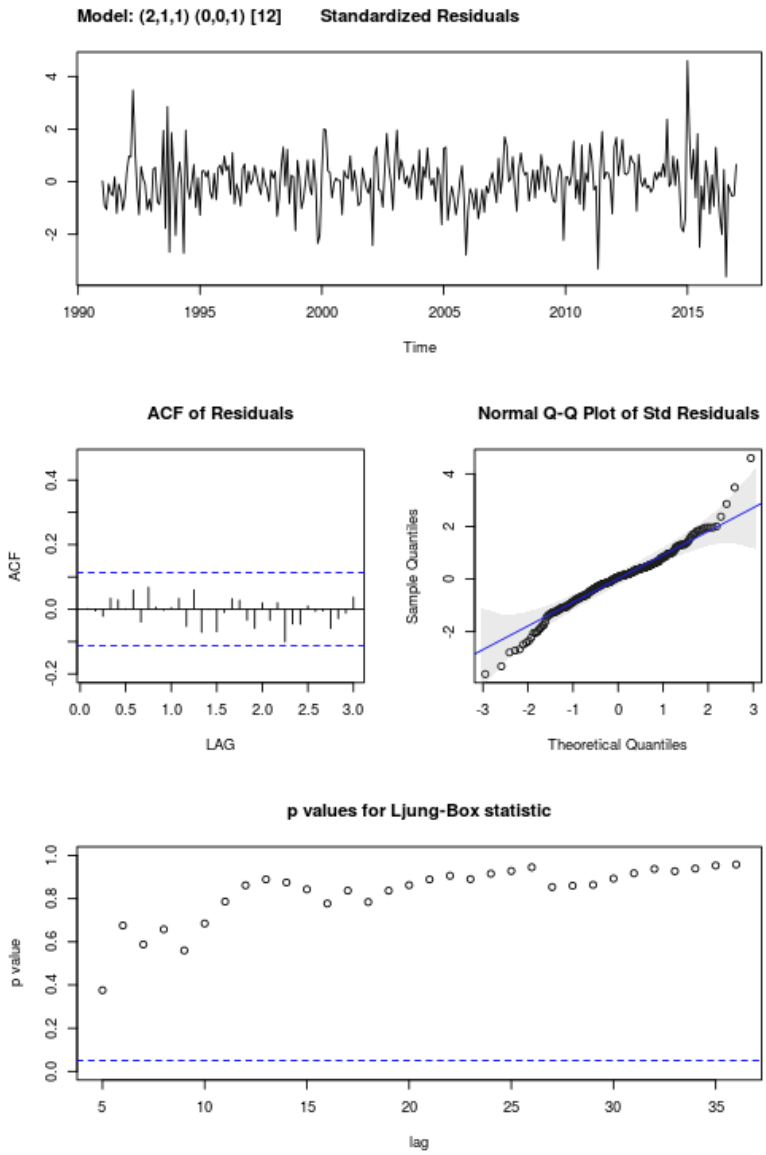

FIGURA 3.11. Diagnóstico del Modelo 2. SARIMA (2,1,1)(0,0,1), población carcelaria sindicada.

Fuente: INPEC Elaboración propia

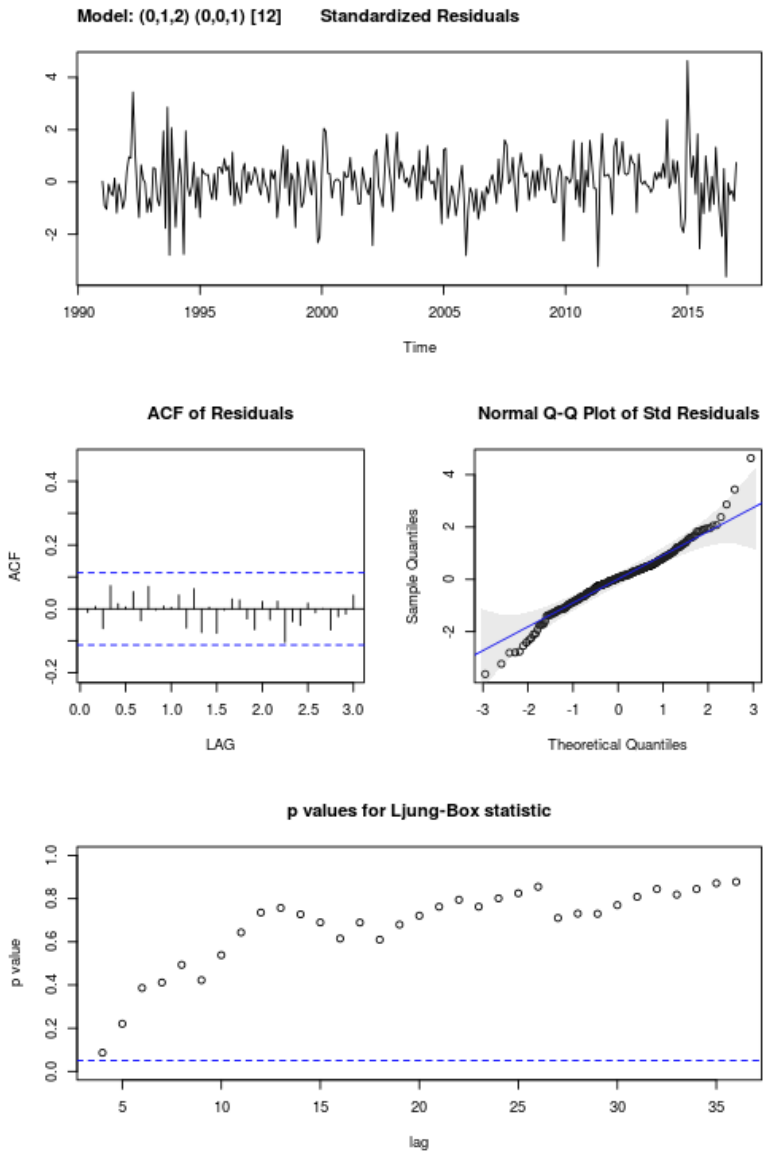

FIGURA 3.12. Diagnóstico del Modelo 3. SARIMA (0,1,2)(0,0,1), población carcelaria sindicada.

Fuente: INPEC Elaboración propia

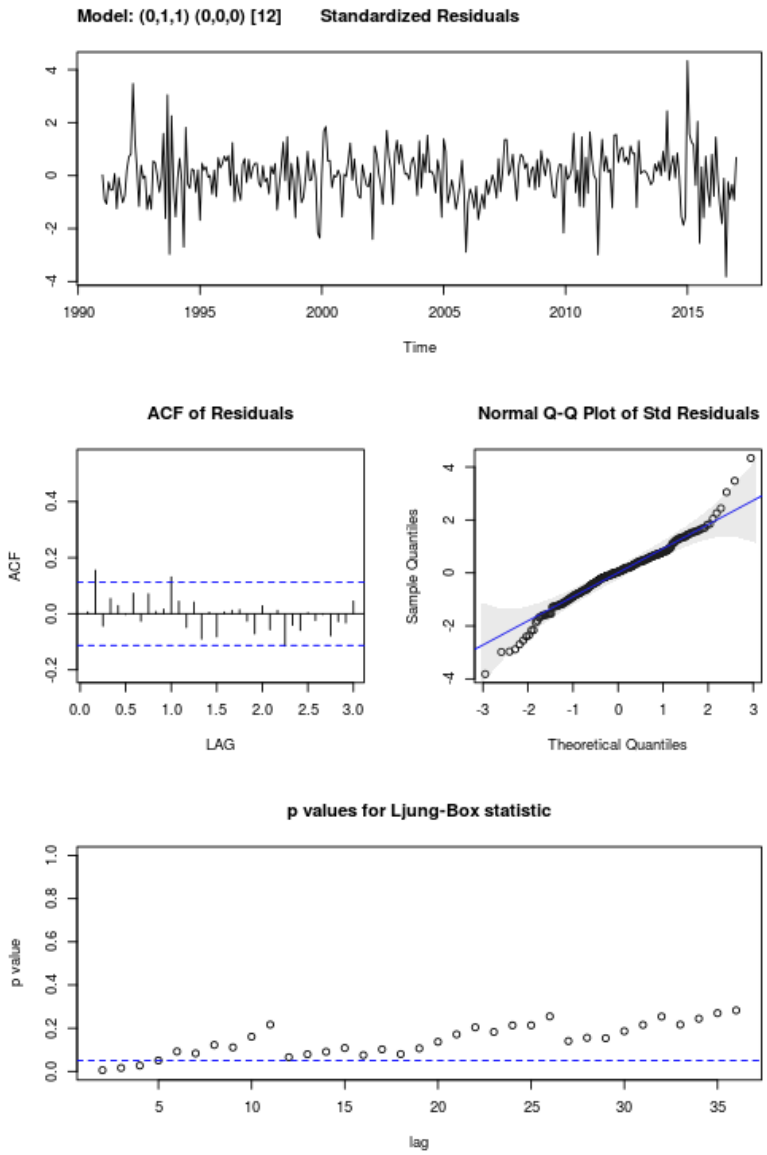

FIGURA 3.13. Diagnóstico del Modelo 4. SARIMA (0,1,1)(0,0,0), población carcelaria sindicada. Fuente: INPEC Elaboración propia

Para la población condenada se utilizaron los siguientes modelos:

1. SARIMA de orden (1,1,1) y estacionalidad (0,0,1).
2. SARIMA de orden (2,1,1) y estacionalidad (0,0,1).
3. SARIMA de orden (0,1,2) y estacionalidad (0,0,1).
4. SARIMA de orden (0,1,1) y estacionalidad (0,0,0).

TABLA 3.7. Comparación de criterios de información

	ORDEN	AIC	BIC
1	(1,1,1)(0,0,1)[12]	4904.08	4922.79
2	(2,1,1)(0,0,1)[12]	4906	4928.46
3	(0,1,2)(0,0,1)[12]	4914.84	4933.55
4	(0,1,1)(0,0,0)[12]	4925.62	4936.85

Se escoge el modelo 1, pues aunque tiene menos parsimonia, además los pvalue de la estadística de Ljung-Box soportan rechazar la hipótesis de dependencia.

Los modelos 3 y 4 se utilizan para comparar con el comportamiento de la población sindicada. En este caso, el comportamiento de la población condenada es diferente y es bien definido por el modelo 1 con los parámetros presentados en la tabla 3.8.

TABLA 3.8. Parámetros del modelo (1,1,1,0,0,1)

	Estimate	SE	t.value	p.value
AR1	0.88	0.06	13.75	0.00
MA1	-0.73	0.09	-7.80	0.00
SMA1	0.09	0.06	1.55	0.12
Constante	198.62	86.28	2.30	0.02

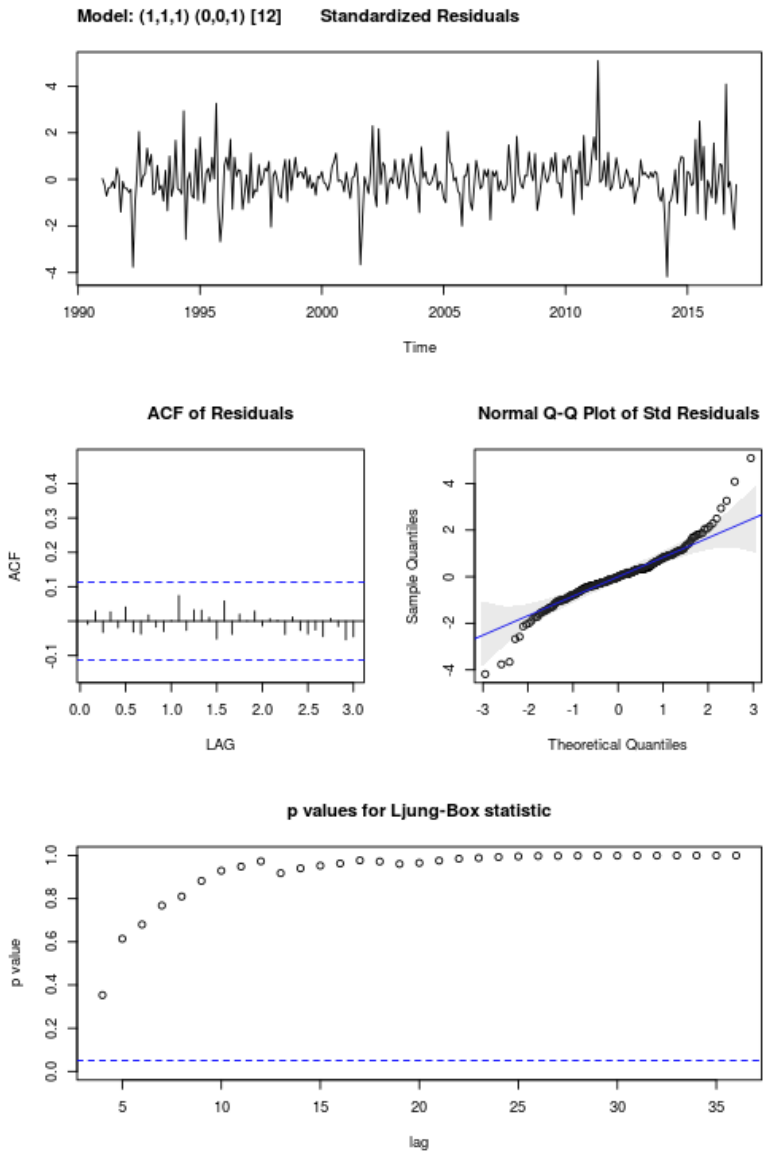

FIGURA 3.14. Diagnóstico del Modelo 1. SARIMA (1,1,1)(0,0,1), población carcelaria condenada.

Fuente: INPEC Elaboración propia

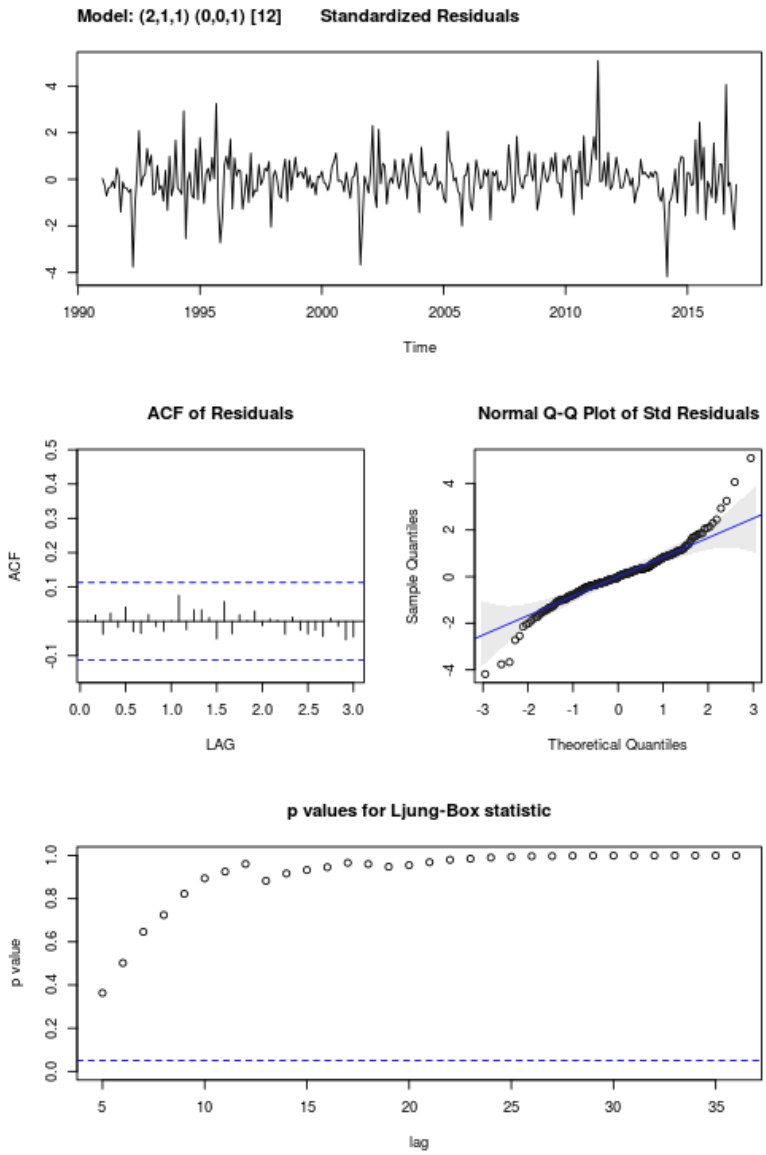

FIGURA 3.15. Diagnóstico del Modelo 2. SARIMA (1,1,1)(0,0,1), población carcelaria condenada.

Fuente: INPEC Elaboración propia

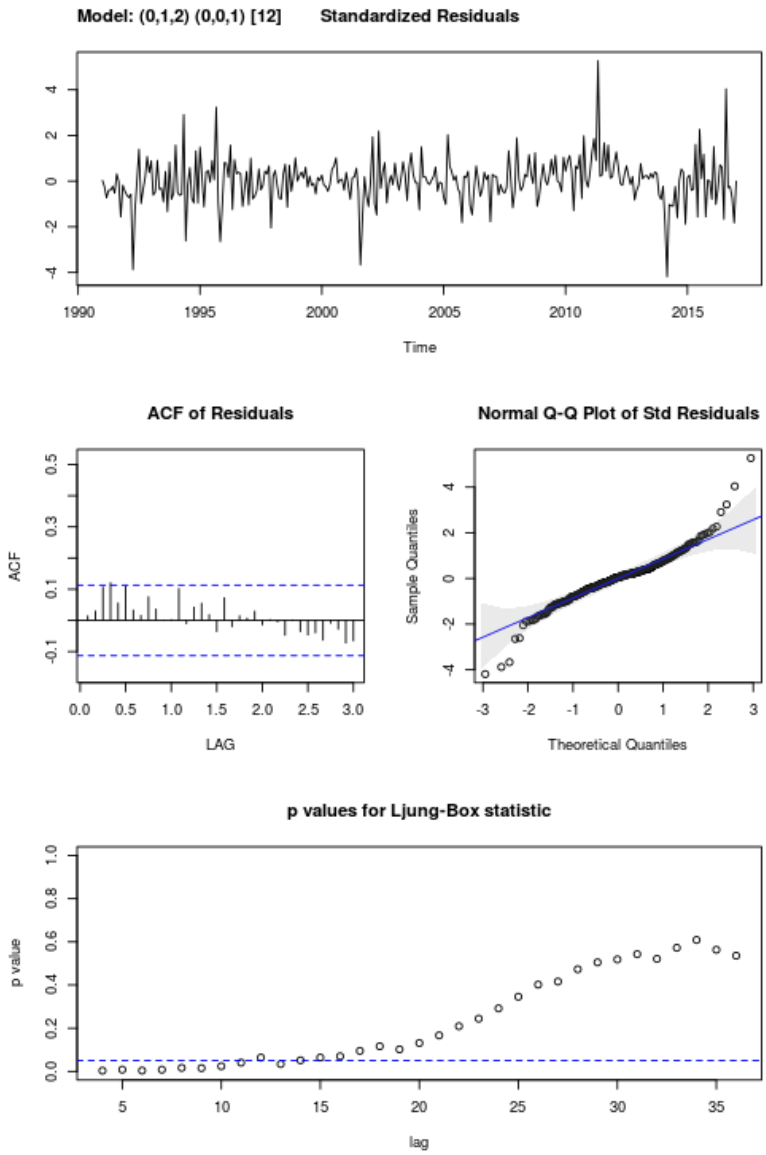

FIGURA 3.16. Diagnóstico del Modelo 3. SARIMA $(0,1,2)(0,0,1)$, población carcelaria condenada.

Fuente: INPEC Elaboración propia

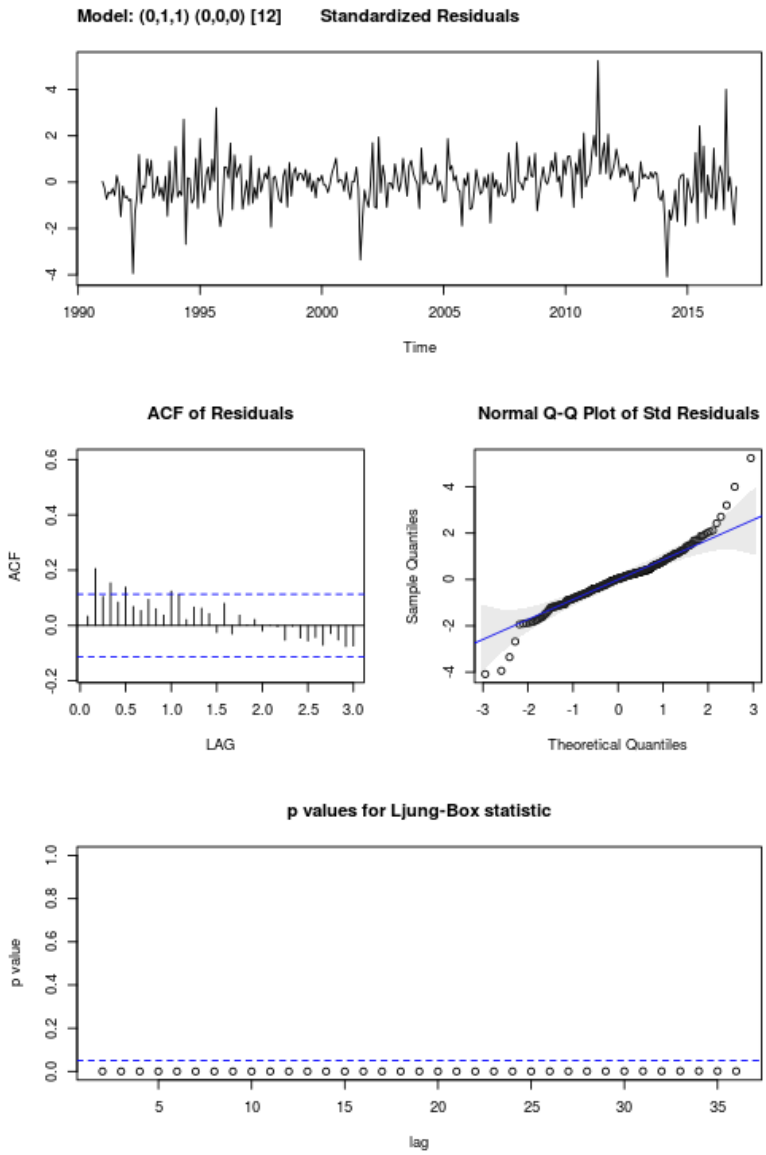

FIGURA 3.17. Diagnóstico del Modelo 4. SARIMA (0,1,1)(0,0,0), la población carcelaria condenada.

Fuente: INPEC Elaboración propia

3.3. Tratamiento de outliers

Para eliminar outliers se realizó el siguiente procedimiento: 1) Se identificaron los outliers, como aquellos instantes donde el error se encuentra a más de tres desviaciones estándar 2) se incluyeron una variable dummy en cada instante identificado como outlier 3) se seleccionó un nuevo modelo, con el menor AIC.

Los diagnósticos de los modelos seleccionados se encuentran en las figuras 3.18, 3.19 y 3.20.

Una vez retirados los outliers, los modelos seleccionados presentan un comportamiento similar al normal, sin muestras de autocorrelación.

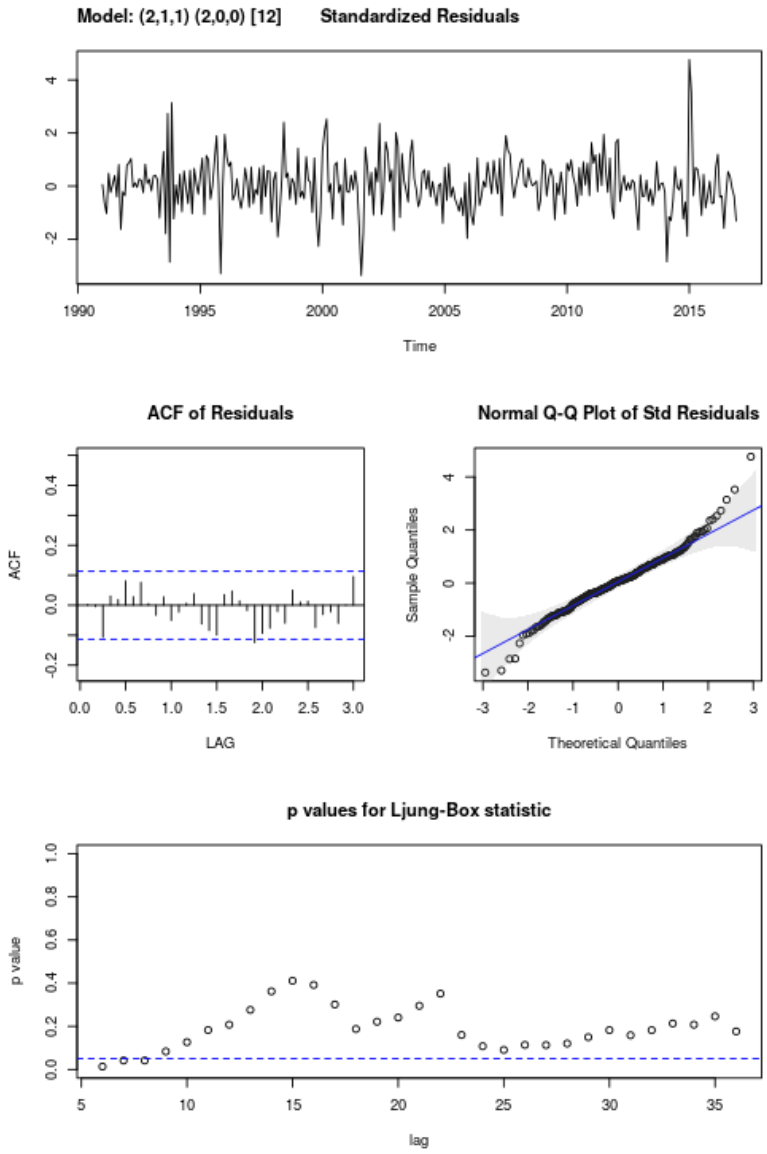

FIGURA 3.18. Diagnóstico del Modelo SARIMA (0,1,1)(0,0,0)[12], población carcelaria total con variables Dummy.

Fuente: INPEC Elaboración propia

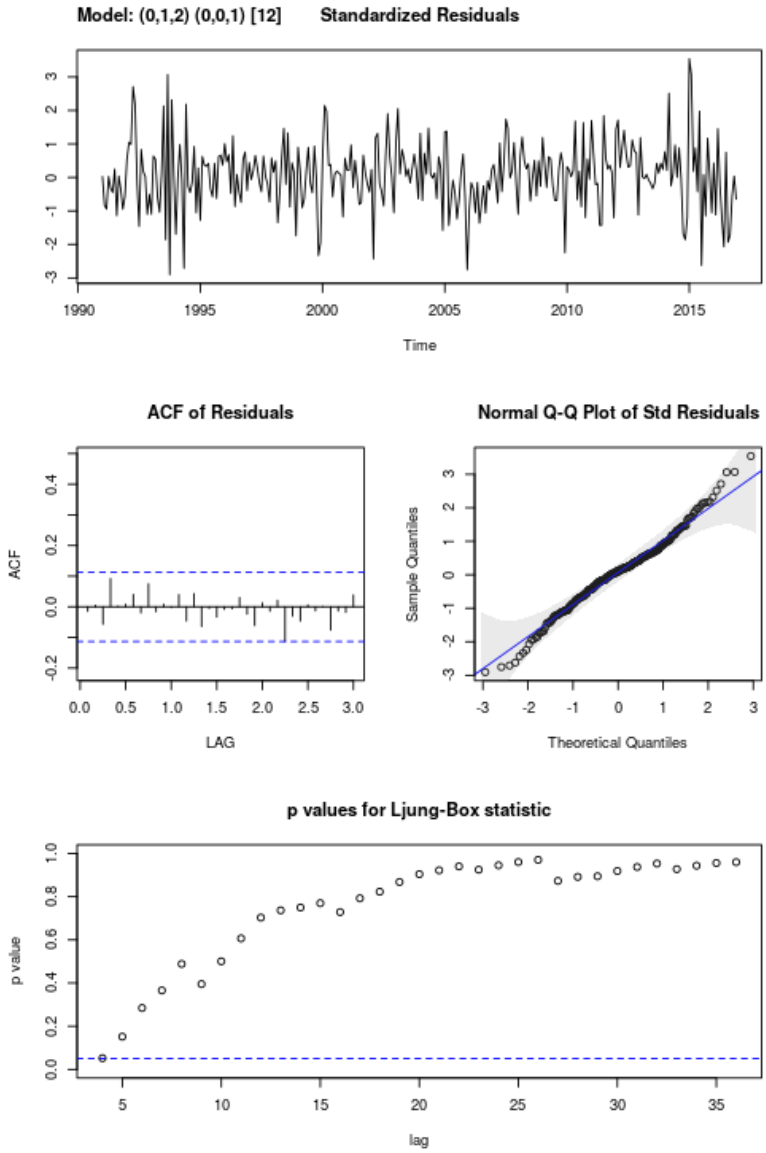

FIGURA 3.19. Diagnóstico del Modelo SARIMA (0,1,2)(0,0,1)[12], población carcelaria sindicada con variables Dummy.

Fuente: INPEC Elaboración propia

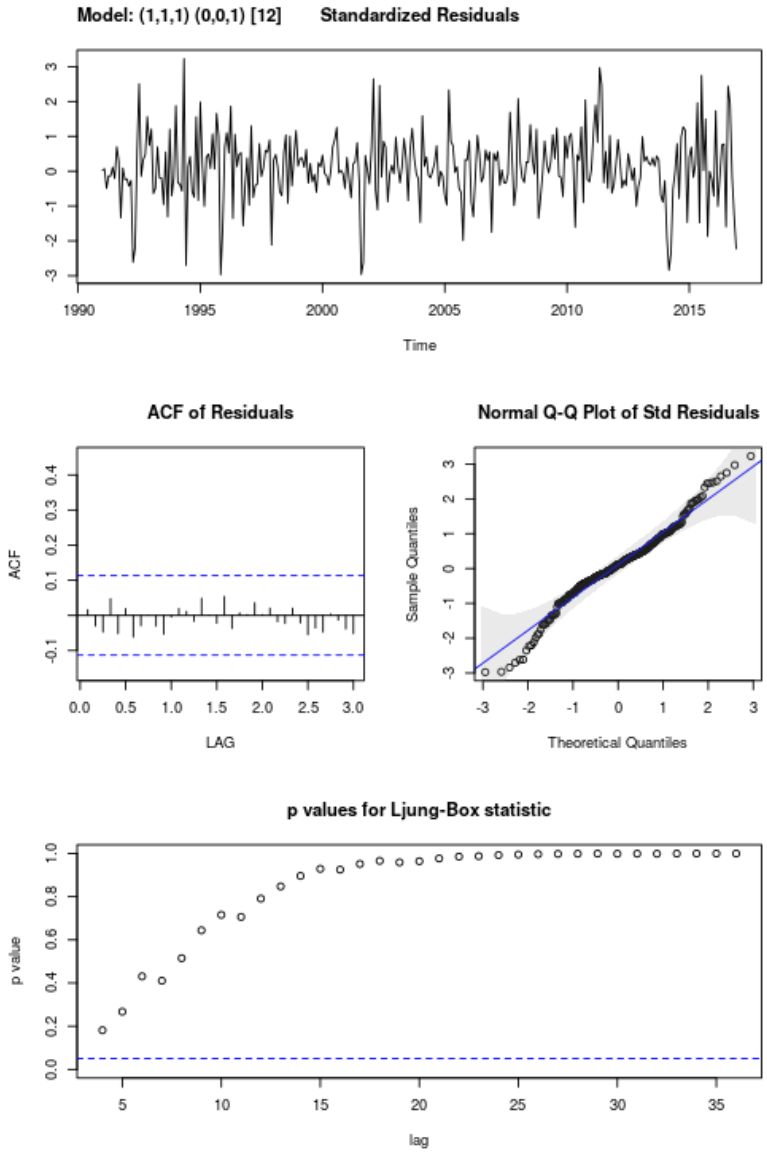

FIGURA 3.20. Diagnóstico del Modelo SARIMA (1,1,1)(0,0,1)[12], población carcelaria condenada con variables Dummy.

Fuente: INPEC Elaboración propia

3.4. Proyección

Proyección de la población total, luego de retirar outliers, se presenta en la gráfica 3.21.

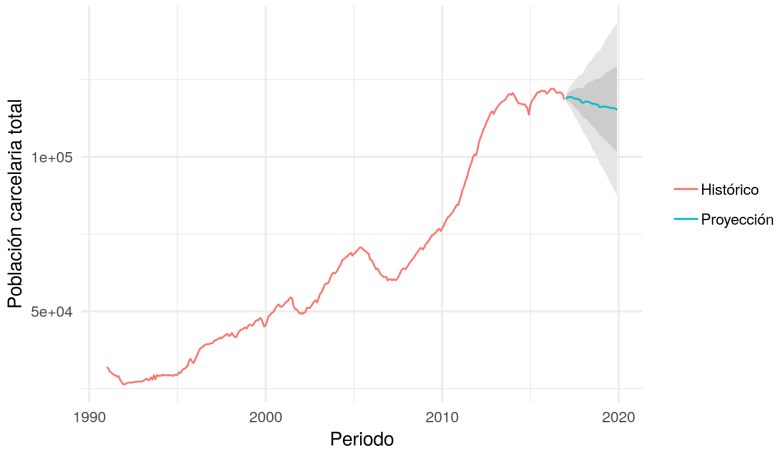

FIGURA 3.21. Proyección de la población carcelaria total, SARIMA (2,1,1,2,0,0)

Fuente: INPEC Elaboración propia

Proyección de la población sindicada, luego de retirar outliers, se presenta en la gráfica 3.22

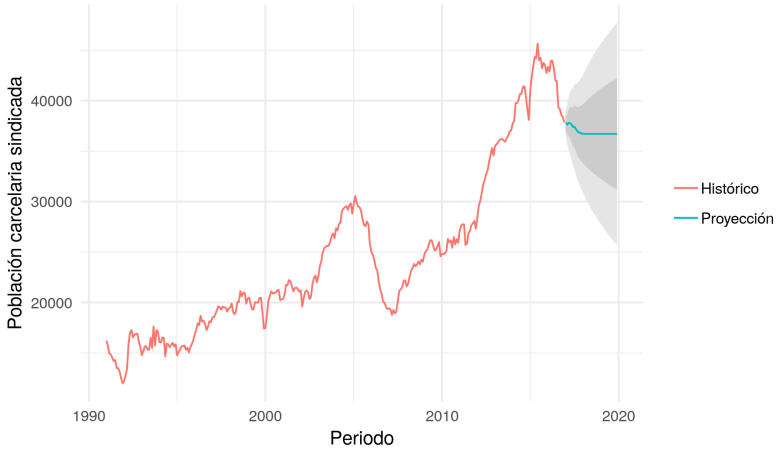

FIGURA 3.22. Proyección de la población carcelaria sindicada, SARIMA (0,1,2,0,0,1)

Fuente: INPEC Elaboración propia

Proyección de la población condenada, luego de retirar outliers, se presenta en la gráfica 3.23

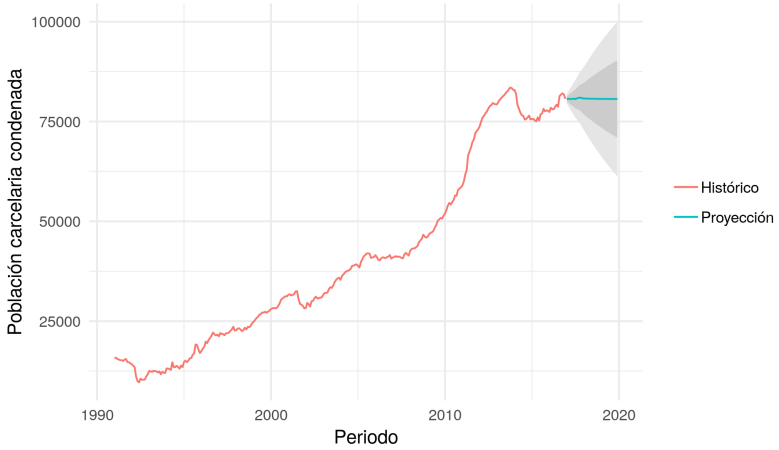

FIGURA 3.23. Proyección de la población carcelaria condenada, SARIMA (1,1,1,0,0,1)

Fuente: INPEC Elaboración propia

3.5. Discusión

Se seleccionó un modelo SARIMA para cada una de las series propuestas: población carcelaria total, población sindicada y población condenada. El residual en cada modelo seleccionado, ya no muestra correlación, y los estimadores de los parámetros son significativos. Con base en el modelo seleccionado para cada serie se realizaron proyecciones de población para 36 meses, la media de la proyección de la población total muestra una tendencia a la baja, mientras las proyecciones de la población sindicada y condenada tienen tendencia al alza.

Puesto que la población carcelaria total esta compuesta en su totalidad por la población sindicada y la población condenada, es de esperar que la proyección de la población total coincida en sentido y magnitud con la suma de las proyecciones de menor escala; sin embargo, la proyección de la población total difiere en sentido de la suma de las proyecciones.

Usar modelos ARIMA permite tener en cuenta la correlación del error y el comportamiento de las últimas observaciones, generando proyecciones precisas de corto plazo; sin embargo, dado que los modelos propuestos tienen un componente integrado, los intervalos de confianza se hacen significativamente más amplios, a medida que se alejan del inicio de la proyección, haciéndolos poco prácticos para proyecciones de largo plazo.

La elaboración de los intervalos de confianza descansa en el supuesto de normalidad del error, sin embargo, los Q-Q plot residuales muestran varias observaciones fuera de los intervalos de confianza para población sindicada y condenada, sugiriendo varios cambios de nivel o una distribución no normal del error. Contrastar los periodos donde el error está fuera de los intervalos de confianza contra los cambios en el sistema penitenciario, permitiría identificar cambios de nivel y controlarlos a través de variables dummy.

Realizar un análisis de impacto en series de tiempo, como el sugerido por Mataris en [23] permitiría medir el impacto que los diferentes cambios en la política criminal han tenido en la población sindicada y condenada. El análisis realizado en la sección 3.3 trata todos los errores como errores aditivos, cuyo efecto se limita al periodo en cuestión, por ejemplo, indultos a un grupo significativo de individuos.

Situaciones como las encontradas en esta serie de tiempo (alta variabilidad, poblaciones compuestas por series de menor escala, procesos de migración inter-grupos, cambios de nivel asociados a cambios de régimen) pueden ser comunes en la proyección de poblaciones pequeñas. Estas situaciones, si se desean abordar usando las herramientas del análisis de series de tiempo pueden incluir: Modelos ARMAX, para los modelos cuya varianza no es estable a través del tiempo, modelos autoregresivos de umbral (TAR) cuando se presenten cambios de régimen y vectores autoregresivos(VAR), para aquellas poblaciones cuyo comportamiento se relaciona con poblaciones pequeñas cercanas o con otros grupos de poblaciones especiales.

CAPÍTULO 4

Métodos demográficos para poblaciones pequeñas

No todas las personas tenemos la misma probabilidad de ser encarceladas. Los hombres tienen tasas de encarcelamiento más altas que las mujeres, y las personas entre 18 y 40 años son más susceptibles que los mayores de 40 [2]. Las cifras de encarcelamiento, presentadas en la figura 4.4 sugieren que estas tasas son estables en periodos cortos de tiempo y proporcionales entre los diferentes rangos de edad (si el encarcelamiento en un rango de edad se eleva sucede un incremento proporcional en los otros).

La población colombiana ha cambiado significativamente desde 1991, tanto en su magnitud como en su composición por edad. Hemos pasado de 35 millones en 1991 a 45 millones en la segunda década del siglo XXI. También hemos presenciado una reducción en la proporción de menores de 14 años, que ha pasado de 35 % en 1991 a 26 % en 2016. Se proyecta que en 2020 tendremos tantas personas entre 15 y 29 años, como entre 0 y 14, cuando en 1991 la diferencia era del 25 millones de personas [7].

El grupo etario con mayor tasa de encarcelamiento está entre los 18 y 29 años [2]. La población total en este grupo crece sostenidamente en el periodo 1991-2020. Como porcentaje de la población total, decrecen en dos puntos porcentuales en el mismo periodo, cediendo espacio a una población de mayor edad. El crecimiento de este grupo etario, particularmente en hombres podría ser un buen predictor de la población carcelaria.

Para incorporar la relación entre la población nacional y la población carcelaria se recurre al método censal-ratio,comúnmente usado en la proyección de poblaciones pequeñas. Este método requiere la inclusión de variables sintomáticas que agreguen información sobre la población pequeña.

En el periodo analizado la población carcelaria, tanto en hombres como en mujeres, vista por rango de edad, tiene consistentemente forma de campana, donde la mayor proporción de personas encarceladas está entre los 25 y 29 años, tanto a nivel agregado como desagregado por regional (figura 4.5). Al controlar por la población nacional, se encuentra que las tasas de encarcelamiento específicas por edad se han incrementado marcadamente entre 2005 y 2016. Una proporción del crecimiento de la población carcelaria se explica por fenómenos demográficos: uno, el cambio en la estructura poblacional, donde la población entre 26 y 40 (quienes tienen tasas de encarcelamiento más altas que el resto de la población) ha crecido como porcentaje de la población total; dos por el crecimiento en la población nacional.

Este capítulo trabaja con los reportes mensuales de población carcelaria publicados por el INPEC, desagregados por rango etario. Se realizó un proceso de integración de los 27 reportes disponibles de marzo 2016 a mayo 2018; con esta serie se realizan los análisis descriptivos presentados en el capítulo. Periodos anteriores a enero 2016 se consideran no disponibles, pues hubo un cambio en la granularidad de la información que pasó de grupos de diez años a grupos quinquenales. Una segunda fuente corresponde a las proyecciones de población del DANE, separadas por edad, de 2005 a 2020, que se agregaron en los mismos grupos disponibles en las cifras del INPEC. Se utilizó en la proyección las tasas de encarcelamiento por rango etario de junio 2016.

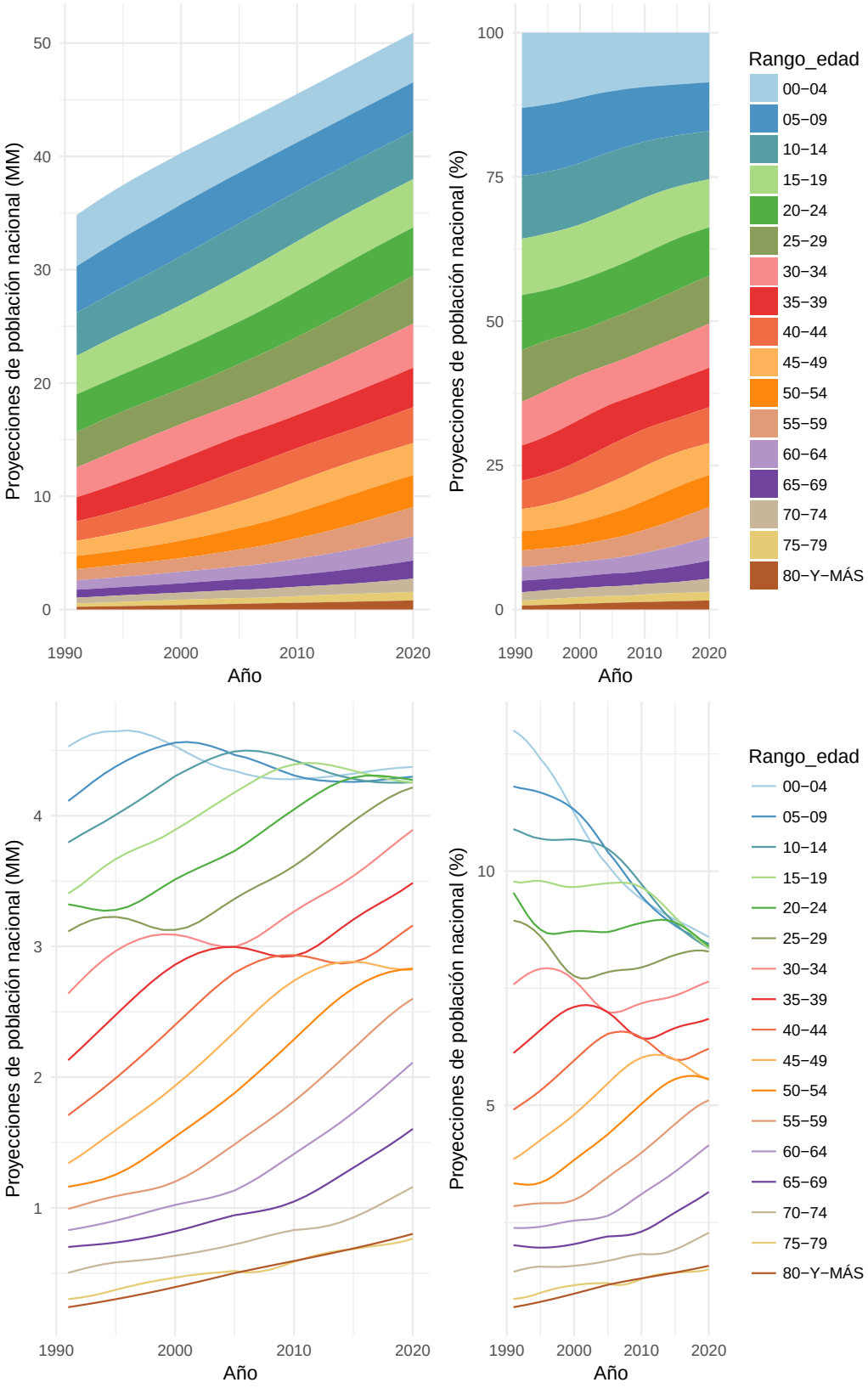

FIGURA 4.1. Proyección de la población nacional por rango de edad, en millones (MM)

Fuente: DANE Elaboración propia

4.1. Censal ratio

Dentro de los métodos para proyectar poblaciones pequeñas se encuentra censal-ratio (razón-censal). Éste método se vale una variable sintomática asociada a la población pequeña, a la que se supone es directamente proporcional. Un ejemplo de variable sintomática es la cantidad de nacimientos en una ciudad, con respecto a su población total; otro ejemplo es la cantidad de niños enrolados en un curso, con respecto a un rango etario en determinado espacio geográfico.

El censal-ratio supone que la relación entre la variable sintomática y la población a estimar es una variable aleatoria cuyo valor esperado no cambia a través del tiempo.

Cuando se cuenta con estimaciones independientes de la población total en periodos inter-censales se realiza una corrección sobre las proyecciones de cada área, tal que la suma de las proyecciones de áreas pequeñas sea igual a la estimada independientemente. La corrección mantiene la proporción entre proyecciones de áreas pequeñas.

Un uso frecuente de los métodos de proyección para poblaciones pequeñas es realizar interpolación en periodos no censales. Uno de los inconvenientes mencionados es la dificultad de realizar extrapolación, puesto que para los periodos post-censales, los indicadores no están necesariamente disponibles. Una posible solución es realizar la estimación usando periodos rezagados o utilizar la última proyección de la variable sintomática como proyección [32].

Siguiendo a Swanson & Tayman en [32] tenemos:

$$R_{A,t,j} = \frac{S_{A,t,j}}{P_{A,t}}$$

R = Razón entre variable sintomática y población pequeña

A = A-ésima área pequeña

S = Valor de la variable sintomática

t = periodo

j = indicador

Si tenemos un R observado en el periodo T tendremos que la proyección del periodo $t+k$ será:

$$\hat{P}_{A,t+k} = \frac{S_{A,t+k,j}}{R_{A,T,j}}$$

Cuando se tiene un estimado independiente para la población padre se realiza un control, de manera que la suma de proyección para las áreas pequeñas sea igual a la proyección del área padre:

$$\hat{Q}_{A,t+k} = \frac{P_{A,t+k,j}}{\sum P_{A,t+k,j}} * P_{t+k} \tag{4.1}$$

$\hat{Q}_{A,t+k}$ = Proyección corregida para el área A

P_{t+k} = Proyección independiente total en el periodo $t+k$

La variable sintomática cumple la tarea de agregar información conocida al modelo para hacerlo más preciso. En un primer paso se considera la población privada de la libertad una población pequeña, dentro de la población nacional. Dentro de las estadísticas publicadas periodicamente por el INPEC no se han encontrado variables sintomáticas que cubran más de dos años [16], razón por la cual se realiza el siguiente proceso: se utiliza una constante en el lugar de la variable sintomática, lo que da lugar a una proyección constante por edad y genero; luego ajustamos la proyección al total nacional por rango de edad proyectado por el DANE. Este modelo es equivalente a proyectar una tasa de encarcelamiento constante para cada rango de edad. Para el periodo 2017 - 2020 se utiliza la tasa de encarcelamiento. Se toma la información del DANE por ser las cifras oficiales y con las que debe trabajar el INPEC también. Para un contraste de estas proyecciones se puede tomar las estimaciones de Naciones Unidas que estima 48.233.000 millones de habitantes para 2015, una cifra muy similar a la del DANE, y con una distribución poblacional por edad y sexo prácticamente igual

Aunque no conocemos el comportamiento de la tasa de encarcelamiento específica por edad, conocemos la población carcelaria total y la tasa de encarcelamiento. Usamos el supuesto de proporción estable por rango etario, para ajustar la proyección 2005 - 2016 a la población carcelaria total.

$R_e = \frac{1}{P_{c,e,2017}}$

$\hat{P}_{c,2017+k} = P_{c,e,2017}$

y

$\hat{Q}_{c,e,2017+k} = (\hat{P}_{c,e,2017} * P_{e,2017}) * P_{e,2017+k}$

donde:

$\hat{Q}_{c,e,2017+k}$ = Proyección de población carcelaria ajustada, para el rango etario, en el periodo $2017+k$

$P_{c,2017}$ = Población carcelaria en 2017, a mitad de periodo, en el rango etario e.

$P_{e,2017}$ = Población estimada 2017 DANE a mitad de periodo en el rango etario e.

$P_{e,2017+k}$ = Población estimada en $2017+k$, del DANE a mitad de periodo en el rango etario e.

Al aplicar esta proyección a periodos anteriores a 2017 tenemos una estimación de lo que habría sido la población carcelaria, en un escenario de tasas de encarcelamiento específicas constantes. Al comparar esta proyección con la población privada de la libertad observada, es posible evidenciar el impacto que ha tenido el proceso de cambio demográfico en la población carcelaria, además de poner en evidencia que no es el principal factor detrás de este cambio. El incremento debe además corresponder con un incremento de las tasas de encarcelamiento específicas.

Para ajustar la proyección de las tasas de encarcelamiento específicas se trata la población carcelaria como un área padre y a las poblaciones por rango etario como poblaciones pequeñas. Usando la ecuación 4.1 tenemos:

$$\hat{V}_{c,e,2017+k} = (\hat{P}_{c,e,2017} * \hat{P}_{c,2017+k}) * P_{c,2017+k}$$

donde:

$\hat{V}_{c,e,2017+k}$ = Población carcelaria ajustada por edad.

$\hat{P}_{c,2017+k}$ = Población carcelaria estimada por género.

$P_{c,2017+k}$ = Población carcelaria observada.

4.2. Población privada de la libertad por rango de edad Febrero 2016 a Febrero 2018

El INPEC publica un reporte mensual con estadísticas detalladas de la población carcelaria al cierre de mes. Los reportes están disponibles en la página del INPEC desde e incluyen la composición de la población privada de la libertad desde Enero 2013. La estructura de los reportes no es uniforme para todos los periodos. La composición por rangos etarios antes de febrero 2016 consolidaba en las categorías: 18 a 29 años, 29 a 54 años, 55 a 64 años, y mayor de 65 años. A partir de marzo de 2016 se agrupa 18 a 24 y grupos quinquenales hasta los 65 años [16].

En la figura 4.2 podemos observar la composición demográfica de la población privada de la libertad en Junio 2017. Es una población compuesta mayoritariamente por hombres. Tanto en hombres como mujeres se observa un patrón similar en la estructura etaria, el punto más alto se encuentra entre 25 y 29 años, 18 a 25 y 29 a 35 tienen un orden de magnitud comparable. Después de los 35 años la población se reduce en cada grupo quinquenal.

En la figura 4.3 se presenta la población privada de la libertad, separada por género y edad desde marzo 2016 hasta marzo 2018. El periodo de análisis se escogió a conveniencia, considerando la disponibilidad de datos del INPEC con grupos etarios uniformes. La figura 4.3 muestra que la población carcelaria mantiene la forma de campana, con pico en 25-29, en todos los meses analizados, tanto en hombres como mujeres. En hombres el rango de edad con mayor variabilidad es el rango 18-24. En mujeres la variabilidad se presetnta en los rangos 18-24, 25-29 y 30-34 [28].

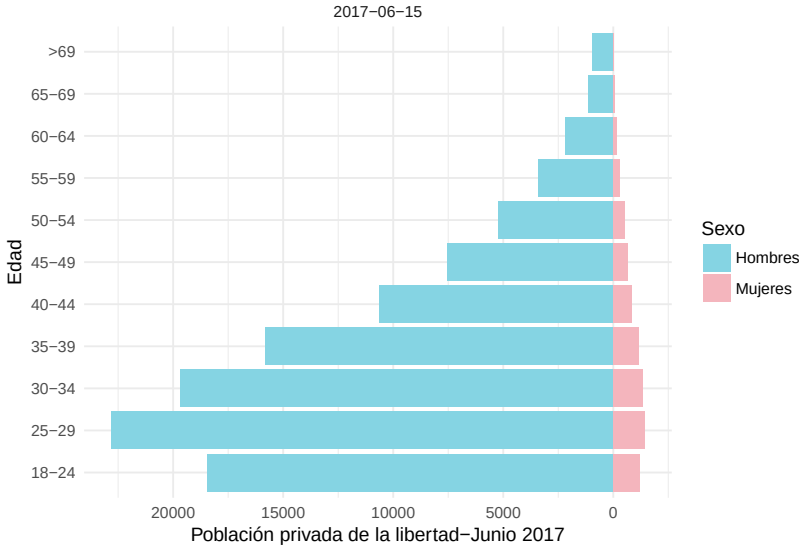

FIGURA 4.2. Pirámide poblacional, población privada de la libertad, Junio - 2017

Fuente: INPEC Elaboración propia

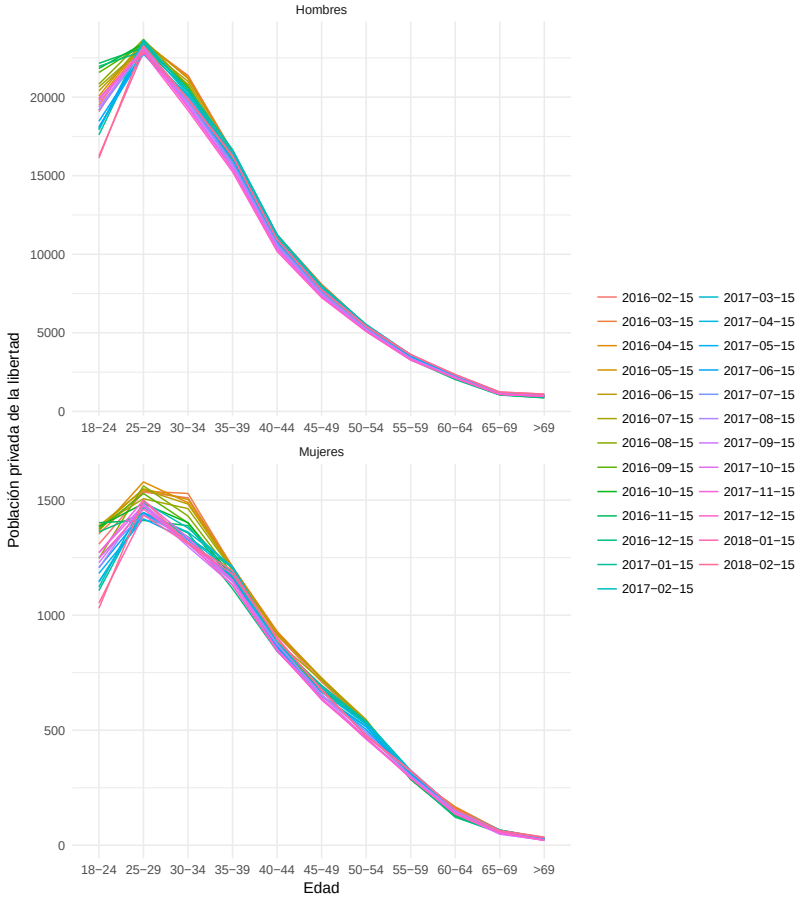

FIGURA 4.3. Población privada de la libertad. Febrero-2016 a Febrero-2018

Fuente: INPEC Elaboración propia

Esta variabilidad podría estar asociada a un componente estacional o a un componente aleatorio. Aunque la longitud de la serie disponible no permite realizar inferencia sobre componentes de estacionalidad y tendencia, la figura muestra que: A través del año el grupo 18-24 crece, para caer abruptamente en enero del año siguiente; comportamiento repetido en enero 2016 y enero 2017. El comportamiento es similar en mujeres una vez se controla por el orden de magnitud. El comportamiento en las demás series es similar, se reduce durante todo el año, para subir nuevamente en enero. Este comportamiento es coherente en hombres y mujeres, una vez se controla la magnitud. 4.3 Este comportamiento implica que no es conveniente trabajar la proyección de la tasa de encarcelamiento a nivel mensual pues no se puede estimar el aparente componente de estacionalidad de la serie. Por conveniencia se trabaja con la proyección de población a mitad de periodo, y se trata la tasa de encarcelamiento como una variable aleatoria cuyo valor esperado se mantiene a través del tiempo.

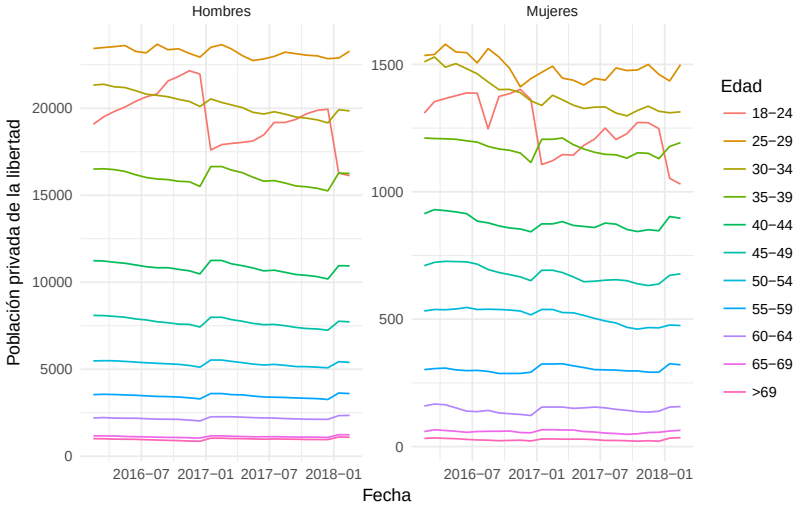

FIGURA 4.4. Población privada de la libertad por rango etario. Febrero-2016 a Febrero-2018

Fuente: INPEC Elaboración propia

Puesto que no se cuenta con información de periodos más extensos se valida análisis a un menor nivel de granularidad. Dentro de la información disponible se encuentra separada la población carcelaria por regional. En la figura 4.5 se observa nuevamente, en todas las regionales, la población carcelaria con forma de campana. Consistentemente las mujeres tienen mayor variabilidad que los hombres en cada rango de edad y poblaciones carcelarias más pequeñas se encuentran asociadas a una mayor variabilidad.

La forma del aparente componente estacional también se replica a nivel regional. Ver figura 4.6.

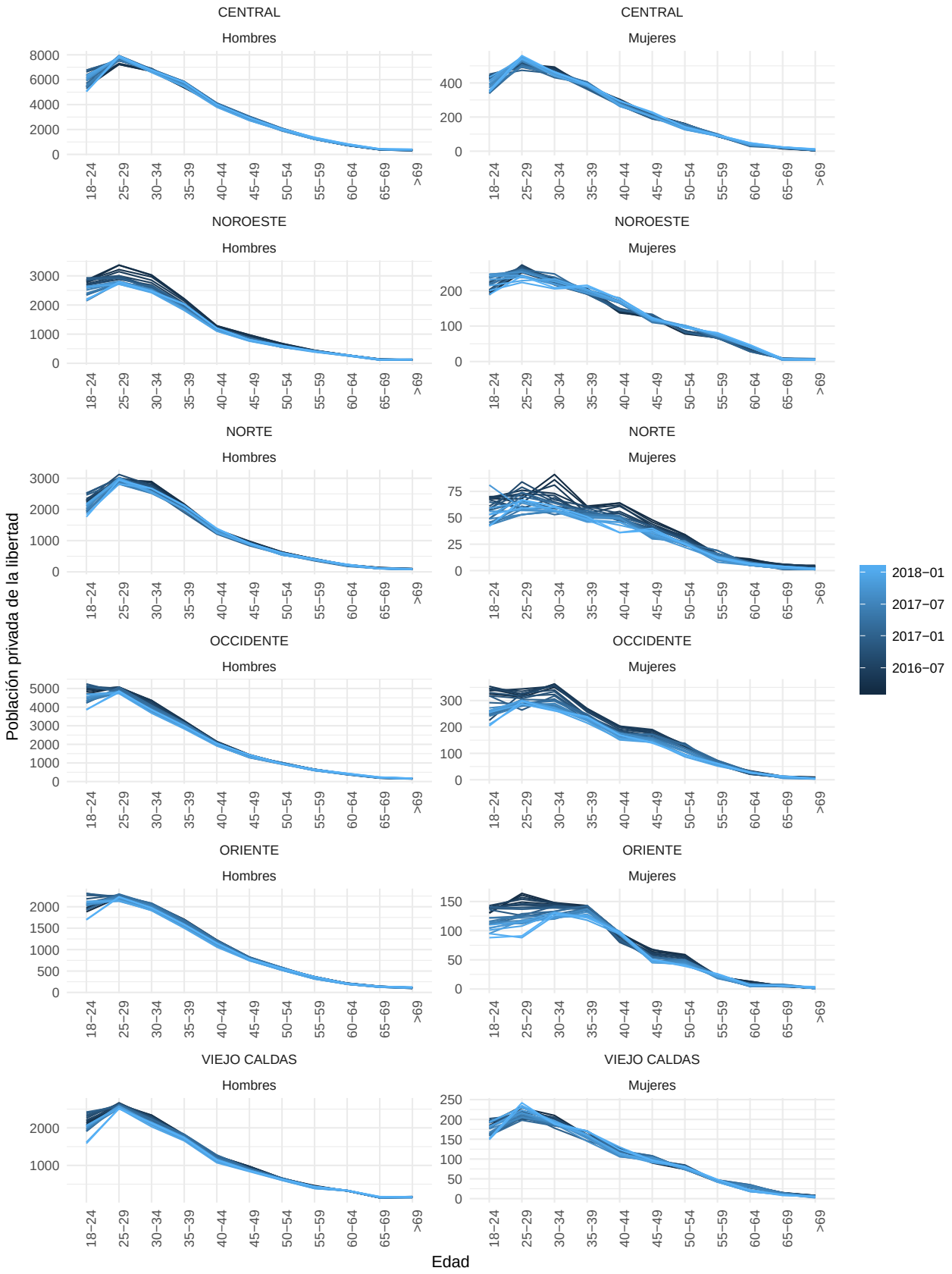

FIGURA 4.5. Población privada de la libertad por rango etario, región. Febrero-2016 a Febrero-2018

Fuente: INPEC Elaboración propia

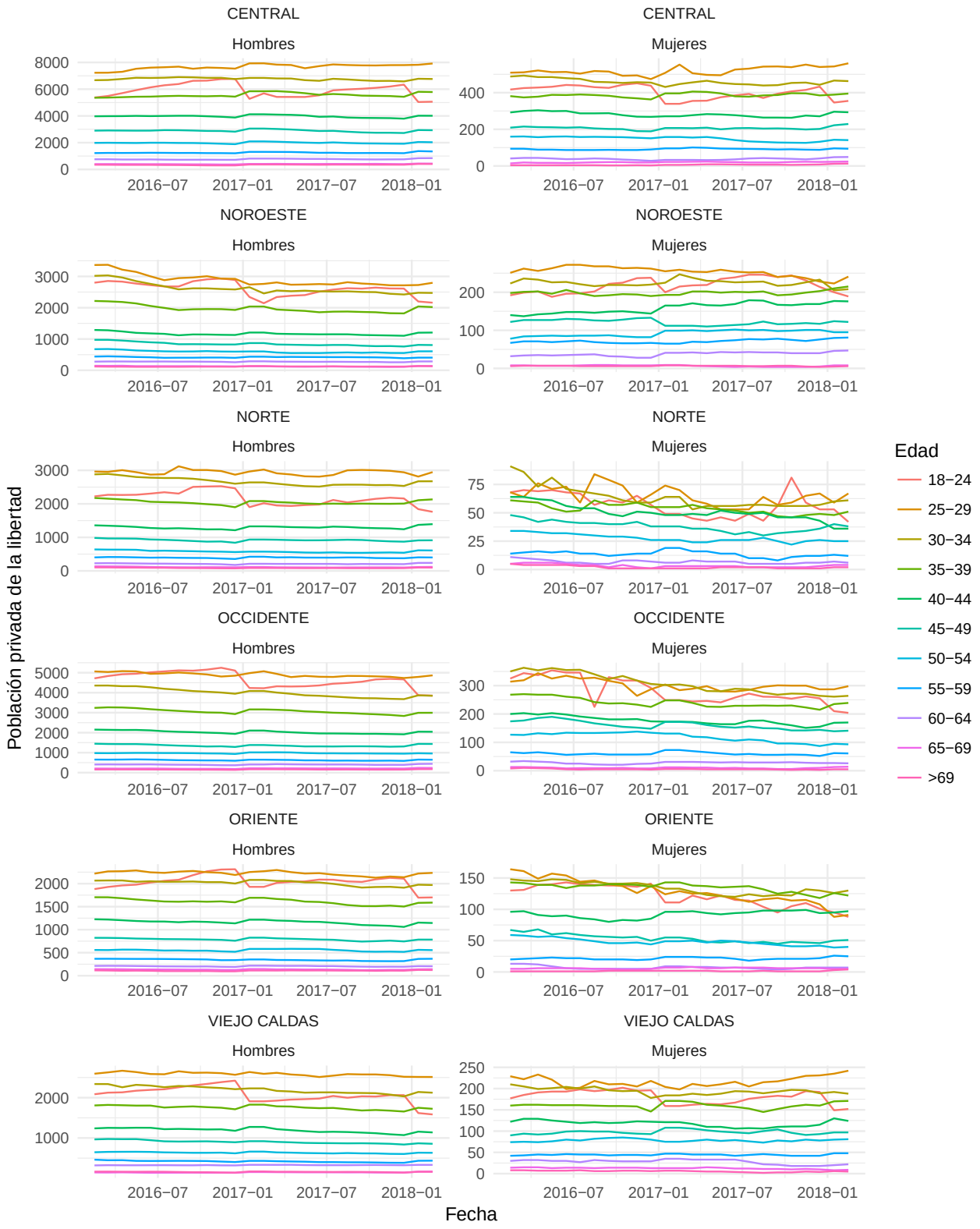

FIGURA 4.6. Población privada de la libertad por rango etario, región. Febrero-2016 a Febrero-2018

Fuente: INPEC Elaboración propia

El comportamiento, cosistente a través del periodo de análisis en cada regional, de la población carcelaria sugiere una tasa de encarcelamiento particular en cada rango de edad.

Con la proyección quinquenal del DANE 2005 - 2020 y la tasa de encarcelamiento en mitad de periodo de 2016 se realiza la proyección de 2017 - 2020 y 2005 a 2015. Si el supuesto de tasas de encarcelamiento estables por rango de edad es coherente la población carcelaria total sería similar a la estimada por rango de edad. La diferencia entre la población observada y la población estimada con este método es un indicador del impacto que el cambio demográfico ha tenido en la población carcelaria. Las tasas de encarcelamiento en Junio 2016 y Junio 2017 se presentan en las Figuras 4.7 y 4.8.

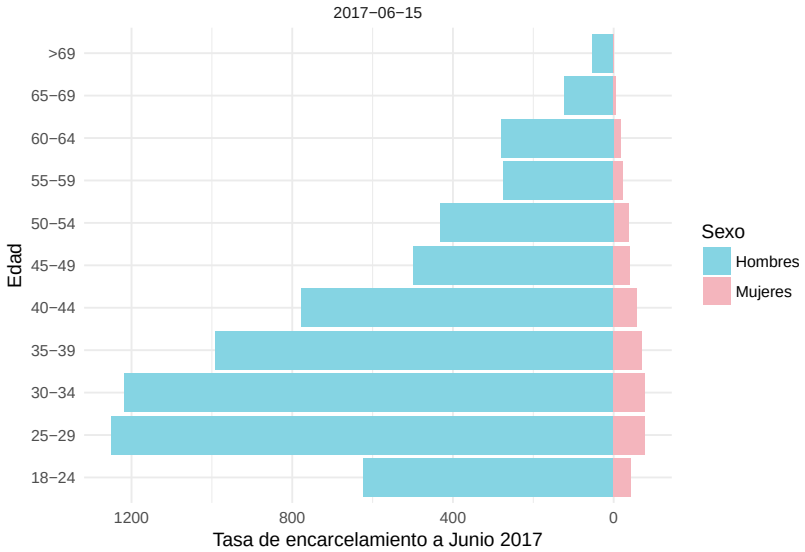

FIGURA 4.7. Tasa de específica de encarcelamiento. Junio-2017

Fuente: INPEC, DANE Elaboración propia

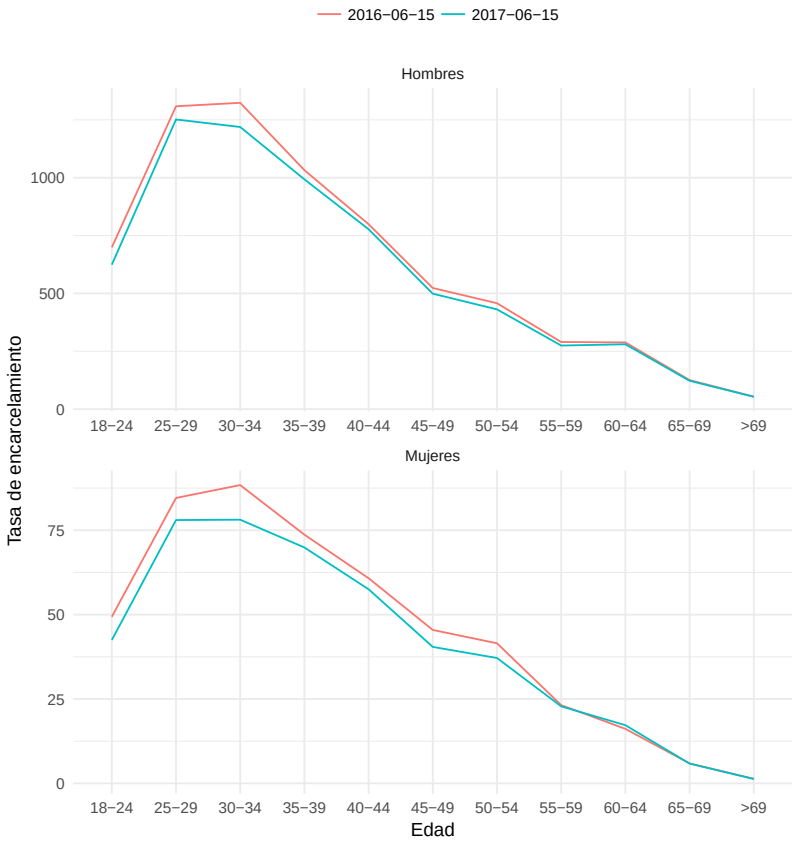

FIGURA 4.8. Tasa específica de encarcelamiento. Junio-2016, Junio -2017

Fuente: INPEC, DANE Elaboración propia

4.3. Proyecciones Censal Ratio

Es inherente a los métodos de poblaciones pequeñas, que la incertidumbre o sesgo de las poblaciones originales, se hereda a las poblaciones pequeñas. En este caso, la calidad de proyección dependerá de la calidad de las proyecciones del DANE, que constituyen la población padre.

La proyección por rango de edad bajo el supuesto de tasas de encarcelamiento estables se presenta en la figura 4.9.

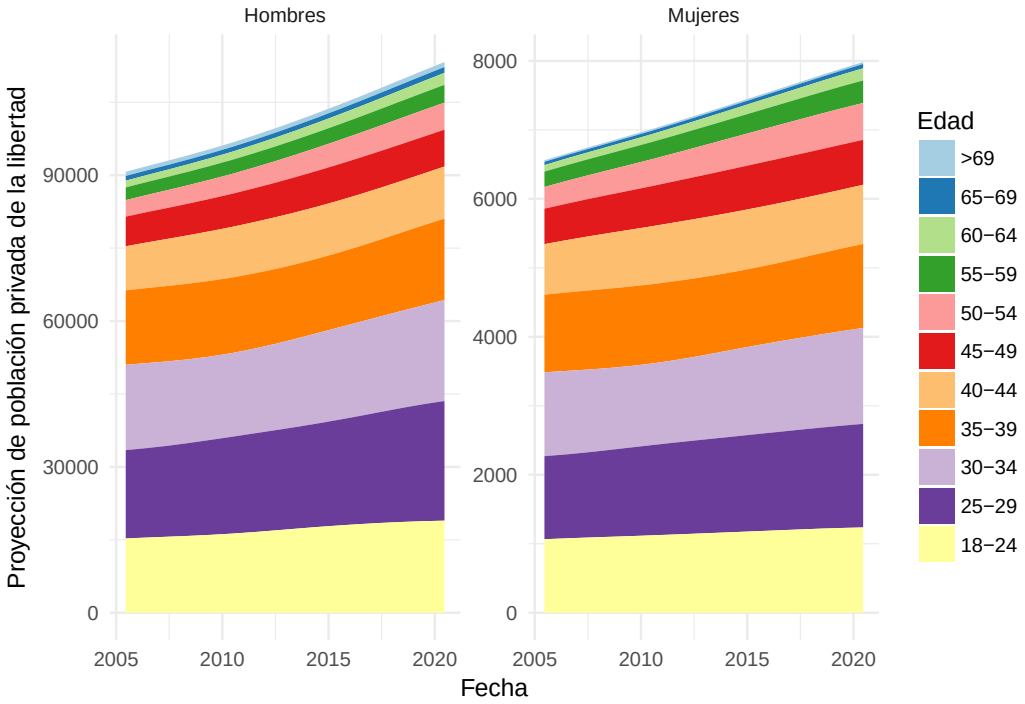

FIGURA 4.9. Proyección de la población privada de la libertad, 2005 - 2020, por rango etario DANE

Fuente: INPEC, DANE Elaboración propia

La proyecciones de población carcelaria a mitad de periodo para el periodo 2005 - 2020 se presentan en la tabla 4.1. Para el periodo 2005 - 2016 se ajustó la proyección total al valor observado. Para proyectar 2017 a 2020 se usó una tasa de encarcelamiento constante.

Con la tasa de encarcelamiento específica observada en 2016, se esperarían alrededor de 96.000 hombres privados de la libertad en 2005. En ese periodo habían 65.885 hombres privados de la libertad. Esto indica que el proceso de cambio demográfico habría explicado el 34 % del incremento de la población carcelaria en este periodo. En las mujeres el porcentaje explicado es 31 %.

	Fecha	Sexo	Poblacion	Pry_ppl_base	Ajuste
1	2005-06-15	Hombres	65885	96463	0.68
2	2005-06-15	Mujeres	4750	7196	0.66
3	2006-06-15	Hombres	59386	97693	0.61
4	2006-06-15	Mujeres	3727	7296	0.51
5	2007-06-15	Hombres	56718	98877	0.57
6	2007-06-15	Mujeres	3457	7392	0.47
7	2008-06-15	Hombres	63562	100110	0.63
8	2008-06-15	Mujeres	4047	7487	0.54
9	2009-06-15	Hombres	69826	101406	0.69
10	2009-06-15	Mujeres	4768	7582	0.63
11	2010-06-15	Hombres	75702	102782	0.74
12	2010-06-15	Mujeres	5078	7679	0.66
13	2011-06-15	Hombres	86698	104287	0.83
14	2011-06-15	Mujeres	6689	7782	0.86
15	2012-06-15	Hombres	101432	105874	0.96
16	2012-06-15	Mujeres	8277	7888	1.05
17	2013-06-15	Hombres	108882	107543	1.01
18	2013-06-15	Mujeres	8981	7996	1.12
19	2014-06-15	Hombres	108929	109277	1.00
20	2014-06-15	Mujeres	8302	8105	1.02
21	2015-06-15	Hombres	112423	111066	1.01
22	2015-06-15	Mujeres	8482	8214	1.03
23	2016-06-15	Hombres	112907	112907	1.00
24	2016-06-15	Mujeres	8323	8323	1.00
25	2017-06-15	Hombres		114790	
26	2017-06-15	Mujeres		8432	
27	2018-06-15	Hombres		116692	
28	2018-06-15	Mujeres		8540	
29	2019-06-15	Hombres		118591	
30	2019-06-15	Mujeres		8646	
31	2020-06-15	Hombres		120461	
32	2020-06-15	Mujeres		8748	

TABLA 4.1. Proyección de la población carcelaria vs real

En la figura 4.10 se presenta la población carcelaria proyectada por rango de edad, luego de ajustar con la población observada. En la figura 4.11 se presenta la tasa de encarcelamiento por edad proyectada para el periodo 1995 - 2020.

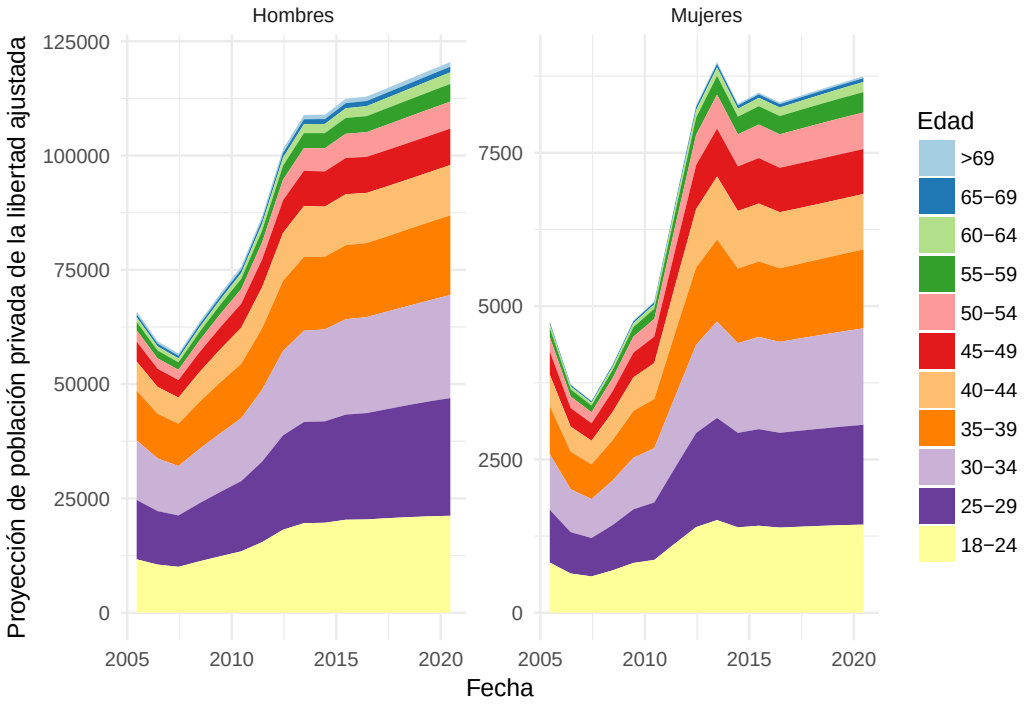

FIGURA 4.10. Proyección de la población privada de la libertad, 2005 - 2020, ajustado por rango etario DANE

Fuente: INPEC, DANE Elaboración propia

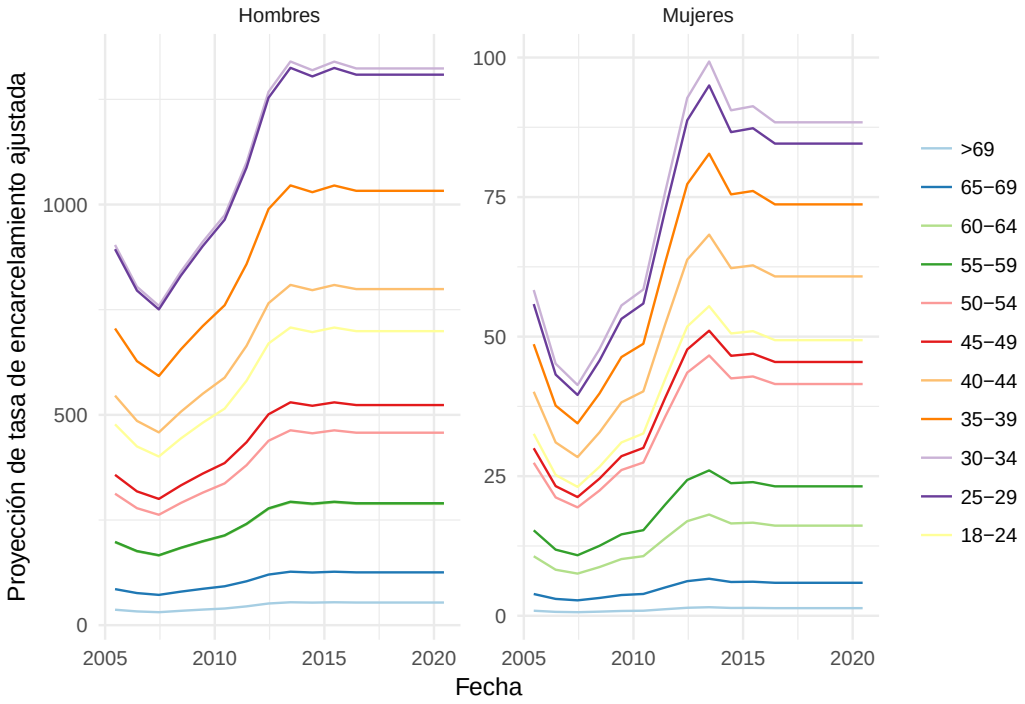

FIGURA 4.11. Proyección de tasa de encarcelamiento específica por rango de edad, por cada 100.000 habitantes, 2005 - 2020

Fuente: INPEC, DANEElaboración propia

4.4. Discusión

Los resultados del análisis descriptivo indican que las tasas de encarcelamiento en hombres son más de diez veces las tasa de encarcelamiento en mujeres, comparadas en el mismo rango etario. En 2016, mientras la tasa de encarcelamiento entre 25 y 35 años en hombres era cercana a 1500 por cada cien mil habitantes, en mujeres era cercana a noventa por cada cien mil habitantes.

Los métodos usuales en la proyección de poblaciones pequeñas se han adaptado en este capítulo, de acuerdo con la información disponible. Los datos evidencian que la tasa de encarcelamiento es diferente según la edad; bajo el supuesto de una tasa de encarcelamiento estable a través del tiempo se pueden realizar proyecciones de mediano plazo de la población carcelaria (3 años). Una proyección para años anteriores permite separar la variación de la población carcelaria asociada con los procesos de cambio demográfico (incremento y envejecimiento) de la variación asociada a otros factores.

La tasa de encarcelamiento más elevada están entre los 25 y 35 años, probablemente asociadas a tasas de ingreso al sistema más elevadas. Políticas públicas orientadas a la prevención del delito en estos rangos de edad resultarían, en principio, las más costo-efectivas, incluso sin incluir el efecto de la reincidencia en el encarcelamiento. Conteos de ingreso al sistema carcelario separados por edad permitirían un análisis más riguroso de costo-beneficio, pues la tasa de encarcelamiento se ve afectada por la duración de la condena y la reincidencia.

El periodo observado sugiere consistentemente a través del género, rango de edad, y regional que existe un comportamiento estacional en la población privada de la libertad. Las tasas de ingreso y salida del sistema permitirían entender mejor las causas de este comportamiento.

Es posible utilizar variables como desempleo nacional, desempleo juvenil o pobreza para mejorar las proyecciones y entender los determinantes de la tasa de encarcelamiento. En este capítulo se utilizo el método de censal-ratio sin una variable sintomática asociada, para separar el efecto del cambio demográfico.

El comportamiento de cada regional tiene componentes diferentes en tendencia, aunque comparte aparentemente el componente estacional. Realizar proyecciones de la tasa de encarcelamiento a nivel regional permitiría realizar un mejor dimensionamiento de la infraestructura carcelaria de cada región asociada a sus propias dinámicas demográficas.

Mejorar la calidad de los datos abiertos del INPEC permitiría desarrollar éstas y otras preguntas de investigación. Medidas que podrían mejorar la calidad de estos datos son:

- Mantener grupos de análisis consistentes a través del tiempo. Dentro de los datos analizados, los rangos de edad observados cambiaron desde febrero 2016 a un menor nivel de granularidad. Este cambio permite análisis de la tasa de encarcelamiento más precisos; sin embargo, no permite comparar con periodos anteriores a ésta fecha.
- Publicar series detalladas con un periodo de análisis más largo. Las estadísticas detalladas por delito, edad, tiempo de condena,...,etc están disponibles desde 2013. Es decir, con los datos abiertos actualmente disponibles no es posible analizar, por ejemplo, el comportamiento de la tasa de encarcelamiento por edad en periodos anteriores.

- Incluir análisis multivariados de la población carcelaria. Los informes publicados actualmente no incluyen análisis multivariados como edad, género, delito, que son pertinentes para la toma de decisiones en política criminal.
- Publicar informes detallados con seguimiento histórico. Parte del trabajo realizado para llegara este capítulo consistió en ensamblar informes históricos en una sola base de datos, publicar las series históricas permitiría acelerar el tiempo de análisis.

CAPÍTULO 5

Modelos Estado-Espacio

El comportamiento de la población carcelaria es diferente según el delito y tiene relación con un componente demográfico y un componente normativo. El componente demográfico se ha tratado en el capítulo anterior, mientras en este capítulo se aborda el comportamiento por delito. La interacción entre edad y delito no ha sido analizada, pues no se cuenta con la información necesaria.

Las personas que se encuentran a la espera de juicio se llaman sindicados, y condenados quienes ya han recibido una sentencia. Cada mes una proporción de las personas sindicadas es sentenciada, puede entonces salir en libertad o cambiar a condenado, y una proporción de los condenados sale en libertad. También se pueden presentar salidas por mortalidad. Un modelo auto-regresivo bivariado para la población total puede modelar estas interacciones; particularmente la población privada de la libertad está compuesta por series separadas de diferentes delitos, cada una con un comportamiento diferente, como lo sugieren la figuras 5.2 y 5.3.

La proyección se separa por género, pues se detectó que el perfil delictivo es diferente. Mientras los delitos con mayor incidencia en hombres son hurto y homicidio, en mujeres es tráfico, fabricación y porte de estupefacientes. Consistentemente en todos los delitos, la población carcelaria masculina es mayor.

Los delitos con mayor incidencia son: hurto; homicidio; fabricación y porte de armas de fuego, y tráfico, fabricación o porte de estupefacientes. Sin embargo, el delito con mayor crecimiento en el periodo analizado es concierto para delinquir, cuya población carcelaria creció en cerca de 6500 personas entre diciembre 2016 y enero 2017. Hurto en el mismo periodo se redujo en 1400 personas. La existencia de una tendencia diferente en cada delito, como se observa en las figuras 5.2 y 5.3, implica que los procesos que generan los datos son diferentes, y se pueden proyectar de forma separada.

Se presentan el comportamiento de la población carcelaria por delito y género. Con esta información se estima un modelo estado-espacio en cada una de las series bivariadas. Los modelos estado-espacio se formulan de forma tal que permitan modelar la interacción entre la población sindicada y condenada.

Se ha trabajado a partir de los reportes de población carcelaria separada por delito, entre Julio 2013 y Diciembre 2016, información anterior no se encuentra disponible como

datos abiertos. La suma de la población por delito es mayor a la población privada de la libertad total, pues un mismo individuo puede contarse en varios delitos.

5.1. Población privada de la libertad 2013-2016 por delito

El INPEC incluye en sus estadísticas mensuales la población privada de la libertad separada por delito, estado (sindicado, condenado) y género. Se ha consolidado la información disponible en reportes individuales entre Julio 2013 y Diciembre 2016, en una sola base de datos. Octubre 2015 se considera un dato faltante, pues a la fecha de descarga el reporte no se encuentra disponible en la página del INPEC [16].

En diciembre de 2016 la base de población privada de la libertad por delito contaba con 184.116, frente a una población carcelaria de 118.532 individuos, en el mismo periodo. Esta diferencia se presenta pues un mismo individuo puede estar privado de la libertad por varios delitos. Las proyecciones se realizan separadamente, se suman como si fueran individuales y se multiplica por un factor de ajuste para agregar las proyecciones al total.

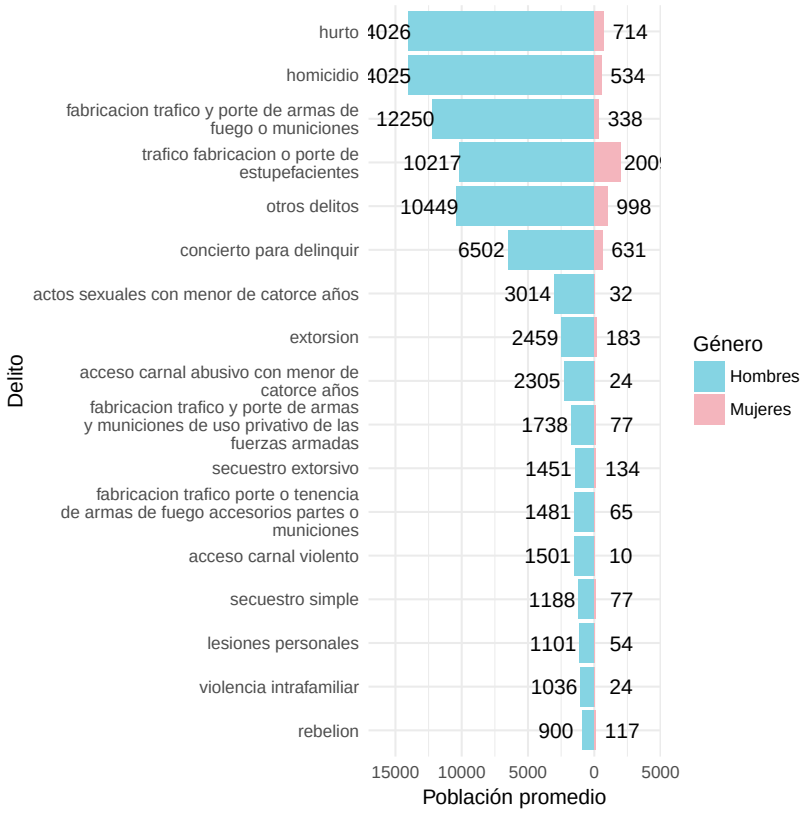

FIGURA 5.1. Población carcelaria promedio por delito y género. Julio 2013-Diciembre 2016

Fuente: INPEC Elaboración propia

En la figura 5.1 se observa la población privada de la libertad promedio entre julio 2013 y diciembre 2016. Los delitos más comunes son: hurto; homicidio; fabricación y porte de armas de fuego o municiones; tráfico, fabricación o porte de estupefacientes y concierto para delinquir. Tanto en hombres como mujeres estos cinco delitos son los más comunes; en los hombres se manifiestan en el orden presentado, pero en las mujeres los delitos más

frecuentes son tráfico, fabricación o porte de estupefacientes y concierto para delinquir, seguidos por hurto, homicidio y fabricación y porte de armas de fuego.

Las figuras 5.2 y 5.3 presentan la evolución de la población carcelaria por delito entre Julio 2013 y Diciembre 2016, para hombres y mujeres respectivamente. Las gráficas evidencian un comportamiento aparentemente independiente de las series: Por ejemplo: mientras las series de hurto, extorsión y secuestro simple presentan un descenso de la población condenada en 2014 y una posterior estabilización, concierto para delinquir presenta una tendencia al alza estable, tanto en sindicados como en condenados.

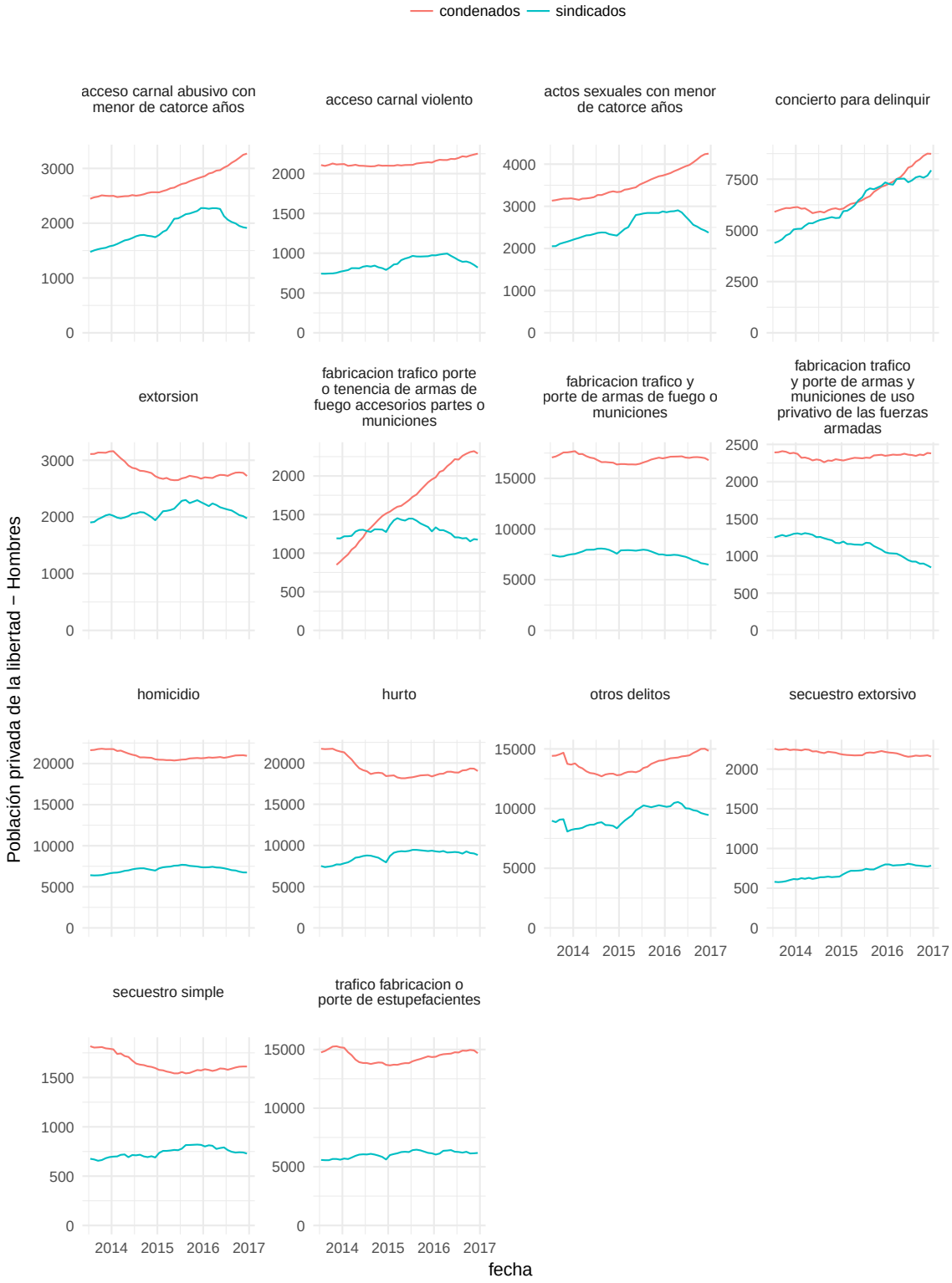

FIGURA 5.2. Evolución de la población carcelaria masculina, Julio 2013-Diciembre 2016

Fuente: INPEC Elaboración propia

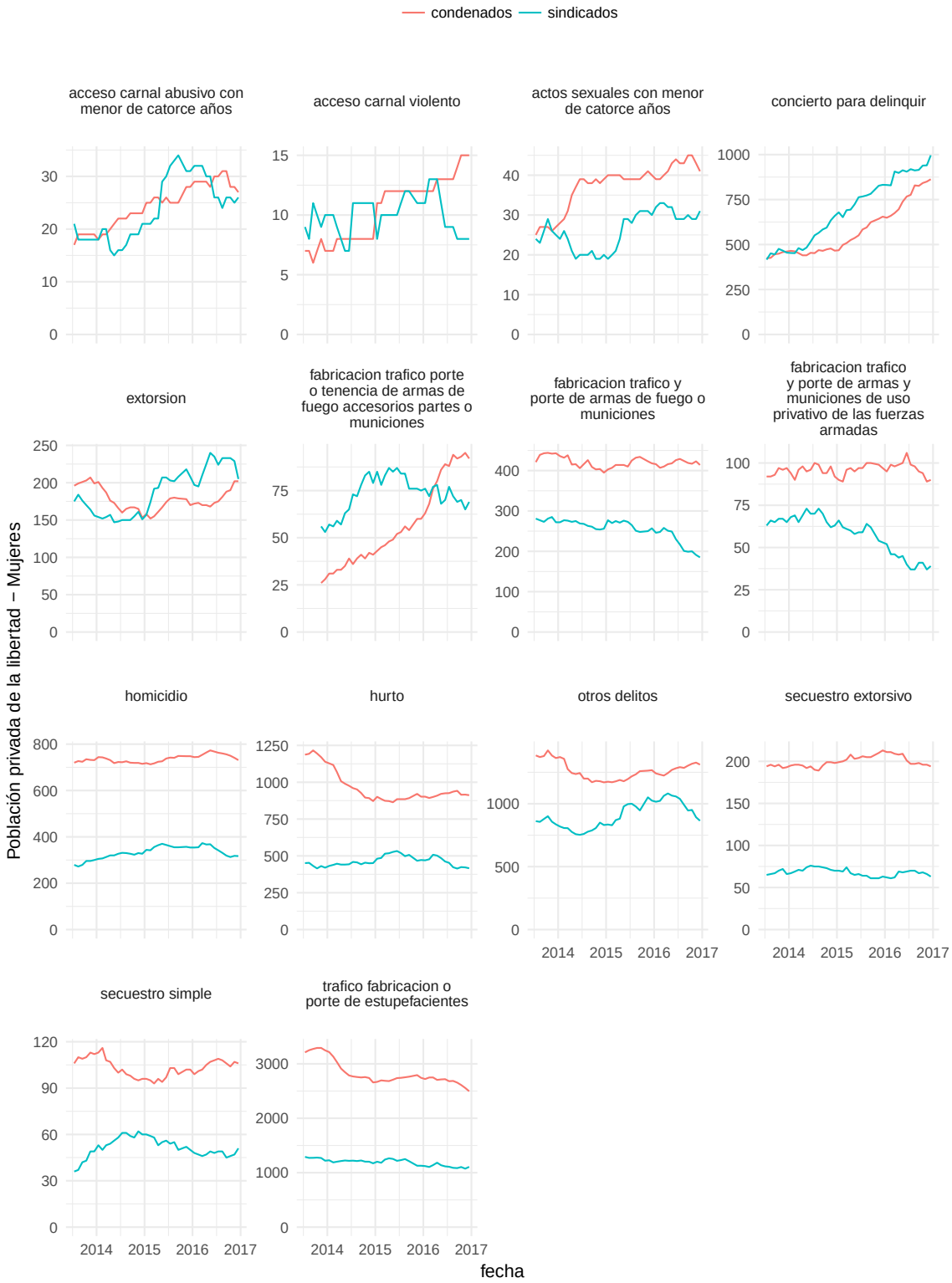

FIGURA 5.3. Evolución de la población carcelaria femenina, Julio 2013-Diciembre 2016

Fuente: INPEC Elaboración propia

5.2. Selección de Software

Para mantener un único ambiente de desarrollo, nos enfocamos en los paquetes desarrollados en R, como los mencionados por Petris en [26], Helske en [11] y Tusell en [34]. Dentro de los paquetes analizados se encuentran *Kfas* por Helske en [11] y *MARSS* de Holmes en [13]. Se utiliza *MARSS* por la flexibilidad en la definición del modelo.

5.3. Identificación del modelo

Holmes & Ward & Willis en [14] definen los modelos MARSS, Modelos auto-regresivos multivariados, de forma general como:

$$\boldsymbol{x_t} = \boldsymbol{B_t x_{t-1}} + \boldsymbol{u_t} + \boldsymbol{C_t c_t} + \boldsymbol{G_t w_t} \quad donde \quad \boldsymbol{w_t} \sim MVN(0, \boldsymbol{Q_t}) \tag{5.1}$$

$$\boldsymbol{y_t} = \boldsymbol{Z_t x_t} + \boldsymbol{a_t} + \boldsymbol{D_t d_t} + \boldsymbol{H_t v_t} \quad donde \quad \boldsymbol{v_t} \sim MVN(0, \boldsymbol{R_t}) \tag{5.2}$$

$$x_1 \sim MVN(\pi, \Lambda) \, o \, x_0 \sim MVN(\pi, \Lambda) \tag{5.3}$$

$\boldsymbol{x}$ es una matriz de $m \times T$, donde cada x_t es una realización de la variable aleatoria X_t en el instante t. m se considera la cantidad de variables latentes.

$\boldsymbol{w}$ es el error de proceso y es un vector aleatorio de dimensión $m \times T$ con distribución normal multivariada (MVN) de media cero y matriz de varianzas y covarianzas $\boldsymbol{Q}$.

$\boldsymbol{y}$ es una matriz de $n \times T$, con T observaciones. Algunas observaciones pueden ser faltantes. n se considera la cantidad de variables observadas.

$\boldsymbol{v}$ es el error de observación de tamaño $n \times T$, con media cero y varianza $\boldsymbol{R}$

$\boldsymbol{B_t}$ y $\boldsymbol{Z_t}$ son parámetros y son de dimensión $m \times m$ y $n \times m$.

$\boldsymbol{u_t}$ y $\boldsymbol{a_t}$ son parámetros de vectores de dimensión $m \times T$ y $n \times T$.

$\boldsymbol{Q_t}$ y $\boldsymbol{R_t}$ son matrices de varianzas y covarianzas de tamaño $m \times m$ y $n \times n$.

π y Λ son parámetros o son fijados con antelación.

$\boldsymbol{C_t}$ y $\boldsymbol{D_t}$ son matrices de parámetros de tamaño $m \times p$ y $n \times q$.

$\boldsymbol{c}$ y $\boldsymbol{d}$ son inputs, p y q variables exógenas, sin datos faltantes, de dimensión $p \times T$ y $q \times T$.

$\boldsymbol{G_t}$ y $\boldsymbol{H_t}$ son inputs, matrices de tamaño $m \times g$ y $n \times h$.

En este capítulo se analizará un modelo bivariado, donde $\boldsymbol{x_t}$ corresponde a la población sindicada y condenada. Se restringe el modelo con las siguientes consideraciones:

- No hay una interacción entre la población de sindicados en el periodo t, y la población de condenados el periodo $t-1$. Es decir se modelo que no hay paso de condenados a sindicados.

- La observación y_t se considera una insesgada frente al valor de la variable x_t. Esto implica $a_t = 0$ y $v_t \sim MVN(0, R)$.
- La población condenada en el periodo t depende solamente de la población en el periodo sindicada y condenada en el periodo t-1. $U2_t = 0$. Las salidas son proporcionales a la población condenada $0 \leq b4 \leq 1$.

$$\begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} = \begin{bmatrix} b1 & 0 \\ b2 & b4 \end{bmatrix} \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} U_1 \\ 0 \end{bmatrix} + \begin{bmatrix} w_{1,t} \\ w_{2,t} \end{bmatrix}, \begin{bmatrix} w_{1,t} \\ w_{2,t} \end{bmatrix} \sim MVN(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{12} & Q_{22} \end{bmatrix}) \quad (5.4)$$

$$\begin{bmatrix} Y_{1,t} \\ Y_{2,t} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix}, \begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix} \sim MVN(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} R_{11} & R_{12} \\ R_{12} & R_{22} \end{bmatrix}) \quad (5.5)$$

Donde

$X_{1,t}$ = Población sindicada en el periodo t

$X_{2,t}$ = Población condenada en el periodo t

b_1 = Efecto en la población sindicada en el periodo $t-1$ en la población sindicada en el periodo t

b_2 = Efecto en la población sindicada en el periodo $t-1$ en la población condenada en el periodo t

b_4 = Efecto en la población condenada en el periodo $t-1$ en la población condenada en el periodo t.

b_4 = Población condenada en el periodo t

U_1 = Variación constante mensual en la población sindicada no asociada a la población sindicada en el periodo anterior.

$Y_{1,t}$ = Población sindicada observada en el periodo t

Y_{2_t} = Población condenada observada en el periodo t

w_1, w_2, v_1, v_2,$\boldsymbol{Q}$, $\boldsymbol{R}$ mantienen el mismo sentido que en las ecuaciones 5.1 y 5.2

Modelar el sistema a partir de este modelo teórico, nos permite estimar sus parámetros usando el paquete MARSS. Tratar la serie como un modelo estado-espacio también permite tratar los faltantes con facilidad.

En el ejercicio con datos reales se ajustó un segundo modelo con una variable Dummy $\boldsymbol{c}$ en 2014, debido al efecto de los cambios normativos en la tendencia de la población. Para incluir la variable Dummy, se incluyó una variable exógena c_t de dimensión $2xT$ y se estimó el parámetroC_t, de la ecuación 5.1 en la ecuación 5.4.

5.4. Estimación de parámetros

La estimación se realiza en series bivariadas por delito, género. Se presenta en detalle procedimiento para la serie de hurto en hombres, en series adicionales se presentan resul-

tados de la proyección. La estimación de los parámetros se realiza con el paquete MARSS, que utiliza el algoritmo EM para realizar la estimación.

Una variable Dummy correspondiente a 2014 se agregó al modelo. En este periodo se detuvo el crecimiento de la población carcelaria gracias a la entrada en vigor de la ley 1709 de 2014. Dos mediciones del impacto de la ley 1709 son realizadas por el INPEC en 2014, señalando el cambio observado en la tendencia de la población carcelaria [25] [1]. Las gráficas 5.2 y 5.3 muestran que en 2015 y 2016 la población carcelaria regresa a su comportamiento usual.

	x
b1	0.94
b2	0.05
b4	0.98
U1	514.58
q11	23317.33
q12	-11076.07
q22	18110.12

TABLA 5.1. Parámetros estimados: hurto, hombres

Fuente: INPEC Elaboración propia

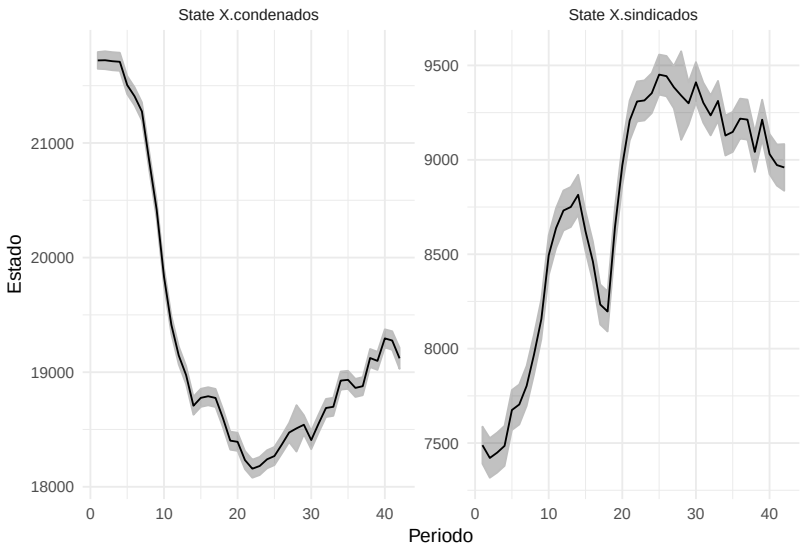

FIGURA 5.4. Población sindicada y condenada por hurto, hombres

Fuente: INPEC Elaboración propia

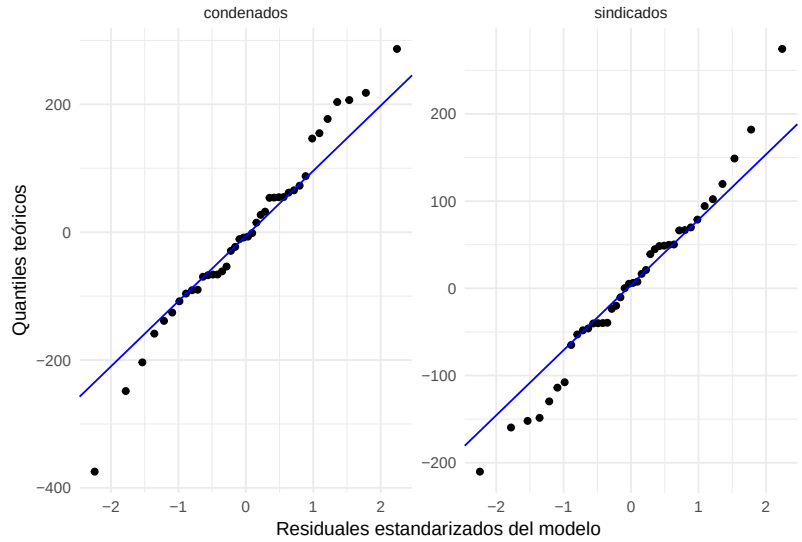

FIGURA 5.5. Población carcelaria promedio por delito y género. Octubre 2013-Diciembre 2016

Fuente: INPEC Elaboración propia

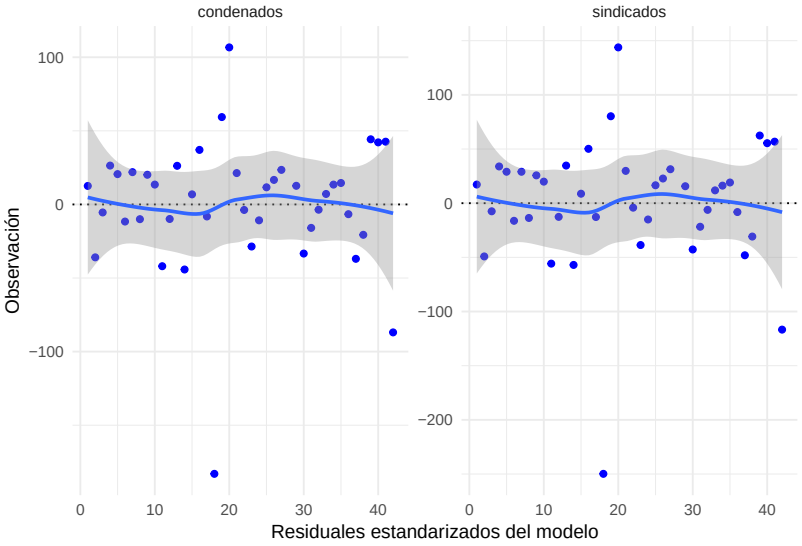

FIGURA 5.6. Población carcelaria promedio por delito y género. Octubre 2013-Diciembre 2016

Fuente: INPEC Elaboración propia

El qqplot sugiere un comportamiento normal del error con presencia de datos atípicos. La gráfica muestra rachas de puntos arriba y abajo de cero, lo que sugiere la presencia de autocorrelación en el error. Esta aparente autocorrelación puede estar asociada a componentes estacionales o cambios estructurales.

5.5. Proyecciones 2017 - 2020

Se realiza proyección de la población media sindicada y condenada en la tabla 5.2

	Periodo	Sindicados	Condenados
V1	2017-01-15	8862.00	19050.00
V2	2017-02-15	8880.00	19067.00
V3	2017-03-15	8897.00	19085.00
V4	2017-04-15	8913.00	19103.00
V5	2017-05-15	8929.00	19121.00
V6	2017-06-15	8943.00	19140.00
V7	2017-07-15	8956.00	19159.00
V8	2017-08-15	8969.00	19178.00
V9	2017-09-15	8981.00	19198.00
V10	2017-10-15	8993.00	19217.00
V11	2017-11-15	9003.00	19237.00
V12	2017-12-15	9014.00	19257.00
V13	2018-01-15	9023.00	19277.00
V14	2018-02-15	9032.00	19297.00
V15	2018-03-15	9041.00	19317.00
V16	2018-04-15	9049.00	19337.00
V17	2018-05-15	9056.00	19357.00
V18	2018-06-15	9064.00	19377.00
V19	2018-07-15	9070.00	19397.00
V20	2018-08-15	9077.00	19417.00
V21	2018-09-15	9083.00	19436.00
V22	2018-10-15	9089.00	19456.00
V23	2018-11-15	9094.00	19475.00
V24	2018-12-15	9099.00	19494.00
V25	2019-01-15	9104.00	19513.00
V26	2019-02-15	9108.00	19532.00
V27	2019-03-15	9113.00	19550.00
V28	2019-04-15	9117.00	19569.00
V29	2019-05-15	9120.00	19587.00
V30	2019-06-15	9124.00	19605.00
V31	2019-07-15	9127.00	19622.00
V32	2019-08-15	9131.00	19640.00
V33	2019-09-15	9134.00	19657.00
V34	2019-10-15	9137.00	19674.00
V35	2019-11-15	9139.00	19690.00
V36	2019-12-15	9142.00	19707.00

TABLA 5.2. Proyección a treinta y seis meses: hurto, hombres

Fuente: INPEC Elaboración propia

	Periodo	Género	Sindicados	Condenados	Total
1	2017-01-15	Hombres	56331	116190	172521
2	2017-02-15	Hombres	56393	116397	172790
3	2017-03-15	Hombres	56447	116603	173050
4	2017-04-15	Hombres	56495	116807	173302
5	2017-05-15	Hombres	56536	117009	173544
6	2017-06-15	Hombres	56570	117208	173778
7	2017-07-15	Hombres	56598	117405	174003
8	2017-08-15	Hombres	56619	117599	174218
9	2017-09-15	Hombres	56634	117790	174424
10	2017-10-15	Hombres	56643	117978	174621
11	2017-11-15	Hombres	56646	118163	174809
12	2017-12-15	Hombres	56642	118344	174987
13	2018-01-15	Hombres	56633	118522	175155
14	2018-02-15	Hombres	56618	118696	175314
15	2018-03-15	Hombres	56596	118866	175462
16	2018-04-15	Hombres	56569	119032	175601
17	2018-05-15	Hombres	56535	119195	175730
18	2018-06-15	Hombres	56496	119353	175849
19	2018-07-15	Hombres	56450	119507	175957
20	2018-08-15	Hombres	56398	119657	176055
21	2018-09-15	Hombres	56340	119802	176142
22	2018-10-15	Hombres	56276	119943	176219
23	2018-11-15	Hombres	56262	120080	176342
24	2018-12-15	Hombres	56251	120212	176463
25	2019-01-15	Hombres	56236	120340	176576
26	2019-02-15	Hombres	56219	120463	176681
27	2019-03-15	Hombres	56197	120581	176779
28	2019-04-15	Hombres	56173	120695	176868
29	2019-05-15	Hombres	56145	120805	176949
30	2019-06-15	Hombres	56114	120909	177023
31	2019-07-15	Hombres	56080	121009	177088
32	2019-08-15	Hombres	56042	121104	177146
33	2019-09-15	Hombres	56001	121194	177196
34	2019-10-15	Hombres	55957	121280	177237
35	2019-11-15	Hombres	55910	121361	177271
36	2019-12-15	Hombres	55860	121437	177296

TABLA 5.3. Proyección de la población carcelaria total, hombres

Fuente: INPEC Elaboración propia

	Periodo	Género	Sindicados	Condenados	Total
1	2017-01-15	Mujeres	4396	7511	11907
2	2017-02-15	Mujeres	4413	7528	11941
3	2017-03-15	Mujeres	4429	7544	11973
4	2017-04-15	Mujeres	4446	7560	12005
5	2017-05-15	Mujeres	4461	7575	12037
6	2017-06-15	Mujeres	4477	7591	12068
7	2017-07-15	Mujeres	4492	7606	12098
8	2017-08-15	Mujeres	4507	7621	12128
9	2017-09-15	Mujeres	4523	7635	12158
10	2017-10-15	Mujeres	4538	7650	12187
11	2017-11-15	Mujeres	4552	7664	12217
12	2017-12-15	Mujeres	4567	7679	12246
13	2018-01-15	Mujeres	4582	7693	12275
14	2018-02-15	Mujeres	4597	7707	12304
15	2018-03-15	Mujeres	4612	7721	12334
16	2018-04-15	Mujeres	4628	7736	12363
17	2018-05-15	Mujeres	4643	7750	12393
18	2018-06-15	Mujeres	4658	7764	12423
19	2018-07-15	Mujeres	4674	7779	12452
20	2018-08-15	Mujeres	4689	7793	12483
21	2018-09-15	Mujeres	4705	7808	12513
22	2018-10-15	Mujeres	4721	7823	12544
23	2018-11-15	Mujeres	4737	7838	12575
24	2018-12-15	Mujeres	4754	7852	12606
25	2019-01-15	Mujeres	4770	7868	12638
26	2019-02-15	Mujeres	4787	7883	12670
27	2019-03-15	Mujeres	4804	7898	12703
28	2019-04-15	Mujeres	4822	7914	12736
29	2019-05-15	Mujeres	4839	7930	12769
30	2019-06-15	Mujeres	4857	7946	12803
31	2019-07-15	Mujeres	4875	7962	12837
32	2019-08-15	Mujeres	4894	7978	12872
33	2019-09-15	Mujeres	4913	7995	12907
34	2019-10-15	Mujeres	4932	8011	12943
35	2019-11-15	Mujeres	4951	8028	12979
36	2019-12-15	Mujeres	4971	8045	13016

TABLA 5.4. Proyección de la población carcelaria total, mujeres

Fuente: INPEC Elaboración propia

La proyección de la población total se obtiene como la suma de las proyecciones por delito. Proyecciones par hombres y mujeres se presentan en las tablas 5.3 y 5.4. Las gráficas 5.7 y 5.8 presentan la proyección por delito. Las proyecciones son coherentes con el comportamiento histórico de cada serie. En fabricación, tráfico y porte de armas y municiones de uso privativo de las fuerzas armadas no resulta coherente la proyección, pues presenta una serie de sindicados que cae a cero y una serie de condenados que continua creciendo luego, para resolver este inconveniente se podrían restringir los parámetros del modelo, tal que $0 \leq b1, b2, b3 \leq 1$.

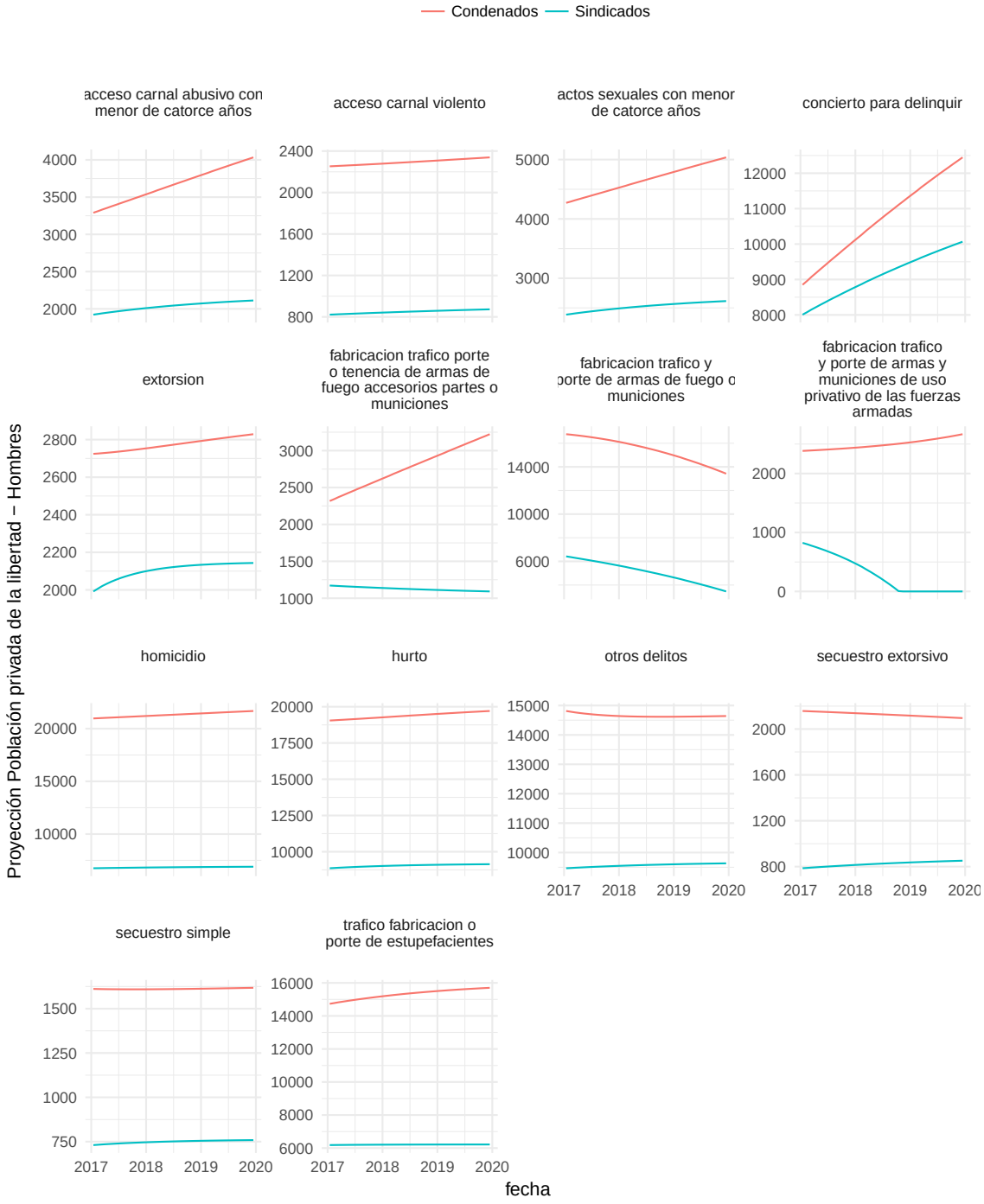

FIGURA 5.7. Proyección de población carcelaria o por delito, hombres 2017 - 2020

Fuente: INPEC Elaboración propia

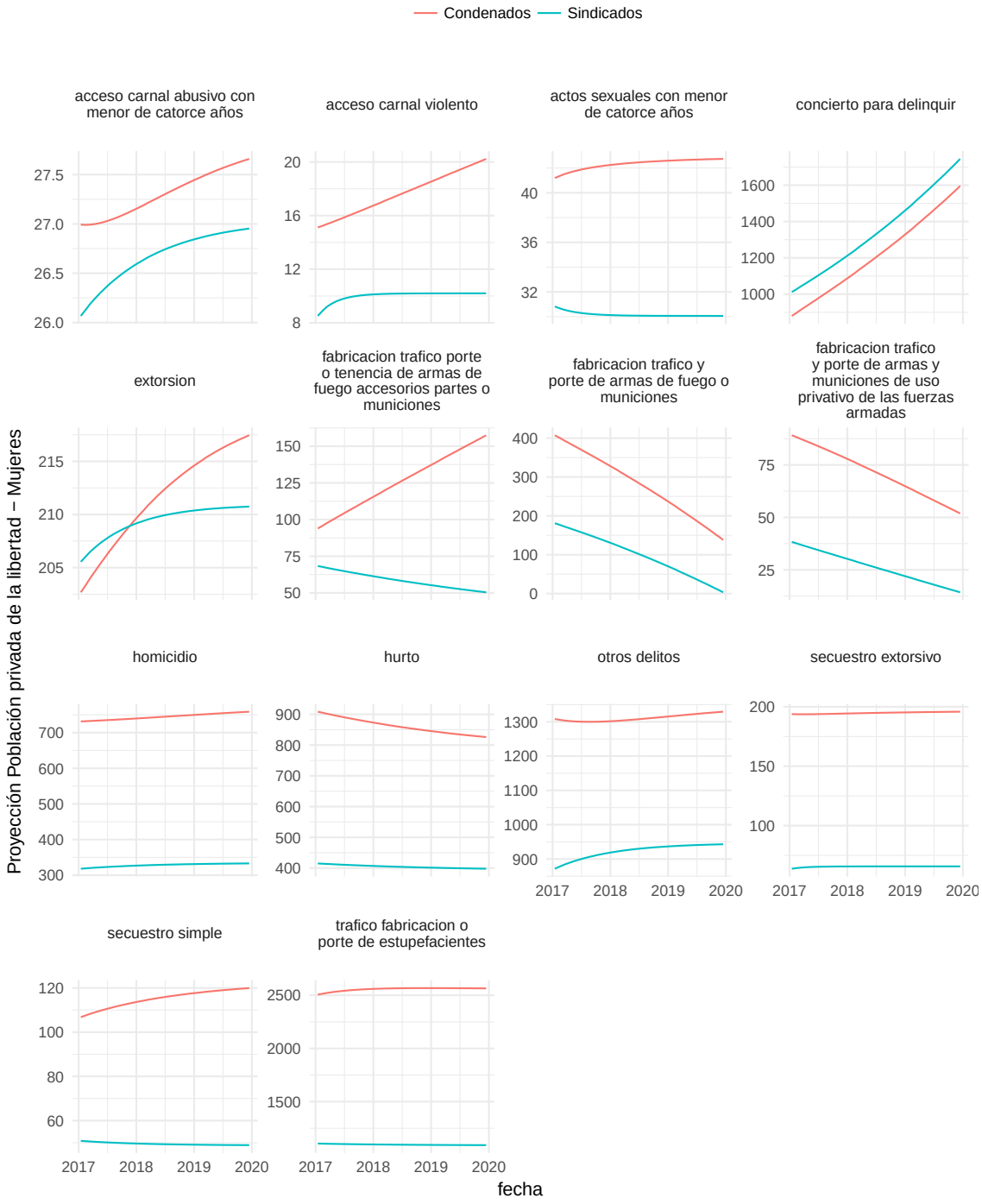

FIGURA 5.8. Proyección de población carcelaria o por delito, mujeres 2017 - 2020

Fuente: INPEC Elaboración propia

5.6. Discusión

El análisis descriptivo nos permite observar que los comportamientos por delito son diferentes en hombres y mujeres, incluso si la estructura etaria es similar, como observamos en el capítulo 4. Cada delito presenta una tendencia diferente, que puede ser de crecimiento acelerado como en concierto para delinquir o reducción leve como otros delitos. Realizar

análisis separados por género y delito permitiría, teniendo información de periodos más extensos, analizar el impacto de decisiones jurídicas puntuales en cada tipología delictiva.

Los modelos multivariados permiten incluir la interacción entre la población sindicada y condenada. Las variaciones en la población sindicada constituyen predictores de variaciones en la población condenada en el mismo sentido, en los periodos posteriores Proyecciones más precisas de la población sindicada implican un mejor pronóstico de la población condenada y la consecuente población total.

Medidas como la ley 1709 de 2014 han generado impactos sobre la población carcelaria y su tendencia de crecimiento, al intervenir las tasas de ingreso al sistema penitenciario y los tiempos de sentencia en el sistema judicial. Las proyecciones de población carcelaria permiten establecer lineas base, sobre las cuales medir el impacto de estas medidas. Análisis de impacto por delito y población, se podrían realizar, al contar con datos desagregados para periodos más extensos que los disponibles actualmente.

Los modelos estado-espacio son una herramienta potente que permite estimar modelos interpretables en series de tiempo, incluso con variables no observadas, como las sentencias y en presencia de datos faltantes. El desarrollo de paquetes en R que permitan realizar estimaciones de máxima verosimilitud bajo restricciones en los parámetros, en forma simple, como la trabajada en el paquete MARSS quedan dentro de trabajo futuro. Por ejemplo restringir los parámetros $b1$ y $b2$ a ser valores entre cero y uno.

Tratar cada delito como una población separada, considerando la interacción entre la población sindicada y condenada permite analizar la tendencia y variabilidad de cada serie. En este capítulo se evidenció comportamientos diferentes para cada serie en tendencia y variabilidad, con algunos componentes compartidos, como el efecto de los cambios de regulación en 2014 orientados a reducir la población carcelaria.

Al imponer un modelo teórico en todas las series de delitos, se puede perder información específica de las series de delitos. Como trabajo futuro está analizar el comportamiento individual de cada serie, y sus cambios estructurales.

La longitud de la serie disponible no permite estimar un comportamiento estacional, y en consecuencia dificulta separar los efectos estacionales de cambios de régimen. En cada delito se presentan errores atípicos, donde no se puede asignar a un componente estacional, o a cambios en la reglamentación.

El modelo no incluye en principio el efecto de la estructura etaria, así que puede resultar menos efectivo en el largo plazo que los modelos de poblaciones pequeñas.

CAPÍTULO 6

Comparación

Dado que los tres modelos analizados utilizan datos a un nivel de agregación diferente, en ventanas de tiempo distintas, no resulta conveniente comparar los modelos frente al ajuste de los periodos que se usaron como base para la estimación. La comparación se realiza entonces comparando las proyecciones realizadas, con las mediciones reales de población en un periodo de validación. El periodo de validación considerado es desde enero de 2017 hasta mayo de 2018.

Para realizar la comparación se debieron realizar ajustes en las proyecciones de razón-censal y estado-espacio. En el caso de la razón-censal se utilizó suavizamiento a través de una serie de tiempo estructural, para llevar la proyección anual a nivel mensual. En las proyecciones por modelos estado-espacio se realizó un ajuste en el orden de magnitud de las proyecciones, puesto que las proyecciones de delitos no son iguales a la población total; se calculó un factor de ajuste como la razón entre la población total observada y la suma de poblaciones por delito. Las proyecciones ajustadas se observan en la tabla 6.1 y la gráfica 6.1. El error de pronóstico definido como $proyectado - real$ se presenta en la tabla 6.2

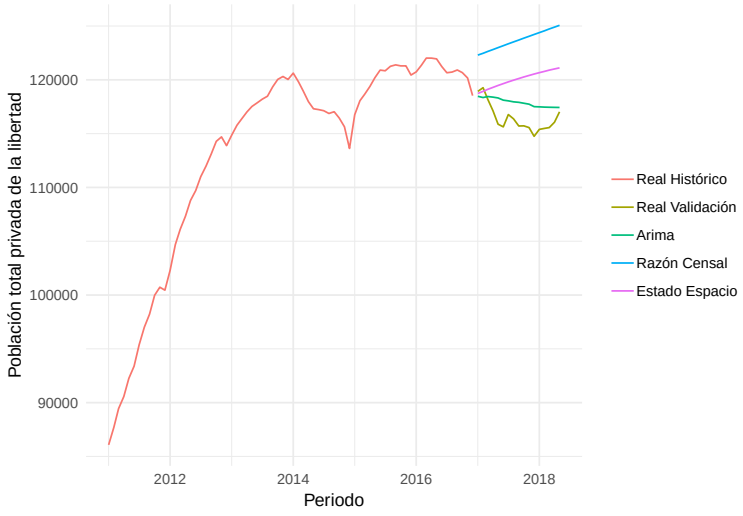

FIGURA 6.1. Proyecciones de Población privada de la libertad Enero 2017 - Mayo 2018

Fuente: INPEC Elaboración propia

	Periodo	Real Validación	Arima	Razón Censal	Estado Espacio
1	2017-01-01	118925	118466	122299	118733
2	2017-02-01	119269	118354	122466	118928
3	2017-03-01	118186	118446	122638	119116
4	2017-04-01	117119	118392	122814	119299
5	2017-05-01	115878	118315	122992	119475
6	2017-06-01	115628	118111	123171	119646
7	2017-07-01	116773	118045	123348	119810
8	2017-08-01	116373	117953	123523	119968
9	2017-09-01	115708	117909	123698	120120
10	2017-10-01	115721	117825	123872	120265
11	2017-11-01	115562	117738	124045	120405
12	2017-12-01	114750	117508	124216	120538
13	2018-01-01	115396	117488	124388	120666
14	2018-02-01	115488	117463	124558	120787
15	2018-03-01	115563	117451	124728	120901
16	2018-04-01	116058	117440	124897	121010
17	2018-05-01	117026	117431	125066	121112

TABLA 6.1. Proyecciones y observado enero 2017 a mayo 2018

	Periodo	Arima	Razón Censal	Estado Espacio
1	2017-01-01	-459	3374	-192
2	2017-02-01	-915	3197	-341
3	2017-03-01	260	4452	930
4	2017-04-01	1273	5695	2180
5	2017-05-01	2437	7114	3597
6	2017-06-01	2483	7543	4018
7	2017-07-01	1272	6575	3037
8	2017-08-01	1580	7150	3595
9	2017-09-01	2201	7990	4412
10	2017-10-01	2104	8151	4544
11	2017-11-01	2176	8483	4843
12	2017-12-01	2758	9466	5788
13	2018-01-01	2092	8992	5270
14	2018-02-01	1975	9070	5299
15	2018-03-01	1888	9165	5338
16	2018-04-01	1382	8839	4952
17	2018-05-01	405	8040	4086

TABLA 6.2. Error de pronóstico enero 2017 a mayo 2018

	Sesgo
Arima	1465
Razón Censal	7253
Estado Espacio	3609

TABLA 6.3. Sesgo de pronóstico enero 2017 a mayo 2018

El modelo ARIMA que considera la suma de la población sindicada y la condenada es el que presenta el mejor desempeño. Este tipo de modelos presentan un desempeño adecuado en el corto plazo; sin embargo, al ser series integradas los intervalos de confianza del pronóstico divergen aceleradamente, haciéndolo menos apropiado para proyecciones de largo plazo.

Tanto el forecast de los los modelos estado espacio como el modelo de poblaciones pequeñas sobre-estiman el crecimiento de la población carcelaria. Este escenario puede ser coyuntural, si se considera que la población carcelaria decreció durante 2016. Dado que el sesgo de ambos modelos era mayor al de los modelos SARIMA no se utilizan otros criterios de comparación como el error absoluto medio.

Conclusiones

Las poblaciones carcelarias como otras poblaciones pequeñas están ligadas a la población total, y a su distribución por edad y género; en este caso existen tasas de encarcelamiento diferenciales por edad y género que influyen en el comportamiento de la población carcelaria en el largo plazo. Una parte del incremento de la población carcelaria total, puede ser asociado al crecimiento de la población nacional y al cambio de su estructura por edad. En este sentido es importante incluir información de la población total, al proyectar poblaciones carcelarias, particularmente en el largo plazo.

La calidad y la disponibilidad de los datos pueden hacer difícil incluir en las proyecciones información de la población total. En el desarrollo de este documento solo se contó con dos años de información sobre la población carcelaria separada por edad, en la granularidad requerida. En casos como este, donde se cuenta con conteos de población en periodos más extensos de tiempo es posible proyectar estos conteos como SARIMA. Trabajar con modelos SARIMA permite estimar los componentes estacionales, para mejorar la precisión. Los modelos ajustados sugieren un buen desempeño de la proyección en el corto plazo.

Al contar con información sobre los conteos de población sindicada y condenada es posible modelar la interacción entre estas, para tener pronósticos más precisos, especialmente si esta proyección se puede realizar por delito. Trabajar por delito permite identificar cambios estructurales asociados a cambios en la política criminal, que no son fácilmente observables en el total de la población privada de la libertad.

Uno de los usos de las proyecciones de poblaciones pequeñas es la proyección de la infraestructura necesaria para atender las necesidades de la población en cuestión. En el caso de poblaciones privadas de la libertad, el uso de infraestructuras separadas hace que sea crítico realizar la proyección separada por género. Dado que los delitos por los que ingresan son diferentes entre hombres y mujeres, esta proyección debería considerar proyecciones separadas por delito.

Los trabajos futuros sobre poblaciones carcelarias en Colombia estarán mediadas por la calidad y la disponibilidad de datos sobre esta población.

Trabajo futuro

Dentro de los posibles trabajos futuros de proyección de poblaciones carcelarias se encuentran:

Proyecciones más precisas se pueden obtener al trabajar con los ingresos, salidas del sistema, duración de las condenas y de los procesos. Así mismo, sería posible estimar el impacto que los cambios normativos tienen sobre las dinámicas de la población carcelaria y la influencia de variables exógenas como el desempleo o los niveles de pobreza en los diferentes componentes.

Proyectar la tasa de encarcelamiento específicas por edad, género y delito, incluso a nivel regional y departamental resultaría en un insumo importante para la planificación de infraestructura y la reducción del hacinamiento.

La estructura de las tasa de encarcelamiento específicas por edad podría permitir el uso de datos funcionales que consideren la interacción entre las tasas de encarcelamiento específicas. Trabajar con tasas de encarcelamiento por edad, delito podría permitir una mejor comprensión de la estructura observada de las tasas de encarcelamiento.

Aunque en este trabajo no se definió como parte de los objetivos, es común analizar: tasas de reincidencia y primera ofensa. Posteriores proyecciones podrían ser enriquecidas con esta información.

El desarrollo de paquetes similares a MARSS [13] , que crean accesos sencillos a paquetes de optimización de la función de máxima verosimilitud que permitan imponer restricciones sobre los parámetros a estimar.

Bibliografía

[1] Oficina Asesora, De Planeación Grupo, Luis Eduardo, and Castro Gil, *Impacto Ley 1709 de 2014 Primer Informe*, (2014).

[2] Alfred Blumstein, Jacqueline Cohen, and Harold D. Miller, *Demographically disaggregated projections of prison populations*, Journal of Criminal Justice **8** (1980), no. 1, 1–26.

[3] George E P Box, Gm Jenkins, and Gc Reinsel, *Time series analysis: forecasting and control*, 2013.

[4] Paolo Brandimarte, *Time Series Models*, Quantitative Methods: An Introduction for Business Management, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 527–579.

[5] Peter J. Brockwell, *Time Series*, International Encyclopedia of Statistical Science, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 1601–1605.

[6] Jacques J.F Commandeur and Siem Jan Koopman, *An Introduction to State Space Time Series Analysis*, (2007), 189.

[7] DANE, *Estimaciones población DANE*, 2009.

[8] Departamento Nacional de Planeación, *Conpes 3828 POLÍTICA PENITENCIARIA Y CARCELARIA EN COLOMBIA*, (2015).

[9] James Durbin and Siem Jan Koopman, *Time Series Analysis by State Space Methods: Second Edition*, vol. 7, 2012.

[10] Andrew C Harvey, *Forecasting, structural time series models and the kalman filter*, Cambridge university press, 1990.

[11] Jouni Helske, *Exponential Family State Space Models in R*, Journal of Statistical Software **78** (2017), no. 10, 1–39.

[12] E E Holmes, E J Ward, and M D Scheuerell, *Analysis of multivariate time- series using the MARSS package*, (2018).

[13] Elizabeth Holmes, Eric Ward, and Kellie Wills, *MARSS: Multivariate Autoregressive State-Space Modeling*, 2018.

[14] Elizabeth E Holmes, Eric J Ward, and Kellie Wills, *MARSS: Multivariate Autoregressive State-space Models for Analyzing Time-series Data.*

[15] Rob J Hyndman, *{forecast}: Forecasting functions for time series and linear models*, 2017.

[16] INPEC, *Estadisticas - Estadística - INPEC*, 2018.

[17] Institute for Criminal Policy Research, *Colombia | World Prison Brief*, 2016.

[18] Instituto Nacional Penitenciario y Carcelario, *Plan de direccionamiento estratégico 2015-108, Misión y Visión*, 2016.

[19] Ministry Justice, *Prison Population Projections 2014 - 2020, England and Wales*, Home Office Statistical Bulletin (2014), no. November, 31.

[20] Ronald Demos Lee and Shripad Tuljapurkar, *Stochastic Population Forecasts for the United States: Beyond High, Medium, and Low*, Journal of the American Statistical Association **89** (1994), no. 428, 1175.

[21] G. M. Ljung and G. E. P. Box, *On a Measure of Lack of Fit in Time Series Models*, Biometrika **65** (1978), no. 2, 297.

[22] Helmut Lütkepohl, *New Introduction to Multiple Time Series Analysis*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.

[23] Florance Matarise, *Intervention analysis in time series*, pp. 682–685, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[24] Todd D Minton, Bjs Statistician, Scott Ginder, Susan M Brumbaugh, Hope Smiley-Mcdonald, and Harley Rohloff, *Census of Jails: Population Changes, 1999 - 2013*, (2015).

[25] Oficina Asesora de Planeación and Grupo Estadistica, *Impacto Ley 1709 de 2014 Segundo Informe*, Tech. report, 2014.

[26] Giovanni Petris and Sonia Petrone, *State Space Models in R*, Journal of Statistical Software **41** (2011), no. 4, 128–129.

[27] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2017.

[28] Karthik Ram and Hadley Wickham, *wesanderson: A Wes Anderson Palette Generator*, 2018.

[29] Robert H Shumway and David Stoffer, *Time Series Analysis and Its Applications*, Springer Texts in Statistics, vol. 45, Springer New York, New York, NY, 2011.

[30] Robert H Shumway and David S Stoffer, *Time Series Analysis and Its Applications With R Examples*, second edi ed., Springer, 2006.

[31] David Stoffer, *astsa: Applied Statistical Time Series Analysis. R package version 1.8*, 2017.

[32] David A Swanson and Jeff Tayman, *Subnational Population Estimates*, The Springer Series on Demographic Methods and Population Analysis, vol. 31, Springer Netherlands, Dordrecht, 2012.

[33] Ruey S. Tsay, *Analysis of Financial Time Series*, John Wiley & Sons, Inc., Hoboken, NJ, USA, jan 2002.

[34] Fernando Tusell, *Kalman Filtering in R*, Journal of Statistical Software **39** (2011), no. 2, 1–27.

[35] Wai-Yin Wan1, Steve Moffatt2, Zachary Xie3, Simon Corben, and Don Weatherburn, *Forecastin prison populations using sentencing and arrest data*, Tech. report.