



UN ACERCAMIENTO A LOS MODELOS ARIMAX Y LINEAL DINÁMICO  
PARA ESTIMAR NATALIDAD EN COLOMBIA EN EL PERIODO ENTRE ENERO  
2011 A DICIEMBRE 2017

JOAQUIN SANTIAGO ROMERO MORENO

FUNDACIÓN UNIVERSITARIA LOS LIBERTADORES  
DEPARTAMENTO DE CIENCIAS BÁSICAS  
ESPECIALISTA EN ESTADÍSTICA APLICADA  
BOGOTÁ, D.C., 2019





UN ACERCAMIENTO A LOS MODELOS ARIMAX Y LINEAL DINÁMICO  
PARA ESTIMAR NATALIDAD EN COLOMBIA EN EL PERIODO ENTRE ENERO  
2011 A DICIEMBRE 2017

JOAQUIN SANTIAGO ROMERO MORENO

Trabajo de grado para optar al título de Especialistas en Estadística  
Aplicada

Heivar Yesid Rodriguez Pinzon

Director de trabajo de grado

FUNDACIÓN UNIVERSITARIA LOS LIBERTADORES

DEPARTAMENTO DE CIENCIAS BÁSICAS

ESPECIALISTA EN ESTADÍSTICA APLICADA

BOGOTÁ, D.C., 2019

Nota de Aceptación

---

---

---

---

---

---

Presidente del Jurado

---

Jurado

---

Jurado

Bogotá, D.C., Junio de 2019

Las directivas de la Fundación Universitaria los Libertadores, los jurados calificadores y el cuerpo docente no son responsables por los criterios e ideas expuestas en el presente documento, estos corresponden únicamente a las al autor.

## ÍNDICE

1. INTRODUCCIÓN.....	10
2. PLANTEAMIENTO DEL PROBLEMA.....	12
3. JUSTIFICACIÓN DEL PROBLEMA .....	14
3.1. Proyecciones DANE.....	15
3.2. Estimación Fecundidad en Colombia .....	16
4. OBJETIVOS.....	18
4.1. Objetivo general .....	18
4.2. Objetivos específicos .....	18
5. MARCO DE REFERENCIA .....	19
5.1. Transición demográfica.....	19
5.2. Factores que influyen en la natalidad.....	20
5.3. Fecundidad .....	20
5.4. Modelos ARIMAX.....	21
5.4.1. ARIMA .....	22
5.4.2. ARIMA Estacional.....	22
5.4.3. ARIMAX.....	23
5.5. Modelos Dinámicos Lineales.....	24
5.5.1. Filtro de Kalman.....	27
6. METODOLOGÍA .....	28
6.1. Datos.....	28
6.2. Tratamiento de datos .....	28
6.3. Variables: .....	29

6.3.1. Nacimientos .....	29
6.3.2. Edad madre .....	30
6.3.3. Edad padre .....	31
6.3.4. Número de hijos.....	32
6.3.5. Defunciones fetales .....	33
7. ANÁLISIS DE RESULTADOS .....	34
7.1. Modelo ARIMAX Estimación .....	35
7.1.1. Identificación.....	35
7.1.2. Estimación .....	37
7.1.3. Diagnóstico .....	38
7.1.4. Estimación Modelo .....	39
7.2. Pronóstico Modelo ARIMAX.....	40
7.3. Modelo Lineal Dinámico .....	43
8. CONCLUSIONES .....	45
9. REFERENCIAS .....	46
10. ANEXOS.....	47

## Índice de gráficas

Gráfica 1. Serie de tiempo nacimientos en Colombia .....	30
Grafica 2 Serie de tiempo edad de la madre .....	31
Grafica 3 Serie de tiempo edad promedio del padre .....	32
Grafica 4 Serie de tiempo número de hijos promedio .....	33
Grafica 5 Serie de tiempo defunciones fetales .....	34
Grafica 6 . ACF y PACF de la serie natalidad estacionalizada .....	36
Grafica 7 Residuales.....	39
Grafica 8 Estimación modelo ARIMAX .....	40
Grafica 9 Datos reales 2017 .....	41
Grafica 10 Predicción 6 meses año 2018 .....	42
Grafica 11 Estimación modelo lineal dinámico .....	44

## Índice de tablas

Tabla 1. Tratamiento de datos .....	29
Tabla 2. Datos base para análisis.....	41
Tabla 3. Proyecciones .....	42
Tabla 4. Estimaciones modelo ARIMAX .....	47
Tabla 5. Estimacion Modelo Lineal Dinamico .....	48



## RESUMEN

En este documento se presenta un acercamiento a los modelos ARIMAX y lineal dinámico para estimar natalidad en Colombia en el periodo entre enero 2011 a diciembre 2017, se parte de la necesidad que tiene el estado de contar con información poblacional veraz y confiable. Para el análisis exploratorio de los datos se realizan series de tiempo de los nacimientos en Colombia usando las variables edad de la madre y del padre, número de hijos y defunciones fetales; posteriormente, se analiza la serie nacimientos con las variables de apoyo, usando inicialmente todas las establecidas (edad madre, edad padre, hijos y defunciones fetales), quedando dos como variables exógenas significativas en el modelo ARIMAX. El modelo de predicción realizado contó con una partición de 60 datos (meses, enero 2011 a diciembre 2016) de entrenamiento y 12 datos (meses, enero 2017 a diciembre 2018) de prueba para contrastar proyecciones del modelo. Una vez realizadas las pruebas de estimación se plantea una predicción a 6 meses a través del modelo ARIMAX. Los resultados en las estimaciones de los modelos ARIMAX y lineal dinámico son buenos pues explican el comportamiento de los datos, siendo mejor el lineal dinámico solo usando tendencias polinomiales, las estimaciones y predicciones aparecen en tablas como anexos.

***Palabras clave: Modelo ARIMAX, Modelo Lineal Dinámico, Natalidad***

## 1. INTRODUCCIÓN

El estudio relacionado con el acercamiento a los modelos ARIMAX y lineal dinámico para estimar natalidad en Colombia en el periodo entre enero 2011 a diciembre 2017 que se presenta en este escrito, cuenta con los apartados de planteamiento del estudio, marco de referencia, metodología, análisis de resultados y conclusiones.

En el planteamiento del estudio se delimita la importancia de contar con modelos de predicción de natalidad como estrategia para que el estado colombiano logre determinar estrategias que faciliten la regulación de la población en los diferentes departamentos que integran en el territorio nacional, descentrando la mirada de aspectos económicos, los cuales han sido el principal factor para el planteamiento de las estrategias.

El referente del estudio estuvo sustentado en la comprensión de las proyecciones demográficas realizadas por el Departamento Administrativo Nacional de Estadística (DANE), la estimación de la fecundidad en Colombia, la transición demográfica, los tres factores que influyen en la natalidad: socioeconómicos, políticos y demográficos, la fecundidad, los modelos ARIMAX y los modelos dinámicos lineales.

En el tercer apartado del documento, metodología, se presenta información relacionada con los datos utilizados para el estudio, enunciado la procedencia de estos y el tratamiento realizado para su análisis a través de los modelos ARIMAX y dinámico lineal; así mismo, se realiza una descripción a través de series de tiempo de las variables nacimientos en Colombia, edad de la madre, edad del padre, número de hijos y defunciones fetales. El análisis de resultados se centró en los modelos ARIMAX, aplicado a través de la metodología de Box – Jenkins, con los criterios de estacionaridad de Dickey -Fuller y Philips-Perron, dados sus resultados

se escoge una diferencia de estacionalización a 12 meses; así mismo, la estimación se realiza con las variables de apoyo o covariables de tiempo, edad de la madre y el padre, número de hijos y defunciones fetales. El modelo ARIMAX permite una predicción de la natalidad a seis meses, repartiendo los datos en entrenamiento y prueba para revisar confiabilidad. En el modelo lineal dinámico se establece una estimación de los datos a partir de la tendencia polinomial del los datos, evidenciando una mejor aproximación a los datos de natalidad.

## **2. PLANTEAMIENTO DEL PROBLEMA**

La información demográfica de un país resulta de gran importancia para la toma de decisiones por parte del estado, es así como se pueden plantear análisis relacionados con el crecimiento de la población, donde intervienen variables como la natalidad, variabilidad en un periodo de tiempo determinado. Entendiendo la natalidad como la cantidad de nacidos en un periodo específico.

Los países desarrollados, en su mayoría, generan estrategias orientadas al control poblacional, como resultado de los análisis realizados en torno a las variables poblacionales, ejemplo de esto tipos de control son los realizados por países como China, Japón, Canadá entre otros en su mayoría estados desarrollados, que tienen control en estos factores.

Lo señalado implica la necesidad de tener estimaciones y proyecciones sobre la población que permitan generar planes a futuro en políticas públicas o distribución de recursos, no solamente con información recompilada en los censos, los cuales no está garantizado se realicen de forma periódica.

En Colombia, como es común en casi todos los países en desarrollo, se tiene un control por parte del estado de la tasa de natalidad, buscando la disminución de nacidos, la estrategia procede principalmente de la economía, pues, se entiende que la natalidad es un factor directamente relacionado con el crecimiento de la población, lo cual implica mayor estabilidad y desarrollo económico del país, facilitando la disminución de efectos sociales y económicos derivados de los altos niveles poblacionales, como es la pobreza.

Es así como las políticas públicas en el país han tenido como finalidad la reducción de la tasa de fecundidad en pro del desarrollo social y económico, especialmente la de Salud Sexual y Reproductiva (SSR). Burgos (2015) plantea que la implementación de esta política se ha concretado varios servicios y programas, entre ellos el de Servicios de Salud Amigables para Adolescentes y Jóvenes, realizada por el Ministerio de Salud y Protección Social entre los años 2007 y 2012, con el propósito de prevenir del embarazo adolescente; el programa

de educación para la sexualidad y construcción de ciudadanía, implementado desde 2008 por el Ministerio de Educación para niños y adolescentes, y la política Nacional de Salud Sexual y Reproductiva y del Plan Nacional de Salud Pública, dirigida a toda la población, con un período de vigencia entre 2012 y 2022, a cargo de la Alta Consejería para la Equidad de la Mujer. A través de esta política se incorpora el control prenatal, acciones de prevención, salud sexual para adolescentes y jóvenes y planificación familiar para hombres y mujeres, además incluye medicamentos anticonceptivos hormonales y el condón masculino en el Plan Obligatorio de Salud, como prestaciones de promoción y prevención.

En este sentido, la aplicación de técnicas estadísticas predictivas como los modelos ARIMAX y lineal dinámico, permiten contar con estimaciones y proyecciones de la variabilidad poblacional, que contribuyen en la comprensión de los cambios de la natalidad y mortandad en el país.

### **3. JUSTIFICACIÓN DEL PROBLEMA**

Para una nación es importante la obtención y análisis de datos confiables respecto a su población, dado que de estos dependen políticas fiscales, cobertura en salud y educación, planeación y ejecución presupuestal y evaluación del desarrollo del país. De manera más específica, los departamentos, ciudades y municipios deben contar con datos fiables que permitan la planeación de estrategias para garantizar el transporte público, las proyecciones de construcción de vivienda o zonas comerciales, prestación de servicios de salud, educación, etc.

En la dinámica poblacional de Colombia para el siglo XX (periodo desde el cual existen datos confiables de población) se estima el aumento de la población en dos comportamientos diferentes, en la primera mitad del siglo se evidencia crecimiento de la población de 2,9 veces, mientras posterior a 1950 muestra un aumento de 3,6. Es evidente la tendencia al crecimiento de la población a nivel mundial, aunque la curva reduce su pendiente en el presente siglo, esto por diferentes tipos de factores.

Toda nación cuenta con datos sobre natalidad y de proyecciones a mediano plazo, es importante reconocer los posibles factores que pueden incidir en cambios y tendencias de la misma, y que tipos de modelos pueden ser más fiables en la predicción.

Luego se quiere corroborar la consistencia y fiabilidad de modelos alternos a los ya establecidos que puedan generar estimaciones y proyecciones confiables, además de comparables con los planteados en estudios del DANE. El presente proyecto busca establecer modelos para series de tiempo, que representen una opción a los ya establecidos por la teoría demográfica para Latinoamérica y aplicados en el país.

### 3.1. Proyecciones DANE

Un estado depende de información demográfica para la toma de decisiones, para cuantificar una eficiente distribución de recursos. Esta información procede en el caso de Colombia de los CENSOS, los cuales se realizan cada 10 o más años, luego es necesaria la realización de proyecciones a futuro sobre la población, sexo, rangos de edad, vivienda las cuales son realizadas por el DANE. (DANE, 2009)

En su última publicación de proyecciones de población, del año 2009 usa los censos de los años 1985, 1993, 2005. Con esta información realiza proyecciones desde el año 2005 hasta el año 2020. El método realizado para las proyecciones fue el método de los componentes, esto siguiendo recomendaciones del centro latinoamericano de demografía CELADE (organismo de la naciones unidas).

El método de los componentes se basa en el trabajo por módulos: población base, hipótesis de mortalidad, hipótesis de fecundidad, hipótesis de migración. En cada uno de los módulos se usan tanto planteamientos matemáticos como técnicas estadísticas, donde además se proyecta respecto a los espacios de tiempo intercensales, según la ecuación

$$N^{t+a} = N^t + B^{t,t+a} - D^{t,t+a} + M^{t,t+a}$$

Donde

$a$  = Tiempo entre dos censos.

$N^t$  = Población en el instante  $t$ .

$B^{t,t+a}$  = Nacimientos ocurridos entre el periodo intercensal.

$D^{t,t+a}$  = Defunciones ocurridas durante el periodo intercensal.

$M^{t,t+a}$  = Migrantes netos (inmigrante menos emigrantes), llegados al país durante el periodo intercensal.

Cabe señalar que ninguna de estas variables es conocida, luego se utilizan valores observados para el desarrollo de los cambios demográficos y los censos.

Es importante el cumplimiento del objetivo en estas proyecciones para el gobierno pues de esto depende una distribución hacia toda la población de los planes y programas ejecutados. Luego se valora tanto el cálculo de la estimación como de la incertidumbre que existe en la misma por dinámicas demográficas particulares.

### **3.2. Estimación Fecundidad en Colombia**

El índice sintético de fecundidad hace referencia al número de hijos que en promedio tendría una mujer hipotética al final de su vida fecunda, índice el cual señalan desde la teoría de transición demográfica no puede ser a menor a 2,1 para que las futuras generaciones puedan sustituir a las presentes. La fecundidad representa una de las bases demográficas para explicar el crecimiento poblacional, es necesario contar con una información confiable para los diferentes planes y programas de desarrollo en una nación.

Como fuente de información el DANE (2009) utilizó para las proyecciones 2005-2020 principalmente el censos entre los años 1985 y 2005, como apoyo se usaron las encuestas nacionales de demografía pues contienen información acerca de los determinantes de la fecundidad.

Según documento de proyecciones de población del DANE, como factores de medición para la fecundidad se usan principalmente las tasas de fecundidad específicas por edad y la tasa global de fecundidad. Adicional estas también se usan como apoyo las tasas brutas de natalidad, de reproducción y edad media de fecundidad que se encuentran dentro de las estadísticas vitales publicadas por el mismo organismo.

DANE (2009) señala que "de acuerdo con estudios realizados por el DANE, el comportamiento de la fecundidad en Colombia a partir de la década de los 60 ha estado caracterizada por un importante y continuo proceso de reducción", de igual manera, algunos autores como Florez (2000) califican de "espectacular" la



reducción en la fecundidad para un país del tamaño y características como el de Colombia. Esta situación ha conllevado a cambios en la estructura, adelantando significativamente la edad media de la fecundidad, pasando de 28.02% en 1985 a 26.54% en el 2005, en la actualidad el 50.8% del total de la fecundidad nacional la aportan las mujeres entre los 20 y los 29 años de edad y el 70% de la fecundidad total es aportada por mujeres menores de 30 años (Flórez, 2000; DANE, 2009).

El DANE utiliza la función logística para proyectar la tasa global de fecundidad, esto mediante la información establecida por quinquenios desde 1985 hasta 2005 y mediante la función logística descendente para proyectar el periodo 2005-2020.

Entre los resultados mostrados en el documento cabe señalar el siguiente: "El análisis del último periodo intercensal, 1993 – 2005, ratifica que la fecundidad adolescente se viene incrementando de forma importante. Efectivamente, si se tiene en cuenta la estructura de la fecundidad para los años 1993 y 2005, se tiene que el descenso de la fecundidad en Colombia de los últimos doce años está explicado por la reducción participación de las mujeres mayores de 20 años, mientras que la contribución del grupo de 15 a 19 años se ha incrementado en un 19,7%" (p, 10).

Las proyecciones muestran condiciones parecidas a nivel mundial, que evidencia una preocupación sobre los controles que existen de natalidad mediante políticas de gobiernos y dinámicas poblacionales, pues han llegado a datos con tasas de fecundidad inferiores al 2,1 en ciertas regiones del mundo. Según experiencia de varios países una tasa inferior a este valor implica el no reemplazo de la población. Esto ha llevado a diversos tipos de debates teóricos sobre las políticas públicas sobre la fecundidad, y sus implicaciones a plazos específicos.

## **4. OBJETIVOS**

### **4.1. Objetivo general**

Estimar la natalidad bruta en Colombia con el apoyo de variables alternas que puedan explicar su comportamiento, usando métodos estadísticos de modelos ARIMA con covariables (ARIMAX) y modelos lineales dinámicos.

### **4.2. Objetivos específicos**

Estimar un modelo ARIMAX la Natalidad en Colombia, con apoyo de variables alternas.

Estimar un modelo lineal dinámico de la natalidad en Colombia.

Presentar pronóstico de natalidad seis meses adelante usando el modelo ARIMAX.

## **5. MARCO DE REFERENCIA**

En América Latina, los índices de población han disminuido su progresión en el último tiempo, principalmente por políticas de control de natalidad y migración comunes en toda la región, pero también hay factores específicos en cada país que pueden explicar esta reducción de forma particular.

### **5.1. Transición demográfica**

Para las dinámicas poblacionales se plantean estudios que plantean modelos que intentan explicar su comportamiento, esto a partir de etapas transitivas que dependen de cambios históricos. Adicional permite relacionar estos cambios con otras variables sociales.

Aquí se establece tres etapas, esto teniendo por referencia el caso europeo (Mesa y Junca, 2011):

- I. Etapa 1: Alta tasa de natalidad y alta tasa de mortalidad, por lo que hay un bajo crecimiento poblacional. Se presenta en regiones donde la mortalidad es alta, especialmente en niños, luego en los núcleos de familias tienden a reemplazar esa mortandad es decir se tiene una natalidad alta.
- II. Etapa 2: Se mantiene alta natalidad, la mortalidad comienza a descender. El crecimiento poblacional es alto y esta etapa se conoce como de explosión demográfica. Se explica por la mejora en cobertura en salud y alimentación, lo que tiende a mejorar la esperanza de vida y bajar la mortandad infantil; pero las familias no se acoplan a esta tendencia luego la natalidad se mantiene de la etapa 1.
- III. Etapa 3: La tasa de natalidad desciende, la tasa de mortalidad baja. El crecimiento poblacional desciende. Al aumentar el número de años de vida que puede tener una persona por el aumento de la esperanza de vida, el núcleo familiar ajusta su tamaño con menos nacimientos. en esta etapa se incluyen políticas de control demográfico.

## **5.2. Factores que influyen en la natalidad**

Usualmente se plantean cuatro factores en demografía que influyen en los cambios de natalidad en una población en el tiempo, estos son los culturales, los socioeconómicos, demográficos y políticos.

Culturales: según región, incluir razones de creencias religiosas, la importancia de las relaciones de pareja frente a las de familia, tendencia al individualismo.

Socioeconómicos: en el contexto social y económico pueden influir el aumento de zonas urbanas, costo de vivienda, aumento de la oferta laboral para la mujer, afectaciones económicas.

Demográficos: características de la población tales como la mortalidad infantil, el retraso de la edad para tener pareja estable, separación de parejas.

Políticos: Referente a los deberes del estado sobre la población tenemos apoyo del estado (educación, permiso maternal, disponibilidad de guarderías), legislación.

## **5.3. Fecundidad**

Hay diferentes áreas de conocimiento que intentan explicar y usar el índice de fecundidad, por ejemplo, desde una perspectiva económica en algunas teorías muestran a los hijos como fuente de placer, ingreso familiar y seguro de ancianidad. Además se tiende a propiciar la reducción del índice de fecundidad buscando un ajuste poblacional (controlar su crecimiento) lo cual propicia un bienestar en la población, además de garantizar modernización y desarrollo; en las familias se sustenta el control de natalidad como la intención de tener hijos de mejor calidad (respecto a alimentación, crianza, educación, etc. ), lo cual es posible si la cantidad de hijos es menor.

Esta teoría sobre la fecundidad tiene críticas por tener en cuenta la fecundidad como un seguro de los padres o un bien de consumo durable, y no por ser también una situación espontánea que además responde a factores culturales y no solo a razonamientos económicos. También hay perspectivas que intentan explicar ambas posiciones, Heer (1966) señala que un rápido crecimiento de los indicadores económicos en una población implica un aumento en la fecundidad por el estatus social que tendrán las familias, aunque a largo plazo el índice de fecundidad tiende a reducirse. Es decir "en el largo plazo, las fuerzas de las variables que reducen la fecundidad causan un mayor peso que aquellas que la incrementan. Las variables en cada sentido parecen no estar lo suficientemente claras, a razón que se toman variables como la densidad poblacional, la circulación de periódicos, la mortalidad infantil, el producto per cápita y el consumo de energía. En el estudio las dos primeras tienen efecto negativo sobre la fecundidad, mientras que las restantes son positivas" (Mesa y Junca, 2011).

Según análisis de determinantes próximos realizado por Bongaarts (1978) los factores de la fecundidad se pueden señalar en estas tres categorías:

1. Factores de exposición: proporción de casadas.
2. Factores de control de la fecundidad: anticoncepción y aborto inducido.
3. Factores naturales de la fecundidad: esterilidad, frecuencia de relaciones sexuales, la duración del periodo de fertilidad y la mortalidad intrauterina.

#### **5.4. Modelos ARIMAX**

Por características de los datos empleados los cuales se definen por una línea de tiempo en meses y condiciones de estacionalidad, se escogen los modelos ARIMAX para la estimación que explique estas características y permita crear predicciones a corto plazo. Adicional estos modelos permiten incluir covariables en la estimación de partir de cálculos por regresión, que pueden mejorar la estimación si las variables adicionales muestran correlación con la variable de la serie.

A continuación se explica el modelo ARIMAX a partir modelos base que lo pueden describir.

#### 5.4.1. ARIMA

Un proceso estocástico  $z_t$  es integrado de orden  $d$  ( $d \geq 0$  entero) si y sólo si  $z_t$  sigue un modelo autoregresivo integrado de media móvil de orden  $(p, d, q)$  ( $ARIMA(p, d, q)$  , de manera simplificada), donde  $p$  es el orden de la parte autoregresiva estacionaria,  $d$  es el numero de raíces unitarias y  $q$  es el orden de las medias móvil. Podemos describir el proceso por:

$$(1 - \phi_1 B - \dots - \phi_p B^p)(1 - B)^d z_t = c + (1 - \theta_1 B - \dots - \theta_q B^q) a_t$$

Ahora escribimos el proceso anterior utilizando el operador de diferencia,  $\nabla = 1 - B$ ,

$$\phi_p(B) \nabla^d z_t = c + \theta_q(B) a_t$$

donde  $\phi_p(B)$  es el polinomio autorregresivo,  $\theta_q(B)$  es el polinomio de medias móviles (Peña, 2010. p,187) .

#### 5.4.2. ARIMA Estacional

En caso de estacionalidad en un modelo ARIMA podemos construir un modelo donde ambas formas (la dependencia regular y la estacional) se expresen de manera multiplicativa. (Peña, 2010. p, 207)

$$\Phi_p(B^s) \phi_p(B) \nabla_s^D \nabla^d z_t = \theta_q(B) \Theta_Q(B^s) a_t$$

Donde

$\Phi_p(B^s) = (1 - \Phi_1 B^s - \dots - \Phi_p B^{sP})$  es el operador AR estacional de orden  $P$

$\phi_p = (1 - \phi_1 B - \dots - \phi_p B^P)$  es el operador AR regular de orden  $p$

$\nabla_s^D = (1 - B)^D$  son las diferencias estacionales

$\nabla^d = (1 - B)^d$  las diferencias regulares

$\Theta_Q(B^s) = (1 - \theta_1 B^s - \dots - \theta_Q B^{sQ})$  el operador de media móvil estacional de orden  $Q$

$\theta_q(B) = 1 - \theta_1 B - \dots - \theta_q B^q$  es el operador de media móvil regular de orden  $q$

$a_t$  es un proceso de ruido blanco.

De forma simplificada el modelo se escribe de la forma  $ARIMA(P, D, Q) \times (p, d, q)$

### 5.4.3. ARIMAX

El modelo de media móvil autorregresivo que incluye covariables exógenas  $ARMAX(p, q)$ , extiende el modelo  $ARMA(p, q)$  al incluir el efecto lineal que una o más series exógenas tiene en la serie de respuestas estacionarias  $z_t$ . La forma general del modelo  $ARMAX(p, q)$  es

$$z_t = \sum_{i=1}^p \phi_i z_{t-i} + \sum_{k=1}^r \beta_k x_{tk} + a_t + \sum_{j=1}^q \theta_j a_{t-j}$$

y la forma condensada de operador lag

$$\phi(B)z_t = c + x_t' \beta + \theta(B)a_t$$

El vector  $x_t'$  contiene el valor de los  $r$  predictores exógenos.

El modelo  $ARIMAX$  es una generalización del modelo  $ARIMA$  en el que incluye una o más covariables exógenas al modelo con una respuesta estacionaria. La forma general de los modelos  $ARIMAX(p, d, q)$  es

$$\phi(B)\nabla^d z_t = c + x_t' \beta + \theta_q(B)a_t$$

donde  $\phi(B)$ ,  $\theta_q(B)$ ,  $\nabla^d$  son análogamente al modelo  $ARIMA$ , el polinomio regresivo, polinomio de medias móviles y operador de diferencia respectivamente.

Tenemos adicional que  $x_t$  es la variable explicativa o covariable exógena y  $\beta$  su coeficiente. Cabe señalar  $\beta$  solamente puede interpretarse condicionalmente sobre el valor de los valores previos de la variable de respuesta.

Se puede también incluir componentes estacionales en un modelo *ARIMAX*, el cual se generaliza a un modelo *ARIMAX*  $(p, d, q) \times (P, D, Q)$ . Suponiendo que la serie de respuestas  $y_t$  es estacionaria, el modelo tiene la forma

$$\Phi_P(B^S)\phi_p(B)\nabla_s^D\nabla^d z_t = \sum_{i=1}^n x_{it}\beta + \theta_q(B)\Theta_Q(B^S)a_t$$

donde  $\Phi_P(B^S)$  es el operador *AR* estacional de orden  $P$ ,  $\phi_p(B)$  es el operador *AR* regular de orden  $p$ ,  $\nabla_s^D\nabla^d$  son las diferencias estacionales y regulares,  $\Theta_Q(B^S)$  el operador de media móvil estacional de orden  $Q$ ,  $\theta_q(B)$  es el operador de media móvil regular de orden  $q$ , y por último  $x_t\beta$  es la variable exógena con su coeficiente.

## 5.5. Modelos Dinámicos Lineales

El uso de los modelos dinámicos lineales para datos de natalidad se decide por su capacidad de estimación de series de tiempo complejas en su análisis, que además en muchos casos puede mejorar estimaciones de una *ARIMA* clásico.

Los modelos dinámicos tienen por característica tener dinámicas en los parámetros del modelo, se refiere a tener parámetros cambiantes en el tiempo o parámetros dependientes del tiempo. Se pueden describir como ventajas de los modelos dinámicos: son modelos simples en su planteamiento, son fuertes en la realización de ajustes y predicciones, es posible realizar suavizamientos de la variable, la inclusión de variables explicativas se realiza de manera simple teóricamente.

Los modelos dinámicos lineales (MLD) son un caso particular de los dinámicos, estos tienen aplicación directa sobre en el modelamiento y análisis de series de tiempo o en modelos de regresión con errores autocorrelacionados. Son



muy útiles en métodos secuenciales donde la actualización de los parámetros debe hacerse con datos obtenidos secuencialmente.

El modelo lineal dinámico general con una ecuación de observación y una ecuación modelo es especificado a priori por una distribución normal para el vector de estado  $p$ -dimensional en  $t = 0$ .

$$\theta_0 \sim N_p(m_0, C_0)$$

$$Y_t = F_t \theta_t + v_t, \quad v_t \sim N(0, V_t),$$

$$\theta_t = G_t \theta_{t-1} + w_t, \quad w_t \sim N(0, W_t),$$

donde

- $Y_t$  es un vector de observaciones de longitud  $m$  en el tiempo  $t$ , con  $t = 1, \dots, n$ .
- $\theta_t$  contiene estados no observados del sistema que cambian en función del operador de sistema lineal  $G_t$ , una matriz de orden  $p \times p$ , además  $\theta_t$  tendrá elementos correspondientes a varios componentes del proceso de series de tiempo como la tendencia o la estacionalidad.
- $F_t$  es una matriz de orden  $m \times p$  es el operador de observación que transforma los estados del modelo en observaciones.
- $V_t$  son matrices de covarianza que representan los errores gaussianos de las observaciones.
- $W_t$  son matrices de covarianza que representan los errores gaussianos de los estados.

Para el análisis de series univariadas tendremos que  $p = 1$ , en el caso multivariado las matrices del sistema  $F_t, G_t, V_t, W_t$  se usan para definir correlaciones entre los componentes observados (Laine, 2019).

La primera ecuación define el modelo observacional para una respuesta  $Y_t$  y su relación con  $p$  covariables o variables explicativas  $\theta_t$ . La segunda ecuación conocida como la ecuación de estado o de sistema, especifica la evolución de los parámetros a través del tiempo. Los errores del modelo lineal dinámico son  $v_t$  y  $w_t$

con varianzas dependientes del tiempo  $V_t$  y  $W_t$  que definen la varianza observacional y la varianza del sistema respectivamente.

Para estimar el vector de estados se calculan las densidades condicionales  $\pi(\theta_s, y_{1:t})$ .

Si  $G_t, F_t, V_t, W_t$  son constantes, decimos que el modelo es invariante con respecto al tiempo.

Se clasifican en tres tipos de problemas la aplicación de los modelos dinámicos.(Petris, 2009. p, 41):

1. Filtrado: con datos secuenciales en el tiempo para estimar el valor actual del vector del estado, y actualizar nuestra estimaciones y pronósticos a medida que se actualizan datos en  $t + 1$ .  
El filtro de Kalman proporciona las formulas para actualizar en la inferencia actual el vector de estado, para pasar a la densidad de filtrado.
2. Suavizado: o análisis retrospectivo consiste en estimar la secuencia de estados en  $1, \dots, t$  dados  $y_1, \dots, y_t$ . Se puede estudiar retrospectivamente el comportamiento del sistema que subyace a las observaciones. El problema de suavizado se resuelve calculando la distribución condicional de  $\theta_{1:t}$  dado  $y_{1:t}$ , En cuanto al filtrado, el suavizado se puede implementar como un algoritmo recursivo.
3. Predicción de estados: a través de la función de densidad de probabilidad, un estimado optimo del valor de  $Y_{t+1}$  es la esperanza condicional del error de predicción al cuadrado,  $E(Y_{t+k} | y_{1:t})$  es conocido como función de predicción. Por la naturaleza Markoviana del modelo, la distribución filtrada al tiempo  $t$ , sirve como distribución inicial para la futura evolución del modelo, es decir, la distribución conjunta del presente y futuros estados  $(\theta_{t+k})_{k \geq 0}$  y futuras observaciones  $(Y_{t+k})_{k \geq 1}$  es un modelo de

espacios de estados con distribuciones condicionales  $\pi(\theta_{t+k}|\theta_{t+k-1})$  y  $\pi(y_{t+k}|\theta_{t+k})$  y distribución inicial  $\pi(\theta_t|y_{1:t})$ .

La obtención de distribuciones condicionales relevantes en los procesos no es tarea sencilla, en el caso general, pero en los modelo dinámicos lineales esta labor se reduce por la suposición de distribuciones normales.

### 5.5.1. Filtro de Kalman

El filtro de Kalman es un algoritmo creado por Rudolf Kalman se diseño con el propósito de identificar el estado oculto en un sistema lineal dinámico, el algoritmo tienen características recursivas luego puede correr únicamente usando las entradas actuales. Es de gran importancia pues reduce los cálculos necesarios para la filtración, suavizado y predicción de los modelos lineales dinámicos.

Con el modelo lineal dinámico ya especificado, suponemos que  $\theta_{t-1}|y_{1:t-1} \sim N(m_{t-1}, C_{t-1})$ , se cumple:

1. La distribución predictiva de  $\theta_t$  dado  $y_{1:t-1}$  es Gaussiana con parámetros:

$$a_t = E(\theta_t|y_{1:t-1}) = G_t m_{t-1}$$

$$R_t = Var(\theta_t|y_{1:t-1}) = G_t C_{t-1} G_t' + W_t$$

2. La distribución predictiva de  $y_t$  dado  $y_{1:t-1}$  es Gaussiana con parámetros

$$f_t = E(Y_t|y_{1:t-1}) = F_t a_t$$

$$Q_t = Var(y_t|y_{1:t-1}) = F_t R_t F_t' + V_t$$

3. La distribución filtrada de  $\theta_t$  dado  $y_{1:t}$  es Gaussiana con parametros

$$m_t = E(\theta_t|y_{1:t}) = a_t + R_t F_t' Q_t^{-1} e_t$$

$$C_t = Var(\theta_t|y_{1:t}) = R_t - R_t F_t' Q_t^{-1} F_t R_t$$

con  $e_t = y_t - f_t$ , es el error de predicción. (Petris, 2009. p, 53)

## **6. METODOLOGÍA**

### **6.1. Datos**

Se usan los datos publicados por el DANE en el repositorio de datos abiertos. Son datos de estadísticas vitales, los cuales se encuentran divididos en datos de nacimientos, defunciones fetales y no fetales. Estos aparecen en archivos individuales por años, con aproximadamente 700.000 registros en promedio desde el año 1998, datos que el DANE recopila por medio de los certificados de nacido vivo sobre todo el área nacional. Aparece subdividida la información según geografía, grupos de población.

Inicialmente los datos cuentan con 33 variables aproximadamente, variables las cuales pueden cambiar según el año; se incluyen datos sobre características del nacido, características de la madre y del padre, información del lugar de nacimiento, entre otros.

Para este trabajo se usarán datos desde el año 2011, hasta el año 2017, usando datos de nacimientos y defunciones fetales.

### **6.2. Tratamiento de datos**

Para el objetivo del presente trabajo fue necesario realizar un tratamiento sobre los datos, pues la información está distribuida por características de cada nacimiento:

- Escoger variables necesarias a usar.
- Se decidió seleccionar las principales regiones del país.
- Se revisó la información ya filtrada en busca de posibles datos faltantes.
- Se generaron agrupamientos de los nacimientos y defunciones fetales por cada mes, .
- Al tener información por cada nacimiento, se decidió promediar cada una de las variables adicionales.

Muestra de datos resultantes

Mes	Nacimientos	Edad madre	Edad padre	N hijos	Muerte Fetal
2011m01	48420	25,1606608	29,3436076	1,94603326	3648
2011m02	43533	25,1564385	29,3942581	1,94904952	3440
2011m03	51717	25,1408296	29,3917033	1,95640281	3735
2011m04	49600	25,0950251	29,3302887	1,95640857	3437
2011m05	52073	25,1346854	29,3298613	1,94202557	3411
2011m06	50939	25,0561208	29,2291196	1,93695355	3264
2011m07	52199	25,054511	29,2231192	1,9416206	3417
2011m08	55317	25,0183558	29,2264617	1,93036412	3362
2011m09	58296	25,0303902	29,2338857	1,91396597	3307
2011m10	53721	25,0291286	29,1594383	1,9192404	3163

Tabla 1. Tratamiento de datos

### 6.3. Variables:

Las variables planteadas para la predicción de nacimientos mediante modelos ARIMAX y dinámicos son:

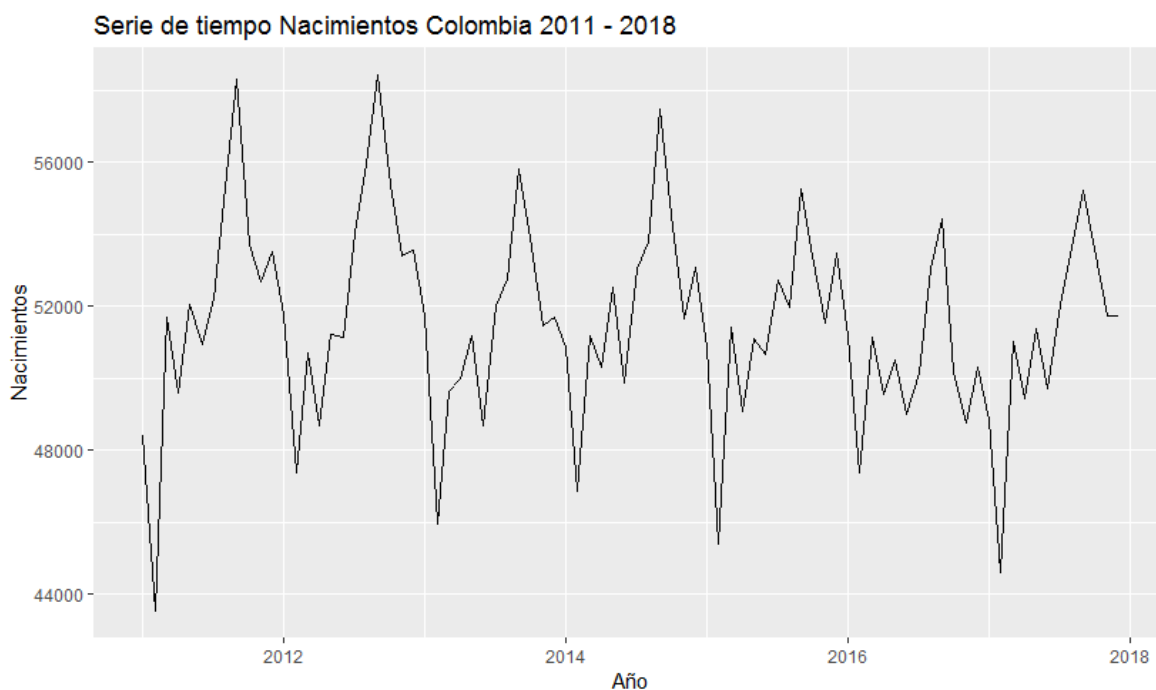
**Año - mes:** Descrita en cada archivo de datos por año y mes enumerado del 1 al 12, representa la temporización de las series de tiempo.

#### 6.3.1. Nacimientos

La variable nacimientos esta descrita de forma frecuencial (cantidad de nacimientos) en periodos mensuales.

Nacimientos sobre las principales regiones del país, gráficamente se evidencia una fuerte estacionalidad aparente a 12 meses, donde los picos altos de nacidos se presentan usualmente en los meses de agosto - septiembre (octubre en algunos años), con un máximo de nacimientos alcanzado en el año 2012 con 58497. Los picos mínimos se presentan en los meses de enero - febrero, alcanzo el valor mínimo de 43533 en el año 2011.

Cabe mencionar la tendencia poco marcada de decrecimiento sobre la cantidad de nacimientos en los años que se está analizando.



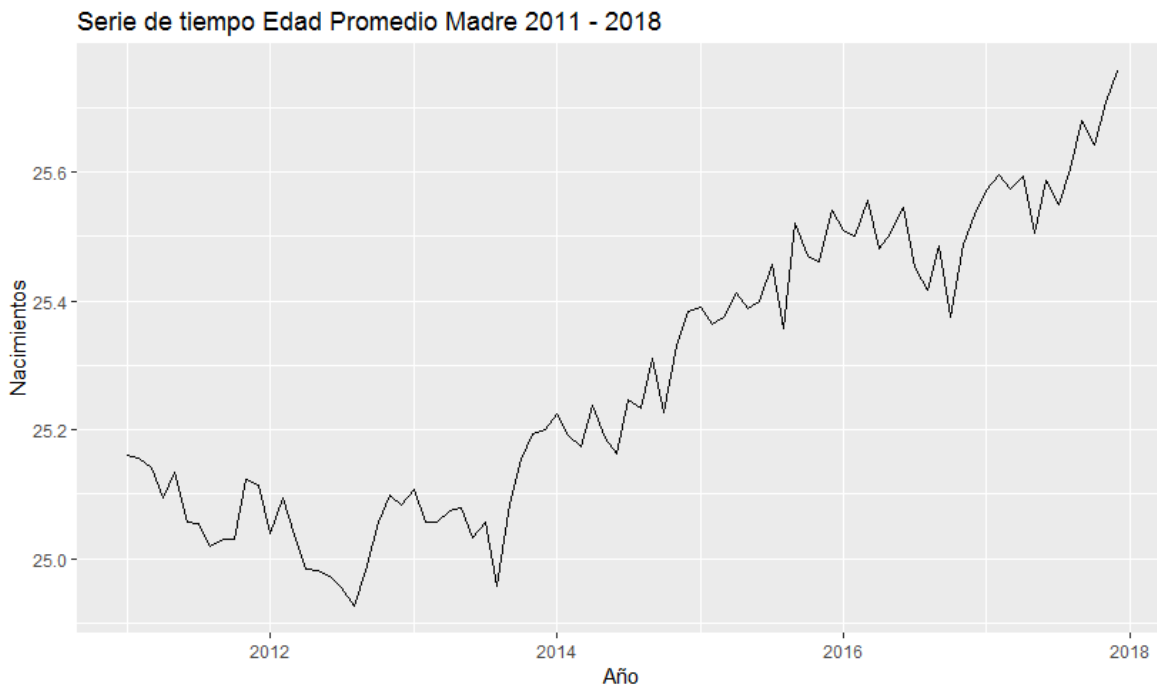
Gráfica 1. Serie de tiempo nacimientos en Colombia

### **6.3.2. Edad madre**

Es la edad promedio calculada de la madre por mes, según intervalos descritos en los datos del DANE: inicia de 10 - 14 Años, hasta 50 - 54 Años.

La serie de tiempo de edad promedio de la madre, se encuentra entre un mínimo de 24.93 años y un máximo de 25.76 años, no presenta algún tipo de estacionalidad (aunque existen picos mínimos por año, que parecen comunes). Cabe mencionar la tendencia de crecimiento que tiene, en solo 6 años muestra un el aumento de edad promedio en casi un año, tendencias comunes a nivel mundial.

En los datos recogidos por el DANE en esta variable, cabe mencionar los datos de nacimientos de madres adolescentes, aunque es difícil estipularlo dentro de un valor promedio como el descrito, en el rango de 10 a 14 años es de destacar la cantidad de nacimientos que existen. Cabe mencionar que los rangos entre 10 a 14 años, y 15 a 19 años de las madres lastran el promedio de edad.

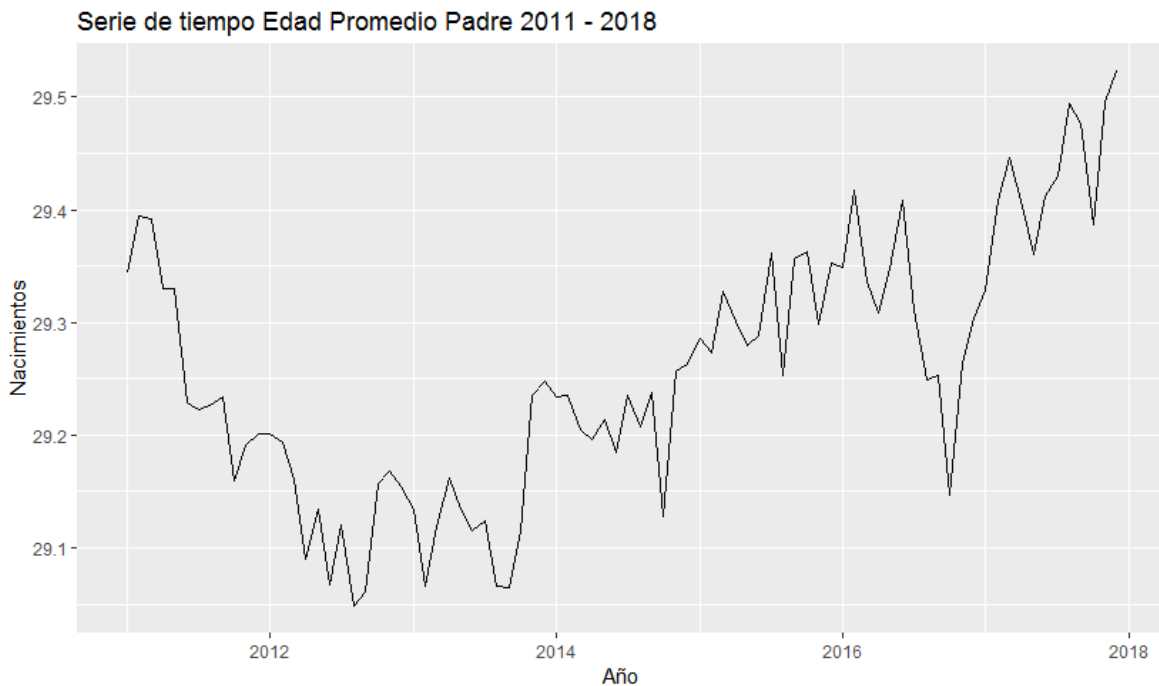


Grafica 2 Serie de tiempo edad de la madre

### **6.3.3. Edad padre**

Edad promedio del padre (en años) según valores descritos en los datos del DANE, por meses. En este caso los datos no están agrupados por rangos de edad.

La edad promedio de edad del padre es de 29.26, es cual es bastante representable de la información, los mínimos y máximos se encuentran en 29.05 y 29.52. No se evidencia estacionalidad en la serie. La serie muestra una tendencia marcada hacia el crecimiento, al aumento de la edad, aunque no pronunciado.



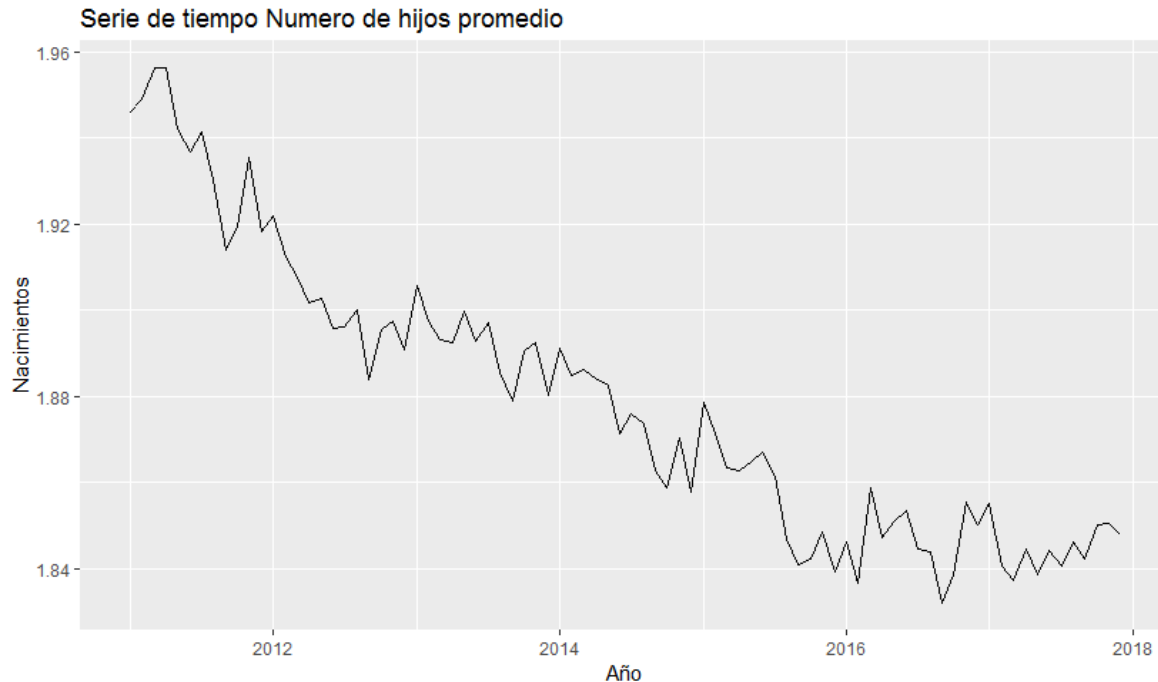
Grafica 3 Serie de tiempo edad promedio del padre

#### **6.3.4. Número de hijos**

Promedio de hijos de la madre contando al gestante, los datos no se encuentran agrupados por parte del DANE.

La variación de la serie está entre 1.832 y 1.956, lo cual muestra una tendencia a la disminución de la cantidad de hijos de las madres en Colombia, cabe señalar que este dato no es equivalente al de tasa de fecundidad en Colombia pero podría ser contrastable con este. La serie no evidencia ningún tipo de estacionalidad.



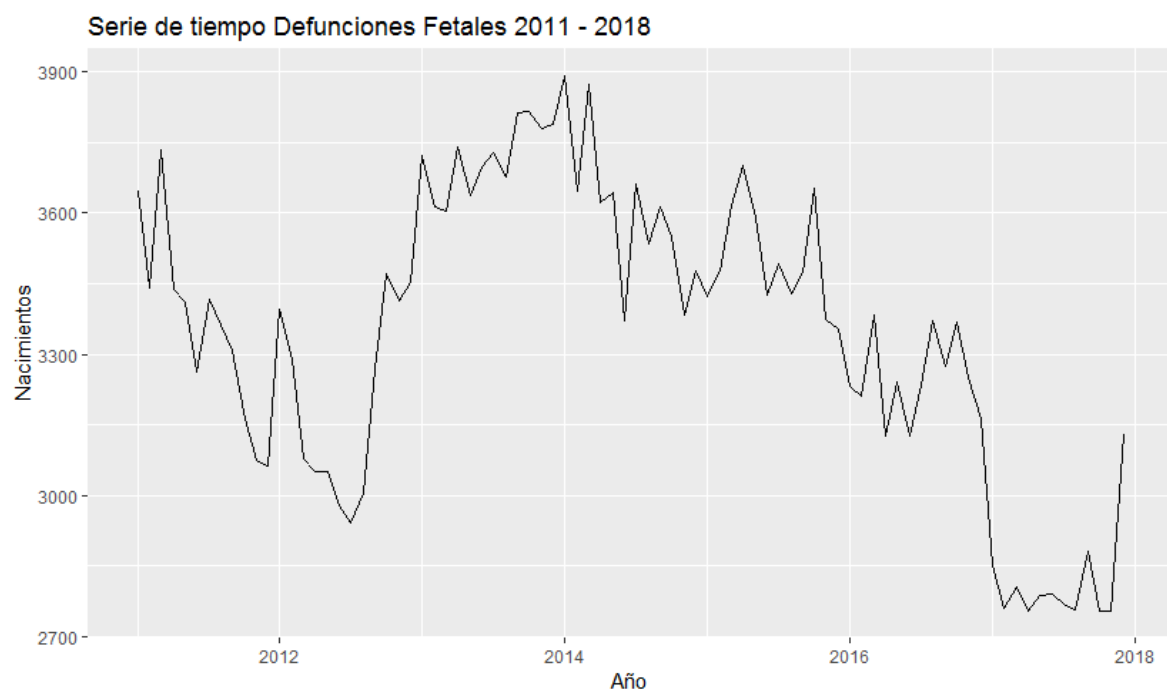


Grafica 4 Serie de tiempo número de hijos promedio

### **6.3.5. Defunciones fetales**

La variable de muertes fetales esta descrita de forma frecuencial en periodos mensuales entre los años 2011 y 2017. La serie tiene particularidades, se observan dos caídas marcadas en los datos, en el año 2012 entre marzo y agosto los datos alcanzan las 2942 muertes fetales el más bajo entre los años 2011 y 2016, y el año 2017 tiene una constante en la cual se reducen las muertes a niveles no vistos en los años anteriores alcanzando su mínimo en octubre con 2754 muertes. Es difícil de explicar la razón de este comportamiento en la serie, pues podría depender de factores tales como los biológicos y la cobertura de atención medica en regiones especificas del país.

Es de señalar la diferencia en la tendencia que tiene esta serie con la serie de nacimientos, pues a priori se contempló una relación entre ellas, con un comportamiento parecido en la característica estacional de los nacimientos.



Grafica 5 Serie de tiempo defunciones fetales

## 7. ANÁLISIS DE RESULTADOS

Se procede con el análisis de la serie nacimientos con variables de apoyo (variables exógenas), usando inicialmente todas las establecidas (edad madre, edad padre, hijos y defunciones fetales).

Recordando que tenemos por intención estimar los datos de natalidad, en el modelo ARIMAX con las covariables anteriores, en el modelo lineal dinámico se establecen condiciones iniciales para establecer estimación de los datos (sin el apoyo de variables exogenas).

También se utiliza el modelo ARIMAX para realizar una predicción, luego se decide realizar una partición de 60 datos (meses, enero 2011 a diciembre 2016) de entrenamiento y 12 datos (meses, enero 2017 a diciembre 2018) de prueba para contrastar proyecciones del modelo. Una vez realizadas las pruebas de estimación se plantea una predicción a 6 meses.

Para toda prueba de hipótesis realizada en el modelo ARIMAX se establece un valor de significancia del 10%.

### 7.1. Modelo ARIMAX Estimación

siguiendo metodología Box - Jenkins establecida en 4 pasos: identificación, estimación, diagnostico, estimación o pronostico.

#### 7.1.1. Identificación

Realizando criterios de estacionaridad para la serie natalidad:

*Test Dickey – Fuller = -6.1463 con p – valor < 0.01*

*Test Phillips – Perron = -38.09 con p – valor < 0.01*

Ambos test señalan la estacionaridad de la serie de tiempo de Nacimientos, por los estadísticos de las pruebas de test Dickey - Fuller y test Phillips - Perron que generan p -valores inferiores a 0.01. La serie no requiere diferenciación por estacionaridad.

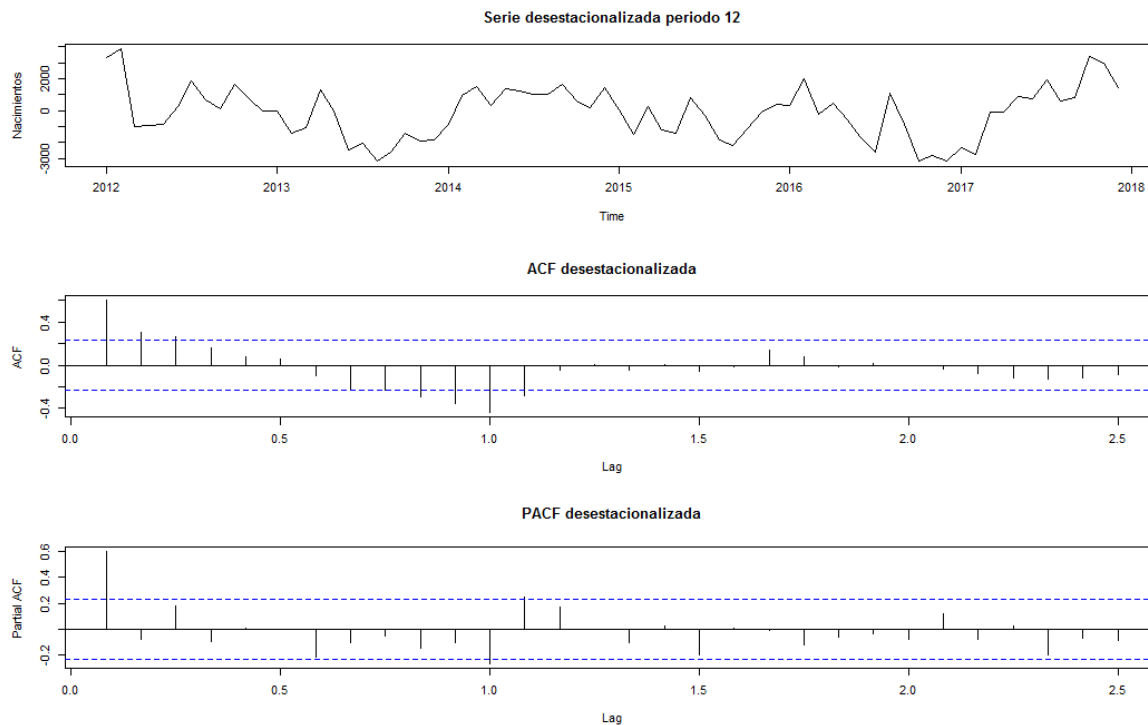
Se evidencia la fuerte condición de estacionalidad en la serie, aunque es necesario probar estacionalidad a 6 periodos o 12 periodos (6 o 12 meses). también se evidencia un AR(1) para aplicar en el modelo ARIMAX.

Se realizan test de estacionaridad en la serie con una diferencia de estacionalidad con periodicidad de 12 meses esto por características de la serie:

$$\text{Test Dickey} - \text{Fuller} = -4.8754 \text{ con } p - \text{valor} = 0.0397$$

$$\text{Test Phillips} - \text{Perron} = -27.086 \text{ con } p - \text{valor} < 0.01$$

Según análisis de ACF y PACF se evidencian como parámetros autoregresivos un AR(1) pero sin parámetros de medias móviles, con SMA(1) y sin SAR para el caso estacional.



Grafica 6 . ACF y PACF de la serie natalidad estacionalizada

### 7.1.2. Estimación

Para la estimación del modelo se decidió transformar todas las variables por medio de la función logaritmo, pues los datos resultan muy desiguales respecto a la escala de cada una de ellos.

Según el modelo planteado las variables exógenas de número de hijos y defunciones fetales no fueron significativas, luego después de varias pruebas se excluyeron del modelo.

Tenemos como covariables, o variables exógenas:

$$x_1 = \log(\text{Edad madre}_t)$$

$$x_2 = \log(\text{Edad Padre}_t)$$

$$x_3 = \log(\text{Numero de hijos}_t) \text{ Variable no significativa}$$

$$x_4 = \log(\text{Defunciones fetales}_t) \text{ Variable no significativa}$$

el modelo *ARIMAX* escogido es:  $ARIMAX(1, 0, 0) \times (0, 1, 1)_{12}$ , con:

$$z_t = \log(\text{Nacimientos}_t)$$

$$\phi(L)\nabla_s^{12}z_t = \sum_{i=1}^2 x_{i,t}\beta_i + \theta(L^{12})a_t$$

donde:

$$\phi(L) = 1 - 0.5505L \quad AR(1) \text{ con significancia del } 7.473 \times 10^{-7}$$

$$\theta(L^{12}) = 1 + 0.9999L^{12} \quad SMA(1) \text{ con significancia del } 9.630 \times 10^{-6}$$

$$\beta_1 = -1.3391 \quad x_1 \log(\text{edad madre}) \text{ con significancia de } 0.05862$$

$$\beta_2 = -1.2844 \quad x_2 \log(\text{edad padre}) \text{ con significancia de } 0.05800$$

los valores de autoregresivos y medias móviles estacionales fueron significativos con un  $BIC = -329.6902$  y  $\sigma^2 = 0.0003213$ .

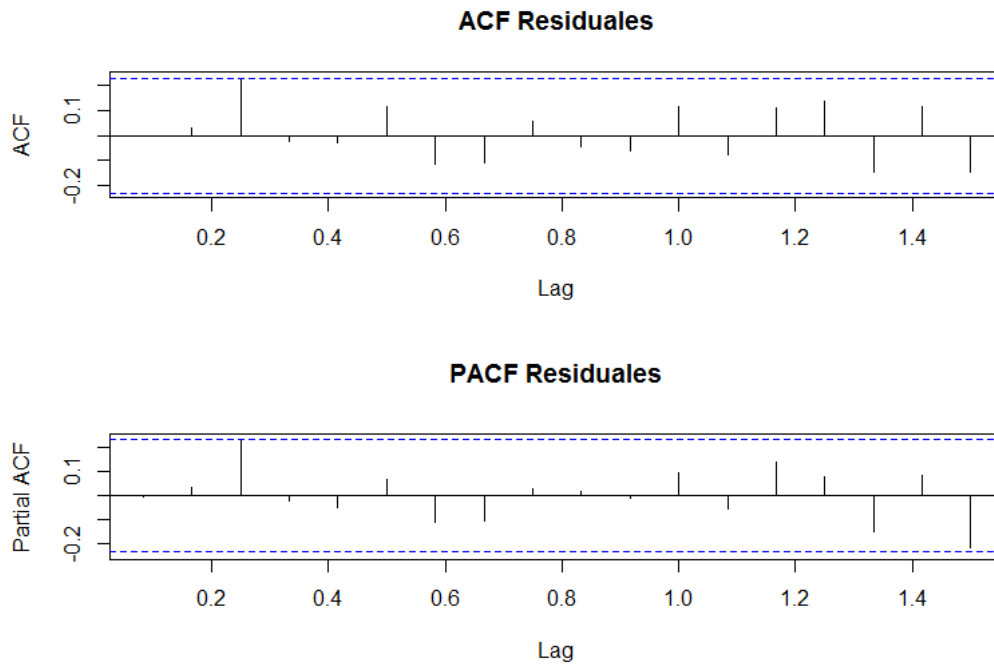
Es importante señalar la falta de unidad de los coeficientes de las variables exógenas, esto se debe a las transformaciones necesarias que se realizaron sobre los datos, la aplicación de la función logaritmo, lo cual representa un cambio de escala de los datos conocidos. Luego se tiene perdida de interpretación sobre el modelo con respecto a los datos originales.

### **7.1.3. Diagnóstico**

Para el diagnóstico del modelo se realizaron pruebas de autocorrelaciones nulas, de normalidad y aleatoriedad.

Mediante la prueba de Lung - Box cuya hipótesis inicial es existencia de autocorrelaciones nulas, que con un p-valor de 0.4306 tenemos evidencia estadística de autocorrelaciones nulas. Con el test Jarque Bera que obtuvo un p-valor de 0.3078, con su hipótesis inicial asegura evidencia estadística de la normalidad en la distribución de la serie. Para el test de aleatoriedad obtenemos un p-valor de 0.6858, con lo obtenemos evidencia estadística sobre la aleatoriedad de los residuales. Recordando, esto con un nivel de significancia del 10% para las pruebas.

Adicional tenemos el análisis de los residuales por medio del ACF y el PACF, los cuales muestran un buen comportamiento al estar entre las cotas de nivel de confianza.

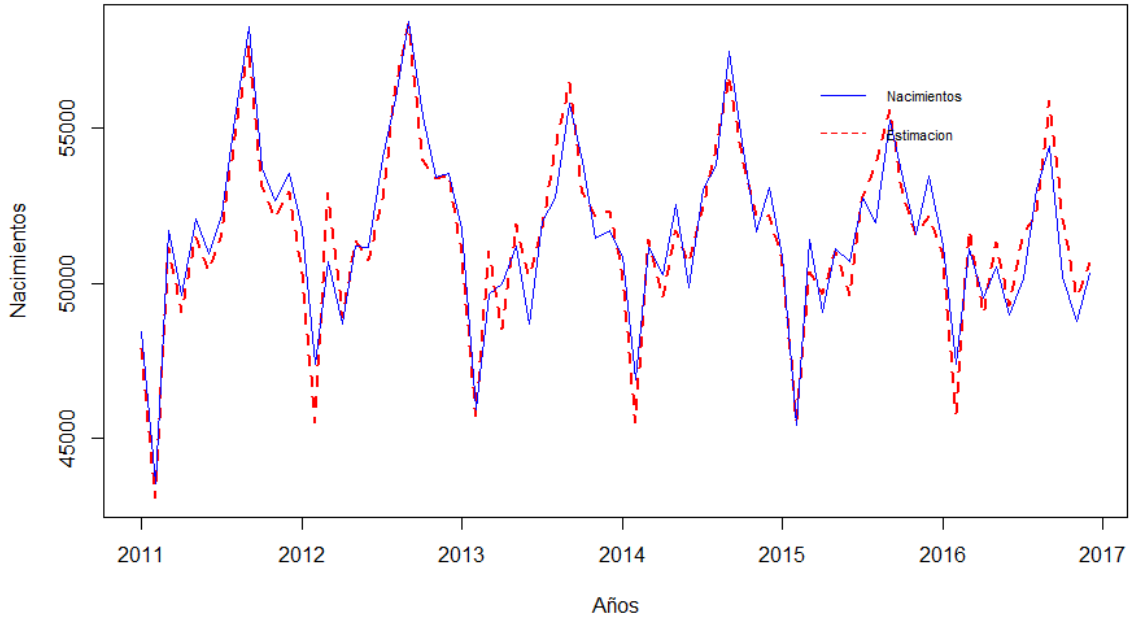


Grafica 7 Residuales

#### **7.1.4. Estimación Modelo**

El modelo establecido calcula una estimación buena a los datos de natalidad, lo que permite explicar los datos a partir del modelo, se deja como anexo todos los datos de estimación.

### Estimación natalidad modelo ARIMAX



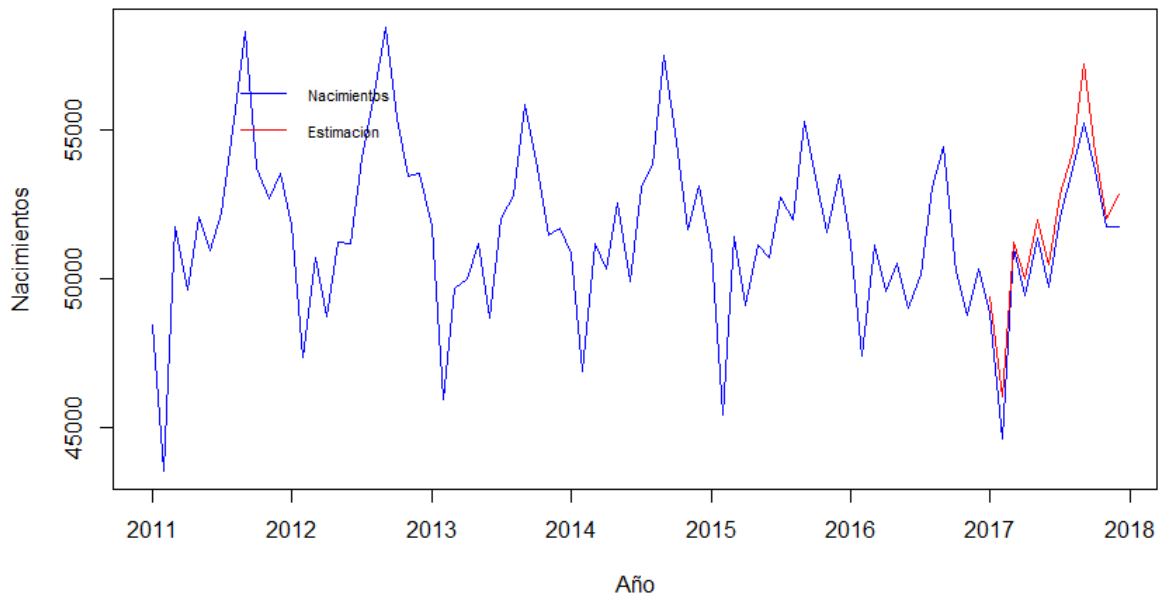
Grafica 8 Estimación modelo ARIMAX

## 7.2. Pronóstico Modelo ARIMAX

Se realiza un pronóstico inicial para comparar con los datos de prueba, y poder comparar la predicción con datos reales del año 2017 y comprobar su fiabilidad, teniendo el siguiente resultado:



### Predicción - Datos reales 2017



Grafica 9 Datos reales 2017

El pronóstico es bueno comparados en los datos reales del año 2017, incluso pronostican el pico mínimo que alcanza en febrero que es de los datos más bajos en todo el periodo analizado. La estimación del año 2017 se presenta a continuación con los datos de prueba del mismo año, incluyendo las cotas de los intervalos de confianza del 95%.

	colombia_2017	Proyección	proy_low	proy_upp
2017m01	48845	49375,27649	47386,84259	51447,14852
2017m02	44623	46005,79637	43898,53566	48214,21187
2017m03	51044	51231,37679	48804,12793	53779,34366
2017m04	49429	49968,78404	47577,56255	52480,18699
2017m05	51378	51990,96658	49495,36388	54612,39991
2017m06	49726	50467,49567	48042,74172	53014,6288
2017m07	52058	52804,95175	50267,16847	55470,85729
2017m08	53653	54248,7138	51641,35889	56987,71318
2017m09	55230	57184,55996	54436,19339	60071,68567
2017m10	53606	54230,44295	51624,55521	56967,86985
2017m11	51738	52039,01589	49540,04418	54664,04439
2017m12	51749	52828,09199	50296,54446	55487,05847

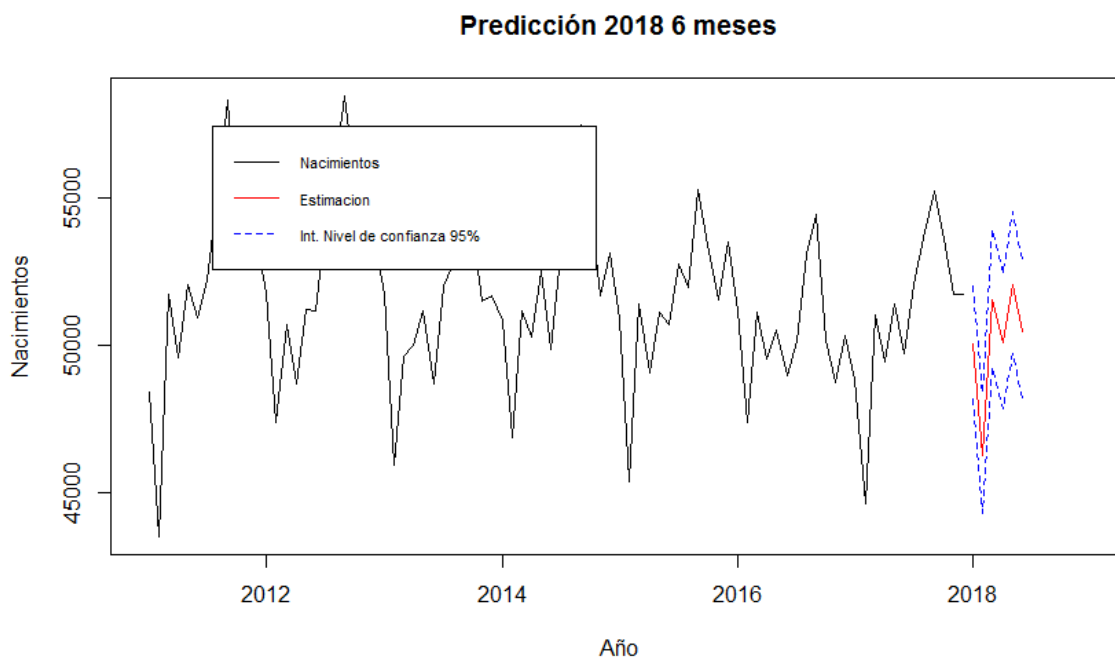
Tabla 2. Datos base para análisis

Después de corroborar la fiabilidad del modelo, se procede a realizar la predicción de los primeros 6 meses del año 2018. Los datos se presentan en la siguiente tabla:

	proy2018	proy_low2018	proy_upp2018
2018m01	50049,08807	48161,52576	52010,62834
2018m02	46284,03906	44301,43815	48355,36634
2018m03	51534,7572	49253,63887	53921,52256
2018m04	50087,86188	47849,59123	52430,8326
2018m05	52063,23927	49730,06241	54505,8814
2018m06	50470,02874	48206,30823	52840,05133

Tabla 3. Proyecciones

en la siguiente grafica se muestra la predicción señalada en la tabla, con incluida sobre el total de los datos de la serie de tiempo de natalidad.



Grafica 10 Predicción 6 meses año 2018

### 7.3. Modelo Lineal Dinámico

Se realiza la estimación de los datos de natalidad a partir del modelo lineal dinámico sobre condiciones presentes, es decir realizando el filtrado de los datos, esto por medio del Filtro de Kalman sobre el que se basan los algoritmos del paquete dlm de R. (Petris, 2010)

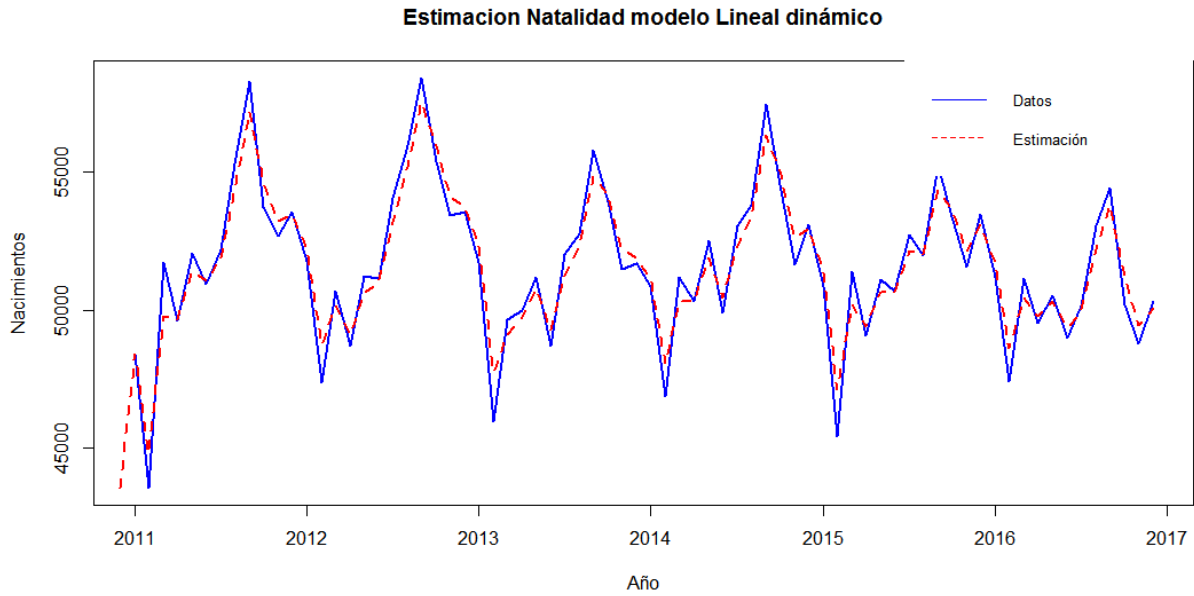
Inicialmente se establece a  $m_0$  de la distribución  $N_p(m_0, C_0)$  como el mínimo de los datos de natalidad, para evitar que el modelo inicie desde el valor 0. Para el planteamiento inicial del modelo, se requieren los valores de  $V$  y  $W$  de las ecuaciones del modelo general, que son las covarianzas que representan los errores gaussianos de las observaciones y los estados. Para esto se utiliza la función dlmMLE que por medio de estimaciones de máxima verosimilitud calcula dichos datos:

$$V = -9.214583 \text{ y } W = 17.020294$$

Con esta información se establecen condiciones y características para establecer el modelo a partir de la tendencia polinomial de la serie, que en este caso es de grado 1.

Una vez planteada la manera de estimar la tendencia se filtra la serie usando como base el modelo polinomial, genera un conjunto de matrices para la filtración de la serie. Esta filtración genera una estimación sobre los valores actuales de la serie.

## Gráfica estimación



Gráfica 11 Estimación modelo lineal dinámico

Por la manera en que se estableció el modelo dinámico, la generación de las matrices y valores fueron constantes, luego es un modelo invariante.

## 8. CONCLUSIONES

Luego del planteamiento de los modelos ARIMAX y el modelo lineal dinámico, como primera conclusión puedo señalar la fiabilidad de ambas estimaciones, los modelos se mostraron suficientes para explicar características de la serie de natalidad, en la cual pudieron estimar su tendencia y estacionalidad.

Respecto al modelo ARIMAX, la estimación de los datos permitió generar una buena predicción, la cual fue acertada para el año 2017 (que se estableció como prueba del modelo), lo cual genera una predicción estadística confiable para el primer semestre del año 2018. Es válido afirmar que las covariables escogidas apoyaron la estimación del modelo, pero es importante continuar con la búsqueda de variables adicionales que expliquen los datos de natalidad.

En el modelo lineal dinámico el objetivo de estimación que se planteo género una estimación más cercana a los datos que el ARIMAX, a falta de agregar características adicionales. También cabe señalar la dificultad que represento el planteamiento del mismo, por ejemplo en la agregación de las configuraciones de autoregresivos y de medias móviles, así como la agregación de la estacionalidad de la serie. Estas dificultades se presentaron tanto por conocimientos teóricos del modelo como también su correcta ejecución en código de R.

Cabe señalar la generalización que los modelos dinámicos lineales representan para los modelos ARIMA clásicos, lo cual permite a estos un mejor modelamiento de los datos en caso de natalidad, es válida la pregunta si en todo contexto esto se mantiene.

## 9. REFERENCIAS

- Bongaarts. (1978). A framework for analysing the proximate determinants of fertility. *Population and Development Review*, 4, 105-132. doi:10.2307/1972149
- Burgos, A. (2015). Políticas públicas en América Latina para la reducción de la mortalidad materna, 2009-2014. Serie: Población y Desarrollo. Publicación de Naciones Unidas, Santiago de Chile.
- DANE. (2009). Proyecciones de población y estudios demográficos. Consultado de [https://www.dane.gov.co/files/investigaciones/fichas/Ficha\\_Proyecciones\\_poblacion.pdf](https://www.dane.gov.co/files/investigaciones/fichas/Ficha_Proyecciones_poblacion.pdf)
- Florez, C. (2000). Las transformaciones sociodemográficas en Colombia. Bogotá: Banco de la república de Colombia.
- Heer, D. (1966). Economic development and fertility. *Demography*, 3(2), 423-444.
- Laine, M. (2019). Introduction to Dynamic Linear Models for Time Series Analysis
- Mesa, S y Junca, R. (2011). Análisis de reducción de la fecundidad en Colombia: modelo de determinantes próximos. *Cuadernos de economía*. 30,54. p. 127-150
- Peña, D. (2007). Análisis de series temporales. España: Alianza editorial
- Petris, G. (2009). *Dynamic Linear Models with R*. London: Springer.
- Petris, G. (2010). An R Package for Dynamic Linear Models. *Journal of Statistical Software*. Volume 36, Issue 12.

## 10. ANEXOS

### Estimación Natalidad Modelo ARIMAX

MES	Natalidad	Estimacion	MES	Natalidad	Estimacion
2011M01	48420	47900,47215	2014M01	50848	50420,65345
2011M02	43533	43070,1985	2014M02	46879	44861,25822
2011M03	51717	51158,69336	2014M03	51174	50717,95071
2011M04	49600	49066,5812	2014M04	50297	50574,67239
2011M05	52073	51510,68244	2014M05	52531	50993,28179
2011M06	50939	50389,65871	2014M06	49880	50051,90186
2011M07	52199	51114,55272	2014M07	53039	53059,18848
2011M08	55317	51969,00858	2014M08	53799	54665,29962
2011M09	58296	58731,89784	2014M09	57476	56788,20915
2011M10	53721	53295,19947	2014M10	54547	54739,53143
2011M11	52681	52862,67618	2014M11	51664	52544,70657
2011M12	53539	52123,70626	2014M12	53102	51659,49641
2012M01	51760	49989,9426	2015M01	50864	51436,57714
2012M02	47373	46582,16161	2015M02	45400	46611,00564
2012M03	50706	53528,22563	2015M03	51404	50166,08046
2012M04	48697	49059,07735	2015M04	49071	50097,44882
2012M05	51217	51173,17337	2015M05	51118	50700,93013
2012M06	51153	51284,33882	2015M06	50678	49732,75714
2012M07	54047	52955,42356	2015M07	52743	53164,36025
2012M08	55972	57008,72274	2015M08	51968	52465,93221
2012M09	58437	56941,48933	2015M09	55261	56625,58874
2012M10	55383	53215,63848	2015M10	53384	52194,35597
2012M11	53417	53549,34669	2015M11	51559	51089,8753
2012M12	53542	53442,62447	2015M12	53478	53372,03561
2013M01	51747	52181,28746	2016M01	51185	50805,28028
2013M02	45937	46617,63776	2016M02	47397	44984,7833
2013M03	49640	50364,82642	2016M03	51133	51901,74389
2013M04	49996	48940,43955	2016M04	49543	49775,62621
2013M05	51186	51787,72398	2016M05	50523	51902,52917
2013M06	48681	50944,03389	2016M06	48999	50211,31738
2013M07	52005	51824,27586	2016M07	50154	51746,97077
2013M08	52769	53567,07759	2016M08	53057	51887,95333
2013M09	55818	55267,28624	2016M09	54426	55488,45543
2013M10	53934	54293,3093	2016M10	50206	52602,38242
2013M11	51470	51912,00846	2016M11	48763	48626,95927
2013M12	51696	50621,49651	2016M12	50331	50796,22184

### Estimación Natalidad Modelo Lineal Dinamico

Año	Natalidad	Estimacion	Año	Natalidad	Estimacion
2010m12		43533	2013m12	51696	51846.3524388348
2011m01	48420	48419.9954968451	2014m01	50848	51127.9064819143
2011m02	43533	44803.305285673	2014m02	46879	48070.2591376282
2011m03	51717	49789.6594840805	2014m03	51174	50303.8091217426
2011m04	49600	49653.1507293436	2014m04	50297	50298.9090626093
2011m05	52073	51394.5745840481	2014m05	52531	51905.19222181
2011m06	50939	51066.7283667471	2014m06	49880	50447.7999150967
2011m07	52199	51881.546876016	2014m07	53039	52312.5093625376
2011m08	55317	54353.8075001573	2014m08	53799	53382.234989423
2011m09	58296	57190.7337741113	2014m09	57476	56328.237629397
2011m10	53721	54693.8037274307	2014m10	54547	55046.4027548826
2011m11	52681	53245.32657267	2014m11	51664	52612.3188688769
2011m12	53539	53456.6632492915	2014m12	53102	52964.7088818129
2012m01	51760	52235.6907707429	2015m01	50864	51452.972400689
2012m02	47373	48736.3448603222	2015m02	45400	47097.0622236156
2012m03	50706	50153.7709280953	2015m03	51404	50196.4707208424
2012m04	48697	49105.4325430346	2015m04	49071	49386.5464320157
2012m05	51217	50624.9831979039	2015m05	51118	50632.5551230605
2012m06	51153	51004.9607708488	2015m06	50678	50665.2586922911
2012m07	54047	53194.108310297	2015m07	52743	52160.4669814524
2012m08	55972	55193.1669348841	2015m08	51968	52021.9616607998
2012m09	58437	57527.5316941646	2015m09	55261	54352.8759838268
2012m10	55383	55984.258932735	2015m10	53384	53655.6422152088
2012m11	53417	54136.7782947909	2015m11	51559	52146.8322358555
2012m12	53542	53708.757042441	2015m12	53478	53104.7826158323
2013m01	51747	52297.0146949719	2016m01	51185	51723.2463918904
2013m02	45937	47720.1471822419	2016m02	47397	48609.9428049093
2013m03	49640	49101.7339257076	2016m03	51133	50425.6144759354
2013m04	49996	49745.2760462001	2016m04	49543	49790.4572137411
2013m05	51186	50782.0665215699	2016m05	50523	50317.618146681
2013m06	48681	49270.0726715776	2016m06	48999	49368.6988679234
2013m07	52005	51238.2127843717	2016m07	50154	49933.8263733603
2013m08	52769	52339.8156284108	2016m08	53057	52181.3607886004
2013m09	55818	54842.8269919218	2016m09	54426	53796.674082228
2013m10	53934	54188.8063750658	2016m10	50206	51212.7115689427
2013m11	51470	52232.2674096323	2016m11	48763	49449.8217277735
			2016m12	50331	50083.9454524035