



Los Libertadores
Fundación Universitaria

**MODELAMIENTO ESTADÍSTICO DEL TRÁFICO AÉREO ILEGAL EN
COLOMBIA**

**HAROLD RANDOLPH DEVIA DÍAZ
IVAN MAURICIO FERREIRA CARDENAS**

**FUNDACIÓN UNIVERSITARIA LOS LIBERTADORES
DEPARTAMENTO DE CIENCIAS BÁSICAS
ESPECIALIZACIÓN EN ESTADÍSTICA APLICADA
BOGOTÁ, D.C.
2015**



Los Libertadores
Fundación Universitaria

**MODELAMIENTO ESTADÍSTICO DEL TRÁFICO AÉREO ILEGAL EN
COLOMBIA**

**HAROLD RANDOLPH DEVIA DÍAZ
IVAN MAURICIO FERREIRA CARDENAS**

**Trabajo de Grado para optar por el título de Especialista en Estadística
Aplicada**

**Director:
DIEGO FERNANDO LEMUS POLANIA
Magister en Ciencias - Estadística**

**FUNDACIÓN UNIVERSITARIA LOS LIBERTADORES
DEPARTAMENTO DE CIENCIAS BÁSICAS
ESPECIALIZACIÓN EN ESTADÍSTICA APLICADA
BOGOTÁ, D.C.
2015**

Nota de aceptación

Firma del presidente del jurado

Firma del jurado

Firma del jurado

Bogotá 30-05-2015

DEDICATORIA

A Dios todo poderoso, por iluminar nuestros caminos y guiarnos correctamente para alcanzar los objetivos propuestos. A nuestros seres queridos por brindarnos su apoyo incondicional, sus buenos ejemplos y motivación para ser mejores personas cada día. Así mismo a los docentes y amigos que participaron con su conocimiento y sabiduría.

AGRADECIMIENTOS

Agradecemos el apoyo de los docentes que compartieron sus conocimientos y en especial al señor Diego Lemus que incondicionalmente asesoró en el desarrollo del trabajo de grado. Así mismo agradecemos a nuestras familias por brindarnos el apoyo y la confianza para mejorar a nivel profesional.

Salvedad:

Las directivas de la Universidad, los jurados calificadores y el cuerpo docente no son responsables por los criterios e ideas expuestas en el presente documento. Estos corresponden a los autores únicamente.

CONTENIDO

	pág.
LISTA DE TABLAS	vii
LISTA DE FIGURAS	viii
GLOSARIO.....	1
RESUMEN	2
ABSTRACT.....	2
1. INTRODUCCIÓN.....	3
1.1. Objetivo.....	3
2. MARCO DE REFERENCIA.....	4
3. METODOLOGÍA DE ADQUISICIÓN DATOS	7
4. MARCO TEÓRICO	10
4.1. Modelo ARIMA (p,d,q)	10
4.1.1 Metodología Box Jenkins para identificación de procesos ARIMA (p,d,q)	11
4.1.2 Pruebas de raíz unitaria	12
4.2. Modelo ARFIMA (p,d,q)	15
4.2.1 Pruebas de Raíz Fraccional.....	15
4.2.2 Prueba Recursiva Cusum – Nyblom – Hansen	18
5. IDENTIFICACIÓN DEL MODELO DE PRONÓSTICO.....	19
5.1 Pruebas de raíz unitaria	22
5.2 Pruebas de raíz fraccional	25
5.3 Prueba Cusum - Nyblom Hansen para residuales.....	29
5.4 Pruebas de incorrelación	29
5.5 Pruebas de normalidad.....	30

5.6 Prueba de bondad de ajuste a una t- student con 16 grados de libertad	31
6. RESULTADOS	33
6.1. Modelo identificado.....	33
6.2. Validación cruzada	34
6.3. Pronósticos.....	37
7. CONCLUSIONES	39
8. REFERENCIAS.....	40
ANEXO 1. DATOS ORIGINALES E IMPUTADOS	42
ANEXO 2. EL PROCESO DE IDENTIFICACIÓN DE UN MODELO ARIMA CUANDO EXISTEN OBSERVACIONES FALTANTES (Castaño, E. 1997).....	44

LISTA DE TABLAS

	pág.
Tabla 1. Características del parámetro d	15
Tabla 2. Hipótesis para las pruebas consideradas en un modelo ARFIMA.....	16
Tabla 3. Pruebas de raíz unitaria.....	25
Tabla 4. Parámetros del modelo ARFIMA (1,d,1)	33

LISTA DE FIGURAS

	pág.
Figura 1. Trazas de vuelos ilegales en Colombia en el año 2003	5
Figura 2. Trazas de vuelos ilegales en Colombia y Caribe en el 2013	6
Figura 3. Flujo de identificación y clasificación de trazas – FAC	8
Figura 4. Flujo de identificación y clasificación de trazas – FAC (Continuación)	9
Figura 5. Proceso iterativo de Box-Jenkins para construir modelos (Guerrero, 1991, p. 102).....	12
Figura 6. Serie de datos original de vuelos ilegales.	19
Figura 7. Transformación Box - Cox	20
Figura 8. Serie original Vs. Serie estabilizada.....	21
Figura 9. Número de vuelos ilícitos con imputación de datos y su logaritmo	21
Figura 10. ACF y PACF de la serie con datos imputados	22
Figura 11. ACF y PACF de la serie Transformada y Diferenciada	27
Figura 12. Muestra extendida de la ACF para la serie fraccionalmente diferenciada	28
Figura 13. Estabilidad estructural de los residuales.....	29
Figura 14. ACF y PACF de los residuales estandarizados	30
Figura 15. Q-Q plot y densidad de los residuales comparados con la distribución Normal	31
Figura 16. Q-Q plot y densidad de residuales con distribución T-student con 16 gl	32
Figura 17. ACF y PACF de los residuales estandarizados al cuadrado	33
Figura 18. Logaritmo de las estimaciones intramuestra con bandas de confianza del 95%	34
Figura 19. Validación método exacto con 95% de confianza	35
Figura 20. Validación Método Bootstrap con 95% de confianza	36
Figura 21. Pronóstico Método Exacto con 95% de confianza	37
Figura 22. Pronóstico Método Bootstrap con 95% de confianza	38

GLOSARIO

AERONAVE: es cualquier vehículo capaz de navegar por el aire, o toda máquina que puede desplazarse en la atmósfera por reacciones del aire que no sean las reacciones del mismo contra la superficie de la tierra.

FUERZAS PÚBLICA: cada una de las Fuerzas y la policía Nacional ejercen control y adelantan operaciones en contra del tráfico de la cocaína tanto en territorio nacional como a nivel internacional con los diferentes convenios que se tienen con los otros países cooperantes.

ORGANIZACIÓN DE AVIACIÓN CIVIL INTERNACIONAL (OACI): es la entidad encargada del desarrollo de la aviación civil internacional y la investigación de los medios aptos para lograr este fin, sean ellos jurídicos, económicos o técnicos.

MEDIO AÉREO: aeronave o máquina que puede sustentarse y desplazarse en el aire, que cuente con propulsión propia o por aprovechamiento de corrientes de aire y que sea apta para transportar cualquier tipo de carga o peso.

RUTA AÉREA: es por donde circulan las aeronaves, comprendiendo una ruta nominal y un área de protección. Es el camino virtual predefinido (tanto en altura como en trazado) que sigue una aeronave que sale desde un punto A hasta un punto B. Es la sucesión de puntos de escala regular o auxiliar, comprendidos en la ruta aérea.

VUELO ILÍCITO EN COLOMBIA: toda aeronave que vuele en Colombia debe cumplir con los requisitos establecidos en el reglamento aeronáutico Colombiano, establecidos en coordinación con la Fuerza Aérea Colombiana, por ejemplo el no cumplir con una ruta aérea establecida o tener activado el transponder (Equipo abordo de una aeronave que permite su identificación.)

RESUMEN

Este trabajo propone un modelo de pronóstico estadístico basado en la teoría de series de tiempo, cuya variable de estudio es la cantidad de vuelos ilegales en Colombia entre el año 2003 y principios del año 2014. Durante la construcción de dicho modelo, se aplican diferentes pruebas de raíces unitarias y fraccionarias para concluir la especificación de un Modelo ARFIMA (p,d,q), al que finalmente se le realiza una validación de datos para la calidad del pronóstico.

Palabras clave: series de tiempo, modelo ARFIMA, pronóstico estadístico, tráfico aéreo ilegal, vuelos ilegales.

ABSTRACT

This project proposes a statistical forecasting model by using Time Series Analysis whose variable is the number of illicit flights in Colombia between 2003 and beginnings of 2014. During the making of this model, different unit roots and fractional roots tests were applied to conclude an ARFIMA (p,d,q) model. The quality of the forecast is guaranteed by a final data validation applied to the model.

Keywords: time series, ARFIMA, statistical forecasting, illegal air traffic, illegal flights.

1. INTRODUCCIÓN

El narcotráfico como crimen transnacional se considera una problemática global, que radica en la producción, distribución, comercialización, consumo y una generación de fenómenos asociados, relacionados principalmente con el tráfico de estupefacientes. Tal es el caso de altos índices de criminalidad, tráfico de armas, terrorismo, grupos delincuenciales, trata de personas, entre otros delitos que principalmente son financiados por recursos provenientes del narcotráfico y que fortalecen los espirales de violencia que permean la sociedad en general. El narcotráfico al ser un negocio que ofrece tanta rentabilidad en tan corto plazo incentiva el aumento de personas y/u organizaciones ilegales interesadas en promover e incrementar este delito.

Los delincuentes dedicados al negocio del narcotráfico han diseñado diferentes estrategias para lograr movilizar sus mercancías a los puntos estratégicos de distribución y comercialización, logrando así continuar con la cadena de esta estructura criminal; el medio más usado en la década de los 80's (inicios del auge del narcotráfico) fue el terrestre, pero al ser el más controlado por la fuerza pública en esa época, tener la limitante de cobertura internacional y el tiempo excesivo de movilización, los narcotraficantes incursionaron en los medios marítimo y aéreo para efectuar sus movimientos ilegales, siendo este último el más rápido y de mayor cobertura. En consecuencia nace la preocupación de realizar un estudio de los movimientos aéreos ilegales concentrados en la región Caribe.

Ante esta situación y con el interés de aplicar algún método estadístico que estudie el transporte aéreo ilegal en Colombia y el Caribe se empleó la teoría en series de tiempo sobre los datos históricos extraídos del sistema de Defensa Aérea de la Fuerza Aérea Colombiana. Específicamente, se propone como **objetivo de investigación** identificar un modelo de pronóstico que sirva como herramienta fundamental y estratégica en el proceso de toma de decisiones para la construcción de una política pública de regulación y control del espacio aéreo colombiano, y así contribuir efectivamente en la lucha contra el narcotráfico.

Este trabajo de grado, de manera general, se estructura en ocho partes. La primera consta de la *Introducción* para contextualizar al lector en la problemática de los vuelos ilícitos y exponer la pregunta de investigación; en segundo lugar se encuentra el *Marco de referencia*, el cual hace un análisis histórico del comportamiento de los vuelos ilegales; en tercer lugar la *Metodología* para la adquisición de los datos empleados; posterior, el *Marco teórico* que enmarca el sustento del tratamiento estadístico aplicado; y las últimas partes son la *Identificación del modelo*, los *Resultados*, las *Conclusiones* y las *Referencias*.

2. MARCO DE REFERENCIA

El tráfico de estupefacientes es toda tenencia, transporte y comercio de drogas ilegales, que varían de acuerdo a las normatividades de cada país, para el caso colombiano se reglamenta con la Ley 30 de 1986 con la cual se adoptó el Estatuto Nacional de Estupefacientes (Rocha, 2011, p. 143) y la Ley 599 de 2000. Para combatir este flagelo, el Estado colombiano ataca cada uno de los eslabones de la cadena del narcotráfico, iniciando por la siembra, producción, distribución, comercialización, consumo y lavado de activos, con el ánimo de neutralizar el negocio y detener este sistema de enriquecimiento ilícito.

La Fuerza Aérea Colombiana (FAC) tiene como deber misional dentro de esta lucha contener y controlar el eslabón de la distribución de los estupefacientes, evitando todos los movimientos ilícitos que emplean el espacio aéreo nacional para movilizar dicha mercancía, siendo este escenario motivo de análisis del presente proyecto.

Según García (2010), el tráfico aéreo ilegal de cocaína comenzó entre los años 80s y 90s, desde entonces los países de América Latina ganaron la reputación de principales productores y exportadores de estupefacientes. Las organizaciones al margen de la ley relacionadas con el narcotráfico han empleado entre otros, el medio aéreo para llevar a cabo sus acciones ilícitas; una de las razones por la cual los delincuentes utilizan el medio aéreo es el corto tiempo en el cual se logra movilizar la mercancía y así satisfacer rápidamente la demanda del producto en el mercado internacional.

Emplear el espacio aéreo para acciones ilegales no es tan fácil, ya que además de una supervisión directa de la Fuerza Aérea existen entes de control civil que contribuyen en el control de los tránsitos aéreos nacionales e internacionales, con el ánimo de garantizar seguridad en el medio. La Organización de Aviación Civil Internacional (OACI), es la encargada de emitir y promover los reglamentos y normas que rigen los temas aeronáuticos, de los cuales se desprenden los reglamentos internos de cada Estado que permitan determinar la legalidad de los vuelos, ya que cada piloto al mando de una aeronave debe cumplir con una serie de requisitos previos al vuelo, durante el vuelo y al llegar a su destino, trámites correspondientes que deben adelantarse ante las autoridades aeronáuticas competentes que validan su legalidad; de no ser así se obtiene el primer indicio de que la aeronave se encuentra haciendo uso ilegal del espacio aéreo y está siendo empleada para el tráfico ilegal de cualquier tipo de carga.

En Colombia la entidad responsable de garantizar el cumplimiento del convenio de la Aviación Civil Internacional¹ es la Aeronáutica Civil de Colombia, institución que da a conocer a todo el personal del medio aeronáutico el Reglamento Aeronáutico Colombiano (RAC), en el cual se establecen las normas y procedimientos para desempeñar acciones relacionadas con la aviación en Colombia, normas y procedimientos que garantizan legalidad y seguridad en el espacio aéreo colombiano.

¹ Convenio de Chicago 1944, cuya función es establecer y actualizar las normas sobre aviación civil internacional.

La FAC ejerce control del espacio aéreo por medio del Sistema de Defensa Aérea cuya función principal es la vigilancia de la totalidad del territorio Nacional. Este sistema garantiza el buen uso de la zona aérea, correlacionando los datos e información de cada uno de los vuelos inusuales para verificar su legalidad y de ser necesario, lanzando operaciones de interdicción aérea cada vez que se presentan vuelos ilegales.

En la Figura 1, se observa el registro de vuelos ilegales despegando y aterrizando en territorio colombiano en el año 2003. Se registró un promedio de casi dos vuelos diarios (para un total de 639) en los cuales traficaban estupefacientes movilizándolos a los países fronterizos, convirtiendo el fenómeno en un crimen transnacional.



Figura 1. Trazas de vuelos ilegales en Colombia en el año 2003
Fuente: Dirección Defensa Aérea de la FAC.

Desde entonces la FAC desarrollo una estrategia de interdicción aérea que redujo considerablemente el número de vuelos ilegales en el territorio nacional disminuyendo la frecuencia a 11 en el año 2013, de los cuales se lograron neutralizar varios de ellos en operaciones militares. Dado el accionar militar y el aumento del control y efectividad de la Fuerza Pública, los narcotraficantes diseñaron nuevas rutas para continuar con su accionar delictivo, trazando una ruta directa a Centroamérica que les permitió transportar los estupefacientes con éxito.

Las nuevas rutas movilizan las drogas ilícitas desde territorios aledaños a las fronteras colombianas hacia el Caribe, afectando el territorio Nacional de los países fronterizos, sin embargo las aeronaves que realizan esa actividad logran sobrevolar parte del espacio aéreo colombiano, dado

que al cruzan el Mar Caribe, atravesando parte del mar patrimonial². Dado que por normas internacionales³ no se puede hacer uso de las armas en aguas patrimoniales, el Estado colombiano soporta su lucha contra el Narcotráfico con la implementación de convenios para combatir este delito de manera conjunta con los demás países de la región que se ven afectados por este delito.

En la Figura 2, se muestra la transformación del comportamiento del tráfico ilícito aéreo de drogas para el año 2013. Al comparar la Figura 1 y la Figura 2 se observa una disminución en los vuelos provenientes o con destino Colombia y un aumento considerable de los vuelos partiendo de las aéreas cercanas a las fronteras Nacionales hasta llegar a los países de Centroamérica.



Figura 2. Trazas de vuelos ilegales en Colombia y Caribe en el 2013
Fuente: Dirección Defensa Aérea de la FAC.

Actualmente Colombia mantiene convenios y procedimientos para operar con algunos de los países que conforman la región Caribe a nivel de las Fuerza Aéreas⁴, permitiendo de esta forma compartir información en tiempo real para el desarrollo de operaciones de interdicción en contra de las aeronaves que sobrevuelan el espacio aéreo, con el fin de ser neutralizadas, así mismo controlando el flagelo desde su origen hasta su destino y cerrando el corredor por el cual

² Es una franja marítima que se extiende desde el límite exterior del mar territorial.

³ Derecho internacional Humanitario, Convenios de Ginebra 1949.

⁴ Procedimientos operacionales Vigentes. Fuerza Aérea Colombiana en convenios con EE.UU, Honduras, República Dominicana, Guatemala, Perú, Brasil, Venezuela, Ecuador y Panamá.

movilizan los estupefacientes hacia Centroamérica.

Para lograr determinar un modelo estadístico que explique la cantidad de vuelos ilegales esperados semanalmente, es necesario abordar el tema de series de tiempo, con el fin de lograr pronosticar la cantidad de vuelos ilegales y contribuir a la toma de decisiones que ejecuten operaciones de la Fuerza Pública. Tema que se abordará desde la perspectiva estadística para su correspondiente análisis.

3. METODOLOGÍA DE ADQUISICIÓN DATOS

Los datos necesarios para la realización de este trabajo se obtuvieron por medio de la vigilancia del espacio aéreo colombiano en el Centro de Comando y Control de la Fuerza Aérea Colombiana (CCOFA), donde se aplican unos procedimientos de análisis de información para clasificar los vuelos. Es importante aclarar que para determinar la legalidad de un vuelo éste debe cumplir con ciertas características fundamentales, las cuales están sujetas a una verificación por el personal del CCOFA. Para establecer la ilegalidad de un vuelo se deben considerar ciertas características particulares, que se estipulan en la Reglamentación Interna COFAC 04-20-444 de normas y procedimientos para el empleo de aeronaves de la FAC contra aeronaves que violen el espacio aéreo Nacional.

Se clasifica una traza sospechosa (TS) que por sus características de vuelo, se encuentra fuera de las normas o reglamentación de aeronavegación o de Seguridad Nacional, contemplando una o varias de las siguientes condiciones a saber:

1. Una traza que este volando a bajo nivel (menor a 3000 pies) en un radio menor o igual a 60 millas náuticas de una cabeza radar o una Base Aérea, sin identificar.
2. Una aeronave que vuela el Espacio Aéreo Nacional sin la debida autorización de plan de vuelo antela la Autoridad ATS competente.
3. Una aeronave que ingresa a un Área Restringida sin autorización.
4. Una aeronave que aterrice y/o despegue desde un aeródromo no autorizado.
5. Una aeronave sobre la cual, a pesar de agotar los recursos de correlación con la agencia de control de tránsito aéreo, no se logra su identificación.
6. Que incumpla sin causa justificada los procedimientos especiales dictados para las áreas restringidas, entrando o saliendo de ellas, aun mediando autorización de sobrevuelo.
7. La aeronave clasificada así, de acuerdo a información suministrada por la Jefatura de Inteligencia de la FAC.

A continuación se muestra gráficamente el procedimiento de detección, identificación y validación de las trazas empleado por la Fuerza Aérea Colombiana en su función del control del espacio aéreo

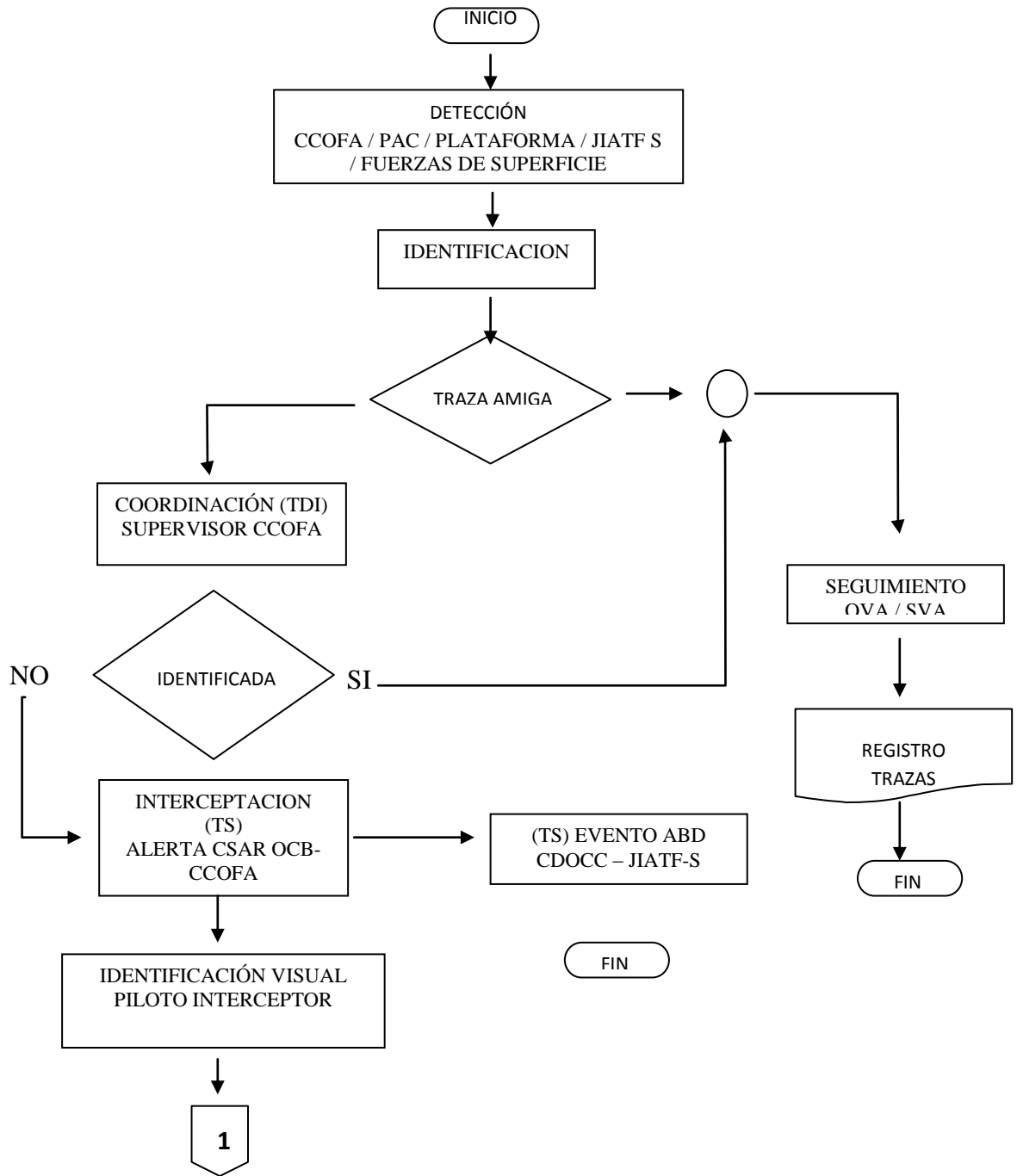


Figura 3. Flujo de identificación y clasificación de trazas – FAC

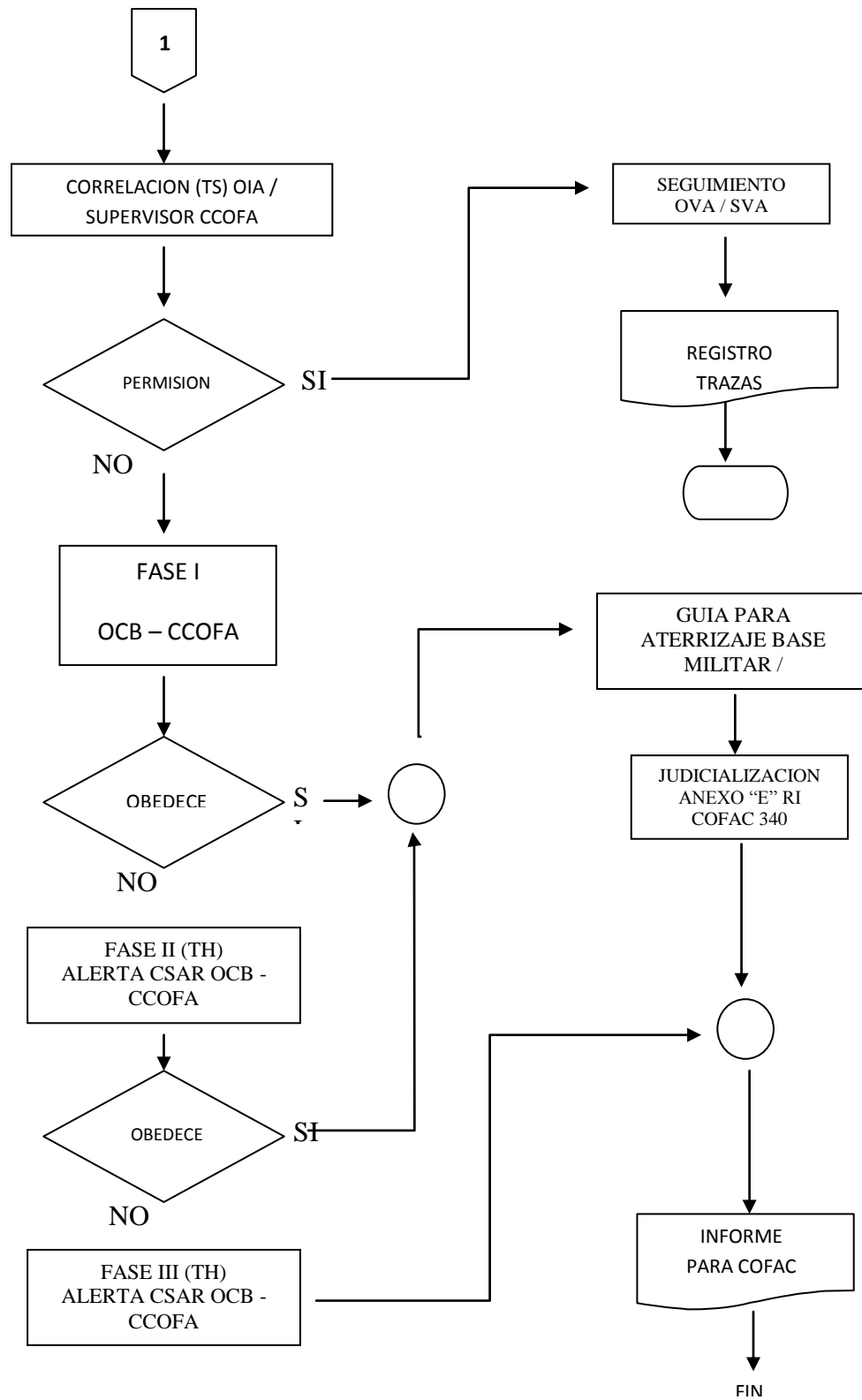


Figura 4. Flujo de identificación y clasificación de trazas – FAC (Continuación)

Simultáneamente se relaciona la información con los centros de control civil, de Fuerza Pública y los sistemas de Defensa Aérea de otros países dependiendo de la ubicación del vuelo, para correlacionar la información del comportamiento de la traza. Una vez se han agotado todos los medios para confirmar la legalidad del tránsito, sin obtener resultados que indiquen la normalidad del vuelo, se clasifica la traza como sospechosa y se inician los procedimientos de interceptación para el control del vuelo ilegal.

El CCOFA, como procedimiento estándar, efectúa un reporte diario del comportamiento del tránsito en el espacio aéreo nacional, para evidenciar y analizar novedades que permitan mitigar el número de vuelos ilegales a mediano y largo plazo. Con el protocolo anteriormente descrito, se agruparon los reportes por semana desde inicio del año 2003 hasta la tercera semana de febrero del año 2014, obteniendo un total de 535 datos.

Cabe resaltar que toda la información que se evidencie en el trabajo es de clasificación secreta según el protocolo del Ministerio de Defensa, debido a que es de interés nacional y tiene un impacto a nivel estratégico en la defensa de la Nación. Por lo anterior esta información debe tener un manejo específico de confidencialidad.

4. MARCO TEÓRICO

A continuación se presenta un resumen de la teoría estadística considerada para el desarrollo del presente trabajo de grado.

4.1. Modelo ARIMA (p,d,q)

El proceso autoregresivo y de media móvil integrado, usualmente conocido como modelo ARIMA (p,d,q), es definido por medio de la siguiente ecuación:

$$\phi(B)\nabla^d Y_t = \theta_0 + \theta(B)a_t \quad \text{Ecuación (1)}$$

Donde

- $\phi(B)$ es el polinomio autorregresivo estacionario de orden p . Lo anterior implica que las raíces de la ecuación $\phi(B) = 0$ se encuentran fuera del círculo unitario.
- $\theta(B)$ es el polinomio de media móvil invertible de orden q . Lo anterior implica que las raíces de la ecuación $\theta(B) = 0$ se encuentran fuera del círculo unitario.
- $\phi(B) = 0$ y $\theta(B) = 0$ no tienen raíces comunes entre ellos.
- $\nabla^d = (1 - B)^d$ es el operador de diferenciación entera. d el parámetro de diferenciación del proceso y toma valores en los enteros no negativos.
- θ_0 es la parte determinística del proceso.
- a_t es una variable aleatoria, independiente e idénticamente distribuida, con media cero y varianza finita.

Cuando $d=0$ la Ecuación (1) representa un proceso ARMA (p,q) estacionario e invertible. Las condiciones de estacionariedad e invertibilidad se aplican independientemente y en general los operadores $\phi(B)$ y $\theta(B)$ no serán del mismo orden (George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, 2008).

4.1.1 Metodología Box Jenkins para identificación de procesos ARIMA (p,d,q)

La metodología de Box y Jenkins explicada por Guerrero (1991) para la construcción de series de tiempo univariadas es especificada en cuatro etapas:

1. Identificación de un posible modelo dentro de la clase de modelos ARIMA; cuyo objetivo es determinar los órdenes de los polinomios autorregresivos (AR) y medias móviles (MA), así como el número de veces que deberá aplicarse el operador diferencia (pruebas de raíz unitaria) para cancelar la no-estacionariedad homogénea.
2. Estimación de los parámetros involucrados en el modelo, a través de técnicas de estimación no-lineal. Esta etapa presupone que ya está identificado un modelo teniendo que encontrar los mejores valores de los parámetros para que dicho modelo represente apropiadamente a la serie de tiempo.
3. Verificación de que el modelo proporciona un ajuste adecuado y de que los supuestos básicos implícitos en el modelo, se satisfacen; de no cumplirse los supuestos, se determinan las modificaciones necesarias y de hecho, se repiten las etapas anteriores hasta que la verificación indique los resultados aceptables.
4. Uso del modelo para los fines que el investigador haya tenido en mente al construir; dichos fines son por lo general de pronóstico, control, simulación o explicación del fenómeno en estudio.

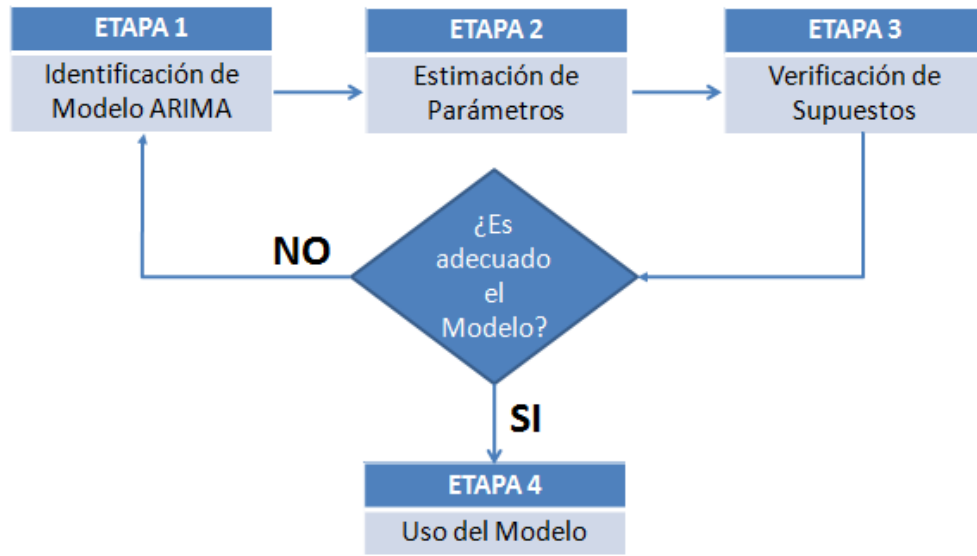


Figura 5. Proceso iterativo de Box-Jenkins para construir modelos (Guerrero, 1991, p. 102)

4.1.2 Pruebas de raíz unitaria

Las pruebas de raíz unitaria permiten evaluar el grado de no estacionariedad homogénea de una serie de tiempo. En el presente trabajo se consideran las pruebas de raíz unitaria de Dickey Fuller (DF), Dickey Fuller Aumentada (DFA), Phillips-Perron (PP).

4.1.2.1 Prueba de Dickey Fuller (DF)

La prueba DF evalúa la estacionariedad de una serie de tiempo Y_t con base en el estadístico τ (Tau) construido a partir de la hipótesis nula de existencia de una raíz unitaria y por lo tanto la NO estacionariedad; mientras que la de la hipótesis alterna considera la estacionariedad de la serie (Gujarati y Porter, 2010).

Esta prueba considera tres modelos en los cuales la serie de tiempo podría comportarse según la presencia de deriva, tendencia determinista y estocástica.

Y_t es una caminata aleatoria $\Delta Y_t = \delta Y_{t-1} + a_t$ *Ecuación (2)*

Y_t es una caminata aleatoria con deriva $\Delta Y_t = \beta_1 + \delta Y_{t-1} + a_t$ *Ecuación (3)*

Y_t es una caminata aleatoria con deriva, alrededor de una tendencia determinista $\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + a_t$ *Ecuación (4)*

El estadístico de prueba es $\tau = \hat{\delta}/se(\hat{\delta})$ donde $\hat{\delta}$ es el estimador de mínimos cuadrados de δ y $se(\hat{\delta})$ es el error estándar en las Ecuaciones (2), (3) y (4), donde a_t es un ruido blanco Gaussiano de media cero y varianza constante $-RBG(0, \sigma_a^2) -$.

El contraste de hipótesis de la prueba es el siguiente

- $H_0: \delta = 0$. La serie de tiempo posee una raíz unitaria y es NO Estacionaria homogénea.
- $H_a: \delta < 0$. La serie de tiempo NO posee raíz unitaria y el proceso es estacionario.

La distribución de τ bajo H_0 no es la t -Student tradicional, pertenece a una clase de distribuciones no estándar. Si $\tau(\alpha, n)$ es el percentil α -inferior de dicha distribución se debe rechazar H_0 si el valor observado de τ es $\tau < \tau(\alpha, n)$ (Castaño, 2011).

Debido a que los modelos de la prueba DF no consideran la posible autocorrelación de las variables a_t , la eficiencia de los estimadores puede ser sesgada obteniendo un modelo inadecuado. Por lo tanto, Dickey-Fuller plantearon una extensión de su prueba inicial considerando el término de error como un modelo ARMA.

4.1.2.2 Prueba de Dickey Fuller Aumentada (DFA)

En Gujarati y Porter (2010) se especifica la diferencia entre las pruebas DF y DFA, la cual radica en la suposición de que el término a_t no estaba correlacionado. En cambio la DFA sí considera la correlación de estos términos, lo cual implica aumentar la sumatoria de valores rezagados del proceso Y_t en las ecuaciones de la prueba DF.

Y_t es una caminata aleatoria
$$\Delta Y_t = \delta Y_{t-1} + \sum_{i=1}^m \alpha_i \Delta Y_{t-i} + a_t \quad \text{Ecuación (5)}$$

Y_t es una caminata aleatoria con deriva
$$\Delta Y_t = \beta_1 + \delta Y_{t-1} + \sum_{i=1}^m \alpha_i \Delta Y_{t-i} + a_t \quad \text{Ecuación (6)}$$

Y_t es una caminata aleatoria con deriva, alrededor de una tendencia determinista
$$\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \sum_{i=1}^m \alpha_i \Delta Y_{t-i} + a_t \quad \text{Ecuación (7)}$$

Donde a_t es un ruido blanco Gaussiano de media cero y varianza constante $-RBG(0, \sigma_a^2) -$ y $\Delta Y_{t-1} = Y_{t-1} - Y_{t-2}$, $\Delta Y_{t-2} = Y_{t-2} - Y_{t-3}, \dots$ hasta lograr determinar el número de rezagos, logrando que el a_t no este serialmente relacionado y se logre obtener una estimación insesgada de δ . Los estadísticos y criterios de rechazo en la prueba DFA son los mismos que se aplican en la prueba DF.

4.1.2.3 Prueba de Raíz Unitaria Phillips-Perron (PP)

La prueba PP es un variante de la prueba DF, donde se permite que los a_t no sean exactamente ruido blanco. Para el desarrollo de la prueba es necesaria una estimación del espectro a_t (estimación de las autocorrelaciones del error). La prueba PP utiliza la DF, donde a_t puede estar autocorrelacionada. En la prueba Phillips – Perron el estadístico no es un error estándar, como se refleja en la DFA, es decir, el error estándar es ajustado por el efecto de la autocorrelación. (Montenegro A., p. 171-177). Finalmente la prueba PP está bajo las mismas hipótesis de la prueba DF así como su estimador y los criterios de rechazo.

- $H_0: \delta = 0$. La serie de tiempo posee una raíz unitaria y es NO Estacionaria homogénea.
- $H_a: \delta < 0$. La serie de tiempo NO posee raíz unitaria y el proceso es estacionario.

4.1.2.4 Prueba de Zivot-Andrews

La prueba de Zivot y Andrews es formulada contemplando la poca confianza de las pruebas DF, DFA y PP porque no detectan los posibles cambios estructurales de una serie y en consecuencia sesgan su resultado; la prueba consiste en la detección de raíces unitarias independientes de la existencia de una ruptura o cambio estructural, siendo estas en tendencia, intercepto o en ambos casos. De acuerdo a Rangel. A (2007) la prueba de Zivot Andrews denomina tres modelos A, B y C.

Modelo A: las variables explicatorias incluyen una posible tendencia determinística lineal, la variable endógena rezagada un periodo, una variable indicadora para el cambio estructural y finalmente j términos rezagados de las primeras diferencias de la variable endógena, teniendo

$$Y_t = \beta_0 + \beta_1 t + \tau Y_{t-1} + \theta DU_t + \sum_{j=1}^k c_j \Delta Y_{t-j} + a_t \quad \text{Ecuación (8)}$$

Donde la variable DU_t es la variable dicotómica que captura cambio estructural en intercepto y T_B es el periodo en que sucede la ruptura, quedando definida así: $DU_t: \begin{cases} 1 & t > T_B \\ 0 & \text{en otro caso} \end{cases}$

Modelo B: permite la modelación de un cambio estructural en tendencia mediante la inclusión de una variable indicadora que defina una nueva tendencia después de la ruptura, definiendo

$$Y_t = \beta_0 + \beta_1 t + \tau Y_{t-1} + \lambda DT_t + \sum_{j=1}^k c_j Y_{t-j} + a_t \quad \text{Ecuación (9)}$$

La variable que capta la ruptura estructural en tendencia es DT_t , la cual se define como $DT_t: \begin{cases} 1 & t > T_B \\ 0 & \text{en otro caso} \end{cases}$, donde T_B es el periodo en que sucede la ruptura.

Modelo C: este modelo tiene en cuenta el cambio tanto en intercepto como en tendencia empleando las variables indicadoras definidas en los modelos A y B de la siguiente manera:

$$Y_t = \beta_0 + \beta_1 t + \theta DU_t + \lambda DT_t + \tau Y_{t-1} + \sum_{j=1}^k c_j \Delta Y_{t-j} + a_t \quad \text{Ecuación (10)}$$

La cual consiste en estimar las anteriores ecuaciones por mínimos cuadrados ordinarios de manera secuencial para los valores de $T_B = 2, \dots, T-1$ y T es el número de observaciones.

4.2. Modelo ARFIMA (p,d,q)

A continuación se realiza una descripción de un proceso autorregresivo y de medias móviles fraccionalmente integrados denominado modelo ARFIMA, con el fin de determinar la capacidad con la que cuenta para la obtención de memoria larga en la serie de tiempo.

Según Castaño (2008), un proceso estocástico $\{Y_t\}$ sigue un proceso ARFIMA (p,d,q) si es una solución a la Ecuación (11).

$$\phi(B)(1-B)^d Y_t = \theta_0 + \theta(B)a_t, \quad t = 1, \dots, T \quad \text{Ecuación (11)}$$

Donde $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ y $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ son, respectivamente, los polinomios autorregresivo y de medias móviles de orden p y q en términos del operador de rezago B de un proceso ARMA, cuyos ceros están fuera del círculo unidad y sin raíces comunes; d y θ_0 son números reales, y a_t son variables aleatorias no observables, independientes e idénticamente distribuidas con media cero y varianza finita σ_a^2 y $(1-B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k$ es el operador de diferencia fraccional. Cabe aclarar que por ser $d \in \mathcal{R}$, el coeficiente $\binom{d}{k}$ se calcula con la función Gamma.

En la Tabla 1 se especifican las propiedades de un modelo ARFIMA (p,d,q) de acuerdo al valor del parámetro de diferenciación fraccional d

Tabla 1. Características del parámetro d

Valor del parámetro d	Características
$d = 0$	Es un proceso ARMA (p,q) estacionario de memoria corta.
$d \in \mathbb{Z}^+$	Es un proceso ARIMA (p,d,q) , con d raíces unitarias.
$d > -\frac{1}{2}$	Es un proceso ARFIMA (p,d,q) invertible.
$d < \frac{1}{2}$	Es un proceso ARFIMA (p,d,q) estacionario.
$d \in \left(-\frac{1}{2}, \frac{1}{2}\right)$	Es un proceso ARFIMA (p,d,q) estacionario e invertible.
$d \in \left(-\frac{1}{2}, 0\right)$	Es un proceso estacionario que exhibe una fuerte reversión a la media y su función de autocovarianza decrece más rápidamente a cero que la de un proceso ARMA (p,q) . En este caso, el proceso es llamado antipersistente (Lemus y Castaño, 2013).
$d \in \left(0, \frac{1}{2}\right)$	Es un proceso estacionario de memoria larga (Lemus y Castaño, 2013).

4.2.1 Pruebas de Raíz Fraccional

Las pruebas de raíz fraccional evalúan la estacionariedad e invertibilidad de una serie de tiempo de acuerdo al valor del parámetro d . En el presente trabajo se consideran las pruebas de Geweke Y

Porter – Hudak – GPH (1983), Robinson (Rob95) y Castaño (2008). Las pruebas mencionadas permiten identificar procesos de memoria larga no asociables a la existencia de raíces unitarias, siendo estas demasiado restrictivas para el grado de persistencia de las innovaciones de la serie.

Para las pruebas del modelo ARFIMA que se desarrollarán a lo largo de la tesis, se consideran las siguientes hipótesis para evaluar los respectivos procesos fraccionalmente integrados, teniendo en cuenta las consideraciones de la Tabla 2.

Tabla 2. Hipótesis para las pruebas consideradas en un modelo ARFIMA

PROCESO	HIPOTESIS	RESULTADO
BILATERAL	$H_0: d = 0$	Es un proceso ARMA (p,q).
	$H_a: d \neq 0$	Es un proceso ARFIMA (p,d,q) estacionario e invertible.
UNILATERAL	$H_0: d = 0$	Es un proceso ARMA (p,q).
	$H_a: d < 0$ Con $d \in \left(-\frac{1}{2}, 0\right)$	Es un proceso ARFIMA (p,d,q) antipersistente.
	$H_0: d = 0$	Es un proceso ARMA (p,q).
	$H_a: d > 0$ Con $d \in \left(0, \frac{1}{2}\right)$	Es un proceso ARFIMA (p,d,q) estacionario e invertible con memoria larga.

4.2.1.1 Prueba de Geweke y Porter - Hudak (GPH)

De acuerdo a (Castaño, E, Gómez, K y Gallón, S, 2008) la prueba de Geweke y Porter-Hudak (1983) consiste en la inferencia sobre el estimador del parámetro de diferenciación fraccional d en la ecuación de regresión de mínimos cuadrados en el dominio de la frecuencia

$$\log I_z(w_j) = c + d \log(4 \sin^2(w_j/2)) + \varepsilon_j \quad \text{Ecuación (12)}$$

donde $I_z(w_j)$ es el periodograma de Z_t en la frecuencia $w_j = 2\pi j/T$ con $j = 1, \dots, m = \lfloor T^{1/2} \rfloor$, donde $\lfloor x \rfloor$ denota el mayor entero que es menor o igual a x .

Geweke y Porter-Hudak (1983) también argumentaron que cuando $-0.5 < d < 0$ y se cumplen algunas condiciones adicionales, se obtiene que el estadístico de prueba converge a una normal estándar, es decir:

$$t_{GPH} = \frac{\hat{d} - d}{Var(\hat{d})^{1/2}} \xrightarrow{d} N(0,1) \quad \text{Ecuación (13)}$$

donde $Var(\hat{d})^{1/2} = \pi^2/6 \left(\sum_{j=1}^m (R_j - \bar{R})^2 \right)^{-1}$ con $R_j = \log(4 \sin^2(w_j/2))$. La prueba se realiza empleando el contraste de hipótesis de interés (Ver Tabla 1), empleando el estadístico de prueba enunciado y el criterio de rechazo adecuado.

4.2.1.2 Prueba de Robinson (ROB95)

Robinson (1995) realizó una modificación de la prueba GPH demostrando que sólo necesita de la densidad espectral del proceso, para verificar la siguiente condición en las bajas frecuencias $f_x(w) \sim C|w|^{-2d}$ cuando $w \rightarrow 0^+$, donde C puede ser cualquier constante positiva. De igual manera comprueba que el modelo de regresión propuesto por Geweke & Porter-Hudak (1983) es asintóticamente equivalente al modelo:

$$\log\{I_x(w_j)\} = \beta_0 + 2d \log(|w_j|) + v_j \quad \text{Ecuación (14)}$$

Donde $\{I_x(w_j)\}$ es la j -ésima ordenada del periodograma de X_t , $w_j = 2\pi j/T$ es la j -ésima frecuencia de Fourier y $v_j = \log\left\{\frac{f_u(w_j)}{f_z(0)}\right\}$ es el término de error en la regresión, el cual se asume independiente e idénticamente distribuido con media cero y varianza constante $\pi/6$, $\forall_j = 1, \dots, m$. De esta manera el estadístico de prueba propuesto es $t_{ROB} = \sqrt{2m}(\hat{d} - d) \xrightarrow{d} N(0, \frac{\pi^2}{6})$.

4.2.1.3 Prueba Castaño

De acuerdo al procedimiento propuesto por Castaño (2008) en el cual prueba el parámetro de diferenciación fraccional de un modelo ARFIMA, basado en la aproximación autorregresiva de su componente a corto plazo, permitiendo probar la existencia de memoria larga en presencia de errores dependientes, por ello, plantea que si $d > -0.5$ el proceso es invertible, partiendo de la Ecuación (12)

$$(1 - B)^d \pi(B) Z_t = a_t \quad \text{Ecuación (15)}$$

Donde $\pi(B) = \frac{\phi(B)}{\theta(B)} = 1 - \pi_1 B - \pi_2 B^2 - \dots$ es la componente de corto plazo del modelo ARFIMA, y conforme a Said & Dickey (1984), para cierto orden p^* suficientemente grande, el proceso anterior puede ser aproximado a

$$(1 - B)^d \pi^*(B) Z_t = a_t \quad \text{Ecuación (16)}$$

Donde $\pi^*(B) = 1 - \pi_1^* B - \dots - \pi_{p^*}^* B^{p^*}$, obteniendo un modelo ARFIMA($p^*, d, 0$) aproximado. Los autores proponen contrastar la hipótesis nula de memoria corta planteada en la Tabla 1, $H_0: d = 0$ contra la hipótesis de memoria larga $H_a: d > 0$, empleando el siguiente estadístico de prueba

$$t_d = \frac{\hat{d}}{se(\hat{d})}$$

donde \hat{d} es el estimador máximo verosímil de d , $se(\hat{d})$ es la estimación de su error estándar. Los autores demostraron que bajo la hipótesis nula, el estadístico de prueba converge a una normal estándar, es decir, $t \xrightarrow{d} N(0,1)$.

Las anteriores pruebas permiten determinar la evidencia de posibles raíces fraccionales en el proceso, ahora bien, para evaluar el ajuste y pertinencia que tiene un modelo determinado a la serie de tiempo se emplea la Prueba de Cusum – Nyblom – Hansen, que tiene el valor agregado de examinar la existencia de posibles puntos de quiebre.

4.2.2 Prueba Recursiva Cusum – Nyblom – Hansen

Según Brown et al. (1975, citado en Vasco et al., 2002) la prueba CUSUM examina la suma recursiva de los residuos acumulados, dada por:

$$W_r = \frac{\sum_{k+1}^r W_t}{\hat{\sigma}} \quad r = k + 1, \dots, T \quad \text{Ecuación (17)}$$

La cual se obtiene a partir de la desviación estándar estimada de los $\sigma^2 =$

$$\frac{\sum_{k+1}^T (w_t - \bar{w})^2}{T - k - 1}, \bar{w} = \frac{\sum_{k+1}^T w_t}{T - k}, \text{ como } \bar{w} = \frac{\sum_{k+1}^T w_t}{T - k} \quad \text{Ecuación (18)}$$

Bajo la hipótesis nula de que W_r tiene media cero y varianza aproximada de $t - k$, porque sus elementos son independientes y tienen varianza unitaria. La prueba sirve para evaluar la importancia de las desviaciones de W_r con respecto una línea definida por $E(W_r) = 0$, Por ello, es necesario construir bandas de “significancia” o líneas críticas por encima y por debajo de la línea de $W_r = 0$ por lo que la probabilidad de W_r de sobrepasar las bandas sea α , o el nivel de significancia deseado.

Las Bandas de significancia se construyen a través de las líneas que pasan por los puntos $(k, \pm\alpha\sqrt{T - k})$ e $(T, \pm 3\alpha\sqrt{T - k})$, donde α es el parámetro que controla el nivel de significancia. Por lo tanto, la hipótesis nula de coeficientes estables se rechaza cuando W_r excede los límites críticos definidos por las líneas rectas $W_r^* = \pm a\sqrt{T - k} \pm 2a \frac{r-k}{\sqrt{T-k}}$, o cuando el estadístico $M = \max_{k+1 < r \leq T} \left| \frac{W_r}{\sqrt{T-k}} \right| / (1 + 2 \frac{r-k}{T-k})$ excede el valor crítico determinado dado por a . también se escribe como $M = \sup_{0 \leq \lambda \leq 1} |W^{(T)}(\lambda) / (1 - 2\lambda)|$ con $W^{(T)}(\lambda) = \frac{1}{\sigma\sqrt{T}} \sum_{k+1}^{k+\lambda(T-k)} w_t, 0 \leq \lambda \leq 1$, de esta manera el camino W es la ruta aproximada de elementos estocásticos de funciones reales y continuas. Dado que $w_t \sim N(0, \sigma^2)$ siendo $W^{(T)}$ un proceso de sumas parciales de W_t .

La aplicación de la función del teorema del límite central (TLC) es inmediata. Así, $W^{(T)}(\lambda) \Rightarrow W(\lambda)$, es decir, el límite de $W(\lambda)$ esta dado por un movimiento Browniano, que significa "=>" convergencia débil.

Se puede entender entonces, que esta prueba se basa en la suma acumulada de los residuos normalizados. Si sus valores cambian en el tiempo de manera sistemática, será evidencia de inestabilidad, bajo la hipótesis nula de estabilidad, el estadístico tiene media cero y una varianza aproximadamente igual al número de residuos acumulados. La evaluación se realiza graficando el estadístico CUSUM a lo largo de la serie. Si la gráfica permanece dentro de las bandas de confianza significa que los coeficientes son estables; pero si cruza las bandas, se rechaza la hipótesis nula y se reconoce la existencia de un cambio estructural en el modelo (Liquitaya B., 2007).

5. IDENTIFICACIÓN DEL MODELO DE PRONÓSTICO

Para evaluar el comportamiento del tráfico aéreo ilegal en la región Caribe entre la primera semana de enero de 2003 hasta la tercera semana de febrero de 2014, fueron registradas 535 observaciones para construir el modelo estadístico. Todos los resultados y gráficos que se presentan en este apartado se obtuvieron con ayuda del software estadístico R y las siguientes librerías: strucchange, car, urca, fracdiff, arfima, rugarch, TSA, pracma, fArma, lmtest y fBasics.

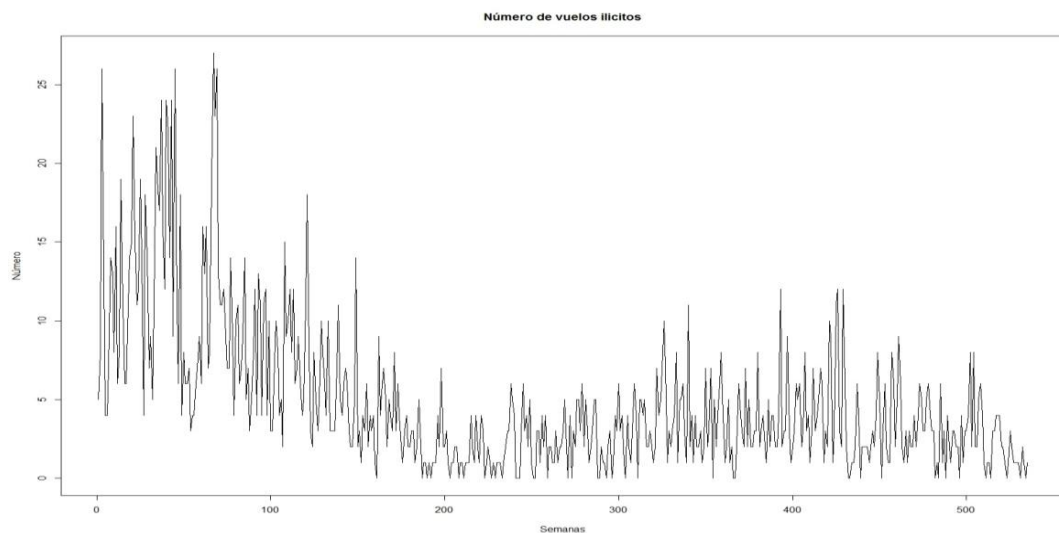


Figura 6. Serie de datos original de vuelos ilegales.

Comparado con una media constante $\theta_1 \in (10, 15)$, en la Figura 6 se observa inicialmente una tendencia decreciente de la cantidad de vuelos ilegales hasta la semana 200 aproximadamente, con una varianza bastante amplia. Luego de dicha semana, el comportamiento de la varianza disminuye y la cantidad de vuelos parece estar oscilando alrededor de una media constante $\theta_2 \in (0, 5)$.

Debido a que existen unos registros iguales a cero, fue necesario sumar una constante a todos ellos para determinar la mejor forma de estabilizar la varianza del proceso con ayuda de la transformación Box – Cox para identificar el valor más adecuado de λ , implementada como powerTransform en el programa R, y así optar $\lambda \in \mathbb{Z}^+$ ó aplicando el operador de logaritmo natural.

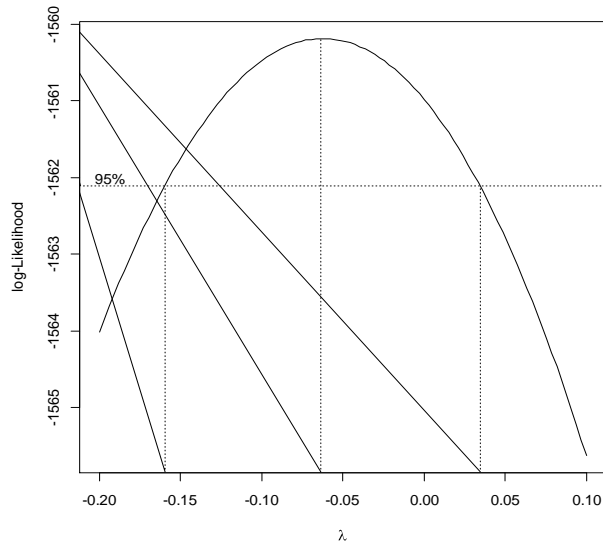


Figura 7. Transformación Box - Cox

La λ para el cual la verosimilitud se maximiza es $\lambda = 0.003309268$ y se logra verificar visualmente con ayuda de la Figura 7, que corresponde a la abscisa del vértice de la parábola. Por aproximación se asume $\lambda = 0$ y por tal razón se decide estabilizar varianza con el operador Logaritmo. Se descarta un segundo intento de estabilización a la serie transformada al obtener $\lambda_2 = 0.9616908 \approx 1$. La Figura 8 muestra que los valores de la serie con el operador logaritmo natural disminuyen, conservando buena parte del comportamiento de la serie original.

A partir de esta transformación se hizo un proceso de identificación del modelo con toda la validación de los supuestos que son requeridos según el capítulo 4, sin embargo no fue posible hallar un modelo que obtuviera bajos valores en los criterios AIC y BIC, teniendo en cuenta que la Función de Autocorrelación Extendida proponía un modelo ARFIMA (4,d,3) y al compararlo con otros de menor grado se mejoraban los criterios pero, ya no eran significativos los coeficientes de los polinomios AR y MA.

Debido a esta situación se optó por considerar la metodología de tratamiento de datos de Castaño (1997), la cual se aplica para tres tipos de efectos: información faltante, intervenciones y casos atípicos. En nuestro caso particular, asumimos los 41 datos iguales a cero como casos atípicos en los vuelos ilegales de Colombia, para así mostrar el proceso de identificación de un modelo en series de tiempo que se describe en las posteriores secciones. Los datos originales y tratados están en el Anexo 1.

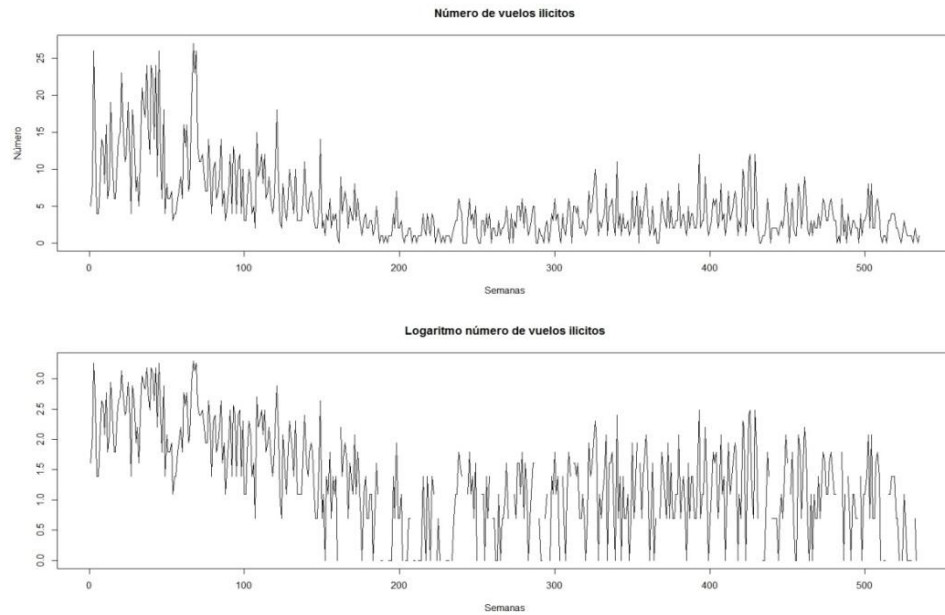


Figura 8. Serie original Vs. Serie estabilizada

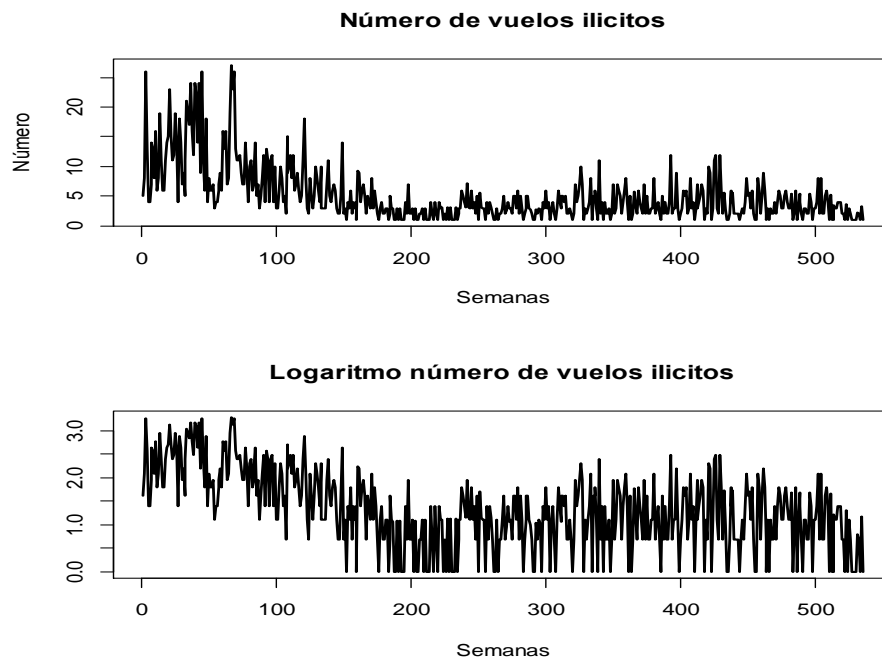


Figura 9. Número de vuelos ilícitos con imputación de datos y su logaritmo

En la Figura 9 se observa la serie con los datos imputados junto con su transformación a partir del operador logaritmo natural, conforme se había descrito en el párrafo anterior. La Figura 10

muestra una ACF que disminuye lentamente en el tiempo, lo que indica que la serie presenta una fuerte correlación entre datos muy distantes en el tiempo. Lo anterior parece indicar que la serie bajo análisis presenta memoria larga pero no es claro el grado de estacionariedad del proceso. A continuación se presentarán las pruebas de hipótesis necesarias para establecer con claridad el grado de persistencia del proceso bajo análisis.

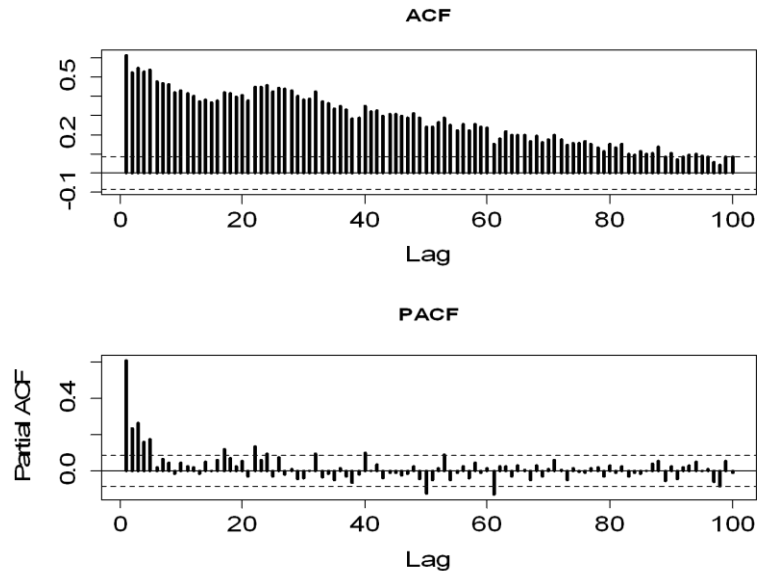


Figura 10. ACF y PACF de la serie con datos imputados

5.1 Pruebas de raíz unitaria

Con el fin de indagar sobre la posibilidad de obtener un proceso estacionario con ayuda de la diferenciación de la serie, se aplican las pruebas de raíz unitaria de la sección 4.1.2.

Prueba de Dickey-Fuller Aumentada

En la Prueba de Dickey-Fuller se determina la existencia o no de una raíz unitaria, evaluándola con tendencia y sin tendencia determinística, para ello se consideran las ecuaciones de la sección 4.1.2.1 con el fin de contrastar las siguientes hipótesis:

- $H_0: \delta = 0$. La serie de tiempo posee una raíz unitaria, es decir, el proceso bajo estudio es NO estacionario homogéneo.
- $H_a: \delta < 0$. La serie de tiempo NO posee raíz unitaria y el proceso bajo estudio es estacionario.

A continuación se presenta la prueba de DFA sin considerar la tendencia determinística. No se presentan los resultados con dicha tendencia porque el coeficiente asociado a la pendiente de la recta no resultó significativo.

Variable	Coefficiente	Error Estándar	Estadístico t	Valor P
Intercepto	0.43521	0.10845	4.013	6.88e-05 ***
z.lag.1	-0.19227	0.04534	-4.241	2.63 e-05 ***
z.diff.lag1	-0.57803	0.05570	-10.378	< 2e-16 ***
z.diff.lag2	-0.46846	0.05689	-8.234	1.47e-15 ***
z.diff.lag3	-0.31823	0.05305	-5.999	3.72e-09 ***
z.diff.lag4	-0.13401	0.04346	-3.084	0.00215 **
Residual standard error: 0,6288 con 520 grados de libertad				
R ² :	0,3819	R ² ajustado:		0,376
Estadístico F:	64,27 con 5 y 520 GL	Valor P		< 2.2e-16

Estadísticos de Prueba	Valor obtenido	1pct	5pct	10pct
tau2	-4,2411	-3.43	-2.86	-2.57
phi1	9.057	6.43	4.59	3.78

Al contrastar los resultados de la evaluación sin tendencia determinística, se puede concluir que existe suficiente evidencia para rechazar H_0 , dado que en las pruebas el estadístico es menor en cualquiera de los percentiles, ósea $\tau < \tau(\alpha, n)$, logrando concluir la no existencia de una raíz unitaria, indicando que el proceso es estacionario.

Prueba de Phillips-Perron

En esta prueba se contrastan las mismas hipótesis que en la prueba anterior.

- $H_0: \delta = 0$. La serie de tiempo posee una raíz unitaria y es NO Estacionaria homogénea.
- $H_a: \delta < 0$. La serie de tiempo NO posee raíz unitaria y el proceso es estacionario.

A continuación se presenta la prueba de PP sin considerar la tendencia determinística. No se presentan los resultados con dicha tendencia porque el coeficiente asociado a la pendiente de la recta no resultó significativo.

Variable	Coefficiente	Error Estándar	Estadístico t	Valor P
Intercepto	1,1398E+05	0.09199	12.39	< 2e-16 ***
y.l1	0.50765	0.03748	13.55	<2e-16 ***
Error Estándar: 0.6915 con 532 grados de Libertad				
R ²	0,2565	R ² Ajustado		0,2551
Estadístico F	183,5 con 1 y 532 GL	Valor P		< 2.2e-16

Valor de la prueba estadística tipo Z-tau: -18,8057			
Z-tau-mu		17,7595	
Estadístico Z	1pct	5pct	10pct
Valores críticos	-3.444837	-2.867257	-2.569809

Los resultados de la Prueba Phillips—Perron también rechazan H_0 , manteniendo la no existencia evidente de una raíz unitaria, confirmando que el proceso generador de los datos es estacionario.

Prueba de Zivot-Andrews

De igual forma se aplicó la Prueba de Zivot-Andrews para lograr una posible detección de raíces unitarias independientes de la existencia de una ruptura o cambio estructural, teniendo en cuenta la siguiente $H_0: \delta = 0$, donde la serie de tiempo posee una raíz unitaria y es NO Estacionaria homogénea y la $H_1: \delta < 0$, donde la serie de tiempo NO posee raíz unitaria y el proceso es estacionario.

A continuación se presenta la prueba de Zivot-Andrews sin considerar la tendencia determinística. No se presentan los resultados con dicha tendencia porque el coeficiente asociado a la pendiente de la recta no resultó significativo.

Variable	Coficiente	Error Estándar	Estadístico t	Valor P
Intercepto	2,3393864	0.1821468	12.843	< 2e-16 ***
y.l1	0.2652546	0.0543374	4.882	1.40e-06 ***
Tendencia	-0.0001034	0.0002622	-0.394	0.694
y.dl1	-0.0662545	0.0434262	-1.526	0.128
du	-0.8093422	0.1107917	-7.305	1.03e-12 ***
Error Estándar: 0.6203 con 528 grados de Libertad				
R ²	0,4051		R ² Ajustado	0,4006
Estadístico F	89,9 con 4 y 528 GL		Valor P	< 2.2e-16

Estadístico de Prueba	-13.5219
Valores Críticos	
0,01	-5,34
0,05	-4,8
0,1	-4,58
Punto de ruptura potencial en la posición: 133	

Luego de realizar las pruebas de Dickey-Fuller, Phillips-Perron y Zivot-Andrews se concluye que NO hay raíz unitaria en el proceso y el modelo es estacionario. En este punto se procede a realizar pruebas de raíz fraccional con el fin de identificar un modelo ARFIMA (p,d,q) estacionario e invertible.

Tabla 3. Pruebas de raíz unitaria

Prueba	Con tendencia		Sin tendencia	
	\hat{t}	$\tau_{0.05}$	\hat{t}	$\tau_{0.05}$
DFA	-6.28	-3.41	-4.36	-2.86
PP	-18.91	-3.42	-16.66	-2.86
Zivot - Andrews	-12.59	-4.8	-13.19	-4.42

Como se observa en la Tabla 3, los valores de los estadísticos de prueba \hat{t} y sus valores teóricos han sido resumidos para cada una de las pruebas de raíz unitaria. Debido a que en todos los casos se cumple $\hat{t} < \tau_{0.05}$, existe suficiente evidencia en el rechazo de la hipótesis nula de existencia de raíz unitaria en la serie de los vuelos ilícitos en Colombia. Hasta el momento no existe una razón fuerte para afirmar que el proceso se comporta como un modelo ARMA (p,q) y tampoco como un modelo ARIMA (p,d,q); por un lado se debe a lo observado en la ACF y la PACF, y por otro lado los resultados de las pruebas efectuadas de raíz unitaria no tienen el suficiente peso para continuar con la metodología de Box-Jenkins. Debido a lo anterior se plantea evaluar la serie con un modelo intermedio ARFIMA (p,d,q), cuya raíz es fraccional y está entre -0,5 y 0,5 utilizando la metodología de Castaño, Gallón y Gómez (2010).

5.2 Pruebas de raíz fraccional

Con la implementación de estas pruebas se pretende buscar un número real d que esté comprendido entre -0.5 y 0.5 para analizar el ajuste de la familia ARFIMA(p, d, q) a la serie de tiempo bajo análisis.

Para las pruebas de raíz fraccional se establecen las siguientes hipótesis:

- $H_0: \delta = 0$. Es un proceso ARMA (p,q) estacionario de memoria corta.
- $H_a: \delta \neq 0$ con $-1/2 < d < 1/2$. Es un proceso ARFIMA (p,d,q) estacionario e invertible.

Prueba GPH

Estimador d	Error Estándar	Estadístico t	Valor P
0.4383	0.0933	46.954	0.0001

Prueba ROB 95

Estimador d	Error Estándar	Estadístico t	Valor P
0.2879	0.1403	20.517	0.0245

Los resultados de las pruebas GPH y ROB 95 rechazan la hipótesis nula, dado que el valor p en ambas pruebas es menor a un nivel de significancia del 5%, concluyendo que hay evidencia de raíz fraccional estacionaria e invertible en la serie de vuelos ilícitos.

A continuación se ilustra la implementación de la prueba propuesta por Castaño (2008), el primer paso es la identificación y estimación del modelo adecuado para la serie transformada de la cantidad de vuelos ilícitos. Inicialmente se aproxima el componente de memoria corta con un proceso autorregresivo ARFIMA ($p^*, d, 0$) para obtener el valor de la raíz fraccional. El orden obtenido para el ajuste del modelo es: $p^* = T^{1/4} = 5$, según Castaño (2010, p. 147).

Prueba CASTAÑO - Paquete ARFIMA

Variable	Coefficiente	Error Estándar Th	Error Estándar	Valor Z	Valor P
phi (1)	-0.2605258	0.0633798	0.1095445	-4.11055	3.947e-05 ***
phi (2)	-0.1833995	0.0574021	0.0891684	-3.19500	0.0013983 **
phi (3)	-0.0790990	0.0544059	0.0787909	-1.45387	0.1459826
phi (4)	0.0305175	0.0499099	0.0653529	0.61145	0.5409005
phi (5)	0.0596811	0.0447549	0.0507642	1.33351	0.1823648
d.f	0.4591215	0.0441135	0.0997747	10.40773	< 2.22e-16 ***
Media Ajustada	1.3764539	0.6788225	NA	2.02771	0.0425901 *

Los resultados obtenidos del modelo aproximado de la anterior tabla, se logra observar que hay tres coeficientes autorregresivos que no son significativos para un nivel de significancia del 5%, junto con un parámetro de diferenciación fraccional de $d = 0.4591215$ que sí es significativo, lo cual indica que el modelo está sobreparametrizado. Se reespecifica el modelo aproximado a un ARFIMA (2,d,0) obteniendo los siguientes resultados:

Re especificando el modelo en la prueba de Castaño

Variable	Coefficiente	Error Estándar Th	Error Estándar	Valor Z	Valor P
phi (1)	-0.2233072	0.0564358	0.0664906	-3.95684	7.5949e-05 ***
phi (2)	-0.1560035	0.0482390	0.0528394	-3.23397	0.0012208 **
d.f	0.4355482	0.0387298	0.0551996	11.24581	< 2.22e-16 ***
Media Ajustada	2.3754743	0.4977951	NA	4.77199	1.8241e-06 ***
Sigma^2			Log-likelihood		
0.386072			255.014		
AIC			BIC		
-500.027			2164.97		

Considerando un nivel de significancia del 5%, se concluye que el parámetro estimado de diferenciación fraccional $d = 0.4355482$ resulta significativo, al igual que los dos coeficientes autorregresivos obtenidos.

Continuando con el proceso de Castaño (2008), el segundo paso consiste en diferenciar la serie transformada como sugiere la Ecuación 19, y sobre los resultados aplicar las técnicas de identificación de un modelo ARMA (p,q) para el proceso de memoria a corto plazo.

$$W_t = (1 - B)^{0.4355482} Y_t \quad \text{Ecuación (19)}$$

En la Figura 11 se observa que el comportamiento característico de un proceso MA puro no está presente en la Función de Autocorrelación ni en la Función de Autocorrelación Parcial. Debido a esto, fue necesario aplicar la Función de Autocorrelación Extendida (EACF).

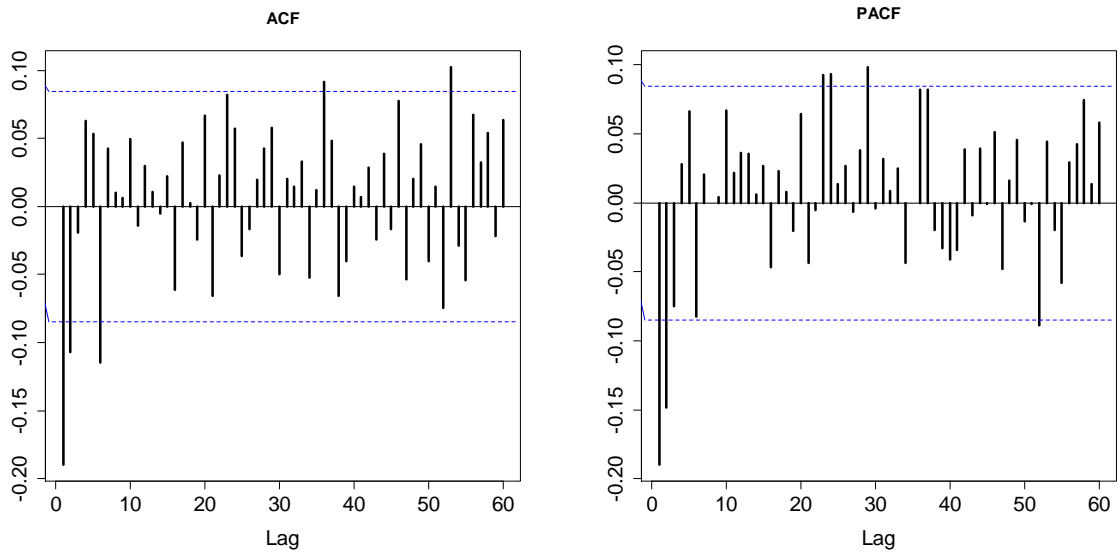


Figura 11. ACF y PACF de la serie Transformada y Diferenciada

En la Figura 12 se observan los resultados de (EACF) en una matriz cuyos campos están con X ó 0, en la cual se resalta con ayuda del triángulo de ceros, específicamente con el vértice izquierdo, que un proceso ARMA (1,1) logra ajustarse mejor al comportamiento de la serie a corto plazo.

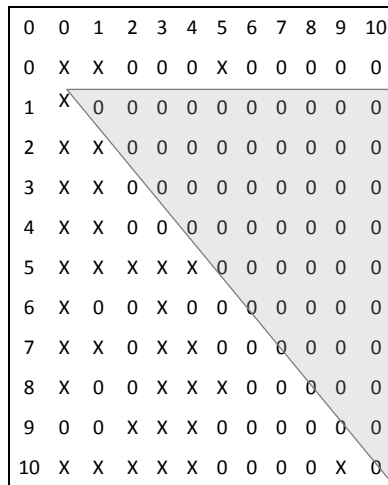


Figura 12. Muestra extendida de la ACF para la serie fraccionalmente diferenciada

El tercer paso del procedimiento consiste en estimar conjuntamente los parámetros del modelo identificado ARFIMA $(1,d,1)$ en la serie transformada y por último validar los supuestos en el modelo.

Modelo Ajustado

Variable	Coficiente	Error Estándar	Th	Error Estándar	Valor Z	Valor P
phi(1)	0.2908901	0.1063485		0.1192476	2.73525	0.0062332 **
theta(1)	0.5802192	0.0952050		0.1535252	6.09442	1.0984e-09 ***
d.f	0.4783202	0.0283972		0.1083513	16.84393	< 2.22e-16 ***
Media Ajustada	2.3782955	0.8813058		NA	2.69860	0.0069631 **
Sigma^2				Log-likelihood		
0,384556				255.529		
AIC				BIC		
-501.058				2163,94		

El modelo ajustado permite observar que el parámetro de diferenciación fraccional y las componentes AR y MA son significativos, para explicar el comportamiento de los vuelos ilícitos.

5.3 Verificación del modelo

Con el fin de validar el ajuste de este modelo ARFIMA $(1, d, 1)$ con $d = 0.4783202$, a la serie de tiempo de los vuelos ilícitos en el territorio nacional, se procede a verificar el cumplimiento de los supuestos en esta familia.

5.3.1 Prueba Cusum - Nyblom Hansen para residuales

En la Figura 13 los residuales permanecen dentro de las bandas de confianza a lo largo de la serie, lo que significa que los coeficientes son estables, por lo cual no existen cambios estructurales en el modelo.

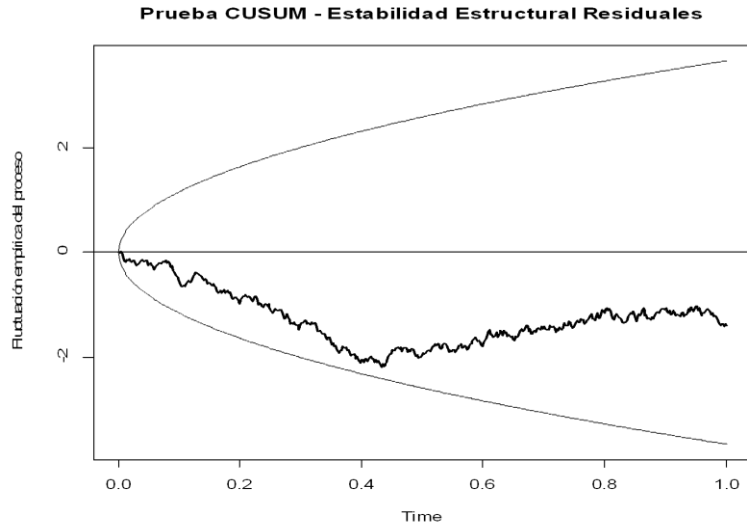


Figura 13. Estabilidad estructural de los residuales

De igual forma se valida la hipótesis nula de la estabilidad de los parámetros de los coeficientes con cualquier significancia inferior al 10% como se muestra a continuación:

Estadístico de Prueba	Valor P
0,854	0,10905

5.3.2 Pruebas de incorrelación

En esta sección se utilizan las pruebas de Ljung-Box y Durbin-Watson cuyo objetivo es observar la incorrelación serial en la secuencia de residuales en el modelo ajustado, por ello se establecen las siguientes hipótesis para su evaluación:

- H_0 : No existe evidencia de correlación serial entre los residuales.
- H_a : Existe evidencia de correlación serial entre los residuales.

Prueba de Ljung-Box			Prueba de Durbin-Watson	
X-cuadrado	df	Valor P	DW	Valor P
32,048	30	0,3653	1,9989	0,4947

Los resultados de las pruebas para cualquier grado de significancia menor al 10% aceptan H_0 y se concluye que los residuales no están correlacionados serialmente entre sí, demostrando que el modelo ajustado logra capturar la correlación serial de corto y largo plazo del proceso. Esta conclusión se logra visualizar en la Figura 14, donde la ACF y la PACF no muestran evidencia de alguna correlación entre rezagos de los residuales.

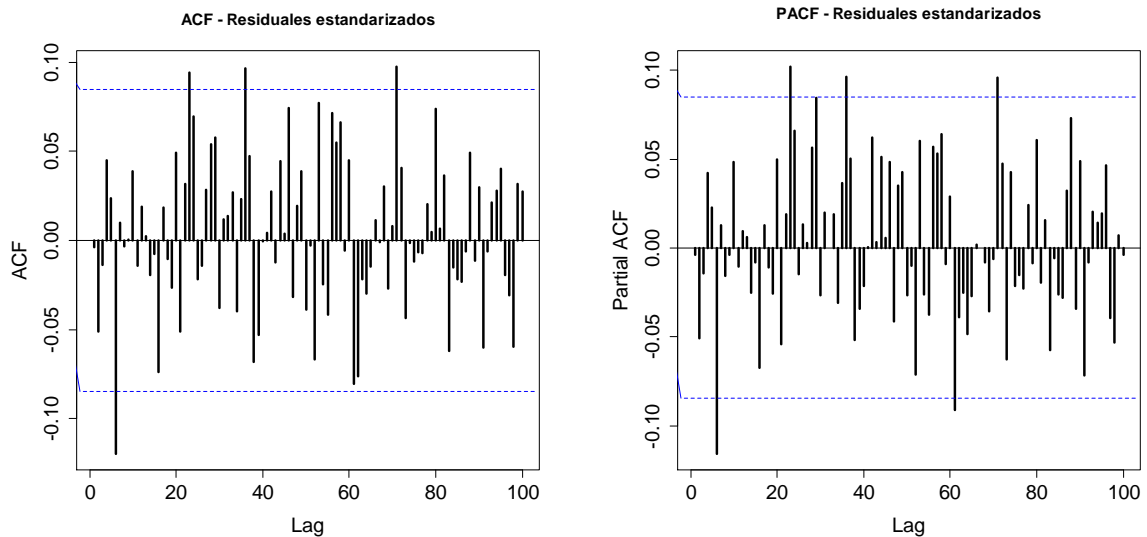


Figura 14. ACF y PACF de los residuales estandarizados

5.3.3 Pruebas de normalidad

En esta sección se aplican las pruebas de Shapiro-Wilk, Jarque-Bera, Kolmogorov-Smirnov y D'Agostino, cuyo objetivo es verificar que los residuales del modelo se ajusten a una distribución normal. Las hipótesis de evaluación son:

- H_0 : Existe normalidad en los residuales del modelo ajustado.
- H_a : No existe normalidad en los residuales del modelo ajustado.

Prueba de Shapiro-Wilk		Prueba de Jarque - Bera	
W	Valor P	χ^2	Valor P
0,9869	9,544e-05	12,071	0,002392

Estas dos pruebas rechazan el ajuste de los residuales a una distribución normal, con cualquier significancia superior al 1%.

Prueba de Kolmogorov-Smirnov		Prueba de D'Agostino	
Estadísticos de prueba		Estadísticos de prueba	
D	0,0466	Chi^2 General	21,0729
Valor P (Dos Colas)	0,1954	Z3 Asimetría	-1,8951
Valor P (Menor)	0,09781	Z4 Curtosis	-4,1811
Valor P (Mayor)	0,3083	Valor P General	2,655e-05
		Valor P Asimétrico	0,05808
		Valor P Curtosis	2,901e-05

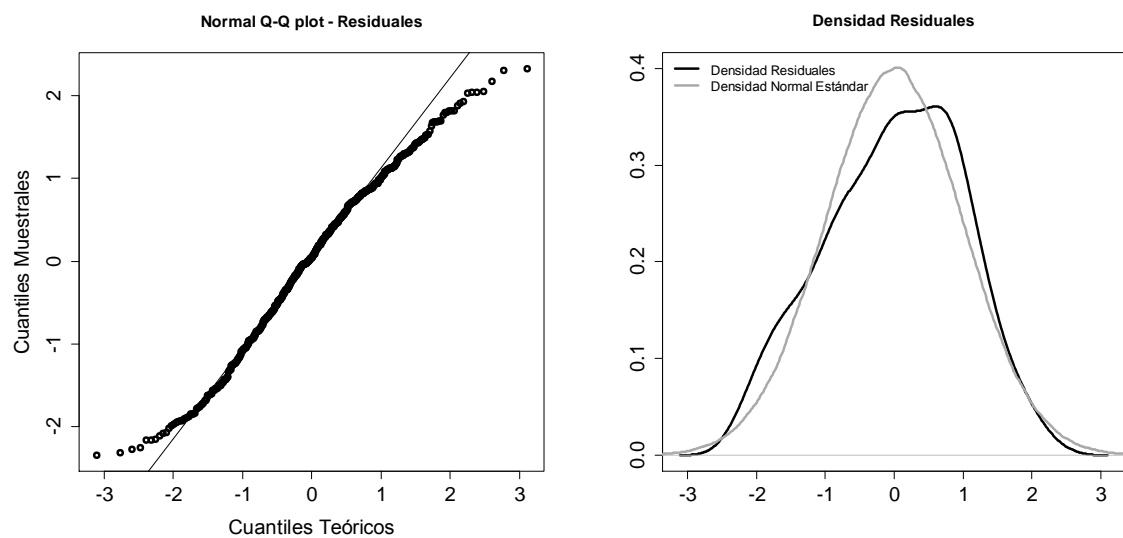


Figura 15. Q-Q plot y densidad de los residuales comparados con la distribución Normal

La prueba de Kolmogorov-Smirnov acepta H_0 , mientras que la D'Agostino la rechaza, es decir que tres de las cuatro pruebas de normalidad, informan que no hay un ajuste de los residuales a la distribución Normal.

5.3.4 Prueba de bondad de ajuste a una t- student con 16 grados de libertad

En la sección 5.3.3 se evidenció que tres pruebas aplicadas a los residuales del modelo rechazaron la evidencia de un comportamiento normal, lo que visualmente se identifica en el Q-Q plot de la Figura 15 en los múltiples puntos alejados de la recta en sus extremos. Por esta razón se toma como alternativa examinar el ajuste de los residuales a una distribución T-student con 16 grados de libertad, estableciendo las siguientes hipótesis.

- H_0 : Los residuales del modelo se ajustan a una distribución T-student con 16 grados de libertad.
- H_a : Los residuales del modelo NO se ajustan a una distribución T-student.

D	Valor P
0,0449	0,6546

De acuerdo al valor-P obtenido hay suficiente evidencia para aceptar el ajuste indicado en H_0 . En la Figura 16 se muestra el ajuste con la distribución T-student y el Q-Q plot correspondiente.

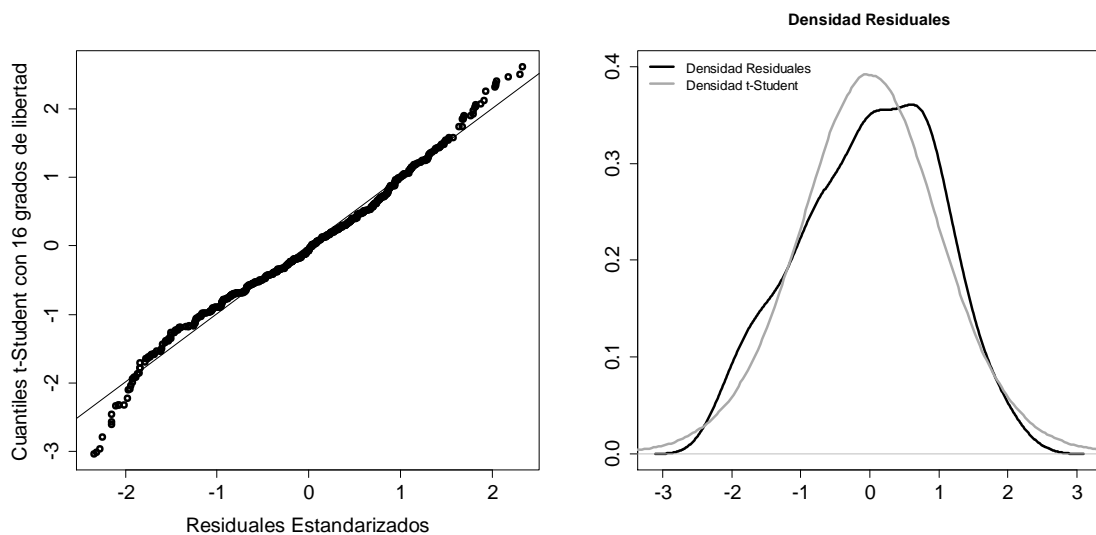


Figura 16. Q-Q plot y densidad de residuales con distribución T-student con 16 gl

En la Figura 17, la función de autocorrelación estandarizada de los residuales al cuadrado, no se evidencian efectos de heterocedasticidad condicional porque en estas gráficas no se visualiza que los primeros pulsos de correlación sean significativos, debido a que no salen de las bandas de confianza.

Es decir que los residuales presentan un comportamiento adecuado a una distribución t – Student con 16 grados de libertad y el modelo ajustado de la sección 5.2 es pertinente para describir la cantidad de vuelos ilícitos en Colombia.

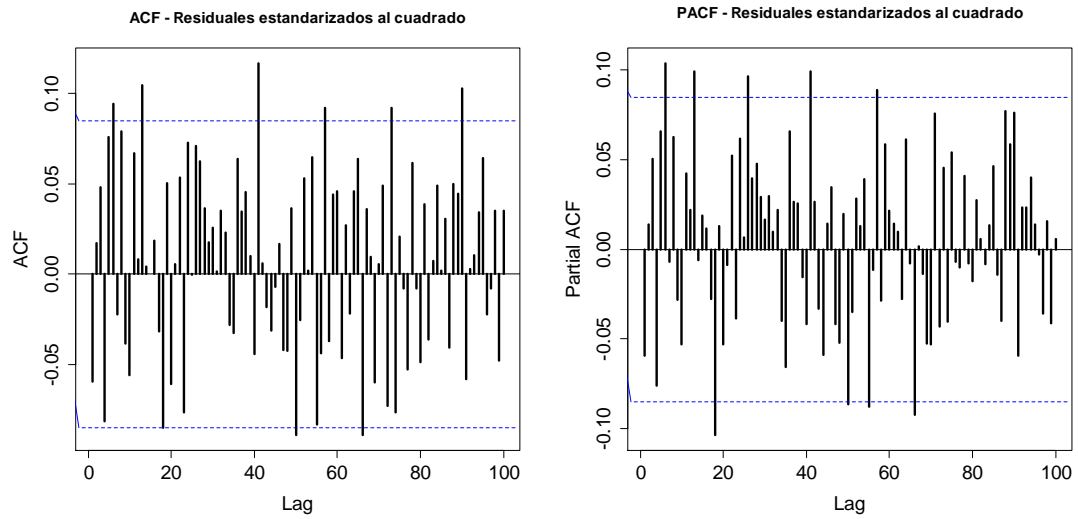


Figura 17. ACF y PACF de los residuales estandarizados al cuadrado

6. RESULTADOS

6.1. Modelo identificado

Una vez se identificaron los parámetros del posible modelo que explica el comportamiento de los vuelos ilícitos en Colombia y se validaron los supuestos relacionados con los residuales, se identificó un modelo ARFIMA (1,d,1) como se resume en la Tabla 4.

Tabla 4. Parámetros del modelo ARFIMA (1,d,1)

Variable	Coficiente	Error Estándar
phi (1)	0.2908901	0.1192476
theta(1)	0.5802192	0.1535252
d.f	0.4783202	0.1083513
Media Ajustada	2.3782955	NA

Con los anteriores parámetros, el modelo se plantea con la siguiente expresión:

$$(1 - \phi_1 B)(1 - B)^d \text{Ln}(Y_t) = \theta_0 + (1 - \theta_1)a_t, \quad \text{donde } a_t \sim t_{16gl} \quad \text{Ecuación (20)}$$

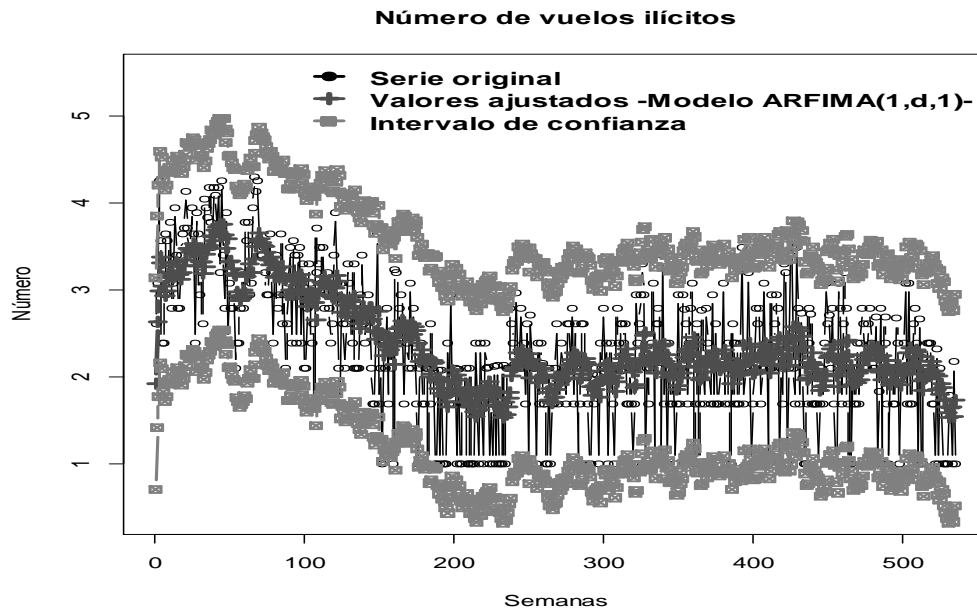


Figura 18. Logaritmo de las estimaciones intramuestra con bandas de confianza del 95%

6.2. Validación cruzada

La validación cruzada es una manera de verificar que un modelo en series de tiempo está realmente describiendo el comportamiento del proceso. Para el caso de las 535 observaciones semanales de los vuelos ilícitos, se tomaron las primeras 515 para que su información pronosticara las 20 siguientes y así poder contrastar el pronóstico con los datos originales.

A continuación se presenta la validación cruzada por dos métodos implementados en el paquete **arfima** de R (función *predict.arfima*), el método exacto y el bootstrap. En ambos casos se muestra el resultado numérico y la visualización con bandas de confianza del 95% de confianza.

Resultados del Método exacto

Semana	516	517	518	519	520	521	522
Pronóstico	1,891424	1,917025	1,930442	1,942768	1,954604	1,965692	1,975915
SD Exacta	0,620075	0,630716	0,635371	0,639575	0,643548	0,647197	0,650492
SD Limite	0,619940	0,630530	0,635146	0,639310	0,643242	0,646847	0,650099
Semana	523	524	525	526	527	528	529
Pronóstico	1,985292	1,993897	2,001819	2,009141	2,015940	2,022279	2,028213
SD Exacta	0,653457	0,656133	0,658562	0,660780	0,662819	0,664702	0,66645
SD Limite	0,653020	0,655653	0,658039	0,660215	0,662211	0,664052	0,665760

Semana	530	531	532	533	534	535
Pronóstico	2,033787	2,039042	2,044010	2,048720	2,053197	2,057463
SD Exacta	0,668085	0,669615	0,671054	0,672412	0,673697	0,674917
SD Limite	0,667351	0,668840	0,670238	0,671555	0,672799	0,673979

En la Figura 19 se observan tres tipos de líneas, aquella con signos + representa el comportamiento del proceso en estudio, la línea descrita con triángulos representa el pronóstico y las líneas unidas con cuadrados son las bandas de confianza de 95%, con la desviación estándar y la media pronosticada que está en los resultados del método exacto.

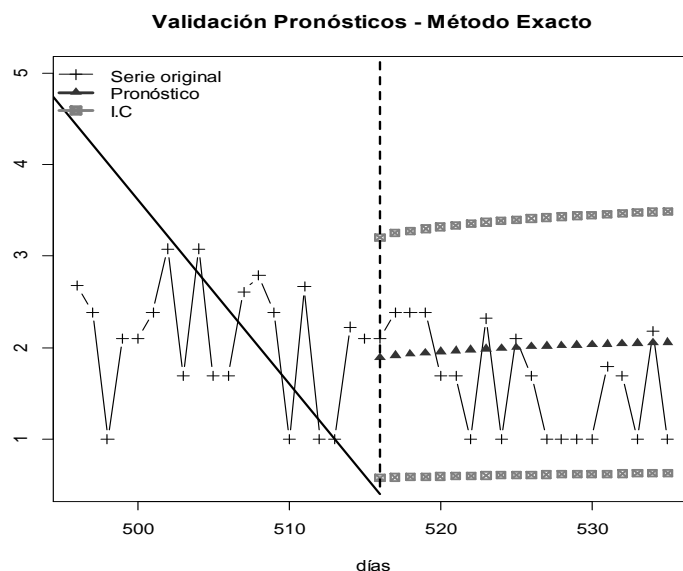


Figura 19. Validación método exacto con 95% de confianza

Resultados del Método Bootstrap

Semana	516	517	518	519	520	521	522
Pronóstico	1,593703	1,655627	1,678959	1,698661	1,717553	1,735339	1,751795
SD Exacta	0,620259	0,631288	0,636536	0,641286	0,645699	0,649691	0,653261
SD Límite	0,620126	0,631104	0,636310	0,641018	0,645386	0,649334	0,652858
Semana	523	524	525	526	527	528	529
Pronóstico	1,766927	1,780841	1,793671	1,805549	1,816588	1,826889	1,836535
SD Exacta	0,656453	0,659322	0,661919	0,664285	0,666457	0,668461	0,670322
SD Limite	0,656005	0,658829	0,661381	0,663703	0,665831	0,667791	0,669608
Semana	530	531	532	533	534	535	
Pronóstico	1,845598	1,854139	1,862211	1,869858	1,877120	1,884030	
SD Exacta SD	0,672057	0,673681	0,675208	0,676649	0,678012	0,679305	
SD Limite SD	0,671300	0,672881	0,674366	0,675764	0,677084	0,678336	

En la Figura 20 se muestra el comportamiento de los vuelos ilícitos con la línea unida por símbolos +, las bandas de confianza del 95% con respecto a la media pronosticada aparecen desde la observación 536 con la unión de cuadrados y finalmente, el pronóstico del modelo ajustado está representado por triángulos.

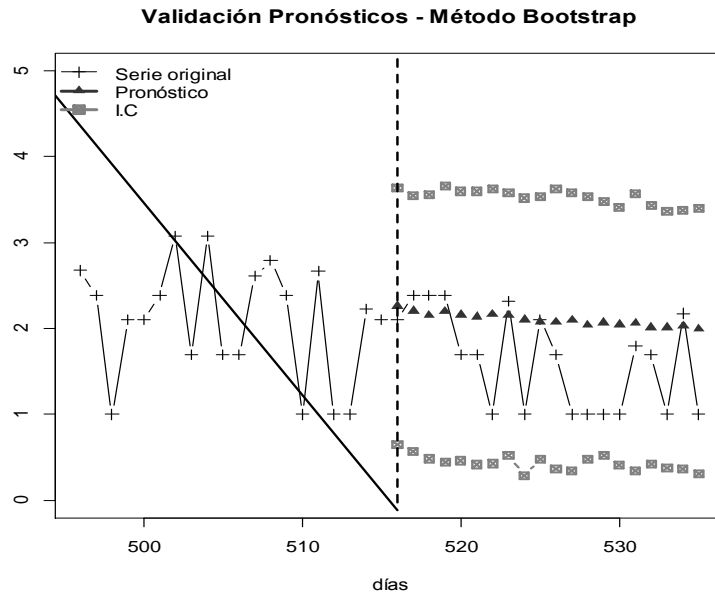


Figura 20. Validación Método Bootstrap con 95% de confianza

Se observa en la Figura 19 y la Figura 20 que en ambos métodos de validación cruzada los pronósticos con las veinte observaciones de prueba, al igual que el proceso están dentro de las bandas de confianza calculadas con un 95%, con lo cual afirmamos que el modelo ajustado está explicando el proceso en cuestión.

Adicional a esto, se calculó el Error Cuadrático Medio (ECM) en la predicción de estas veinte observaciones y se obtuvo para el método Exacto el valor de 1,78803632, y para el Método Bootstrap 1,76904487. A partir de estos resultados parece ser que el último método validación de pronóstico aplicado al modelo ajustado describe mejor la cantidad de vuelos ilegales en Colombia.

6.3. Pronósticos

Una vez validado el funcionamiento del modelo ajustado con los datos de entrenamiento y los de prueba, se procede a realizar el pronóstico de veinte semanas más a partir de la observación 535 empleando el paquete Arfima.

Pronóstico con Método Exacto

Semana	536	537	538	539	540	541	542
Limite Sup. 95%	3.205924	3.254083	3.277368	3.298606	3.318865	3.337688	3.354896
Media pronosticada	1,891424	1,917025	1,930442	1,942768	1954604	1,965692	1,975915
Limite Inf. 95%	0.576924	0.579967	0.583516	0.586930	0.590343	0.593696	0.596933
Semana	543	544	545	546	547	548	549
Limite Sup. 95%	3.370559	3.384837	3.397908	3.409932	3.421054	3.431384	3.441028
Media pronosticada	1,985292	1,993897	2,001819	2,009141	2,015940	2,022279	2,028213
Limite Inf. 95%	0.600025	0.602957	0.605730	0.608350	0.610826	0.613174	0.615398
Semana	550	551	552	553	554	555	
Limite Sup. 95%	3.450064	3.458562	3.466581	3.474170	3.481366	3.488223	
Media pronosticada	2,033787	2,039042	2,044010	2,048720	2,053192	2,057463	
Limite Inf. 95%	0.617510	0.619521	0.621439	0.623270	0.625018	0.626703	

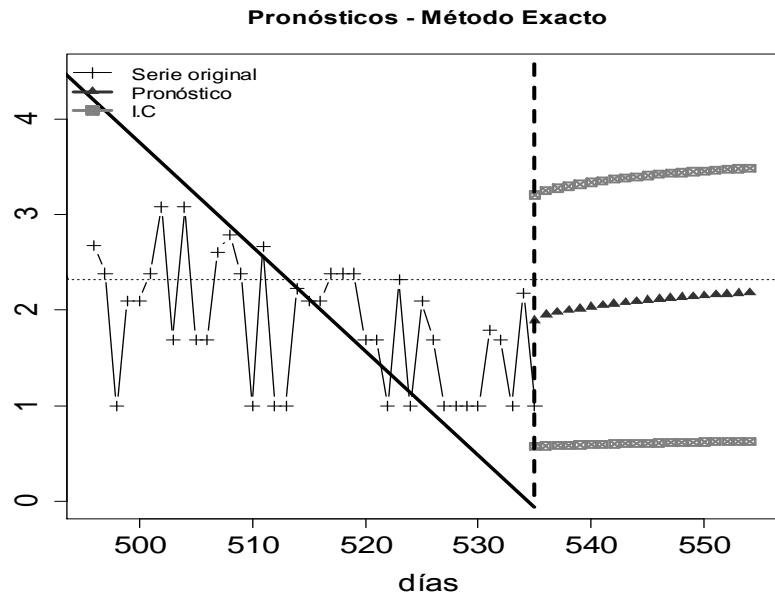


Figura 21. Pronóstico Método Exacto con 95% de confianza

Pronóstico con Método Bootstrap

Semana	536	537	538	539	540	541	542
Limite Sup, 95%	3,63590	3,54444	3,55781	3,65588	3,59593	3,59560	3,62813
Media pronosticada	2,26425	2,20547	2,15495	2,19684	2,15871	2,13642	2,16455
Limite Inf, 95%	0,44630	0,37404	0,28441	0,24976	0,26310	0,21767	0,22898
Semana	543	544	545	546	547	548	549
Limite Sup, 95%	3,57925	3,51670	3,53416	3,62628	3,57538	3,53622	3,47544
Media pronosticada	2,15474	2,10623	2,07959	2,07579	2,10138	2,04271	2,06863
Limite Inf, 95%	0,32041	0,08856	0,27740	0,16599	0,14187	0,27745	0,32598
Semana	550	551	552	553	554	555	
Limite Sup, 95%	3,40965	3,56720	3,43068	3,36101	3,37976	3,39603	
Media pronosticada	2,04888	2,06135	2,01541	2,01251	2,03181	1,99855	
Limite Inf, 95%	0,21246	0,13991	0,22301	0,17905	0,16458	0,10825	

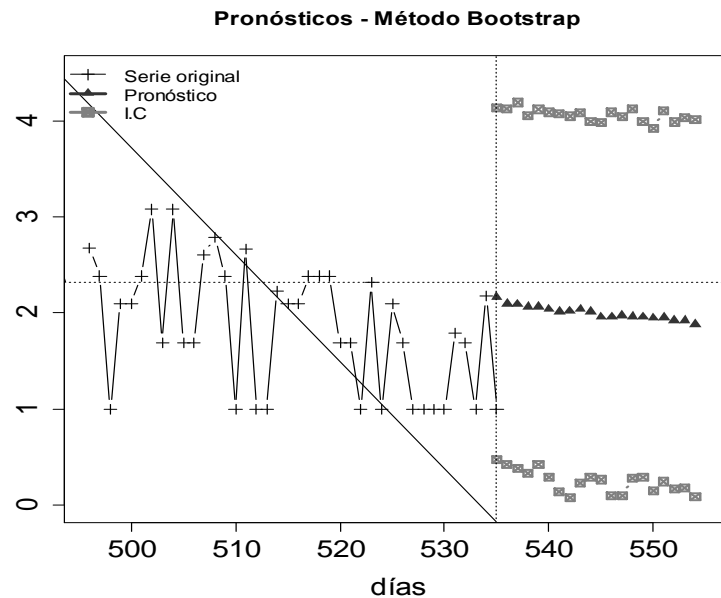


Figura 22. Pronóstico Método Bootstrap con 95% de confianza

En la Figura 21 y la Figura 22 se observa el pronóstico desde la semana 536 hasta la semana 555, donde la cantidad de vuelos ilícitos para cada semana logran permanecer dentro de las bandas de confianza, tanto en el método exacto, como en el método bootstrap; los cuales generan una media aproximada de 2 vuelos ilícitos por semana.

Sin embargo, el pronóstico del método exacto tiende a regresar a la media y el del método bootstrap parece que se aleja de ésta, a pesar de que su ECM es menor. Esto indica que por el

último método de pronóstico mencionado se esperaría que la cantidad de vuelos ilícitos sigan disminuyendo de acuerdo a la estrategia de interdicción aérea implementada.

7. CONCLUSIONES

Las conclusiones logradas luego de realizar el tratamiento de los datos relacionados con los vuelos ilícitos en Colombia, entre los años 2002 y 2014, y registrados en semanas son las siguientes:

- La metodología para el tratamiento de series de tiempo propuesta por Box & Jenkins no logró establecer un modelo adecuado por no capturar memoria a largo plazo, dado que las diferentes pruebas de raíz unitaria siempre indicaron la no existencia de las mismas. En consecuencia no existe un modelo ARIMA (p,d,q) con $d \in \mathbb{Z}^+ \cup \{0\}$ que explique el comportamiento de los vuelos ilícitos en Colombia.
- Después de aplicar algunas pruebas de raíz fraccional a la serie de datos y concluir la existencia de una raíz de este tipo, que captura la memoria a largo plazo del proceso, se utilizó la metodología propuesta por Castaño, Gallón y Gómez (2008), la cual consiste en una aproximación de un modelo autorregresivo puro en primer lugar, junto con la aproximación de un modelo de media móvil en segundo lugar, para finalmente llegar a la construcción del siguiente modelo ARFIMA (1,d,1):

$$(1 - \phi_1 B)(1 - B)^d \text{Ln}(Y_t) = \theta_0 + (1 - \theta_1) a_t, \quad \text{donde } a_t \sim t_{16gl}$$

$$(1 - 0.2908901B)(1 - B)^{0.478202} \text{Ln}(Y_t) = 2.3782955 + (1 - 0.5802192) a_t, \quad \text{donde } a_t \sim t_{16gl}$$

- La serie original de datos contenía una gran cantidad de ceros, los que ocasionaban una contaminación en la estimación del modelo, causando sesgo en los pronósticos de los vuelos ilícitos en Colombia, por tanto fue necesario efectuar un tratamiento de datos para reducir la probabilidad de un pronóstico incorrecto. En la siguiente tabla se evidencia a partir de los criterios de Akaike y Bayesiano que el método de imputación logró obtener un modelo más parsimonioso y de mayor eficiencia.

Serie de Tiempo	d	AIC	BIC	Modelo ARFIMA
Con ceros	0.48309288	-604.295	4725.7	(4,d,3)
Sin ceros	0.4783202	-501.058	2163,94	(1,d,1)

- Finalmente el modelo propuesto hace una estimación de vuelos ilícitos para un total de 20 semanas, pronosticando el comportamiento desde la semana 536 hasta la semana 555 aplicando el método exacto y el método Bootstrap. Así mismo, se evidenció un promedio

real de 1,5 vuelos ilícitos semanales comparado en el mismo periodo de tiempo, logrando permanecer entre los límites establecidos en el pronóstico. Cabe resaltar que los resultados pudieron variar a causa de externalidades, que para el caso en cuestión, responden a los intereses de los Estados implicados y a las estrategias implementadas por los gobernantes de turno, quienes al sentir el impacto de la transnacionalización del tráfico aéreo ilícito, fortalecen sus leyes y operaciones militares como herramientas fundamentales para la defensa de su soberanía.

8. REFERENCIAS

- Aeronáutica Civil. (2009). *Resolución 06565. "Por la cual se adoptan procedimientos y documentos para el Grupo de Normas Aeronáuticas de la Unidad Administrativa Especial de la Aeronáutica Civil"*. Bogotá, Colombia.
- Castaño, E. (1997). Identificación de un modelo ARIMA cuando existen observaciones faltantes. *Lecturas de Economía*, (47), 27 – 45. Universidad de Antioquia, Medellín, Colombia.
- Castaño, E. (2011). *Análisis de series de tiempo lineales*, partes 1 y 2. Notas de clase.
- Castaño, E., Gómez, K. y Gallón, S. (2008). *Una nueva prueba para el parámetro de diferenciación fraccional*. *Revista Colombiana de Estadística*, (31), 67–84.
- Castaño, E., Gómez, K. y Gallón, S. (2010). Sesgos en estimación, tamaño y potencia de una prueba sobre el parámetro de memoria larga en modelos ARFIMA. *Lecturas de Economía*, (73), 131 -148. Universidad de Antioquia, Medellín, Colombia.
- García, A. (2010). *Narcotráfico en Colombia, un problema creado en otra parte, importado y asumido como propio*. Antioquia, Colombia. Disponible en <http://www.leyex.info/magazines/vol67n14911.pdf>
- Box, G., Jenkins, G y Reinsel, G. (2008). *Time Series Analysis. Forecasting and control*. Cuarta Edición.
- Glynn, J., Perera N. y Verma R. (2007). Unit root test san structural breaks: a survey with applications. *Revista de métodos cuantitativos para la economía y la empresa*. (3), 63-79.
- Guerrero, V. (1991). *Análisis estadístico de series de tiempo económicas*. Editorial UAM. México.
- Gujarati, D. y Porter, D. (2010). *Econometría*. Mc Graw Hill, Quinta edición, 745-762.

- Hernández, J. (2012). Procedimientos de Interdicción Aérea FAC. En *Seminario internacional de Interdicción aérea al Narcotráfico para Latinoamérica y el Caribe*. Bogotá, Colombia.
- Lemus, D. y Castaño, E. (2013). Prueba de hipótesis sobre la existencia de una raíz fraccional en una serie de tiempo no estacionaria. *Lecturas de Economía*, (78), 153-184. Universidad de Antioquia, Medellín, Colombia.
- Liquitaya, J. (2007). La demanda de saldos monetarios en México: pruebas de homogeneidad, estabilidad y sensibilidad a la tasa de interés. *Revista Nicolaita de Estudios Económicos*. Vol. II (1). Enero-junio, 53-74.
- Montenegro, A. (2005). *Series de Tiempo*. Pontificia Universidad Javeriana. Cuarta edición. Bogotá.
- Rangel, A. (2007). Consideraciones sobre el uso de los test ADF y Zivot-Andrews. *Revista de Economía & Administración*, Vol. 4 No. 2, 133-150.
- Robinson, P. (1995). Log-periodogram regression of time series with long range dependence. *The Annals of Statistics*, Vol. 23, No. 3, 1048-1072.
- Rocha, R. (2011). *Las nuevas dimensiones del narcotráfico en Colombia*. United Nations Office on Durgs and Crime – UNODC y Ministerio de Justicia y del Derecho. Bogotá. Disponible en http://www.odc.gov.co/docs/publicaciones_nacionales/Dimensiones%20del%20narcotrafico%20en%20Colombia.pdf
- Vasco, J. (2002). Testes de alteracao de estrutura em modelos multivariados: uma visita guiada pela literatura. *Notas Económicas*, (16), 16–33.

ANEXO 1. DATOS ORIGINALES E IMPUTADOS

Datos originales de los vuelos ilegales.

2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
5	4	12	3	1	0	0	6	1	0	3	1
8	8	4	2	1	0	2	3	5	1	0	1
26	6	10	2	1	0	1	1	2	1	1	0
13	6	3	5	4	3	1	11	4	2	0	2
4	7	3	14	2	6	0	2	4	6	6	1
4	3	6	2	7	3	2	4	2	4	1	0
7	4	10	3	2	4	3	1	2	0	3	1
14	4	8	1	2	2	0	4	4	2	0	
13	6	4	4	3	5	1	2	12	2	4	
8	7	5	3	1	1	4	2	2	2	2	
16	9	2	6	0	0	3	3	3	2	1	
6	6	15	2	1	0	6	1	3	1	3	
8	16	9	4	1	3	3	2	9	2	3	
19	13	10	3	2	3	4	7	3	3	2	
11	16	12	4	2	1	1	2	1	2	2	
6	7	8	1	0	4	0	3	2	4	0	
6	8	12	0	1	2	4	7	3	8	4	
11	19	6	9	1	4	2	0	6	5	1	
14	27	7	4	0	0	1	5	5	0	3	
15	23	9	6	1	2	3	2	6	3	3	
23	26	6	7	1	2	6	4	2	6	4	
15	13	4	5	1	1	4	6	3	2	8	
11	11	5	2	4	1	0	8	8	1	2	
12	11	11	5	2	3	5	4	3	1	8	
19	12	18	4	1	1	5	1	4	8	2	
14	9	8	3	4	2	4	2	1	6	2	
4	7	3	8	2	2	5	5	3	2	5	
18	7	2	3	1	3	2	1	7	5	6	
14	14	8	6	4	5	2	2	3	9	4	
7	9	5	4	3	2	3	0	4	6	1	
9	4	3	2	0	0	2	0	5	2	0	
5	10	5	1	1	4	1	2	7	1	1	
11	11	10	3	2	0	2	6	6	3	1	
21	6	8	4	1	3	7	4	1	1	0	
19	7	6	2	0	2	4	3	3	3	3	
17	9	4	2	1	5	5	2	2	2	3	
24	14	10	3	0	5	7	7	10	2	4	
16	5	3	3	1	3	10	2	9	4	4	
12	7	3	1	1	6	7	5	1	2	4	
24	3	3	2	1	2	1	2	3	4	2	
22	5	3	5	0	5	3	2	11	6	2	
14	8	7	3	1	3	2	3	12	5	1	
24	12	11	0	2	1	3	3	3	3	0	
9	4	5	1	3	2	4	8	2	3	1	
26	13	4	1	3	3	8	2	12	5	3	
11	11	6	0	6	5	1	3	6	6	2	
6	4	7	1	5	5	5	4	2	4	1	
18	11	6	0	4	0	5	2	0	3	1	

Datos con el ajuste de los ceros de los vuelos ilegales.

2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
5	4	12	3	1	3,14	2,7	6	1	2,7	3	1
8	8	4	2	1	7,11	2	3	5	1	5,41	1
26	6	10	2	1	3,17	1	1	2	1	1	2,21
13	6	3	5	4	3	1	11	4	2	2,29	2
4	7	3	14	2	6	2,8	2	4	6	6	1
4	3	6	2	7	3	2	4	2	4	1	3,24
7	4	10	3	2	4	3	1	2	5,59	3	1
14	4	8	1	2	2	2,79	4	4	2	5,38	
13	6	4	4	3	5	1	2	12	2	4	
8	7	5	3	1	1	4	2	2	2	2	
16	9	2	6	2,97	5,52	3	3	3	2	1	
6	6	15	2	1	3,18	6	1	3	1	3	
8	16	9	4	1	3	3	2	9	2	3	
19	13	10	3	2	3	4	7	3	3	2	
11	16	12	4	2	1	1	2	1	2	2	
6	7	8	1	2,99	4	2,86	3	2	4	5,34	
6	8	12	9,31	1	2	4	7	3	8	4	
11	19	6	9	1	4	2	5,93	6	5	1	
14	27	7	4	3,01	3,17	1	5	5	5,51	3	
15	23	9	6	1	2	3	2	6	3	3	
23	26	6	7	1	2	6	4	2	6	4	
15	13	4	5	1	1	4	6	3	2	8	
11	11	5	2	4	1	2,92	8	8	1	2	
12	11	11	5	2	3	5	4	3	1	8	
19	12	18	4	1	1	5	1	4	8	2	
14	9	8	3	4	2	4	2	1	6	2	
4	7	3	8	2	2	5	5	3	2	5	
18	7	2	3	1	3	2	1	7	5	6	
14	14	8	6	4	5	2	2	3	9	4	
7	9	5	4	3	2	3	5,85	4	6	1	
9	4	3	2	3,05	2,56	2	3,28	5	2	5,3	
5	10	5	1	1	4	1	2	7	1	1	
11	11	10	3	2	2,89	2	6	6	3	1	
21	6	8	4	1	3	7	4	1	1	3,4	
19	7	6	2	3,08	2	4	3	3	3	3	
17	9	4	2	1	5	5	2	2	2	3	
24	14	10	3	3,1	5	7	7	10	2	4	
16	5	3	3	1	3	10	2	9	4	4	
12	7	3	1	1	6	7	5	1	2	4	
24	3	3	2	1	2	1	2	3	4	2	
22	5	3	5	3,12	5	3	2	11	6	2	
14	8	7	3	1	3	2	3	12	5	1	
24	12	11	2,91	2	1	3	3	3	3	3,73	
9	4	5	1	3	2	4	8	2	3	1	
26	13	4	1	3	3	8	2	12	5	3	
11	11	6	2,93	6	5	1	3	6	6	2	
6	4	7	1	5	5	5	4	2	4	1	
18	11	6	2,94	4	3,03	5	2	5,63	3	1	

ANEXO 2. EL PROCESO DE IDENTIFICACIÓN DE UN MODELO ARIMA CUANDO EXISTEN OBSERVACIONES FALTANTES (Castaño, E. 1997)

Suponga que Z_t es una serie de tiempo que sigue un proceso ARIMA de la forma

$$Z_t = \frac{\theta(B)}{\delta(B)\phi(B)} a_t$$

Donde B es el operador usual de rezagos, $\theta(B)$ es el polinomio de medias móviles con todas sus raíces fuera del círculo unidad, $\phi(B)$ es el polinomio autoregresivo con sus raíces fuera del círculo unidad y que no tiene factores comunes con $\theta(B)$ y $\delta(B)$ es el polinomio de diferencias (que induce estacionariedad) con sus raíces sobre el círculo unidad. En primer lugar se considera el caso donde hay solamente una observación faltante. Para esto, suponga que la serie se observó durante T periodos y que no se encuentra disponible la observación para el periodo $t=t^*$. Una caracterización natural de un valor faltante es describirlo como una observación atípica aditiva. Esta caracterización ha sido empleada por varios autores entre los que se encuentran Peña y Maraval (1990) y Liu y Chen (1990). La razón es la siguiente: si se supone que en el periodo $t=t^*$ ocurre una observación atípica aditiva, se puede representar la serie observada como:

$$\begin{aligned} NZ_t &= Z_t && \text{si } t \neq t^* \\ &= AI_t^{t^*} + Z_t && \text{si } t = t^* \end{aligned}$$

Donde A indica la cantidad de desviación desde el verdadero valor de Z_{t^*} , y la variable $I_t^{(t^*)}$ toma el valor uno cuando $t=t^*$ y de cero en otro caso.

En este caso, Chen y Liu (1990) mostraron que el valor ajustado de NZ_{t^*} (es decir, después de remover el efecto atípico sobre NZ_{t^*}) es:

$$NZ_{t^*}^* = \left\{ \sum_{j=1}^{t-1} \left[\sum_{k=j}^{n-t^*+j} \pi_k \pi_{k-j} \right] NZ_{t^*-j} + \sum_{j=1}^{n-t^*} \left[\sum_{k=j}^{n-t^*} \pi_k \pi_{k-j} \right] NZ_{t^*+j} \right\} / \sum_{j=0}^{n-t^*} \pi_j^2 \quad (1)$$

Donde los coeficientes π son obtenidos del polinomio auto-regresivo $\pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \dots = \phi(B)/\theta(B)$. Si la serie no es estacionaria entonces la parte autorregresiva debe incluir el operador que induce estacionariedad.

De acuerdo con (1), el valor interpolado $NZ_{t^*}^*$ está basado en las observaciones de la serie anteriores y posteriores de NZ_{t^*} , es decir, en los valores anteriores y posteriores de la serie original Z a la observación faltante Z_{t^*} . Por lo tanto el valor ajustado no tiene nada que ver con la observación atípica NZ_{t^*} .

El resultado anterior sugiere que se puede estimar un valor faltante en una serie de tiempo tratándole como si fuera una observación atípica aditiva.

Por tanto, si se conoce a $\theta(B)$, $\phi(B)$ y $\delta(B)$, el procedimiento para estimar el valor desconocido de Z_{t^*} consiste en asignar un valor atípico cualquiera a la observación faltante y estimar el modelo intervenido:

$$NZ_t = A I_t^{(t^*)} + \frac{\theta(B)}{\delta(B)\phi(B)} a_t ,$$

Donde NZ_t es igual a la serie Z_t con la observación faltante reemplazada por un valor atípico, A es el coeficiente que indica el impacto de la observación atípica aditiva sobre el nivel medio de la serie en el periodo $t=t^*$, y donde $I_t^{(t^*)}$ es una variable indicadora con un uno en el periodo $t=t^*$ y ceros en los demás periodos. La estimación óptima de Z_{t^*} es el valor predicho de NZ_{t^*} menos la estimación de A (Box y Tiao, 1975).

De igual forma se procedería si la serie tuviera m observaciones perdidas no consecutivas en los periodos T_1, T_2, \dots, T_m : A cada periodo donde se desconoce la observación se asigna un valor atípico y en el modelo

$$N Z_t = \sum_{j=1}^m A_j I_t^{(T_j)} + \frac{\theta(B)}{\delta(B)\phi(B)} a_t \quad (2)$$

La estimación de NZ_{T_j} se obtiene como la predicción de NZ_{T_j} menos la estimación de A_j en el modelo anterior, para $j=1,2,\dots,m$.

Sin embargo, el problema en la práctica es más complicado pues, en general, no se dispone del conocimiento de $\theta(B)$, $\phi(B)$ y $\delta(B)$ y se debe tratar de identificarlos a partir de la información incompleta propuesta para un modelo donde existen observaciones faltantes:

i) Aproxime a Z_t usando un proceso puro autorregresivo de orden alto. En la elección del orden se deben tener en cuenta la frecuencia del periodo de observación, y la clase de proceso (estacional o no). En otras palabras Z_t puede aproximarse como el proceso autorregresivo puro:

$$Z_t = \frac{1}{\delta(B)\phi'(B)} a_t \quad (3)$$

Donde $\phi'(B)$ es el polinomio autorregresivo del orden seleccionado. Para la aproximación de un proceso ARIMA $(p,1,q)$, Said y Dickey (1984) consiguieron el siguiente resultado: Todo proceso ARIMA $(p,1,q)$ puede ser adecuadamente aproximado por medio de un proceso ARIMA $(n,1,0)$, donde $n \leq T^{1/3}$

ii) A cada una de las observaciones faltantes asigne un valor que sea atípico y ajuste el modelo

$$N Z_t = \sum_{j=1}^m A_j I_t^{(T_j)} + \frac{1}{\delta(B)\phi'(B)} a_t \quad (4)$$

Estime el modelo anterior y haga un análisis de residuales para verificar la buena aproximación a la estructura de Z_t del modelo estimado. Si los residuales se comportan como ruido blanco el orden elegido aproxima adecuadamente la estructura del proceso.

Una estimación preliminar de las observaciones faltantes está definida por los valores predichos del modelo aproximado estimado y corregido por la estimación del respectivo impacto. Esta estimación tiene en cuenta la estructura de autocorrelación aproximada de la serie.

iii) Reemplace los valores atípicos de la serie por los valores estimados preliminares. Sobre esta serie pueden emplearse las técnicas de identificación de Box-Jenkins para los modelos ARIMA.

iv) Suponga que el modelo identificado es de la forma

$$Z_t = \frac{1}{\delta(B)\phi(B)} a_t \quad (5)$$

Para obtener una estimación más refinada de las observaciones faltantes ajuste el modelo

$$N Z_t = \sum_{j=1}^m A_j I_t^{(T_j)} + \frac{1}{\delta(B)\phi(B)} a_t \quad (6)$$

El valor predicho de NZ para el periodo T_j menos la estimación de A_j es una estimación óptima de Z_{T_j} , $j=1,2,\dots,m$.