

*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.  
ABRIL DE 2017

*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.  
ABRIL DE 2017

**Título en español**

Proyección de poblaciones carcelarias en Colombia

**Title in English**

Prison populations projections for Colombia

**Resumen:** Se presentan proyecciones de población carcelaria cuando no se conocen las tasas de ingreso al sistema carcelario, el tiempo de juicio, ni la duración de la pena. Para estimar estas variables no observadas se utilizan modelos estado espacio multivariados, tomando como base la población sindicada y condenada mensual

**Abstract:** Not there yet!

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

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

## Nota de aceptación

Trabajo de tesis

Aprobado

“Mención Meritoria o Laureada”

---

Jurado

Jurado uno

---

Jurado

Jurado dos

---

Director

B. Piedad Urdinola

Bogotá, D.C., Abril 19 de 2017

---

---

## Dedicado a

---

---

El problema es la codificación de los archivos, que cambia.

---

---

## Agradecimientos

---

---

Tildes corregidas.

---

---

# Índice general

---

---

|  |            |
|--|------------|
| <b>Índice general</b>                                      | <b>I</b>   |
| <b>Índice de tablas</b>                                    | <b>III</b> |
| <b>Índice de figuras</b>                                   | <b>IV</b>  |
| <b>Introducción</b>  | <b>V</b>   |
| .1. Objetivos . . . . .                                    | VI         |
| .1.1. Objetivo General . . . . .                           | VI         |
| .1.2. Objetivos Específicos . . . . .                      | VI         |
| <b>1. Antecedentes teóricos</b>                            | <b>1</b>   |
| 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 ARIMA . . . . .                             | 3          |
| 1.2.2. Modelos Estado-Espacio . . . . .                    | 4          |
| <b>2. La población carcelaria en Colombia 1991 - 2017</b>  | <b>5</b>   |
| 2.1. Análisis exploratorio . . . . .                       | 5          |
| 2.2. El sistema penitenciario en Colombia . . . . .        | 6          |
| 2.2.1. Identificación del modelo . . . . .                 | 6          |
| <b>3. Modelos SARIMA</b>                                   | <b>8</b>   |
| 3.1. Identificación del modelo . . . . .                   | 8          |
| 3.1.1. Población total . . . . .                           | 11         |
| 3.1.2. Población sindicada . . . . .                       | 12         |

---

|  |           |
|--|-----------|
| 3.1.3. Población condenada . . . . .     | 12        |
| 3.2. Estimación de parámetros . . . . .  | 12        |
| 3.2.1. Población total . . . . .         | 12        |
| 3.2.2. Población sindicada . . . . .     | 16        |
| 3.2.3. Población condenada . . . . .     | 16        |
| 3.3. Proyección . . . . .                | 16        |
| 3.3.1. Población total . . . . .         | 16        |
| 3.3.2. Población sindicada . . . . .     | 17        |
| 3.3.3. Población condenada . . . . .     | 17        |
| 3.4. Conclusiones . . . . .              | 17        |
| <b>4. Modelos estado-espacio</b>         | <b>18</b> |
| 4.1. Marco teórico . . . . .             | 18        |
| 4.2. Selección de Software . . . . .     | 18        |
| 4.3. Identificación del modelo . . . . . | 19        |
| 4.4. Simulación Monte-Carlo . . . . .    | 19        |
| 4.5. Estimación de parámetros . . . . .  | 19        |
| 4.6. Proyecciones 2017 - 2020 . . . . .  | 19        |
| 4.7. Conclusiones . . . . .              | 19        |
| <b>Conclusiones</b>                      | <b>20</b> |
| <b>Trabajo futuro</b>                    | <b>21</b> |
| <b>Bibliografía</b>                      | <b>22</b> |



---

---

## Índice de tablas

---

---

|  |    |
|--|----|
| 3.1. Comportamiento de la ACF y PACF en modelos ARMA(p,q) [19] . . . . . | 10 |
| 3.2. Comparación de criterios de información . . . . .                   | 13 |
| 3.3. Parámetros del modelo (3,2,1,0,0,2) . . . . .                       | 13 |
| 3.4. Parámetros del modelo (1,2,1,0,0,2) . . . . .                       | 13 |

---

---

## Índice de figuras

---

---

|   |    |
|---|----|
| 2.1. Población privada de la libertad 1991 - 2017 . . . . .                       | 5  |
| 2.2. Tasa de encarcelamiento según genero 1991 - 2017 . . . . .                   | 6  |
| 2.3. Población carcelaria por situación jurídica . . . . .                        | 7  |
| 3.1. Variación inter-mensual de población carcelaria . . . . .                    | 8  |
| 3.2. Descomposición de la variación inter-mensual de población carcelaria total . | 9  |
| 3.3. Descomposición de la variación inter-mensual de población carcelaria total . | 9  |
| 3.4. Descomposición de la variación inter-mensual de población carcelaria total . | 10 |
| 3.5. Autocorrelación parcial de la variación inter-mensual . . . . .              | 11 |
| 3.6. Diagnóstico del sarima (1,1,1)(0,0,2), población total . . . . .             | 14 |
| 3.7. Diagnóstico del sarima (1,1,1)(0,0,1), población total . . . . .             | 15 |
| 3.8. Diagnóstico del sarima (3,2,1)(0,0,2), población total . . . . .             | 16 |
| 3.9. Proyección de la población carcelaria total . . . . .                        | 17 |

---

---

## 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 .<sup>es</sup> la institución pública administradora del sistema penitenciario y carcelario del país"[12], publica periódicamente la serie histórica de población carcelaria, que da cuenta de la población al cierre de cada mes desde 1991, separada por situación judicial (sindicados, condenados) y género.

De esta serie de población El CONPES 3828 anota que .<sup>el</sup> 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"[7]; ambas tasas están muy por encima del 1,18 % anual, crecimiento poblacional nacional estimado por el DANE para el quinquenio 2005-2010 [6]. 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 .<sup>en</sup> el periodo 1993-2014 la proporción de cupos habilitados creció en 173,59 %, mientras la población privada de la libertad (PPL) creció en un 315,39 %."[7]. 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.

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. Este problema se ha abordado anteriormente desde el estudio de poblaciones pequeñas.[21]

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 ". [21]. 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.

Las series de población carcelaria disponibles para Colombia son mediciones de la población al cierre de mes, agrupada por situación judicial y género. Las 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 son libres, 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 busca ofrecer 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.

## **.1. Objetivos**

### **.1.1. Objetivo General**

Comparar el ajuste de los métodos clásicos de proyección de poblaciones pequeñas (ration-correlation, Censal-ration methods) con el ajuste de series de Tiempo ARMAX y de modelos Estado-Espacio, para la población carcelaria de Colombia 1991-2016.

### **.1.2. Objetivos Específicos**

- Proyectar la población carcelaria usando los métodos de ratio-correlation y censal-ratio. (0 %)
- Proyectar el comportamiento de la población carcelaria a través de series de tiempo ARIMA y ARIMAX, suponiendo tasas de transición estables. (PROCESO) ()
- Aplicar modelos estado espacio para estimar las tasas de transición, cuando: son no observadas, varían en el tiempo y tienen correlación con variables exógenas. (PROCESO)
- Comparar el ajuste de los tres métodos a través de simulación Montecarlo. (Nulo)
- Aplicar los métodos descritos en los objetivos anteriores, a los datos de población carcelaria colombiana 1991-2016. ()

# CAPÍTULO 1

---

---

## Antecedentes teóricos

---

---

En 2016 Colombia ocupa 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. [11]

### 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 [21].

Para incluir la incertidumbre en las proyecciones de población Lee enumera los siguientes métodos [14]:

- El enfoque de escenarios alto, medio y bajo: Asume comportamientos fijos para la fertilidad, la mortalidad y las migraciones durante el periodo de proyección, basado en algunos supuestos [14].
- Análisis estocásticos
  - Análisis ex-post: consiste en evaluar el error de pronóstico en proyecciones anteriores y aplicarlo a las nuevas proyecciones [14].
  - Simulación estocástica: Permite hacer proyecciones de población, al asignar una distribución de probabilidad a las tasas vitales (mortalidad, natalidad, migraciones) [14].
  - 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 [14].

- 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 periodo  $i$ . 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 [21]. 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. [21].

### 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 (CONPES 3828) se proyectó la población carcelaria usando la tasa media de crecimiento anual (1993-2014) [7]. Esta proyección no tiene en cuenta la incertidumbre asociada con las variaciones aleatorias en las tasas, ni las asociadas a cambios estructurales.

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 [13].

El departamento de Justicia de los Estados Unidos realizó estimaciones de la población carcelaria por estado para el periodo 2013-2014. Las estimaciones parten del censo de prisiones 1993-2014. Estas proyecciones se puede enmarcar dentro de las proyecciones de áreas pequeñas [16].

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 periodo 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 [23].

Blummstein 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 ARIMA ni el método de Blummstein, pues las tasas de encarcelamiento y sentencia no se publican. La tesis busca proponer un método de proyección, para situaciones donde no se cuenta con el registro de los vitales o su equivalente en la población analizada.

## 1.2. Series de tiempo

Una serie de tiempo es un conjunto de observaciones  $x_t$  asociadas a un instante de tiempo  $t$ . Es usual referirse como series de tiempo, tanto a las realizaciones  $x_t$  como a las variables aleatorias  $X_t$  que las generan.[3]

En una regresión lineal clásica, una variable  $Y$  es explicada o predicha en función de una variable  $X$ . La diferencia entre el valor observado  $y_i$  y el valor predicho se suponen provenientes de un proceso aleatorio con media cero. [5]

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \quad (1.1)$$

Donde  $\{\epsilon_1, \epsilon_2, \dots\}$  son independientes e idénticamente distribuidas (i.i.d.). Es posible representar una serie de tiempo en esta forma, tomando como variable explicativa el tiempo. Sin embargo, el análisis de series de tiempo se ha desarrollado como un área particular de la estadística, pues es este supuesto no suele cumplirse, enfocándose particularmente en variables aleatorias dependientes, que pueden ser representadas en la forma [3]:

$$Y_i = \beta_0 + \beta_1 Y_1 + \beta_2 Y_2 + \dots + \beta_{i-1} Y_{i-1} + \epsilon_i \quad (1.2)$$

Para medir la dependencia entre las observaciones recurrimos a la función de autocovarianza y a la función de autocorrelación. La función de autocovarianza se define como:

$$\gamma(h, h+k) = \text{cov}(x_h, x_{h+k}) = E[(x_h - \mu_h)(x_{h+k} - \mu_{h+k})] \quad (1.3)$$

para todo  $h$  y  $k$  mayor a cero. La autocovarianza mide la dependencia lineal entre dos puntos en la misma serie en diferentes instantes. [19].

La función de autocorrelación se define como [19]:

$$\rho(h, h+k) = \frac{\gamma(h, h+k)}{\sqrt{\gamma(h, h)\gamma(h+k, h+k)}} \quad (1.4)$$

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  $\{X_1, \dots, X_k\}$  tienen la misma distribución conjunta que  $\{X_{h+1}, \dots, X_{h+k}\}$ , para todos los enteros  $h$  y  $k$ . [3]

La estacionaridad débil se presenta cuando  $E[X_j]$  y  $E[X_j X_{j+h}]$  son independientes de  $j$ , es decir: i) Presenta media  $\mu$  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 ( $k$ ). [19]

### 1.2.1. Modelos ARIMA

Box & Jenkins proponen una aproximación iterativa de cuatro etapas para la selección de un modelo [1]:

- Selección de una clase de modelos, con base en la teoría y la práctica.

- Identificación del modelo, donde se seleccionan un conjunto de parámetros que permitan explicar el sistema con parsimonia.
- Estimación de parámetros.
- Chequeo diagnóstico, para detectar fallas en el ajuste. Si se detectan fallas en el ajuste, se regresa al segundo paso.

Los procesos ARMA son procesos aleatorios de la forma [19] :

$$Y_t = \gamma Y_{t-1} + \gamma_2 Y_{t-2} \dots + \gamma_i Y_{t-i} + \epsilon + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} \dots \theta_j \epsilon_{t-j} \quad (1.5)$$

A este modelo se le conoce como ARMA(i,0,j).

Los procesos ARIMA resultan al considerar una serie de la forma [19]:

$$Y_t = \alpha + Y_{t-1} \quad (1.6)$$

Tal que el proceso  $Y_t - Y_{t-1}$  es un proceso ARMA.

Los criterios para identificar el orden de un ARMA son [2]:

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

### 1.2.2. Modelos Estado-Espacio

Tomado de [15] En un modelo estado espacio una serie de tiempo (multiple) observada  $y_1, \dots, y_t$  depende de un estado  $z_t$ , 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*: [15]

$$y_t = H_t z_t + v_t \quad (1.7)$$

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:

$$z_t = B_{t-1} z_{t-1} + w_{t-1} \quad (1.8)$$

La matriz  $B_t$  es una matriz de coeficientes que puede depender de  $t$  y  $w_t$  es un proceso de ruido. [15]

Es posible considerar los modelos estado-espacio una generalización de los modelos ARIMA y VARMA.



---

# 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 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 2017 (pasa de 1633 personas a 7800).

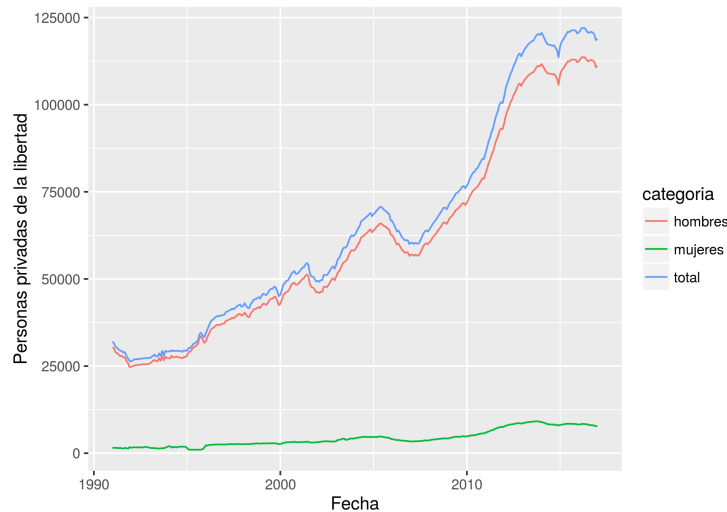


FIGURA 2.1. Población privada de la libertad 1991 - 2017

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.

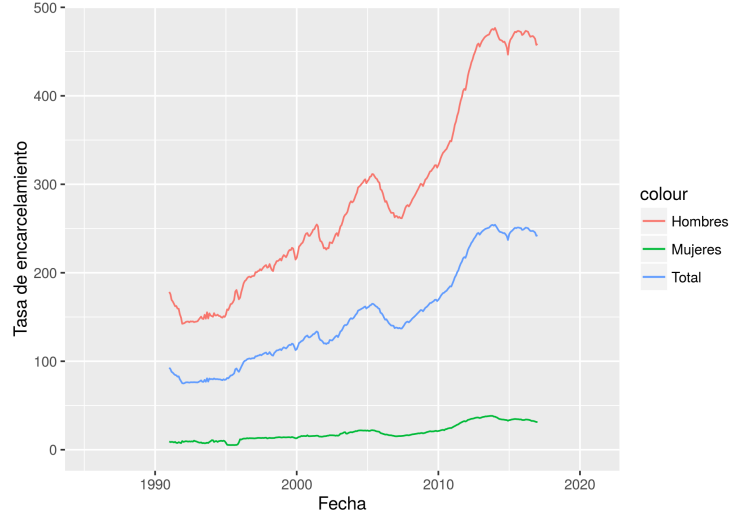


FIGURA 2.2. Tasa de encarcelamiento según género 1991 - 2017

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 (cita pendiente). 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. [7]

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

### 2.2.1. Identificación del modelo

Podemos modelar el sistema penitenciario de la siguiente manera:

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

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

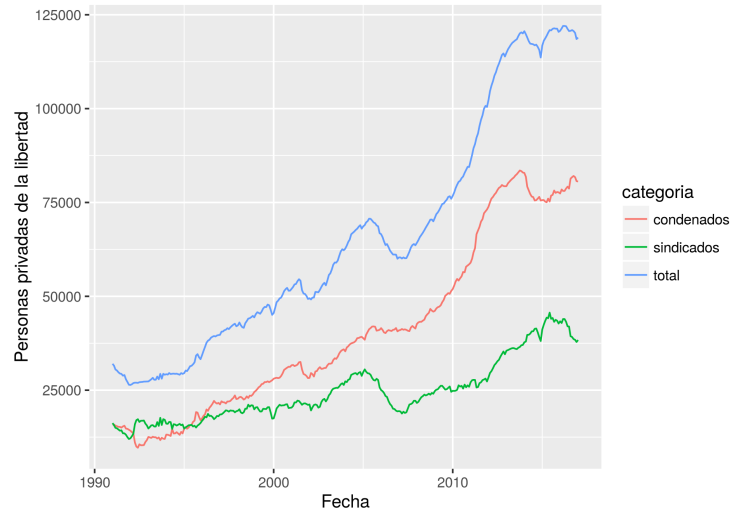


FIGURA 2.3. Población carcelaria por situación jurídica

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

$\alpha$  = proporción de la población libre que ingresa al sistema carcelario

$\gamma$  = proporción de sindicados que es juzgada cada periodo

$\beta$  = proporción de sindicados que han sido encontrados culpables durante el juicio

$\omega$  = proporción de condenados que cumplen su pena cada periodo.

Y esto es un problema interesante porque no tengo las series de tiempo de la transición!

---

### Modelos SARIMA

---

La estimación de los modelos se realiza usando el paquete base del software R [18], y el paquete *astsa* [20], cuyo uso es discutido en detalle por Shumway [19].

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

#### 3.1. Identificación del modelo

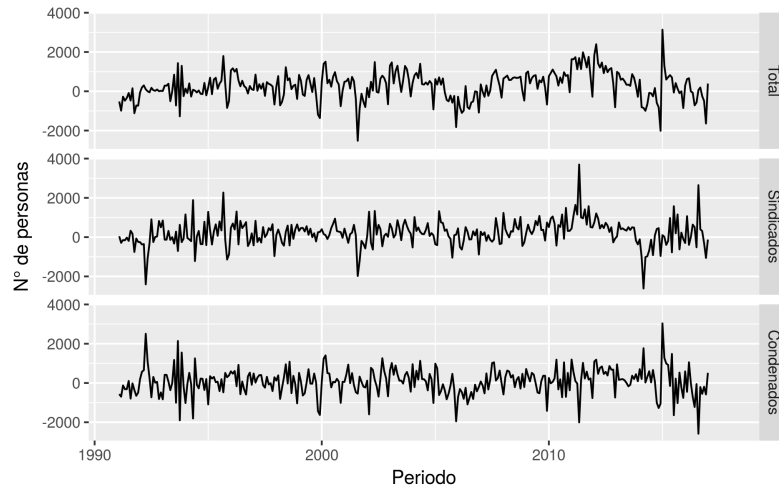


FIGURA 3.1. Variación inter-mensual de población carcelaria, sindicados y condenados

Antes de realizar la proyección de una serie de tiempo, es necesario identificar el modelo que explique adecuadamente su comportamiento.

Aunque conocemos, por la estructura del proceso que genera los datos, que las series de tiempo de población sindicada y condenada no son independientes, podemos simplificar la proyección, tratándolas como independientes. 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 tener una mejor estructura de análisis usamos la función "decompose" de R base.

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.

La mayor parte de la variación de la población total se puede asociar con variaciones en a población sindicada 3.3. La población condenada tiene una tendencia con una tendencia más estable. 3.4

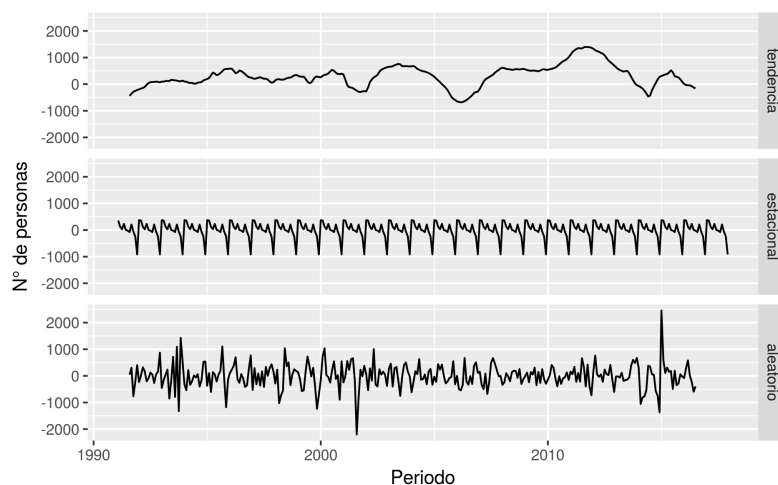


FIGURA 3.2. Variación mensual de población carcelaria descompuesta por tendencia, estacionalidad y componente aleatorio.

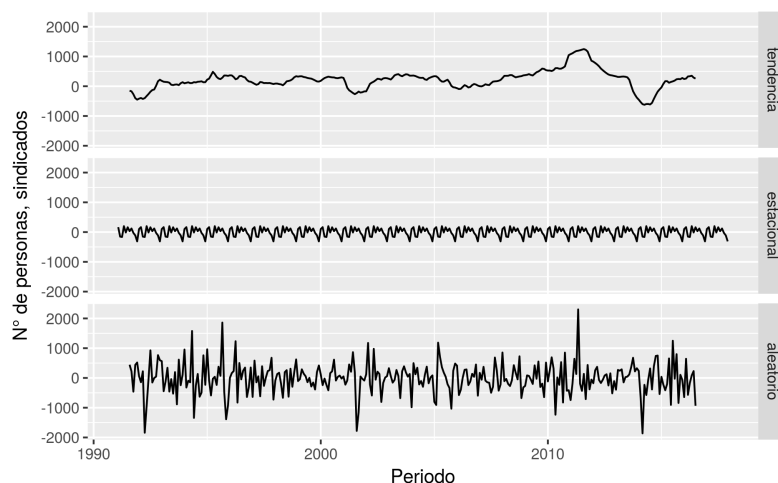


FIGURA 3.3. Variación mensual de población carcelaria sindicada, descompuesta por tendencia, estacionalidad y componente aleatorio.

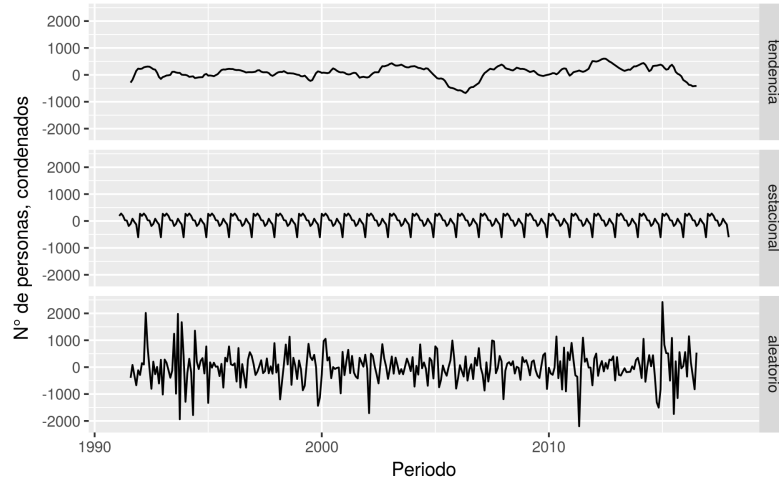


FIGURA 3.4. Variación mensual de población carcelaria condenada, descompuesta por tendencia, estacionalidad y componente aleatorio.

Se trabaja sobre la diferencia de cada serie de tiempo con lag 1 y lag 12, al observar que la serie tiene un crecimiento sostenido entre 1991 y 2017, sugiriendo un proceso estocástico integrado. Sobre cada serie (población total, población sindicada y población condenada) se realiza la función de autocorrelación y autocorrelación parcial y se presenta en la figura 3.5. Estas funciones son usadas como herramienta de diagnóstico, para intuir 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.

TABLA 3.1. Comportamiento de la ACF y PACF en modelos ARMA(p,q) [19]

|             | <b>AR(p)</b>         | <b>MA(q)</b>         | <b>ARMA(p,q)</b> |
|-------------|----------------------|----------------------|------------------|
| <b>ACF</b>  | Tails off            | Cuts off after lag q | Tails off        |
| <b>PACF</b> | Cuts off after lag p | Tails off            | Tails off        |

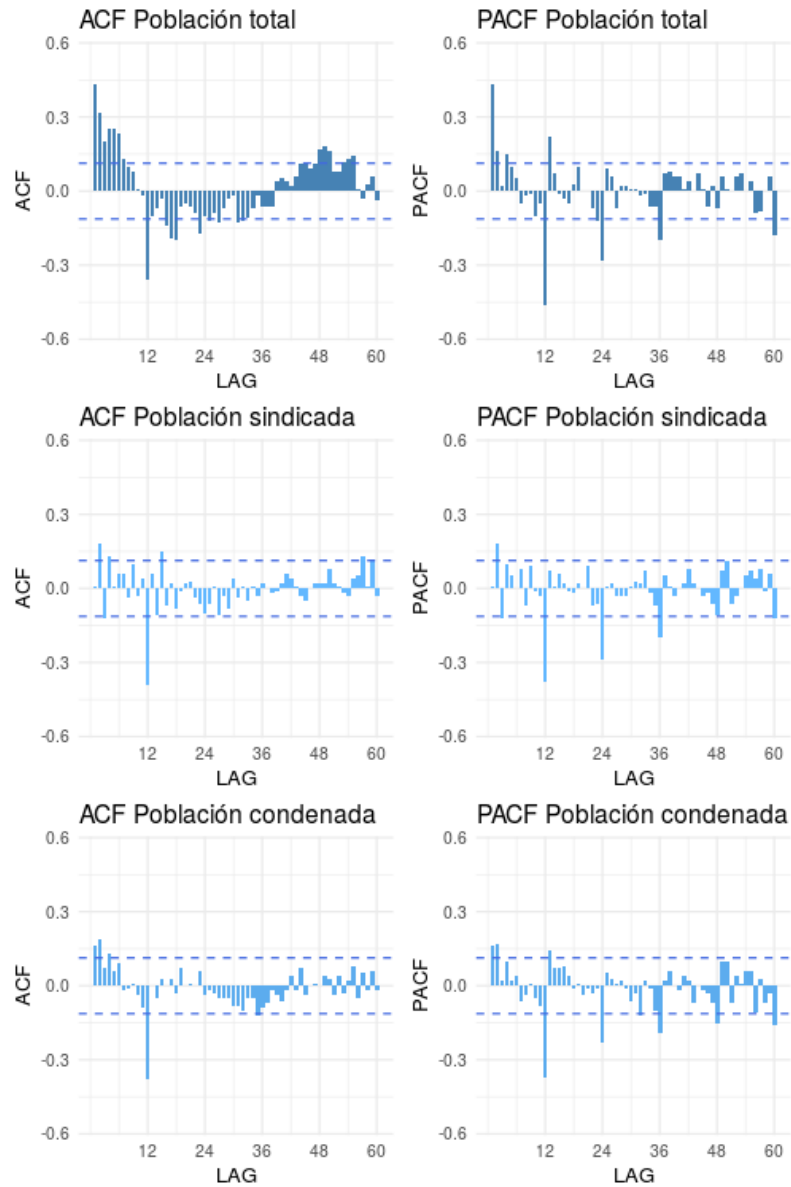


FIGURA 3.5. Autocorrelación parcial de la variación inter-mensual de la población

### 3.1.1. Población total

La tabla 3.1 presenta una guía para la interpretación de la ACF y la PACF. En la 3.5 podemos observar que tanto la ACF como la PACF decaen lentamente. La APCF decae progresivamente en los meses 12, 24, 36, mientras que la ACF se corta en el lag 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)

### 3.1.2. Población sindicada

En la 3.5 podemos observar que la ACF decae lentamente, mientras la PACF cae por debajo del error luego del lag 1. La APCF decae progresivamente en los meses 12, 24, 36, mientras que la ACF se corta en el lag 12, 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 sindicada como SARIMA(1,1,1,0,0,1)

### 3.1.3. Población condenada

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

## 3.2. Estimación de parámetros

El proceso de estimación se realizará para cada serie por separado, luego se validará la efectividad de incluir más parámetros, usando como criterio de comparación el BIC. A modo informativo se usa la función `auto.arima` del paquete `forecast`, para seleccionar el modelo con menor AIC. [9] [10].

### 3.2.1. Población total

De acuerdo con el diagnóstico se evaluaron los siguientes modelos:

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. Ver 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 ( $pvalue > 0.05$ ). Ver 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.

Para el modelo 1 y el modelo 2 la estadística de 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) | 4885.2  | 4907.66 |
| 2 | (1,1,1)(0,0,1) | 4897.79 | 4916.51 |
| 3 | (3,2,1)(0,0,2) | 4874.46 | 4900.64 |
| 4 | (1,2,1)(0,0,2) | 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    |

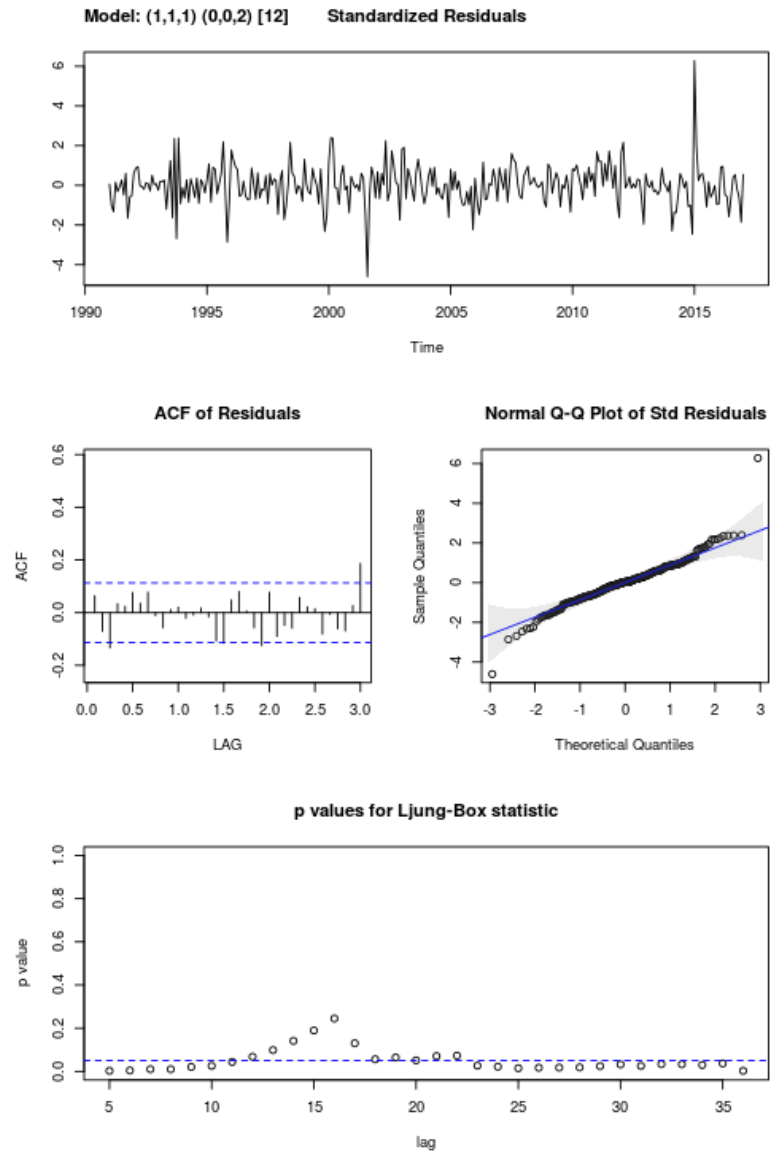


FIGURA 3.6. Diagnóstico del Modelo 1. sarima (1,1,1)(0,0,2), para la población carcelaria total. Generado con astsa [20]

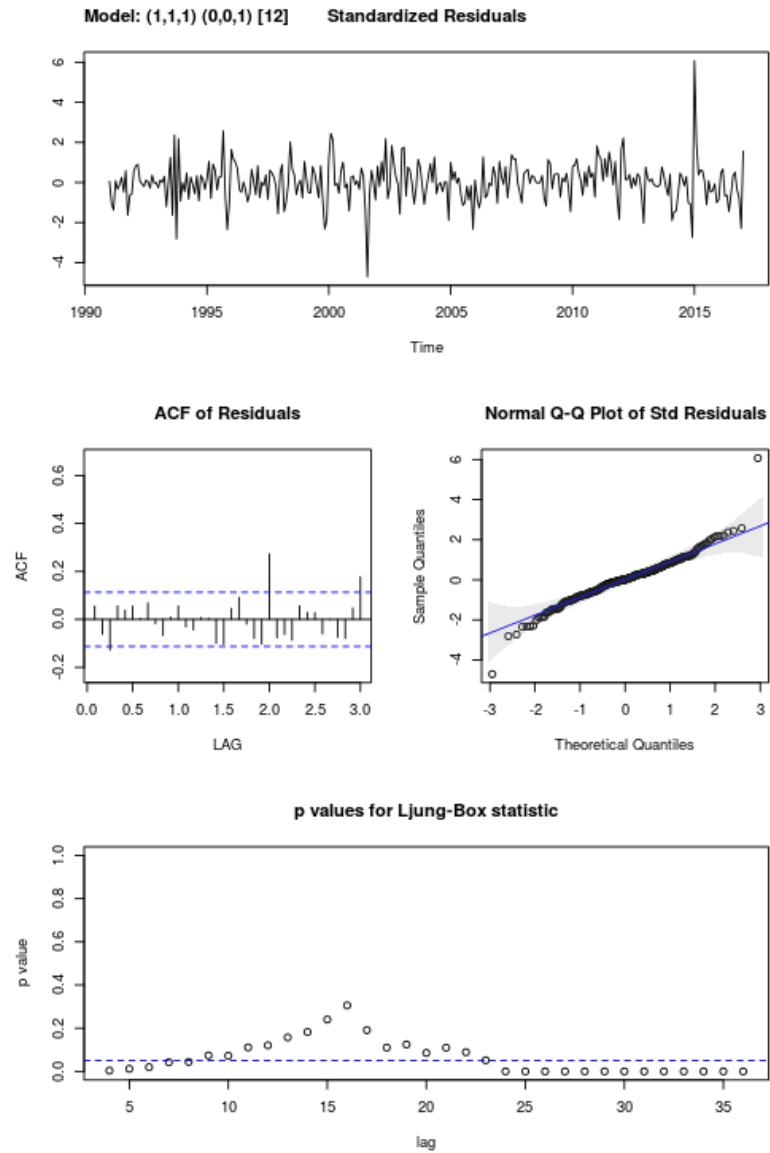


FIGURA 3.7. Diagnóstico del Modelo 2. sarima (1,1,1)(0,0,1), para la población carcelaria total. Generado con astsa [20]

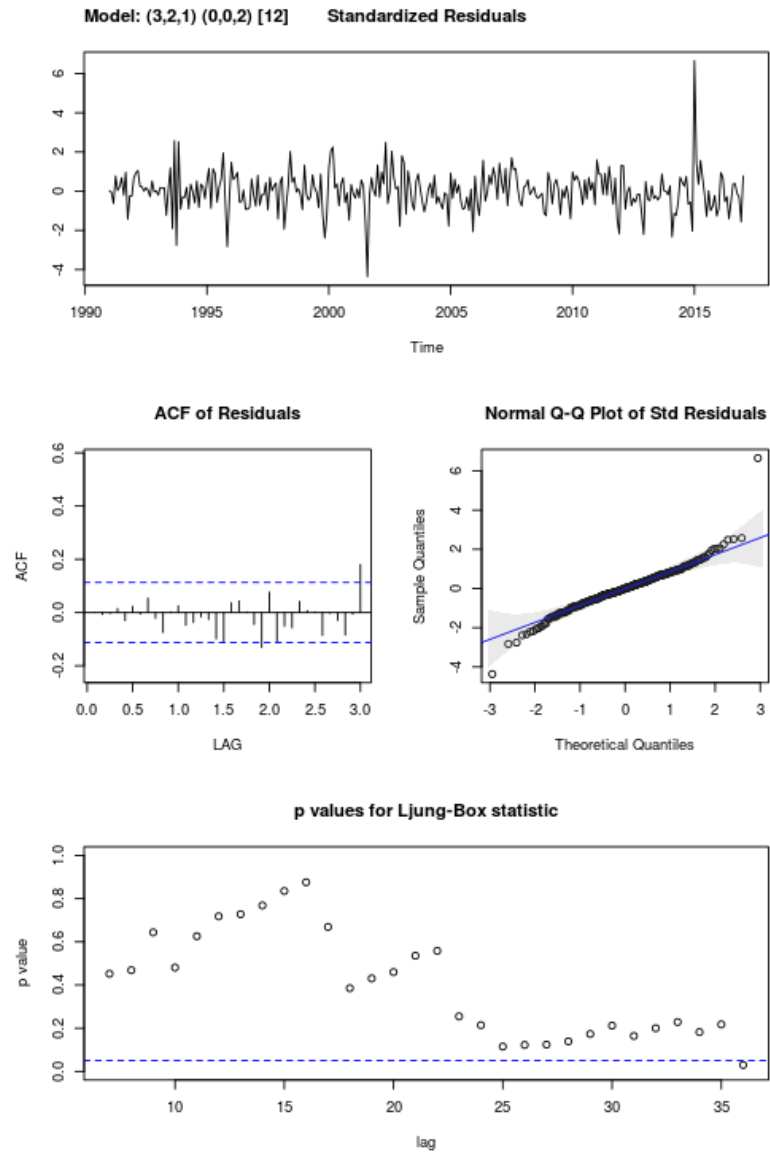


FIGURA 3.8. Diagnóstico del Modelo 3. sarima (3,2,1)(0,0,2), para la población carcelaria total. Generado con astsa [20]

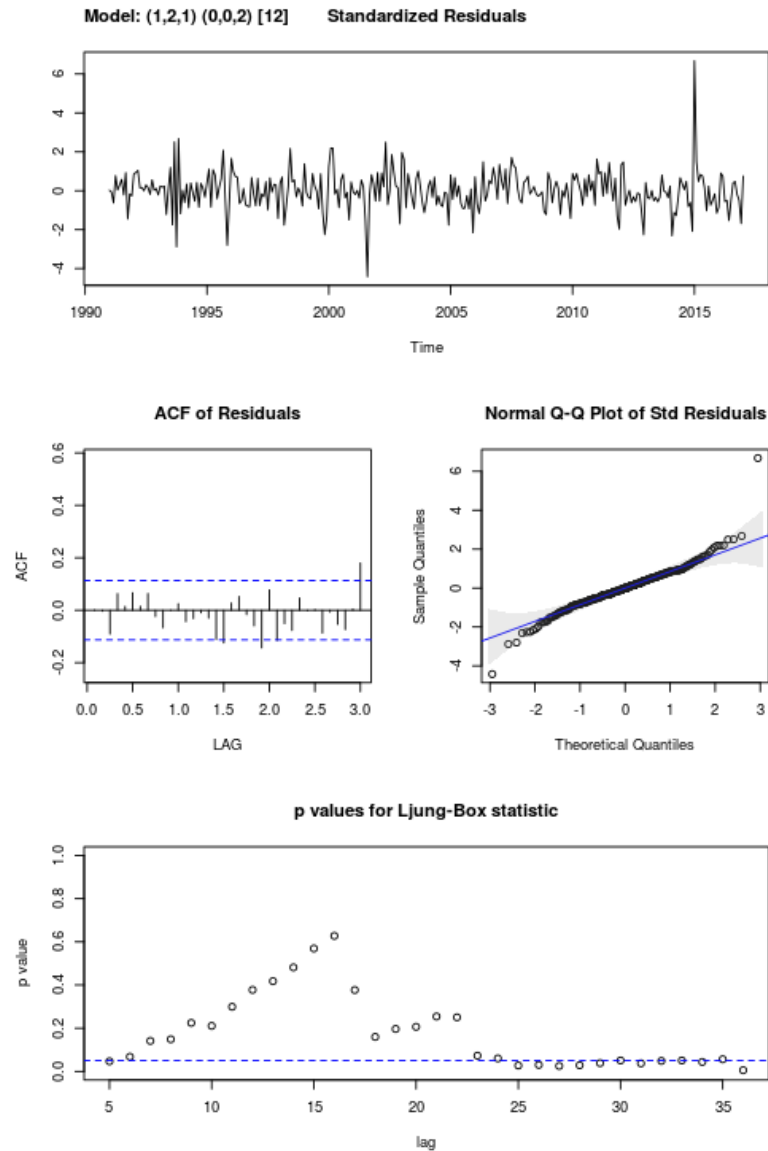


FIGURA 3.9. Diagnóstico del Modelo 4. sarima (1,2,1)(0,0,2), para la población carcelaria total. Generado con astsa [20]

### 3.2.2. Población sindicada

### 3.2.3. Población condenada

## 3.3. Proyección

### 3.3.1. Población total

Proyección de la población total se presenta en la gráfica 3.9

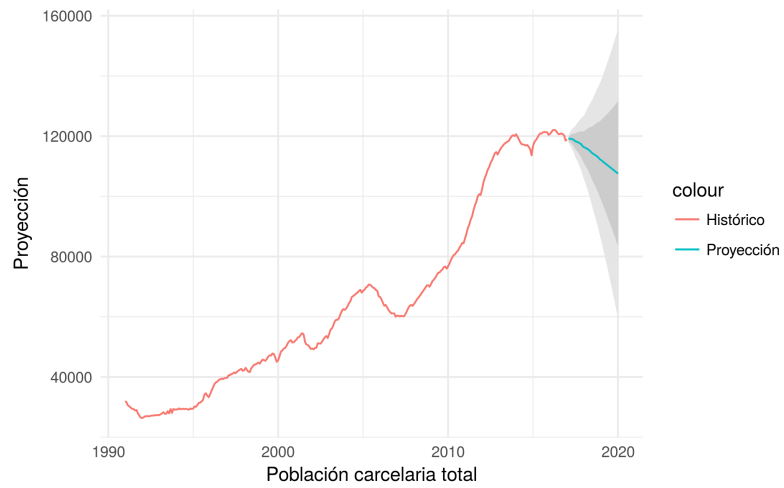


FIGURA 3.10. Proyección de la población carcelaria total, sarima (1,2,1,0,0,2)

### 3.3.2. Población sindicada

### 3.3.3. Población condenada

## 3.4. Conclusiones

Al usar un ARIMA

---

---

## Modelos estado-espacio

---

---

### 4.1. Marco teórico

### 4.2. Selección de Software

Dado que el alcance del trabajo es aplicado, se recurre a software prediseñado para estimar los parámetros de los modelos planteados. En el caso de ARIMA hay varios paquetes que soportan la estimación de los modelos, el análisis de los resultados y la elaboración de los pronósticos. A propósito del software usado al trabajar con modelos estado-espacio [4] presenta una revisión de los principales programas y paquetes, dentro de los que incluye:

- Eview
- gretl
- MATLAB
- R base
- R + KFAS
- RATS
- REGCMPNT
- SAS
- S-PLUS
- SsfPack
- SsfPack
- STAMP
- Stata

Para mantener un único ambiente de desarrollo, nos enfocamos en los paquetes desarrollados en R, como los mencionados por Petris [17], Helske [8] y Tusell [22]

---

### 4.3. Identificación del modelo

### 4.4. Simulación Monte-Carlo

### 4.5. Estimación de parámetros

### 4.6. Proyecciones 2017 - 2020

### 4.7. Conclusiones



---

---

## Conclusiones

---

---

- La escritura de tesis utilizando L<sup>A</sup>T<sub>E</sub>X permite que se obtengan documentos de una presentación elegante, agradable, de una impresión incomparable, de escritura bastante simple en cuanto al texto técnico y de formulas matemáticas, junto con un manejo automático del formato de las partes de un documento y las referencias bibliográficas, desprendiéndose así de los detalles de edición que en otras herramientas, producen tantas frustraciones y dolores de cabeza.

---

---

## Trabajo futuro

---

---

- Implementar y corregir todos aquellos errores que los usuarios de esta plantilla puedan encontrar, así como las sugerencias para la modificación de la plantilla que sean pertinentes.

---

---

## Bibliografía

---

---

- [1] George E P Box, Gm Jenkins, and Gc Reinsel, *Time series analysis: forecasting and control*, 2013.
- [2] Paolo Brandimarte, *Time Series Models*, Quantitative Methods: An Introduction for Business Management, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 527–579.
- [3] Peter J. Brockwell, *Time Series*, International Encyclopedia of Statistical Science, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 1601–1605.
- [4] Jacques J F Commandeur, Siem Jan Koopman, and Marius Ooms, *Statistical Software for State Space Methods*, JSS Journal of Statistical Software **41** (2011), no. 1.
- [5] Jacques J.F Commandeur and Siem Jan Koopman, *An Introduction to State Space Time Series Analysis*, (2007), 189.
- [6] Departamento Administrativo Nacional de Estadística (DANE), *Comunicado de prensa*, 2009.
- [7] Departamento Nacional de Planeación, *Conpes 3828 POLÍTICA PENITENCIARIA Y CARCELARIA EN COLOMBIA*, (2015).
- [8] Jouni Helske, *Exponential Family State Space Models in R*, Journal of Statistical Software **78** (2017), no. 10, 1–39.
- [9] Rob J Hyndman, *{forecast}: Forecasting functions for time series and linear models*, 2017.
- [10] Rob J. Hyndman and Yeasmin Khandakar, *Automatic Time Series Forecasting: The forecast Package for R*, Journal of Statistical Software **27** (2008), no. 3, 1–22.
- [11] Institute for Criminal Policy Research, *Colombia / World Prison Brief*, 2016.
- [12] Instituto Nacional Penitenciario y Carcelario, *Plan de direccionamiento estratégico 2015-108, Misi{ó}n y Visi{ó}n*, 2016.
- [13] Ministry Justice, *Prison Population Projections 2014 - 2020, England and Wales*, Home Office Statistical Bulletin (2014), no. November, 31.
- [14] 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.

- 
- [15] Helmut Lütkepohl, *New Introduction to Multiple Time Series Analysis*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.
  - [16] Todd D Minton, Bjs Statistician, Scott Ginder, Susan M Brumbaugh, Hope Smiley-Mcdonald, and Harley Rohloff, *Census of Jails: Population Changes, 1999 - 2013*, (2015).
  - [17] Giovanni Petris and Sonia Petrone, *State Space Models in R*, Journal of Statistical Software **41** (2011), no. 4, 128–129.
  - [18] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2017.
  - [19] Robert H Shumway and David S Stoffer, *Time Series Analysis and Its Applications*, Springer Texts in Statistics, vol. 45, Springer New York, New York, NY, 2011.
  - [20] David Stoffer, *astsa: Applied Statistical Time Series Analysis. R package version 1.3*, 2014.
  - [21] David A Swanson and Jeff Tayman, *Subnational Population Estimates*, The Springer Series on Demographic Methods and Population Analysis, vol. 31, Springer Netherlands, Dordrecht, 2012.
  - [22] Fernando Tusell, *Kalman Filtering in R*, Journal of Statistical Software **39** (2011), no. 2, 1–27.
  - [23] Wai-Yin Wan<sup>1</sup>, Steve Moffatt<sup>2</sup>, Zachary Xie<sup>3</sup>, Simon Corben, and Don Weatherburn, *Forecastin prison populations using sentencing and arrest data*, Tech. report.