

MODELO DE PRONÓSTICO DE HORAS DE VUELO PARA LAS AERONAVES
B-350 Y SA-237 DE LA FUERZA AÉREA COLOMBIANA

HAROLD ERASO CASTILLO
ANDRÉS EDUARDO PEÑA
MARIBEL TIQUE

FUNDACIÓN UNIVERSITARIA LOS LIBERTADORES
ESPECIALIZACIÓN EN ESTADÍSTICA APLICADA
BOGOTÁ D.C.
2015

MODELO DE PRONÓSTICO DE HORAS DE VUELO PARA LAS AERONAVES B-
350 Y SA-237 DE LA FUERZA AÉREA COLOMBIANA

HAROLD ERASO CASTILLO
ANDRÉS EDUARDO PEÑA
MARIBEL TIQUE

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

Asesor

DIEGO FERNANDO LEMUS POLANÍA
M.Sc. en Estadística
Ingeniero Industrial

FUNDACIÓN UNIVERSITARIA LOS LIBERTADORES
ESPECIALIZACIÓN EN ESTADÍSTICA APLICADA
BOGOTÁ D.C.
2015

Nota de Aceptación

Firma del presidente del jurado

Firma del jurado

Firma del jurado

Bogotá D.C., Junio de 2015

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.
INTRODUCCIÓN	15
1. PLANTEAMIENTO DEL PROBLEMA	17
1.1 FORMULACIÓN	17
1.2 OBJETIVOS	18
1.2.1 Objetivo general	16
1.2.2 Objetivos específicos	18
1.3 JUSTIFICACIÓN	18
2. MARCO DE REFERENCIA	20
2.1 GENERALIDADES	20
2.2 PROCESO DE GESTIÓN DE MATERIALES PARA LAS AERONAVES	21
2.2.1. Primera estrategia:	22
2.2.2. Segunda estrategia	22
2.3 METODOLOGÍA DE ADQUISICIÓN DE DATOS DE HORAS DE VUELO	23
3. METODOLOGÍA ESTADÍSTICA CONSIDERADA	24
3.1 PROCESOS ESTACIONARIOS	24
3.2 PROCESOS DE MEMORIA LARGA	25
3.3 MODELOS AUTORREGRESIVOS INTEGRADOS DE MEDIA MÓVIL ARIMA (p,d,q)	25
3.3.1 Definición y propiedades	25
3.3.2. Metodología Box Jenkins para identificación de modelos Arima (p,d,q)	26
3.4 PRUEBAS DE RAÍZ UNITARIA	27
3.4.1 Prueba Dickey-Fuller	27
3.4.2. Prueba de Phillips – Perrón (PP)	29

3.4.3. Prueba Zivot-Andrews	30
3.5. MODELO ARFIMA (P,D,Q)	31
3.5.1. Propiedades del proceso ARFIMA (p,d,q)	32
3.6. PRUEBAS DE RAÍZ FRACCIONAL ESTACIONARIA	33
3.6.1. Prueba Dickey- Fuller Fraccional Aumentada	33
3.6.3 La prueba de robinson	35
3.6.4. Prueba de Castaño, Gómez Y Gallón	35
3.7 METODOLOGÍA DE CASTAÑO ET. AL (2008) PARA IDENTIFICACIÓN DE MODELOS ARFIMA (p,d,q)	36
3.8 MÉTODOS DE PRONÓSTICO	37
3.8.1 Método Exacto.	37
3.8.2 Método Bootstrap	37
4. ANÁLISIS DE DATOS	38
4.1 AERONAVESA-237	38
4.1.1. Pruebas de raíz unitaria	40
4.1.2. Pruebas de raíz fraccional estacionaria	43
4.1.3. Procedimiento de identificación y estimación - modelos ARFIMA (p,d,q)	44
4.2 AERONAVE B-350	53
4.2.1 Pruebas de raíz unitaria	56
5. CONCLUSIONES	69
BIBLIOGRAFÍA	71

LISTA DE FIGURAS

	Pág
Figura 1. Equipos SA2-37 y B-350	19
Figura 2. Suboficial realizando mantenimiento	21
Figura 3. Secuencia lógica factores críticos	22

LISTA DE GRÁFICOS

	pág.
Gráfico 1. Serie Original Horas de vuelo SA-237	38
Gráfico 2. Curva de comparación de transformaciones de Potencia Maximización de la log-verosimilitud.	39
Gráfico 3. Serie transformada horas de vuelo SA-237	39
Gráfico 4. ACF-Serie transformada Equipo SA-237	40
Gráfico 5. ACF Serie Diferenciada Fraccionalmente	46
Gráfico 6. PACF Serie Diferenciada Fraccionalmente	46
Gráfico 7. EACF Serie Diferenciada Fraccionalmente	47
Gráfico 8. Serie transformada, valores ajustándose Intervalos de confianza del 95% para el equipo SA-237	51
Gráfico 9. Pronósticos obtenidos e intervalos de confianza método exacto	52
Gráfico 10. Pronósticos obtenidos e intervalos de confianza Bootstrap	53
Gráfico 11. Serie Original de vuelo B-350	53
Gráfico 12. Curva de comparación de transformaciones de Potencias Maximización de la log-verosimilitud.	54
Gráfico 13 Serie transformada horas de vuelo B-350	55
Gráfico 14. ACF- Serie original horas de vuelo B-350	55
Gráfico 15. ACF Serie Diferenciada Fraccionalmente	61

Gráfico 16. PACF Serie Diferenciada Fraccionalmente	61
Gráfico 17. EACF Serie diferenciada fraccionalmente	62
Gráfico 18. Serie transformada, valores ajustados e intervalo de confianza del 95% del modelo ajustado para el equipo B-350	66
Gráfico 19. Pronósticos obtenidos e intervalos de confianza método exacto	68
Gráfico 20. Pronósticos obtenidos e intervalos de confianza Bootstrap	68

LISTA DE TABLAS

	pág.
Tabla 1. Valores Críticos Prueba ADF	41
Tabla 2. Regresión prueba ADF sin tendencia Determinística	41
Tabla 3. Valores críticos Prueba PP	41
Tabla 4. Regresión prueba PP sin Tendencia Determinística	42
Tabla 5. Valores Críticos prueba Zivot-Andrews	42
Tabla 6. Regresión prueba de Zivot-Andrews- Modelo A	42
Tabla 7. Prueba de Nyblom-Hansen	43
Tabla 8. Pruebas de Raíz fraccional	44
Tabla 9. Modelo aproximado ARFIMA (6,d,0)	44
Tabla 10. Modelo aproximado ARFIMA (0,d,0)	45
Tabla 11. Prueba de Ljung-Box para residuales modelo ARFIMA (0,d,0)	45
Tabla 12. Resultados de los criterios AIC y BIC selección de modelos	47
Tabla 13. Prueba de Ljung-Box para residuales modelo ARFIMA (0,d,0)	48
Tabla 14. Prueba Durbin-Watson para residuales modelo ARFIMA (0,d,0)	48
Tabla 15. Prueba de Shapiro-Wilk para residuales modelo ARFIMA (0,d,0)	49
Tabla 16. Prueba de Jarque-Bera para residuales modelo ARFIMA (0,d,0)	49
Tabla 17. Prueba de Kolomogorov-Smirnov para residuales modelo ARFIMA (0,d,0)	49
Tabla 18. Prueba de D`Agostino para residuales modelo ARFIMA (0,d,0)	50

Tabla 19. Pronóstico e intervalo de confianza paramétrico	52
Tabla 20. Pronóstico e intervalo de confianza vía Bootstrap	52
Tabla 21. Valores CRÍTICOS PRUEBA ADF	56
Tabla 22. Regresión prueba ADF sin Tendencia Determinística	56
Tabla 23. Valores críticos prueba PP	57
Tabla 24. Regresión prueba PP sin Tendencia Determinística	57
Tabla 25. Valores críticos prueba de Zivot-Andrews	58
Tabla 26.. Regresión prueba de Zivot-Andrews-Modelo A	58
Tabla 27. Prueba de Nyblom-Hansen	58
Tabla 28. Prueba de raíz Fraccional	59
Tabla 29. Modelo aproximado ARFIMA (6,d,0)	60
Tabla 30. Modelo aproximado ARFIMA (0,d,0)	60
Tabla 31. Prueba de Ljung-Box para residuales modelo ARFIMA (0,d,0)	60
Tabla 32. Resultado de los criterios AIC y BIC selección de modelos	62
Tabla 33. Prueba de Ljung-Box para residuales modelos ARFIMA (0,d,0)	63
Tabla 34. Prueba de Durbin-Watson para residuales modelo ARFIMA (0,d,0)	63
Tabla 35. Prueba de Shapiro-Wilk para residuales modelo ARFIMA (0,d,0)	64
Tabla 36. Prueba de Jarque-Bera para residuales modelo ARFIMA (0,d,0)	64
Tabla 37. Prueba de Kolomogorov-Smirnov para residuales modelo ARFIMA (0,d,0)	64
Tabla 38. Prueba de D`Agostino para residuales modelo ARFIMA (0,d,0)	65
Tabla 39. Prueba de Nyblom-Hansen para residuales del modelo ajustado	65

Tabla 40. Pronóstico e intervalo de confianza paramétrico 67

Tabla 41. Pronóstico e intervalo de confianza vía Bootstrap 67

RESUMEN

La inteligencia aérea es la capacidad de recolectar, procesar, explotar y difundir información precisa y oportuna por medio de aeronaves, con la finalidad de planear y conducir operaciones militares exitosas. Debido a la inexactitud actual en los pronósticos del número de horas de vuelo de los equipos de la FAC se presentan una serie de complicaciones a la hora de ejecutar el presupuesto requerido pues generalmente resulta insuficiente. Este trabajo identifica y construye un modelo de series temporales que permita pronosticar el número de las horas de vuelo de las aeronaves de la Fuerza Aérea Colombiana. Como resultado se obtiene un modelo ARFIMA (p,d,q) que resulta adecuado para el cumplimiento del objetivo de esta investigación.

Palabras claves: Equipos, Aeronaves, Modelo ARFIMA, Pruebas de raíz unitaria, pronóstico.

ABSTRACT

Air intelligence is the ability to collect, process, exploit and spread accurate and proper information by aircrafts, in order to plan and conduct successful military operations. Due to current forecasts inaccuracy in the flight time of FAC teams, many complications are presented when executing the required budget which is generally insufficient given an underprediction. This work identifies and builds a time series model to predict flight hours of Colombian Air Force aircrafts. The result is an ARFIMA (p, d, q) model which is suitable for this research objective.

Keywords: Devices, aircrafts, ARFIMA model, unitary root tests, prediction.

INTRODUCCIÓN

Desde que el hombre comenzó a volar apoyado por artefactos estos han hecho presencia en los diferentes conflictos humanos. Por ejemplo, en la guerra civil norteamericana se utilizaron globos para la observación y transmisión de información al frente de batalla, es decir, fueron empleados para hacer inteligencia militar. Al ver los excelentes resultados obtenidos al utilizar equipos aéreos, las diferentes naciones empezaron una carrera tecnológica y armamentista con el fin de establecer dominio aéreo en el globo terráqueo.

La historia aérea Colombiana comenzó cuando algunos empresarios trajeron al país, en 1911, el primer avión (no militar) con el fin de realizar un espectáculo aéreo. La terminación de la primera guerra mundial, marcó el inicio de la aviación comercial en muchos países en vía de desarrollo, ya que se podía comprar aviones usados en la guerra a muy bajos costos y existía una oferta de pilotos cesantes en Europa (Revista credencial, 2011). El 31 de diciembre de 1919, se expide en Colombia la ley “arma de la aviación” dando origen a la Fuerza Aérea Colombiana (cuyo acrónimo es FAC) y su primer equipo fue un Cuadro G-3 (CIHFA,SF)¹. Desde entonces La FAC, ha trabajado para el cumplimiento de su misión:

La Fuerza Aérea Colombiana ejerce y mantiene el dominio del espacio aéreo, conduce operaciones aéreas, para la defensa de la soberanía, la independencia, la integridad del territorio nacional, el orden constitucional y el logro de los fines del Estado.(S.F)

Para la consecución del objetivo mencionado, la FAC cuenta con una serie de equipos diseñados para el cumplimiento de diferentes misiones entre las que se resaltan: Logística, inteligencia y combate. La FAC mide el número de horas de vuelo de sus aeronaves empleando el software denominado “Sistema Integrado de Información Operacional de Comando y Control” (SIIOC2), el cual registra a diario la información de interés en el estudio.

El presente trabajo tiene como finalidad identificar un modelo de series de tiempo que permita pronosticar el número de horas de vuelo de algunos equipos militares de la FAC (puntualmente, los aviones B-350 y SA-237) diseñados con la finalidad de realizar inteligencia aérea en el país y así contribuir en la planeación adecuada en el uso de estas aeronaves.

El documento tiene la siguiente estructura, en la primera sección se presenta el planteamiento del problema, la formulación de la investigación, los objetivos y

¹ Centro de Investigaciones Históricas de la Fuerza Aérea.

justificación. La segunda parte está compuesta por el marco referencial, donde se exponen algunos elementos de la metodología usada por la FAC en sus operaciones de vuelo.

En el tercer apartado se considera la teoría estadística utilizada para la identificación del modelo de pronóstico del número de horas de vuelo. En la cuarta parte se presentan los resultados del modelo estimado y finalizando el documento se presentaran las conclusiones y recomendaciones del presente trabajo.

1. PLANTEAMIENTO DEL PROBLEMA

La inteligencia aérea es la capacidad integrada a recolectar, procesar, explotar y difundir información precisa y oportuna; para planear y conducir operaciones militares exitosas. La búsqueda de la información, referente a la inteligencia aérea se realiza empleando principalmente aeronaves². (Manual de operaciones Aéreas, 2005).

Actualmente el Departamento Planeación y Estadística de la FAC es el encargado de planificar el número de horas de vuelo de sus equipos. Hoy por hoy, la metodología empleada para pronosticar valores futuros de la variable de interés es el promedio de horas de vuelo por aeronave, el cual generalmente no da muy buenos resultados. Debido a la inexactitud de los pronósticos actuales se presentan una serie de complicaciones a la hora de ejecutar el presupuesto requerido pues generalmente resulta insuficiente.

Por lo anterior, la jefatura militar debe solicitar constantemente adiciones presupuestales que desencadenan procesos administrativos dispendiosos (tramitología) que en casos extremos pueden interrumpir las operaciones militares que se están realizando; situación que genera una desventaja estratégica en la lucha contra el narcotráfico y los ejércitos disidentes que operan en el país.

Los equipos escogidos para el desarrollo del modelo adecuado para pronosticar el número de horas de vuelo de aeronaves, son de vital importancia para el desarrollo de operaciones militares colombianas debido al rol que desempeña cada uno de ellos. En primer lugar está el SA-237, este avión desarrolla misiones de vigilancia encubiertas ya que posee tecnología que le permite operar en modo silencioso, y con una serie de sensores puede buscar y almacenar información que posteriormente se utiliza en la inteligencia militar. El segundo es el B-350, este avión desempeña un papel crucial en el tiempo real del combate, apoyando a equipos de tierra con información táctica y estratégica (webinfomil, 2012).

1.1 FORMULACIÓN

¿Existe algún modelo de series temporales que permita pronosticar adecuadamente el número de horas de vuelo de los equipos de inteligencia B-350 y SA-237 de la FAC y de esta manera garantizar el funcionamiento continuo y oportuno de dichas aeronaves?

² O-MAOPA. Manual de Operaciones Aéreas. Colombia:. Departamento de Planeación y Estadísticas, 2005. 120 p.

1.2 OBJETIVOS

1.2.1 Objetivo general. Identificar y construir un modelo en series de tiempo que permita realizar pronósticos de número de horas de vuelo para cada una de las aeronaves de la Fuerza Aérea Colombiana consideradas.

1.2.2 Objetivos específicos

- Verificar si hay evidencia empírica de tendencias deterministas y/o estocásticas.
- Verificar si la varianza del proceso requiere de alguna transformación Box-Cox para estabilizar la varianza incondicional.
- Determinar si hay evidencia empírica de algún componente estacional en la serie.
- Identificar el componente de corto plazo más adecuado para modelar el proceso.
- Validar los supuestos sobre los residuales del modelo identificado.
- Realizar un pronóstico de horas de vuelo para las aeronaves B-350 y SA-237 para un mes.

1.3 JUSTIFICACIÓN

La realización de este trabajo de investigación permite la utilización de conceptos de series temporales con el fin de pronosticar el número de horas de vuelo de los equipos de inteligencia B-350 y SA-237 de la Fuerza Aérea Colombiana (Ver Figura 1). De forma concreta, se identificará un modelo de series de tiempo que le permita al Departamento de Planeación y Estadística de la FAC planificar de manera más precisa los recursos financieros que serán empleados en las operaciones militares aéreas que requiera la entidad para guardar la soberanía aérea de Colombia.

Figura 1. Equipos SA2-37 y B-350



Fuente: Fuerza Aérea Colombiana, Jefatura de Operaciones Aéreas

2. MARCO DE REFERENCIA

2.1 GENERALIDADES

El planeamiento logístico operacional de forma simultánea ha contribuido significativamente en el éxito de las operaciones aéreas durante las últimas décadas. Para evaluar la contribución de este factor en el concepto operacional de la Fuerza Aérea Colombiana (FAC), definido como la realización de misiones aéreas en los diferentes equipos militares, se debe citar la incidencia histórica de la logística en las organizaciones militares.

En los últimos años, la logística ha evolucionado a pasos agigantados en Colombia, sin embargo, es necesario remontarse al 12 de junio de 1819, época de la Campaña Libertadora, cuando en Tame, Arauca, nació la logística militar del país. Al salir de Venezuela y cruzar la llanura colombiana, el Libertador Simón Bolívar se encontró con el General Francisco de Paula Santander a quien asignó el mantenimiento de los equipos y la consecución de intendencia para el cruce del Páramo de Pisba, con lo que se creó el primer batallón de mantenimiento.

Después de obtener la libertad de la Nueva Granada, lo nombró Vicepresidente y encargado del Poder Ejecutivo, con la tarea de soportar el apoyo logístico para la Batalla de Carabobo con la que selló la libertad de Venezuela. Luego dio inicio a la Campaña del Sur con la Batalla de Bomboná en Nariño y posteriormente, con la Batalla de Pichincha para dar lugar a la libertad del Ecuador. Finalmente, con las Batallas de Junín y Ayacucho, Perú obtuvo su libertad, así fue erradicado el ejército español del continente americano.

Curioso es el tránsito logístico que tenían todas las mercancías que llegaban del nuevo mundo, las cuales tocaban tierra granadina en el Puerto de Cartagena y a través del Canal del Dique, iniciaban su navegación por el gran río de la Magdalena hasta el Puerto de Honda, desde allí, a lomo de mula, llegaban las mercancías a la Sabana de Bogotá para dar gusto a las excentricidades de la sociedad santafereña.

Por tanto, al ser un factor crítico dentro del cumplimiento de su misión, la FAC proyecta anualmente el número de horas de vuelo que se necesitan para cada una de sus aeronaves en la siguiente vigencia fiscal. Debido a la gran cantidad de variables bajo análisis se presentan problemas sistemáticos en la estimación de la cifra requerida por aeronave; ya que algunas de estas variables dependen de factores exógenos a la planeación realizada por la institución como son las diferentes situaciones de orden público que se presentan en el país, situaciones económicas y políticas del estado.

La Fuerza Aérea Colombiana (FAC) cuenta en la actualidad con más de 250 aeronaves que requieren de un presupuesto anual para su óptimo

funcionamiento, el cual depende fundamentalmente del número de horas de vuelo pronosticadas para la siguiente vigencia (año fiscal), el número de pilotos disponibles y de los planes de mantenimiento por equipo.

La variable de interés en el estudio (horas de vuelo por aeronave) está compuesta por dos factores: mantenimiento y combustible. La etapa de mantenimiento (Ver Figura 2) consiste en la cantidad de repuestos necesarios para que un avión funcione correctamente, por tanto, los planes de mantenimiento preventivo están asociados directamente a la cantidad de horas que ha volado el equipo. El combustible es la cantidad de libras o galones que gasta un equipo en una hora de vuelo; por lo general este dato es suministrado por la casa fabricante de los aviones.

Figura 2. Suboficial realizando mantenimiento



Fuente: Fuerza Aérea Colombiana, Jefatura de Operaciones Logísticas Aeronáuticas

A la fecha, los métodos utilizados por la jefatura de operaciones aéreas para realizar los pronósticos de la cantidad de horas de vuelo necesarias por aeronave han sido prácticas basadas en la experiencia (Ad Hoc) o utilizando promedios ponderados; lo cual conlleva a pronósticos desfasados e inexactos que se reflejan en la ejecución de los planes operativos institucionales. Lo anterior, se ve reflejado en la escasez de material aeronáutico (repuestos) y en la insuficiente cantidad de combustible de aviación para cumplir con la misión asignada a la FAC por la constitución nacional.

2.2 PROCESO DE GESTIÓN DE MATERIALES PARA LAS AERONAVES

Una vez realizado el pronóstico de la cantidad de horas de vuelo por aeronave, se inicia todo el proceso logístico para la adquisición del material aeronáutico necesario. Para tal fin, la institución diseñó dos estrategias:

2.2.1. Primera estrategia: Sincronización de los procesos logísticos críticos

Consiste en la implementación del proyecto conocido como el “Programa Anual de Soporte Logístico (PASL)”. Este programa permite la estandarización de material para el mantenimiento programado reflejado por medio de:

- Listas maestras³,
- Listas A,
- Listas de kits de reparación,
- Listas de kits TBO,
- Listas de llantas,
- Listas de equipo electrónico, entre otras.

También implica el análisis de todas las variables que hacen parte del planeamiento logístico que va desde la responsabilidad logística hasta el diseño de escenarios de horas a volar.

2.2.2. Segunda estrategia: Unidades bajo factores críticos. Una vez diseñado el proyecto, su implementación activa se realiza por medio de la siguiente secuencia lógica:

Figura 3. Secuencia lógica factores críticos



Fuente: Fuerza Aérea Colombiana, Jefatura de Operaciones Logísticas Aeronáuticas

³**Lista maestra:** Se relacionan todos los repuestos consumibles que se requieren en las inspecciones y cambios mandatorios de las aeronaves.

Listas A: Consolidado de repuestos de consumo que se requieren bajo el análisis de recurrencia y de fallas.

Kits de reparación: Relación de todos aquellos repuestos que se requieren para la reparación de un componente.

Kits TBO: Componentes reparables que requieren mantenimiento.

Listas de llantas: Consolidación de las llantas requeridas para el periodo.

Listas de equipo electrónico: Elementos electrónicos nuevos o reparables necesarios para una aeronave.

Lo anterior, con el fin de agotar el material aeronáutico, obteniendo un alto impacto en aprovisionamiento y optimización del presupuesto mediante la identificación de material.

- **Priorización:** Corresponden al orden y priorización en que son requeridos los elementos.
- **Consolidación:** Recibir y recolectar toda la información reportada por los Jefes de grupo.
- **Costos por equipo:** Son los cargos asociados clara y directamente con la adquisición de repuestos de las aeronaves.
- **Estandarización:** Proceso mediante el cual se realiza una actividad previamente establecida.
- **Integración de pronóstico:** Anticipación de las necesidades logísticas por equipo.
- **Definición de reservas:** Compromisos legalmente contraídos que al cierre de la vigencia fiscal no se han atendido por no haberse completado las formalidades necesarias que hagan exigible el pago.
- **Niveles de stock en centros logísticos:** Es la cantidad de repuestos necesarios con los que debe contar una Basa Aérea para el buen funcionamiento de las aeronaves.

2.3 METODOLOGÍA DE ADQUISICIÓN DE DATOS DE HORAS DE VUELO

Para hacer posible esta investigación se necesitaba un histórico temporal del número de horas de vuelo realizadas por los equipos de inteligencia aérea B-350y SA-237. El registro histórico fue suministrado por la jefatura de operaciones aéreas de la FAC en un documento Excel, el cual fue exportado del software conocido Sistema Integrado de Información Operacional de Comando y Control (SIIOC2). Este sistema de información es alimentado por los pilotos de las diferentes aeronaves quienes ingresan todos los datos de interés de la misión (una vez finalizada) entre los cuales se encuentra la duración del vuelo.

Los datos suministrados por la Fuerza Aérea Colombiana para la presente investigación son a partir del 01 de enero de 2010 hasta el 07 de noviembre de 2013, con una periodicidad diaria (Lunes - Domingo) para un total de 1407 observaciones de cada equipo.

3. METODOLOGÍA ESTADÍSTICA CONSIDERADA

3.1 PROCESOS ESTACIONARIOS

Según (Wei, 2006), si Z_t es un proceso estocástico de valores reales, la función media del proceso se define como sigue⁴:

$$\mu_t = E(Z_t) \quad (3.1.1)$$

la función varianza del proceso

$$\sigma_t^2 = E(Z_t - \mu_t)^2 \quad (3.1.2)$$

la función de covarianza entre Z_{t_1} y Z_{t_2}

$$\gamma(t_1, t_2) = E(Z_{t_1} - \mu_{t_1})(Z_{t_2} - \mu_{t_2}) \quad (3.1.3)$$

y la función de correlación entre Z_{t_1} y Z_{t_2}

$$\rho(t_1, t_2) = \frac{\gamma(t_1, t_2)}{\sqrt{\sigma_{t_1}^2} \sqrt{\sigma_{t_2}^2}} \quad (3.1.4)$$

Para un proceso estacionario, ya que la función de distribución es la misma para todo t , la función media $\mu_t = \mu$ es una constante, por tanto $E(|Z_t|) < \infty$. Igualmente si $E(|Z_t^2|) < \infty$, entonces $\sigma_t^2 = \sigma^2$, para todo t y por tanto es constante.

De otra parte, ya que $F_{Z_{t_1}, Z_{t_2}}(x_1, x_2) = F_{Z_{t_1+k}, Z_{t_2+k}}(x_1, x_2)$ para cualquier entero t_1 , t_2 y k tenemos que:

$$\begin{aligned} \gamma(t_1, t_2) &= \gamma(t_1 + k, t_2 + k) \\ \rho(t_1, t_2) &= \rho(t_1 + k, t_2 + k) \end{aligned} \quad (3.1.5)$$

Dejando $t_1 = t - k$ y $t_2 = t$, obtenemos:

$$\gamma(t_1, t_2) = \gamma(t - k, t) = \gamma(t, t + k) = \gamma_k \quad (3.1.6)$$

⁴ WEI, William. Time Series Analysis, Univariate and Multivariate Methods. México: Ed. Pearson, 2006. 614 p.

$$\rho(t_1, t_2) = \rho(t - k, t) = \rho(t, t + k) = \rho_k$$

En conclusión, para un proceso estrictamente estacionario con el primer y segundo momentos finitos, la covarianza y la correlación entre Z_t y Z_{t+k} depende solo de la diferencia de tiempo k .

3.2 PROCESOS DE MEMORIA LARGA

El proceso de memoria larga en una serie temporal se entiende como la presencia de una dependencia no despreciable entre observaciones que están separadas entre ellas por largos periodos de tiempo. Muchos autores (Beran (1993), Palma (2007), entre otros) han conceptualizado la propiedad de memoria larga en función del comportamiento asintótico de la función de autocovarianza del proceso⁵.

Según Lemus y Castaño(2013): “Si $\{Z_t: t \in \mathbb{Z}^+\}$ es un proceso estacionario con función de autocovarianza $\gamma(k)$, se puede afirmar que este proceso tiene memoria larga si y solo si:

$$\sum_{-\infty}^{\infty} |\gamma(k)| = \infty$$

Hay que subrayar que la conceptualización del proceso de memoria larga es asintótica ya que informa acerca de la tasa de convergencia hacia 0 de las correlaciones cuando k tiende a infinito”.

3.3 MODELOS AUTORREGRESIVOS INTEGRADOS DE MEDIA MÓVIL ARIMA (p,d,q)

3.3.1 Definición y propiedades. Se dice que un proceso estocástico $\{Z_t: t \in \mathbb{Z}^+\}$ sigue un proceso ARIMA (p,d,q) si es una solución a la ecuación:

$$\phi_p(B)(1 - B)^d Z_t = \theta_0 + \theta_q(B)\alpha_t \quad 3.3.1$$

Donde $\phi_p(B)$ es el polinomio autoregresivo estacionario de orden p y $\theta_q(B)$ es el polinomio de media móvil de orden q invertible, ambos en términos del operador de rezagos B ; los mencionados polinomios no tienen raíces comunes entre sí. El término θ_0 tiene diferentes papeles según el valor del parámetro de diferenciación d . Si $d = 0$ el proceso en [3.3.1] corresponde a un proceso ARMA (p,q) estacionario e invertible, en este caso θ_0 está relacionado con la media del proceso. Si $d \in \mathbb{Z}^+$ el proceso en [3.3.1] corresponde a un proceso no estacionario homogéneo con d raíces unitarias, en este caso θ_0 es llamado el

⁵ BERAN, Jan; PALMA, Maximum. Likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. Estados Unidos de América: Journal of the Royal Statistical Society, 1993. 57(4), p. 659–672

término de tendencia determinística del proceso. El proceso a_t es una sucesión de variables aleatorias no observables, con media cero y varianza finita σ_a^2 – Ver Guerrero (1991), Wei (2006), Brockwell y Davis (2006)-.

El resultado del proceso de la diferenciación adecuada de una serie de tiempo homogénea no estacionaria no es necesariamente ruido blando⁶ (Wei, 2006). Normalmente la serie diferenciada $(1 - B)^d Z_t$ continúa un proceso estacionario general ARMA (p,q), ya que sus partes (AR) y (MA) no poseen factores comunes.

3.3.2. Metodología Box Jenkins para identificación de modelos 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 las siguientes cuatro etapas:

- **Identificación de un posible modelo dentro de la clase de modelos ARIMA (p,d,q).** En esta etapa se determinan 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.
- **Estimación de los parámetros involucrados en el modelo identificado.** Nota: Esta etapa presupone que ya se tiene identificado un modelo tentativo.
- **Verificación de que el modelo proporciona un ajuste adecuado y de que los supuestos básicos implícitos en el modelo.** En esta etapa se realizan pruebas de incorrelación serial, de homocedasticidad y de normalidad sobre los residuales del modelo ajustado.
- **Uso del modelo para su finalidad.** En esta etapa se realizan los pronósticos del proceso bajo estudio o se realiza el control del proceso.

Para la identificación del modelo ARIMA, se debe determinar si existe o no raíces unitarias en el proceso, entonces para este fin se emplean las pruebas de raíz unitaria que a continuación se describen:

⁶ WEI, William. Op. Cit. p. 19

3.4 PRUEBAS DE RAÍZ UNITARIA

3.4.1 Prueba Dickey-Fuller. En Dickey-Fuller (1979) se consideran tres pruebas distintas para determinar si hay existencia de una raíz unitaria en el polinomio autorregresivo del proceso, siempre bajo un supuesto de normalidad.

En la primera prueba utiliza el siguiente modelo:

$$Z_t = \rho Z_{t-1} + a_t$$

El cual bajo algunas operaciones algebraicas queda

$$\nabla Z_t = \gamma Z_{t-1} + a_t \quad \text{con } \gamma = (\rho - 1) \quad (3.5.1)$$

Según este modelo, el contraste de hipótesis para probar la existencia de raíz unitaria es el siguiente:

- $H_0: \gamma = 0$. Lo que implica que $\rho = 1$. En este caso la serie posee raíz unitaria y es no estacionaria homogénea.
- $H_1: \gamma < 0$ Lo que implica que $(\rho < 1)$. En este caso la serie no posee raíz unitaria

El estadístico de prueba es $\tau = \hat{\gamma}/se(\hat{\gamma})$ donde $\hat{\gamma}$ es el estimador de mínimos cuadrados de γ y $se(\hat{\gamma})$ es su error estándar estimado. La distribución de τ bajo H_0 no es la t -Student tradicional, pertenece a una clase de distribuciones no estándar y se encuentra tabulada para diferentes valores de n y niveles de significancia α . Si $\tau(\alpha, n)$ es el percentil α -inferior de dicha distribución se debe rechazar H_0 si el valor observado de τ es $\tau < \tau(\alpha, n)$.

En la segunda prueba utiliza este modelo:

$$Z_t = \beta_0 + \rho Z_{t-1} + a_t$$

El cual bajo algunas operaciones algebraicas queda

$$\nabla Z_t = \beta_0 + \gamma Z_{t-1} + a_t \quad \text{con } \gamma = (\rho - 1) \quad (3.5.2)$$

Según este modelo, el contraste de hipótesis para probar la existencia de raíz unitaria es el siguiente:

- $H_0: \gamma = 0$. Lo que implica que $\rho = 1$. En este caso la serie posee raíz unitaria y es no estacionaria homogénea.

- $H_1: \gamma < 0$ Lo que implica que $(\rho < 1)$. En este caso la serie no posee raíz unitaria

El estadístico de prueba es $\tau_\mu = \hat{\gamma}/se(\hat{\gamma})$ donde $\hat{\gamma}$ es el estimador de mínimos cuadrados de γ y $se(\hat{\gamma})$ es su error estándar estimado. La distribución de τ_μ bajo H_0 no es la $t - Student$ tradicional, pertenece a una clase de distribuciones no estándar y se encuentra tabulada para diferentes valores de n y niveles de significancia α . Si $\tau_\mu(\alpha, n)$ es el percentil α -inferior de dicha distribución se debe rechazar H_0 si el valor observado de τ_μ es $\tau_\mu < \tau_\mu(\alpha, n)$.

La tercera prueba considera el siguiente modelo:

$$Z_t = \beta_0 + \beta_1 t + \rho Z_{t-1} + a_t$$

El cual bajo algunas operaciones algebraicas queda

$$\nabla Z_t = \beta_0 + \beta_1 t + \gamma Z_{t-1} + a_t \quad \text{con } \gamma = (\rho - 1) \quad (3.5.3)$$

Según este modelo, el contraste de hipótesis para probar la existencia de raíz unitaria es el siguiente:

- $H_0: \gamma = 0$. Lo que implica que $\rho = 1$. En este caso la serie posee raíz unitaria y es no estacionaria homogénea.
- $H_1: \gamma < 0$ Lo que implica que $(\rho < 1)$. En este caso la serie no posee raíz unitaria

El estadístico de prueba es $\tau_\tau = \hat{\gamma}/se(\hat{\gamma})$ donde $\hat{\gamma}$ es el estimador de mínimos cuadrados de γ y $se(\hat{\gamma})$ es su error estándar estimado. La distribución de τ_τ bajo H_0 no es la $t - Student$ tradicional, pertenece a una clase de distribuciones no estándar y se encuentra tabulada para diferentes valores de n y niveles de significancia α . Si $\tau_\tau(\alpha, n)$ es el percentil α -inferior de dicha distribución se debe rechazar H_0 si el valor observado de τ_τ es $\tau_\tau < \tau_\tau(\alpha, n)$.

Ya visto eso, es necesario aclarar que no siempre es posible representar cualquier serie de tiempo usando los modelos anteriores donde $a_t \sim RBN(0, \sigma_a^2)$. Puede suceder que término de error del modelo este autocorrelacionado ocasionando que los estimadores no serán eficientes. Para solucionar este problema, Dickey Fuller (1979) proponen aumentar la prueba ampliando los modelos anteriormente vistos considerando el término de error como un proceso ARMA(p,q). Los modelos que permiten capturar la correlación serial remanente son:

$$\text{Modelo 1: } \nabla Z_t = \gamma Z_{t-1} + \sum_{i=1}^p \delta_i \nabla Z_{t-i} + a_t$$

$$\text{Modelo 2: } \nabla Z_t = \beta_0 + \gamma Z_{t-1} + \sum_{i=1}^p \delta_i \nabla Z_{t-i} + a_t \quad (3.5.4)$$

$$\text{Modelo 3: } \nabla Z_t = \beta_0 + \beta_1 t + \gamma Z_{t-1} + \sum_{i=1}^p \delta_i \nabla Z_{t-i} + a_t$$

Las pruebas de hipótesis, los estadísticos de prueba y los criterios de rechazo son los mismos que se presentaron previamente.

3.4.2. Prueba de Phillips – Perrón (PP). La prueba de Phillips-Perrón (1998) es una extensión del test de Dickey-Fuller Aumentado (ADF), diferenciándose de la ADF en cómo tratar la correlación serial y la heterocedasticidad en los errores. En particular, cuando la prueba de Dickey-Fuller aumentada usa una autorregresión paramétrica para aproximarse a una estructura ARMA de los errores en la regresión de prueba, el test de Phillips-Perrón (1988) no tiene en cuenta cualquier correlación serial en la regresión de prueba, este test de regresión está dado por:

$$\nabla Z_t = \gamma Z_{t-1} + \sum_{i=1}^p \delta_i \nabla Z_{t-i} + a_t \quad (3.5.5)$$

Donde se puede excluir la constante o incluir un término de tendencia, por lo tanto se encontraran dos estadísticos τ_ρ y τ_τ , calculados como:

$$\begin{aligned} \tau_\rho &= n(\hat{\rho}_n - 1) - \frac{1}{2} \frac{n^2 \hat{\sigma}^2}{s_n^2} (\hat{\lambda}_n^2 - \hat{\gamma}_{0,n}) \\ \tau_\tau &= \sqrt{\frac{\hat{\gamma}_{0,n} \hat{\rho}_n - 1}{\hat{\lambda}_n^2}} \frac{\hat{\rho}_n - 1}{\hat{\sigma}} - \frac{1}{2} (\hat{\lambda}_n^2 - \hat{\gamma}_{0,n}) \frac{1}{\hat{\lambda}_n} \frac{n \hat{\sigma}}{s_n} \\ \hat{\gamma}_{j,n} &= \frac{1}{n} \sum_{i=j+1}^n \hat{u}_i \hat{u}_{i-j} \\ \hat{\lambda}_n^2 &= \hat{\gamma}_{0,n} + 2 \sum_{j=1}^q \left(1 - \frac{j}{q+1}\right) \hat{\gamma}_{j,n} \\ s_n^2 &= \frac{1}{n-k} \sum_{i=1}^n \hat{u}_i^2 \end{aligned} \quad (3.5.6)$$

Donde u_i son los residuales OLS, k es el número de covariables en la regresión, q es el número de rezagos Newey-West usado en el cálculo de $\hat{\lambda}_n^2$, y $\hat{\sigma}$ es el error OLS estándar de $\hat{\rho}$.

Bajo la hipótesis nula $H_0: \gamma = 0$, implica que ($\rho = 1$), los estadísticos de la prueba Phillips-Perrón τ_ρ y τ_τ poseen la misma distribución asintótica que el estadístico de la prueba ADF y sesgo estadístico normalizado.

La fortaleza de la prueba de Phillips-Perrón (1988) a comparación de la de ADF es que el primer test es más robusto para formas generales de heterocedasticidad en el término de error a_t dado en la ecuación general (3.2.8). Otra ventaja es que el usuario no tiene que especificar la longitud de rezagos para el test de regresión.

3.4.3. Prueba Zivot-Andrews. Zivot y Andrews (1992), presentan su propuesta con una variación del test original de Phillips- Perrón en la cual se da por hecho que el tiempo exacto de un cambio estructural es ignorado. Continuando con la estructura de Perrón para cambios estructurales, Zivot y Andrews trabajan con tres modelos para la prueba de raíz unitaria⁷:

- Modelo A, el cual permite un único cambio en el nivel de la serie.
- Modelo B, el único cambio en la pendiente de la función de tendencia de la serie.
- Modelo C, combina el único cambio en el nivel y la pendiente de la tendencia de la serie.

Ahora, para las pruebas de raíz para determinar un único cambio estructural de la serie, Zivot y Andrews (1992) presentan las siguientes ecuaciones de regresión:

$$\nabla Z_t = c + \gamma Z_{t-1} + \beta t + c + \sum_{i=1}^p \delta_i \nabla Z_{t-i} + a_t \quad (3.5.7)$$

$$\nabla Z_t = c + \gamma Z_{t-1} + \beta t + \theta DT_t + \sum_{i=1}^p \delta_i \nabla Z_{t-i} + a_t \quad (3.5.8)$$

$$\nabla Z_t = c + \gamma Z_{t-1} + \beta t + \theta DT_t + \lambda DU_t + \sum_{i=1}^p \delta_i \nabla Z_{t-i} + a_t \quad (3.5.9)$$

Donde DU_t es una variable “dummy” que señala cambios significativos a cada posible variación estructural, y DT_t corresponde a la tendencia significativa variable, por lo tanto de una manera formal observamos:

⁷ ZIVOT, Eric y ANDREWS, Donald. Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis, USA: *Journal of Business and Economic Statistics*, Vol. 10, No.3, p. 251-270, 1992

$$DU_t = \begin{cases} 1 & \text{si } t > TB \\ 0 & \text{otra forma} \end{cases} \quad (3.5.10)$$

$$DT_t = \begin{cases} t - TB & \text{si } t > TB \\ 0 & \text{otra forma} \end{cases}$$

La hipótesis nula para los tres modelos es $\gamma = 0$, lo cual implica que la serie $\{Z_t\}$ contiene una raíz unitaria con deriva que excluye cualquier cambio estructural, mientras que la hipótesis alternativa $\gamma < 0$ implica que la serie es un proceso de tendencia estacionaria con un cambio estructural ocurriendo en la serie temporal.

En el método de Zivot-Andrews (1992) es considerado cada punto como un potencial “*brake-date*” de la serie y ajusta consecutivamente regresión para cada posible “*brake-date*”. De entre todos los posibles puntos de corte, el método escoge el dato de ruptura que minimiza el estadístico $-t$ de un lado para probar que:

$$\hat{\gamma}(\gamma - 1) = 1$$

De acuerdo con Zivot y Andrews (1992), la presencia de extremos causa la distribución asintótica del estadístico que diverge hacia el infinito⁸.

Por otra parte, en muchas series: financieras, económicas, entre otras puede suceder que las raíces no necesariamente son enteras, es decir, la no estacionaridad del proceso puede darse para raíces fraccionales. A continuación se presenta el modelo ARFIMA (p,d,q), el cual puede ser considerado como un proceso intermedio entre los procesos ARMA (p,q) y los ARIMA (p,d,q)

3.5. MODELO ARFIMA (P,D,Q)

Se suele señalar que un proceso estocástico $\{Z_t: t \in \mathbb{z}^+\}$ sigue un proceso ARFIMA (p,d,q) si soluciona la siguiente ecuación:

$$\phi(B)(1 - B)^d Z_t = \theta_0 + \theta(B)a_t, \quad t = 1, \dots, T \quad (1) \quad (3.4.1)$$

Según (Castaño, Gómez & Gallón, 2008.)

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 autorregresivos 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 de unidad y sin raíces comunes; d y θ_0

⁸ ZIVOT y ANDREWS. Op. Cit. p. 24

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.⁹ (P70)

3.5.1. Propiedades del proceso ARFIMA (p,d,q)

En los modelos ARFIMA, el grado de memoria y estacionaridad del proceso se define en el parámetro diferenciación fraccional d , el cual toma valores en un intervalo continuo de números reales (Lemus y Castaño, 2013). El interés para el análisis de series temporales está en el valor del mencionado parámetro por las propiedades características de los procesos que se obtienen a partir de éste:

- Si $d=0$, el proceso $\{Z_t\}$ en (3.4.1) corresponde a una serie de tiempo estacionaria de memoria corta ARMA (p,q)
- Si $d=1$, el proceso $\{Z_t\}$ en (3.4.1) corresponde a una serie de tiempo no estacionaria que posee una raíz unitaria. En este caso, la primera diferencia del proceso $\{Z_t\}$ es estacionaria.
- Si d es otro número entero, el proceso $\{Z_t\}$ en (3.4.1) es no estacionario homogéneo que posee d raíces unitarias. Por lo tanto, la d -ésima diferencia del proceso $\{Z_t\}$ es estacionaria.
- Si d es un número real no entero, el proceso $\{Z_t\}$ en (3.4.1) corresponde a un proceso fraccionalmente integrado ARFIMA (p,d,q). (P. 10)

Se puede afirmar que, si el valor del parámetro de diferenciación se encuentra entre $-1/2 < d < 1/2$, $\{Z_t\}$ en (3.4.1) corresponde a un proceso fraccionalmente integrado ARFIMA (p, d,q), estacionario e invertible¹⁰, en el cual según Lemus y Castaño (2013)

Si $-1/2 < d < 0$, $\{Z_t\}$ en (3.4.1) 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.(P.12)

⁹CASTAÑO, GÓMEZ y GALLÓN. Una nueva prueba para el parámetro de diferenciación fraccional. Colombia: Revista Colombiana de Estadística, 2008. p.70

¹⁰ LEMUS, Diego y CASTAÑO, Elkin. "Prueba de hipótesis sobre la existencia de una raíz fraccional en una serie de tiempo no estacionaria", Lecturas De Economía ISSN: 0120-2596 ; v.78 - p.153 - 184, 2013. Medellín (Colombia): Universidad de Antioquia.

- “Si $0 < d < 1/2$, $\{Z_t\}$ en (3.4.1) es un proceso estacionario de memoria larga.” (P.12)
- Si $d \geq 1/2$ el proceso es en general no estacionario¹¹

En el proceso de identificación de un modelo ARFIMA (p,d,q) se requiere de identificación y estimación del parámetro de diferenciación fraccional, para tal fin se utilizan pruebas de raíz fraccional que se presentan a continuación.

3.6. PRUEBAS DE RAÍZ FRACCIONAL ESTACIONARIA

3.6.1. Prueba Dickey- Fuller Fraccional Aumentada

Según Dolado, Gonzalo y Mayoral (2002) consideran el siguiente modelo de regresión:

$$(1 - B)^{d_0} Z_t = \phi(1 - B)^{d_1} Z_{t-1} + \mu_t \quad (3.6.1)$$

cond₀ = 1 y $d_1 \in (0,1)$. μ_t es un proceso autoregresivo de orden p, $\phi(B)\mu_t = \alpha_t$, donde $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ es el polinomio AR en términos del operador de rezagos B cuyas raíces están todas por fuera del círculo unitario. α_t es un ruido blanco gaussiano de media cero y varianza constante. Reuniendo todos estos parámetros esta prueba se conforma bajo el siguiente modelo de regresión¹²:

$$(1 - B)Z_t = \phi(1 - B)^{d_1} Z_{t-1} + \sum_{i=1}^p \zeta_i (1 - B)Z_{t-i} + \alpha_t \quad (3.6.2)$$

En la hipótesis nula ($H_0: \phi = 0$ en (6.2) donde Z_t , es un proceso ARIMA (p,1,0), los autores demostraron, según Lemus y Castaño (2013) que:

- “La distribución asintótica de $t_{\phi_{mco}}^{\wedge}$ en (3.6.2) es la misma obtenida en el estadístico (6.1), siempre que se incluya un número adecuado de rezagos.
- Los coeficientes de regresión estimados ($\zeta_1 \dots \zeta_p$) son asintóticamente normales para cualquier valor del parámetro de diferenciación fraccional $d_1 \in (0,1)$ usando la regresión de (3.6.2)” (p.23)

Nótese que bajo la hipótesis alternativa se obtiene un proceso ARFIMA (p,1,0).

3.6.2 Prueba de Gewek y Porter – Hudak (GPH). En esta prueba, Gewenk y Porter-Hudak (1983) plantean un estimador semiparamétrico para el parámetro de

¹¹ GEWEKE, John y PORTER-HUDAK, Susan. The estimation and application of long-memory time series models, 1983. Journal of Time Series Analysis 4(4), p. 221–238.

¹² LEMUS, Diego y CASTAÑO. Op. Cit. p. 27

diferenciación d en el dominio de la frecuencia.¹³ Estos autores en el desarrollo teórico de la propuesta consideraron $(1 - B)^d Z_t = X_t$ donde $X_t \sim I(0)$ y este proceso se puede representar por:

$$f_x(\omega) = |1 - \exp(-i\omega)|^{2d} f_z(\omega) \quad (3.6.3)$$

donde $f_x(\omega)$ y $f_z(\omega)$ son las densidades espectrales de los procesos X_t y Z_t . Empleando el logaritmo en (3.6.3) y mediante unas operaciones algebraicas se puede obtener la siguiente expresión:

$$\log\{f_x(\omega_j)\} = \log\{f_z(0)\} - d \log\left\{4 \operatorname{sen}^2\left(\frac{\omega_j}{2}\right)\right\} + \log\left\{\frac{f_u(\omega_j)}{f_z(0)}\right\} \quad (3.6.4)$$

Donde $f_u(\omega)$ es la densidad espectral del proceso ruido blanco en la frecuencia ω . El contraste se basa en la inferencia sobre el parámetro de diferenciación fraccional d partiendo del estimador por mínimos cuadrados ordinarios en el modelo de regresión para el logaritmo del periodograma de Z_t , el cual se presenta en la siguiente formula:

$$\log\{I_z(\omega_j)\} = \beta_0 + \beta_1 \log\left\{4 \operatorname{sen}^2\left(\frac{\omega_j}{2}\right)\right\} + V_j \quad (3.6.5)$$

Donde $I_z(\omega_j)$ es la j -ésima ordenada del periodograma de Z_t , $\omega_j = 2\pi j/T$ es la j -ésima frecuencia de Fourier y $V_j = \log\left\{\frac{f_u(\omega_j)}{f_z(0)}\right\}$ es el termino del error de la regresión, se asume que i.i.d de media cero y varianza constante $\frac{\pi}{6}$, $\forall j = 1, \dots, m$, siendo $m = \lfloor T^{1/2} \rfloor$ el mayor entero es el menor entero o igual a $T^{1/2}$.

Al comparar las formulas (3.6.4) y (3.6.5) se puede observar que el orden de diferenciación fraccional d en (3.6.4) corresponde al coeficiente negativo de β_1 en (7-3), entonces, el contraste de $I(0)$ vs $FI(d)$ - donde $I(0)$ es el proceso estacionario de memoria corta y $FI(d)$ es el proceso fraccionalmente integrado - se realiza con una prueba de dos colas que permita determinar la significancia del estimador de β_1 en (3.6.5), para de esta forma hacer inferencia sobre el parámetro d ¹⁴. El contraste de hipótesis sería el siguiente:

$$\begin{aligned} H_0 : \beta_1 &= 0 \quad (d = 0) \\ H_a : \beta_1 &\neq 0 \quad (d \neq 0) \end{aligned}$$

Geweke y Porter-Hudak, argumentan que para $-1/2 < d < 0$ y asumiendo otras condiciones más el estadístico de prueba converge en una distribución normal estándar:

¹³ GEWEK y PORTER. Op. Cit. p. 27

¹⁴ LEMUS Y CASTAÑO. Op. Cit. p. 28

$$\frac{\widehat{d}_{GPH} - d}{\sqrt{\text{var}(\widehat{d})}} \rightarrow N(0,1)$$

$$\left(\sum_{j=1}^m R_j - \bar{R} \right)$$

Donde, $\sqrt{\text{var}(\widehat{d})} = \frac{\pi^2}{6} (\sum_{j=1}^m R_j - \bar{R})^{-1}$ con $R_j = \log(4 \sin^2 \omega_j / 2)$

3.6.3 La prueba de robinson. Robinson (1994) demuestra que el estimador \widehat{d}_{GPH} es consistente y asintoticamente normal para todo valor del parámetro de diferenciación fraccional donde el proceso ARFIMA(p,d,q) es estacionario e invertible¹⁵ y propone la siguiente modificación del mismo, dada por el modelo de regresion que se presenta a continuación:

$$\text{Log} \{I_x(W_j)\} = \beta_0 + \beta_1 (2 \log\{|W_j|\}) + \varepsilon_j \quad (3.6.6)$$

Donde $I_x(W_j)$, W_j y ε_j tienen la misma definición presentada en la prueba GPH. El contraste $I(0)$ vs $FI(d)$ se realiza por medio de una prueba de dos colas para determinar la significancia del estimador de β_1 en (3.6.6), donde $I(0)$ es el proceso estacionario de memoria corta y $FI(d)$ es el proceso fraccionalmente integrado. El contraste de hipótesis sería el siguiente:

$$H_0 : \beta_1 = 0 \quad (d = 0)$$

$$H_a : \beta_1 \neq 0 \quad (d \neq 0)$$

Robinson determino que el estimador \widehat{d}_{ROB95} obtenido por minimos cuadrados ordinarios, utilizando las $m = (T^{\frac{1}{2}})$ primeras frecuencia en torno al origen en la regresion, es consistente y asintoticamente normal, bajo la hipótesis de normalidad para el término de error en el caso de los procesos ARFIMA(p,d,q) estacionarios e invertibles.

3.6.4. Prueba de Castaño, Gómez Y Gallón. En un proceso ARFIMA si $d > -0.5$ el proceso es invertible y puede reescribirse de la siguiente manera:

$$(1 - B)^d \pi(B) Z_t = \alpha_t$$

¹⁵ ROBINSON, Peter. Efficient tests of nonstationary hypotheses. Newyork: Journal of the American Statistical Association, 1994 89(428), p. 1420-1437.

Donde $\pi(B) = \phi(B)/\theta(B) = 1 - \pi_1 B - \pi_2 B^2 - \dots$, es la componente a corto plazo infinita del proceso Z_t y α_t es un ruido blanco de media cero y varianza constante. Los autores siguiendo Said & Dickey (1984), aproximan el proceso anterior por medio del siguiente modelo

$$(1 - B)^d \pi^*(B) Z_t = \alpha_t$$

Donde $\pi^*(B) = 1 - \pi_1^* B - \dots - \pi_{p^*}^* B^{p^*}$ para cierto orden p^* suficientemente grande. Basados en el modelo ARFIMA($p^*, d, 0$) aproximado, el procedimiento sugiere contrastar la hipótesis nula de memoria corta, $H_0: d = 0$, con la alternativa de memoria larga $H_1: d > 0$, empleando el estadístico de prueba $t_d = \hat{d}/se(\hat{d})$ donde \hat{d} es el estimador máximo verosímil de d y $se(\hat{d})$ es la estimación de su error estándar en el modelo aproximado. En Castaño et.al (2008) se demostró que el estimador del parámetro es consistente y asintóticamente normal bajo la hipótesis nula - bajo H_0 , $-t_d \xrightarrow{d} N(0,1)$. También encontró que la prueba con un orden $p^* = T^{1/4}$ tiene mejores propiedades estadísticas que otras pruebas consideradas en la literatura.

3.7 METODOLOGÍA DE CASTAÑO ET. AL (2008) PARA IDENTIFICACIÓN DE MODELOS ARFIMA (p,d,q)

La metodología de Castaño et. al (2008) para la identificación de procesos AFIRMA (p,d,q) estacionarios e invertibles consiste de cinco etapas:

- Aproximar la componente a corto plazo del proceso ARFIMA(p, d, q) por medio de un proceso autorregresivo de orden $p^* \approx T^{1/3}$ donde T es el número de observaciones de la serie bajo análisis. Estime el modelo aproximado ARFIMA($p^*, d, 0$).
- Utilizar el estimador del parámetro de diferenciación fraccional obtenida en el paso 1 para diferenciar fraccionalmente la serie. Sobre la serie diferenciada determine los órdenes de los polinomios autorregresivos (AR) y medias móviles (MA) para la componente a corto plazo del proceso.
- Estimar simultáneamente todos los parámetros del modelo ARFIMA(p, d, q) identificado.
- Verificar que el modelo proporciona un ajuste adecuado y que los supuestos básicos implícitos en el modelo, se satisfacen; de no cumplirse, determinar las modificaciones necesarias y repetir las etapas anteriores hasta que la verificación indique que los resultados son aceptables.

- Uso del modelo ARFIMA (p, d, q) identificado para los fines que el investigador haya tenido en mente al construir.

3.8 MÉTODOS DE PRONÓSTICO

En este trabajo se presentaran dos técnicas de pronósticos más identificados, conocidos como método exacto y método Bootstrap.

3.8.1 Método Exacto. Hace referencia a la utilización del modelo identificado para el pronóstico de valores futuros utilizando la verosimilitud del mismo (notas de clase, Diego Lemus)

3.8.2 Método Bootstrap. Este método se enmarca entre los procedimientos de remuestreo, consistentes en generar un elevado número de muestras como base para estudiar el comportamiento de determinados estadísticos. (Flores,2003)

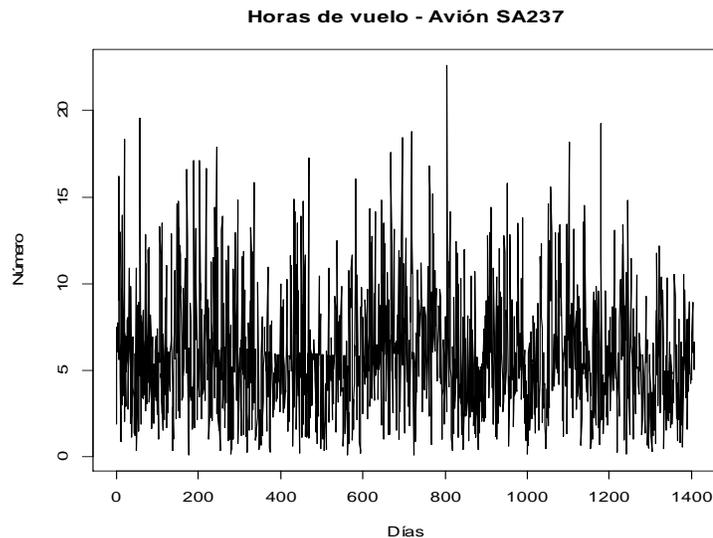
4. ANÁLISIS DE DATOS

Modelo de Pronóstico de Horas de Vuelo Para las Aeronaves B-350 y SA-237 de la Fuerza Aérea Colombiana

4.1 AERONAVESA-237

La serie del equipo SA-237 fue facilitada por el Departamento de Planeación y Estadística de la FAC entidad encargada de registrar, analizar y pronosticar el número de horas de vuelo de cada aeronave. El avión mencionado tiene la misión de contribuir con la realización de inteligencia militar estratégica y se analizó el tiempo de operación del mismo desde el primero de enero de 2010 hasta el 07 de noviembre de 2013, con una periodicidad diaria (Lunes - Domingo) para un total de 1407 observaciones. En el Gráfico 1 se presenta la serie del número de horas de vuelo de la aeronave SA-237.

Gráfico 1. Serie Original Horas de vuelo SA-237



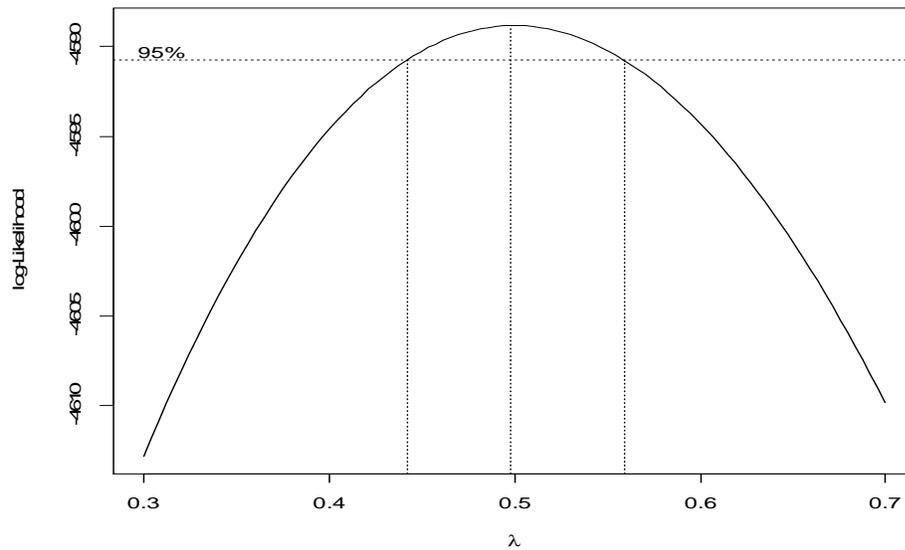
Fuente: Elaboración propia

Del Gráfico 1, se puede observar que la serie del número de horas de vuelo de la aeronave SA-237 al parecer no presenta una tendencia determinística. También se puede apreciar que aparentemente la varianza incondicional del proceso no es constante y depende del nivel de la serie.

En el Gráfico 2 se presentan las posibles transformaciones de potencia Box-Cox de la serie bajo análisis. Bajo la transformación óptima se maximiza la log-verosimilitud

lo cual garantiza que todos los estimadores bajo la serie original o bajo la serie transformada serán consistentes y no va haber un daño en las estimaciones.

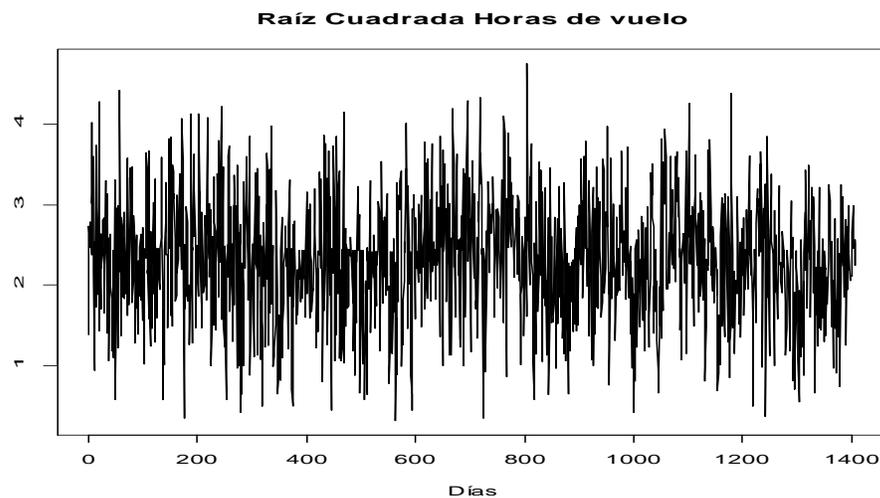
Gráfico 2. Curva de comparación de transformaciones de Potencia Maximización de la log-verosimilitud.



Fuente: Elaboración propia.

Del gráfico 2 de la prueba de transformación de potencia se identificó que la transformación Box-Cox más adecuada para estabilizar la varianza incondicional del proceso es la raíz cuadrada. A partir de este momento se trabajará sobre la serie transformada: $X_t = \sqrt{Z_t}$, la cual se presenta en el grafico 3.

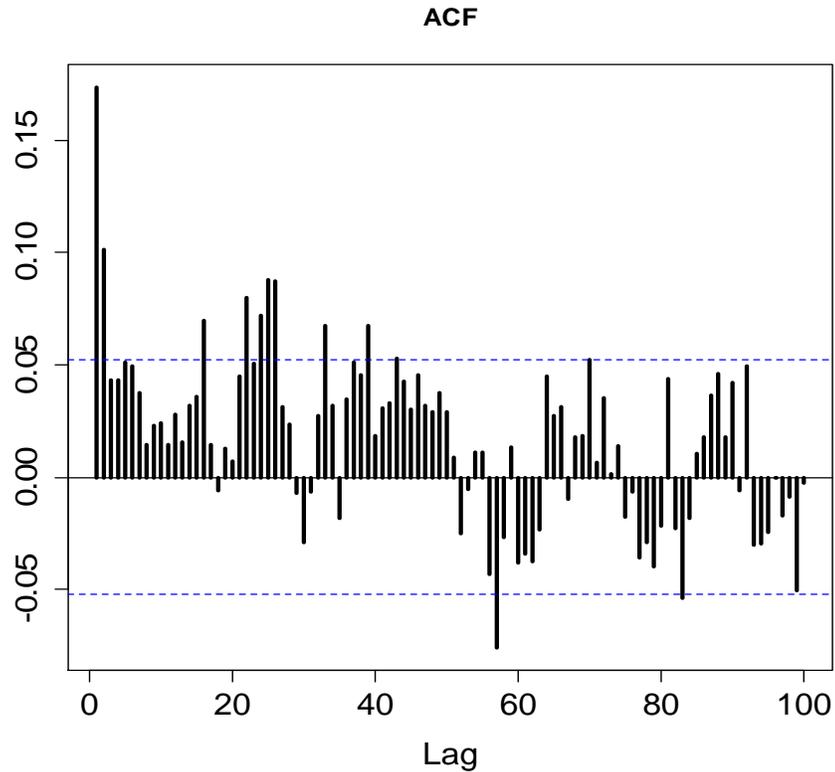
Gráfico 3. Serie transformada horas de vuelo SA-237



Fuente: Elaboración propia.

En el gráfico 4 se presenta el correlograma de la serie transformada

Gráfico 4. ACF-Serie transformada Equipo SA-237



Fuente: Elaboración Propia.

En la función de autocorrelación de la serie transformada se observa un decrecimiento senoidal, el cual parece disminuir de manera acelerada a cero, lo cual puede ser un indicio de que el proceso es posiblemente estacionario. A continuación se presentan pruebas más formales que permitan hacer inferencia de las características del proceso bajo análisis. Primero se realizarán algunas pruebas de raíz unitaria bajo la metodología Box-Jenkins para la identificación de procesos ARIMA (p,d,q).

4.1.1. Pruebas de raíz unitaria

4.1.1.1. Prueba de Dickey-Fuller Aumentada. Como se mencionó en la sección 3.5.1 y ya que no se encontró evidencia empírica de tendencia determinística (al incluir tendencia determinística en la prueba ADF) se consideró el modelo 2 de la ecuación 3.5.4 para realizar el siguiente contraste de hipótesis:

Ho: Hay evidencia de raíz unitaria en el proceso X_t

Ha: No existe raíz unitaria en el proceso X_t

El valor del estadístico de prueba obtenido fue -21.8029.

Tabla 1. Valores Críticos Prueba ADF

Valores críticos			
	1pct	5pct	10pct
Tau 2	-3.43	-2.86	-2.57

Fuente: Elaboración Propia.

Como $\tau_{\mu} < \tau_{\mu}(\alpha, n)$ para todo α , se rechaza la hipótesis nula y de esta manera se puede concluir que no hay raíz unitaria en Z_t . En la Tabla 2 se presentan los resultados del ajuste del modelo 2 en la ecuación 3.5.4, donde se evidencia que la inclusión de rezagos en la prueba ADF resultó significativa.

Tabla 2, Regresión prueba ADF sin tendencia Determinística

	Estimador	Error Estándar	Estadístico t	P valor
Intercepto	1.71515	0.08108	21.155	< 2e-16 ***
z.lag.1	-0.74386	0.03412	-21.803	< 2e-16 ***
z.diff.lag	-0.08982	0.02666	-3.369	0.000774***

Fuente: Elaboración Propia.

4.1.1.2. Prueba de Phillips-Perrón (PP). Empleando la prueba de Phillips-Perrón aumentada sin incluir tendencia determinística se realizó el siguiente contraste de hipótesis:

Ho: Hay evidencia de raíz unitaria en el proceso X_t

Ha: No existe raíz unitaria en el proceso X_t

El valor del estadístico de prueba obtenido fue -33.0575.

Tabla 3. Valores críticos Prueba PP

Valores críticos			
	1pct	5pct	10pct
Tau 2	-3.44	-2.86	-2.57

Fuente: Elaboración Propia.

Como el estadístico Phillips-Perrón es menor a cualquier valor crítico $-\tau_{\mu} < \tau_{\mu}(\alpha, n)$ para todo α , se rechaza la hipótesis nula y de esta manera se puede concluir que no hay raíz unitaria en Z_t . En la Tabla 4 se presentan los resultados del ajuste del modelo de regresión de la prueba.

Tabla 4. Regresión prueba PP sin Tendencia Determinística

	Estimador	Error Estándar	Estadístico t	P valor
Intercepto	1.88591	0.06364	29.633	< 2e-16 ***
y.l1	0.18310	0.02622	6.983	4.46e-12 ***

Fuente: Elaboración Propia.

4.1.1.3. Prueba de Zivot-Andrews. Se empleó la prueba de Zivot y Andrews para realizar el contraste de raíz unitaria contra la alternativa de un único cambio estructural de la serie empleando la ecuación de regresión del modelo A el cual permite un único cambio en el nivel de la serie (pues no hay evidencia de tendencia determinística).

Ho: Hay evidencia de raíz unitaria en el proceso X_t

Ha: No existe raíz unitaria en el proceso X_t

El valor del estadístico de prueba obtenido fue -22.4438

Tabla 5. Valores Críticos prueba Zivot-Andrews

Valores Críticos			
	1pct	5pct	10pct
Tau 2	-5.34	-4.8	-4.58

Fuente: Elaboración Propia.

Como el estadístico de la prueba de Zivot y Andrews es menor a cualquier valor crítico se rechaza hipótesis nula y se concluye que no hay evidencia de raíz unitaria en el proceso bajo estudio. De igual manera se identifica un posible cambio de nivel en la observación 594. En la Tabla 6 se presentan los resultados del ajuste del modelo de regresión de la prueba realizada.

Tabla 6. Regresión prueba de Zivot-Andrews- Modelo A

	Estimador	Error estándar	Estadístico t	P valor
Intercepto	1.877e+00	9.292e-02	20.196	< 2e-16 ***
y.l1	2.290e-01	3.435e-02	6.666	3.76e-11 ***
y.dl1	-7.642e-02	2.659e-02	-2.874	0.00411 **
Du	3.323e-01	7.749e-02	4.289	1.92e-05 ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

Fuente: Elaboración Propia.

A continuación se presenta la prueba de estabilidad estructural realizada sobre la serie transformada con el fin de confirmar lo identificado con la prueba de Zivot y Andrews.

4.1.1.3.1 Prueba de Nyblom – Hansen. En esta prueba se establece el siguiente contraste de hipótesis:

- Ho: Hay estabilidad en la estimación de los coeficientes bajo el modelo considerado.
- Ha: No hay estabilidad en la estimación de los coeficientes bajo el modelo considerado.

En la Tabla 7 se presentan los resultados obtenidos de la prueba

Tabla 7. Prueba de Nyblom-Hansen

Estadístico de prueba	Valor p
0.8241	0.1185

Fuente: Elaboración Propia.

Como el valor p de la prueba de Nyblom -Hansen es mayor que los niveles de significancia nominal considerados del 1%, al 5% y el 10% no se rechaza hipótesis nula y se concluye que hay estabilidad en la estimación de los coeficientes bajo el modelo considerado. Lo anterior implica que no es necesario incluir un cambio de nivel en el modelo considerado para el análisis de proceso bajo estudio. A continuación se presentan las pruebas de raíz fraccional para verificar si en efecto o no hay evidencia de memoria larga en el proceso.

4.1.2. Pruebas de raíz fraccional estacionaria. En las pruebas que se presentan en esta sección se establece el siguiente contraste de hipótesis:

- Ho: $d = 0$. El proceso X_t corresponde a un proceso ARMA (p,q) estacionario e invertible.
- Ha: $d \neq 0$ con $-\frac{1}{2} < d < \frac{1}{2}$. El proceso X_t corresponde a un proceso ARFIMA (p,d,q) estacionario e invertible.

En la Tabla 8 se presentan los resultados de las pruebas de raíz fraccional estacionaria de Geweke y Porter –Hudak (GPH) y de Robinson.

Tabla 8. Pruebas de Raíz fraccional

Prueba	d	Error estándar	Estadístico de prueba	Valor P
GPH	0.3675	0.0684	5.3722	0.0000
ROBINSON	0.3869	0.0949	4.0759	0.0001

Fuente: Elaboración Propia.

Como el valor p de las pruebas mencionadas es menor que el nivel de significancia nominal del 5% se rechaza hipótesis nula y se concluye que hay evidencia de raíz fraccional estacionaria en la serie transformada del número de horas de vuelo del equipo SA-237.

4.1.3. Procedimiento de identificación y estimación - modelos ARFIMA (p,d,q). A continuación se ilustra la implementación del procedimiento propuesto en Castaño et. al. (2008) para la identificación y estimación de un modelo adecuado para la serie transformada del número de horas de vuelo del equipo SA-237 asumiendo que el término de error at en la ecuación 3.1 sigue una distribución normal estándar.

Primer paso: siguiendo la recomendación de Castaño (2008) se aproximó la componente a corto plazo por medio de un proceso autorregresivo de orden $p^* = T^{1/4} = 1407^{1/4} \approx 6$. Los resultados del modelo aproximado ARFIMA (6,d,0) se presentan en la Tabla 9

Tabla 9. Modelo aproximado ARFIMA (6,d,0)

Parámetro	Estimador	Error estándar	t value	Pr(> t)
phi(1)	0.0286	0.0734	0.38996	0.6966
phi(2)	0.0291	0.0419	0.69434	0.4875
phi(3)	-0.0261	0.0332	-0.78538	0.4322
phi(4)	-0.0179	0.0316	-0.56723	0.5706
phi(5)	0.0082	0.0304	0.27144	0.7861
phi(6)	0.0059	0.0289	0.20531	0.8373
d.f	0.1318	0.0682	1.93195	0.0534
Fitted mean	2.3063	0.0510	45.22128	< 2e-16 ***

Fuente: Elaboración Propia.

Los resultados del modelo aproximado ajustado presentados en la Tabla 9 presentaron coeficientes autoregresivos no significativos (valores p mayores al nivel de significancia nominal del 5%) que indican la sobreparametrización del mismo. Se reespecificó el modelo aproximado a un ARFIMA (0,d,0) obteniendo los resultados que se presentan en la Tabla 10:

Tabla 10. Modelo aproximado ARFIMA (0,d,0)

Parámetro	Estimador	Error estándar	t-value	Pr(> t)
df	0.1519	0.0212	7.15619	8.295e-13
Fitted mean	2.3061	0.0575	40.10839	< 2.22e-16

Fuente: Elaboración Propia.

Tabla 11. Prueba de Ljung-Box para residuales modelo ARFIMA (0,d,0)

Estadístico de prueba	Df	Valor p
3.517	12	0.9907

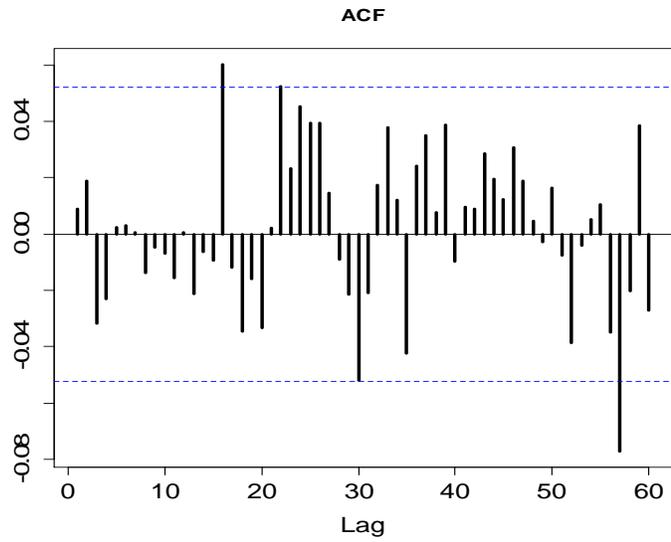
Fuente: Elaboración Propia.

En la Tabla 11 se presenta el resultado de la prueba de correlación serial de Ljung-Box sobre los residuales del modelo aproximado. Como el valor p de la prueba es mayor al nivel de significancia nominal del 5% se rechaza la hipótesis nula de correlación serial en los residuales del modelo aproximándolo anterior indica que el mencionado modelo captura la correlación serial presente en la serie transformada fraccionalmente diferenciada. Por lo enunciado anteriormente, se puede concluir al observar la Tabla 10 que la estimación del parámetro de diferenciación fraccional d en el modelo aproximado es significativa para un nivel de significancia nominal del 5%.

Segundo paso: Usar $\hat{d} = 0.1519$ para diferenciar fraccionalmente la serie transformada y sobre el proceso resultante utilizar las técnicas de identificación de un modelo ARMA (p,q) para el componente a corto plazo del proceso.

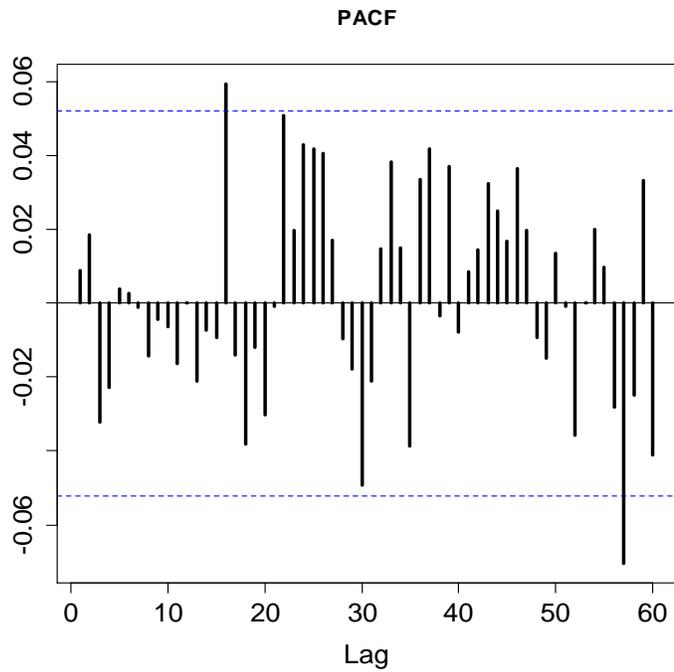
En los Gráficos 5,6 y 7 se presentan la función de autocorrelación, la función de autocorrelación parcial y la función de autocorrelación extendida de la serie transformada diferenciada fraccionalmente.

Gráfico 5. ACF Serie Diferenciada Fraccionalmente



Fuente: Elaboración propia

Gráfico 6. PACF Serie Diferenciada Fraccionalmente



Fuente: Elaboración propia

Gráfico 7. EACF Serie Diferenciada Fraccionalmente

		AR/MA										
		0	1	2	3	4	5	6	7	8	9	10
0	0	0	0	0	0	0	0	0	0	0	0	0
1	X	0	0	0	0	0	0	0	0	0	0	0
2	X	X	0	0	0	0	0	0	0	0	0	0
3	X	X	X	0	0	0	0	0	0	0	0	0
4	X	X	0	0	0	0	0	0	0	0	0	0
5	X	X	0	X	X	0	0	0	0	0	0	0

Fuente: Elaboración propia

Al observar los gráficos 5, 6 y 7, se puede concluir que el modelo ARMA(Paq) para el componente a corto plazo de la serie estudiada parece ser consistente con un modelo ARMA (0,0), lo anterior implica que no hay evidencia de componente a corto plazo en la serie transformada fraccionalmente integrada.

En la Tabla 12 se presenta el resultado de la evaluación de diferentes modelos empleando criterio de información de Amaine (OIC) y el criterio de información bayesiano (BIC) empleando la función Autoimagen R para la serie transformada diferenciada fraccionalmente

Tabla 12. Resultados de los criterios AIC y BIC selección de modelos

Modulo	AIC	BIC
ARIMA(0,0,0)	3126.93	3132.18

Fuente: Elaboración Propia.

Los resultados obtenidos indican que el proceso ARFIMA (0,d,0) parece ser adecuado para modelar la serie transformada del número de horas de vuelo del equipo SA-237.

Tercer paso: Estimar el modelo ARFIMA (0,d,0) identificado y validar los supuestos sobre sus residuales.

A continuación se presenta la verificación de los supuestos básicos implícitos en el modelo. De no cumplirse se determinarán las modificaciones necesarias y se repetirán las etapas anteriores hasta que la verificación indique los son resultados aceptables.

Validación Supuestos del Modelo

Pruebas de incorrelación. Se consideraron las pruebas de Ljung-Box y Durbin-Watson las cuales consideran el siguiente contraste de hipótesis para su implementación:

- Ho: No hay evidencia de correlación serial entre los residuales
- Ha: Si hay evidencia de correlación serial entre los residuales

Tabla 13. Prueba de Ljung-Box para residuales modelo ARFIMA (0,d,0)

Estadístico de prueba	Df	Valor p
3.517	12	0.9907

Fuente: Elaboración Propia.

Tabla 14. Prueba Durbin-Watson para residuales modelo ARFIMA (0,d,0)

Prueba	DW	p- value
Durbin-Watson	3.517	0.3606

Fuente: Elaboración Propia.

En las tabla 13 y 14 se puede observar que el valor p de ambas pruebas es mayor que el nivel de significancia considerado del 5% por lo que se puede concluir que los residuales del modelo ajustado no presentan correlación serial entre sí, lo anterior implica que el modelo ajustado captura el comportamiento del proceso adecuadamente, es decir, el modelo ajustado esta adecuadamente especificado.

PRUEBAS DE NORMALIDAD

Realizar estas pruebas tiene como finalidad el verificar que la secuencia de residuales del modelo ajustado se ajusta a la distribución normal. Para tal fin se consideraron las pruebas de normalidad de Shapiro-Wilk, Jarque – Bera, Kolmogorov - Smirnov y D'Agostino. A continuación se presentan los resultados obtenidos considerando el siguiente contraste de hipótesis:

- Ho: Hay normalidad en los residuales del modelo ajustado.
- Ha: No hay normalidad en los residuales del modelo ajustado.

Prueba de Shapiro-Wilk (S-W)

Tabla 15. Prueba de Shapiro-Wilk para residuales modelo ARFIMA (0,d,0)

Estadístico	Valor p
0.998	0.09235

Fuente: Elaboración Propia.

Según la prueba S-W, para un nivel de significancia nominal del 5% no se rechaza la hipótesis nula, lo cual implica que los residuales del modelo ARFIMA (0,d,0) ajustado siguen una distribución normal.

Prueba de Jarque-Bera(J_B)

Tabla 16. Prueba de Jarque-Bera para residuales modelo ARFIMA (0,d,0)

Estadístico	Valor p
0.4384	0.8032

Fuente: Elaboración Propia.

Según la prueba J-B, para un nivel de significancia nominal del 5% no se rechaza la hipótesis nula, lo cual implica que los residuales del modelo ARFIMA (0,d,0) ajustado siguen una distribución normal.

Prueba de Kolmogorov-Smirnov (K-S)

Tabla 17. Prueba de Kolmogorov-Smirnov para residuales modelo ARFIMA (0,d,0)

Estadístico	Valor p
0.0227	0.4653

Fuente: Elaboración Propia.

Según la prueba K-S, para un nivel de significancia nominal del 5% no se rechaza la hipótesis nula, lo cual implica que los residuales del modelo ARFIMA (0,d,0) ajustado siguen una distribución normal.

Prueba de D'Agostino

Tabla 18. Prueba de D`Agostino para residuales modelo ARFIMA (0,d,0)

	Estadístico	Valor p
Chi2 Omnibus	0.3943	0.8211
Z3 Skewness	-0.5858	0.5580
Z4 Kurtosis	-0.2262	0.8211

Fuente: Elaboración Propia.

Según la prueba de D'Agostino, para un nivel de significancia nominal del 5% no se rechaza la hipótesis nula, lo cual implica que los residuales del modelo ARFIMA (0,d,0) ajustado siguen una distribución normal.

PRUEBA CUSUM - NYBLOM HANSEN para residuales

En esta prueba se establece el siguiente contraste de hipótesis:

Ho: Hay estabilidad en la estimación de los coeficientes del modelo ajustado.

Ha: No hay estabilidad en la estimación de los coeficientes del modelo ajustado.

En la Tabla 19 se presentan los resultados obtenidos de la prueba

Tabla 19. Prueba de Nyblom-Hansen para residuales del modelo ajustado

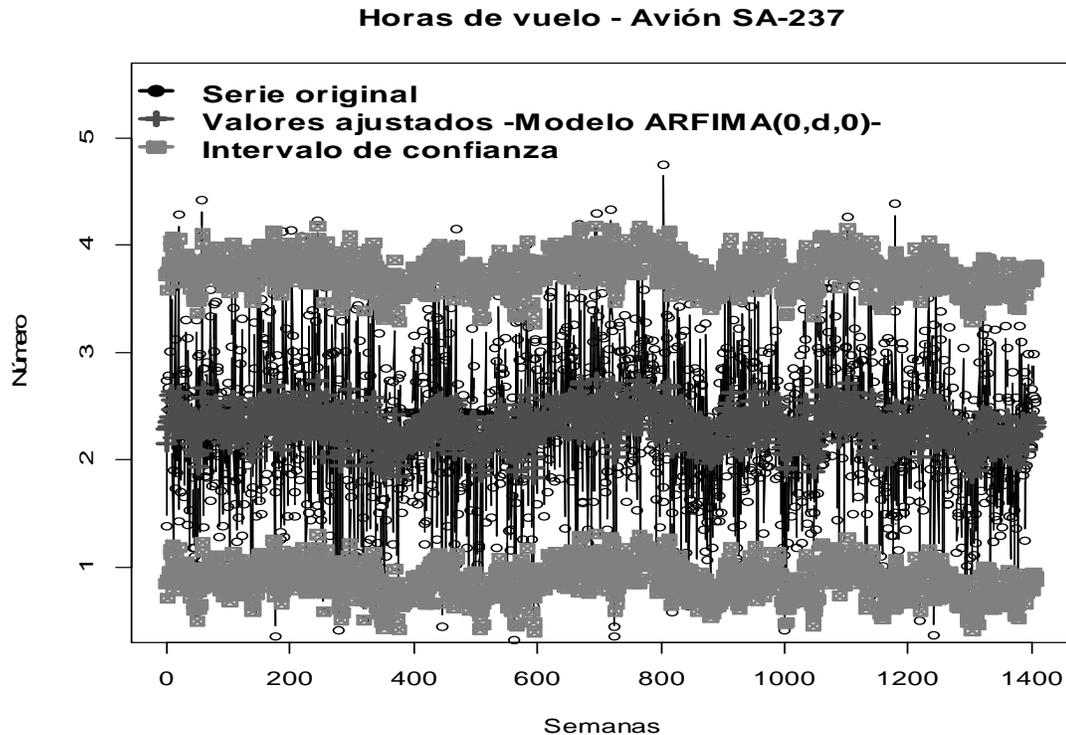
Estadístico de prueba	Valor p
0.4606	0.7163

Fuente: Elaboración Propia.

Como el valor p de la prueba de Nyblom -Hansen es mayor que el nivel de significancia nominal considerado del 5% no se rechaza la hipótesis nula y se concluye que hay estabilidad en la estimación de los coeficientes del modelo ajustado. Con los resultados presentados se han verificado todos los supuestos del modelo y por lo tanto puede ser empleado para generar pronósticos del número de horas de vuelo del equipo SA-237.

En el Gráfico 8 la serie transformada, los valores ajustados y el intervalo de confianza del 95% tipo Wald para el equipo SA-237

Gráfico 8. Serie transformada, valores ajustándose Intervalos de confianza del 95% para el equipo SA-237



Fuente: Elaboración propia

Cuarto paso: Uso del modelo identificado para pronosticar valores futuros del número de horas de vuelo del equipo SA-237.

Teniendo el modelo ajustado: ARFIMA (0,d,0) donde, $X_t = \sqrt{Z_t}$

$$(1 - B)^{0,1519}X_t = 2,3061 + a_t$$

$$\text{Donde, } a_t \sim RBG(0, \sigma_a^2)$$

En la Tabla 19 y la Tabla 20 se presenta el pronóstico de la raíz cuadrada del número de horas de vuelo del equipo SA-237 y su respectivo intervalo de confianza, calculados empleando dos métodos diferentes: el primero es el ajuste paramétrico o tradicional y el segundo es vía Bootstrap.

Tabla 19. Pronóstico e intervalo de confianza paramétrico

	1	2	3	4	5	6	7	8	9	10
Upper 95%	4.544	4.550	4.549	4.547	4.545	4.543	4.542	4.541	4.540	4.540
	699	332	013	124	397	936	720	721	896	216
Forecasts	2.340	2.320	2.310	2.304	2.300	2.297	2.294	2.292	2.290	2.289
	173	531	893	735	362	077	520	483	830	472
Lower 95%	0.135	0.090	0.072	0.062	0.055	0.050	0.046	0.043	0.040	0.038
	647	730	773	346	327	218	320	245	764	728

Fuente: Elaboración Propia

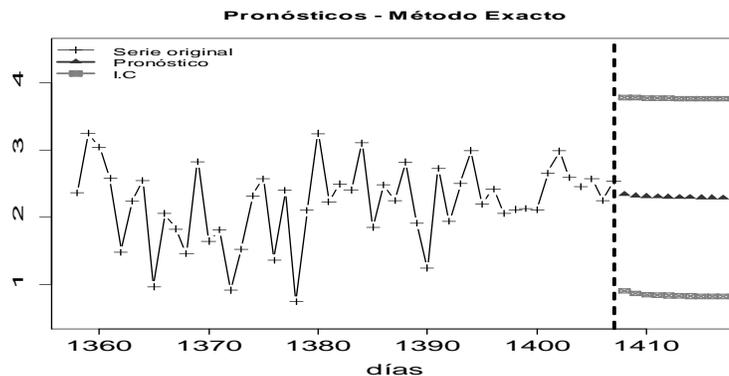
Tabla 20. Pronóstico e intervalo de confianza vía Bootstrap

	1	2	3	4	5	6	7	8	9	10
Upper 95%	6.072	6.082	6.103	6.040	6.039	6.023	6.015	6.058	6.059	6.065
	43	75	46	59	73	58	26	88	67	54
Prediction (Mean)	4.621	4.637	4.611	4.655	4.656	4.620	4.616	4.614	4.668	4.578
	24	55	03	49	59	06	76	74	46	67
Lower 95%	3.097	3.127	3.123	3.150	3.126	3.046	3.228	3.095	3.211	3.063
	77	10	61	53	42	85	78	75	77	76

Fuente: elaboración Propia

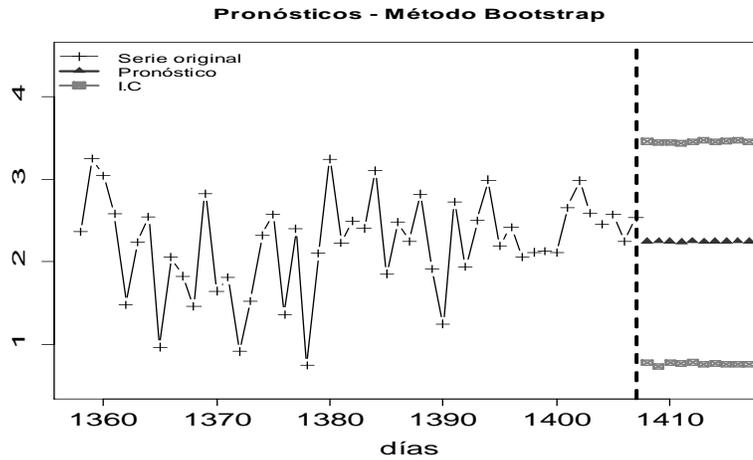
Estos son los pronósticos bajo la transformación Box Cox para el número de horas de vuelo para el equipo SA-237. En el Gráfico 9 y el Gráfico 10 se presenta el gráfico de los pronósticos obtenidos y sus respectivos intervalos de confianza, considerando el método exacto y empleando bootstrap y además el promedio que es el modelo que utiliza la FAC.

Gráfico 9. Pronósticos obtenidos e intervalos de confianza método exacto



Fuente: Elaboración Propia

Gráfico 10. Pronósticos obtenidos e intervalos de confianza Bootstrap

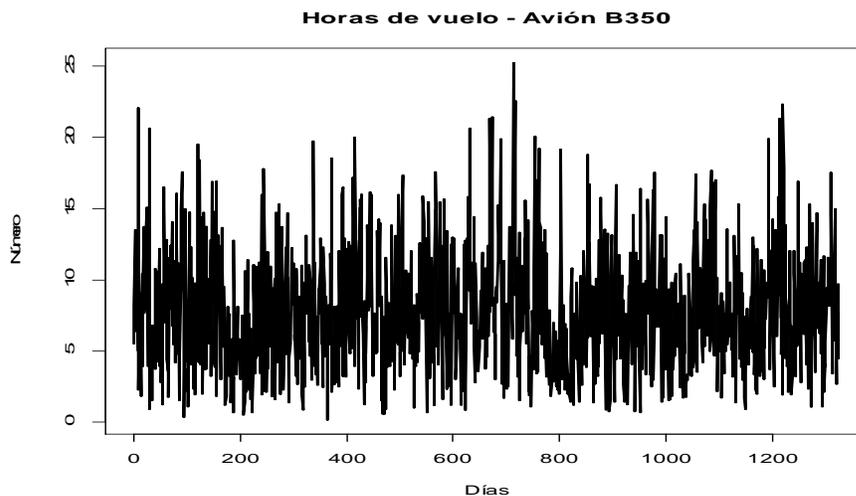


Fuente: elaboración Propia

4.2 AERONAVE B-350

La serie del equipo B-350 fue facilitada por el Departamento de Planeacion y Estadística de la FAC entidad encargada de registrar, analizar y pronosticar el número de horas de vuelo de cada aeronave. El avión mencionado tiene la misión de contribuir con la realización de inteligencia militar estratégica y se analizó el tiempo de operación del mismo desde el primero de Marzo de 2010 hasta el 07 de noviembre de 2013, con una periodicidad diaria (Lunes - Domingo) para un total de 1347 observaciones. En el gráfico 11 se presenta la serie del número de horas de vuelo de la aeronave B-350.

Gráfico 11. Serie Original de vuelo B-350

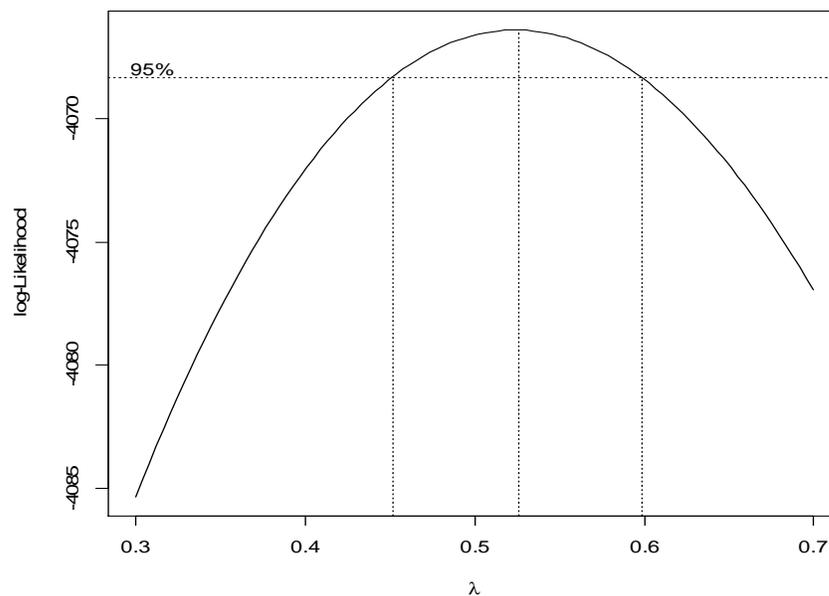


Fuente: Elaboración Propia

Del Gráfico 12 se puede observar que la serie del número de horas de vuelo de la aeronave B-350 al parecer no presenta una tendencia determinística. También se puede apreciar que aparentemente la varianza incondicional del proceso no es constante y depende del nivel de la serie.

En el Gráfico 13, se presentan las posibles transformaciones de potencia Box-Cox de la serie bajo análisis. Bajo la transformación óptima se maximiza la log-verosimilitud lo cual garantiza que todos los estimadores bajo la serie original o bajo la serie transformada serán consistentes y no va haber un daño en las estimaciones.

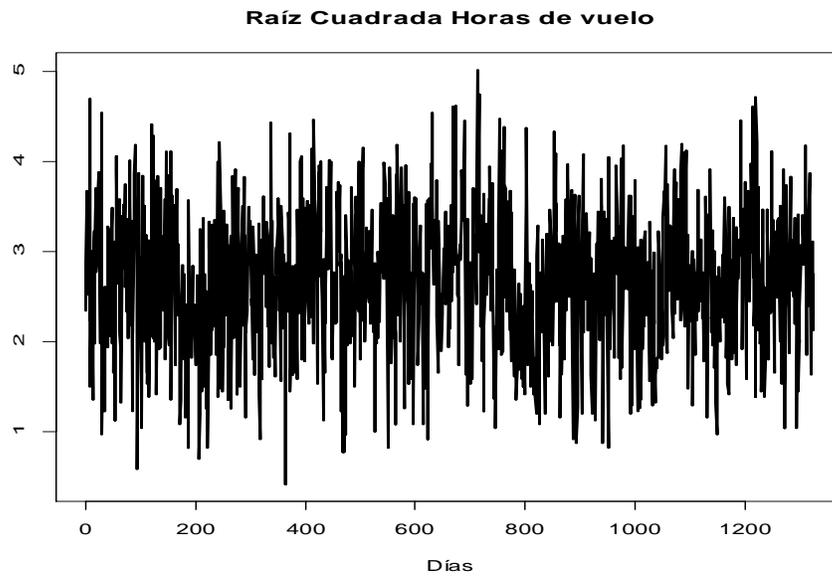
Gráfico 12. Curva de comparación de transformaciones de Potencias Maximización de la log-verosimilitud.



Fuente: Elaboración propia.

Del Gráfico 12 y de la prueba de transformación de potencia se identificó que la transformación Box-Cox más adecuada para estabilizar la varianza incondicional del proceso es la raíz cuadrada. A partir de este momento se trabajará sobre la serie transformada: $X_t = \sqrt{Z_t}$, la cual se presenta en el gráfico 14.

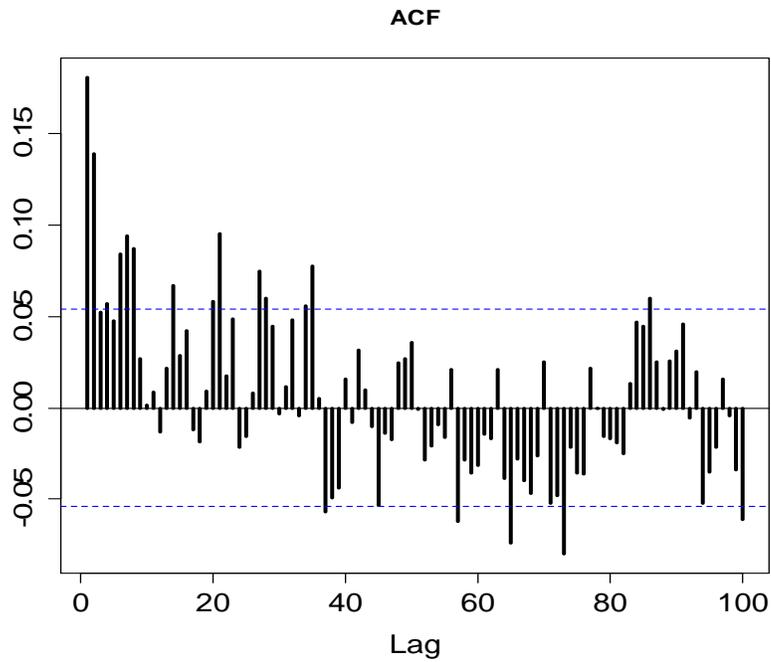
Gráfico 13 Serie transformada horas de vuelo B-350



Fuente: Elaboración propia.

En el gráfico 14 se presenta el correlograma de la serie transformada

Gráfico 14. ACF- Serie original horas de vuelo B-350



Fuente: Elaboración propia.

En la función de autocorrelación de la serie transformada se observa un decrecimiento senoidal, el cual parece disminuir de manera acelerada a cero, lo cual puede ser un indicio de que el proceso es posiblemente estacionario. A continuación se presentan pruebas más formales que permitan hacer inferencia de las características del proceso bajo análisis. Primero se realizarán algunas pruebas de raíz unitaria bajo la metodología Box-Jenkins para la identificación de procesos ARIMA (p,d,q).

4.2.1 Pruebas de raíz unitaria

4.2.1.1. Prueba de Dickey-Fuller Aumentada. Como se mencionó en la sección 3.5.1 y ya que no se encontró evidencia empírica de tendencia determinística (al incluir tendencia determinística en la prueba ADF) se consideró el modelo 2 de la ecuación 3.5.4 para realizar el siguiente contraste de hipótesis:

H_0 : Hay evidencia de raíz unitaria en el proceso X_t

H_a : No existe raíz unitaria en el proceso X_t

El valor del estadístico de prueba obtenido fue -20.8878.

Tabla 21. Valores CRÍTICOS PRUEBA ADF

Valores críticos			
	1pct	5pct	10pct
Tau 2	-3.43	-2.86	-2.57

Fuente: Elaboración Propia.

Como $\tau_\mu < \tau_\mu(\alpha, n)$ para todo α , se rechaza la hipótesis nula y de esta manera se puede concluir que no hay raíz unitaria en Z_t . En la Tabla 22 se presentan los resultados del ajuste del modelo de regresión 2, donde se evidencia que la inclusión de rezagos en la prueba ADF resultó significativa.

Tabla 22. Regresión prueba ADF sin Tendencia Determinística

	Estimador	Error Estándar	Estadístico t	P valor
Intercepto	1.95644	0.09588	20.405	< 2e-16 ***
z.lag.1	-0.73295	0.03509	-20.888	< 2e-16 ***
z.diff.lag	-0.09870	0.02752	-3.586	0.000348 ***

Fuente: Elaboración Propia.

4.2.1.2 Prueba de Phillips-Perrón (PP). Empleando la prueba de Phillips-Perrón aumentada sin incluir tendencia determinística se realizó el siguiente contraste de hipótesis:

Ho: Hay evidencia de raíz unitaria en el proceso X_t

Ha: No existe raíz unitaria en el proceso X_t

El valor del estadístico de prueba obtenido fue -32.7832.

Tabla 23. Valores críticos prueba PP

Valores críticos			
	1pct	5pct	10pct
Tau 2	-3.44	-2.86	-2.57

Fuente: Elaboración Propia.

Como el estadístico de la prueba Phillips-Perrón es menor que los valores críticos para todo α , se rechaza hipótesis nula y se concluye que no hay evidencia de raíz unitaria. En la Tabla 24 se presentan los resultados del ajuste del modelo de regresión de la prueba.

Tabla 24. Regresión prueba PP sin Tendencia Determinística

	Estimador	Error Estándar	Estadístico t	P valor
Intercepto	2.18811	0.07517	29.109	< 2e-16 ***
y.l1	0.18112	0.02707	6.691	3.27e-11 ***

Fuente: Elaboración Propia.

4.2.1.4. Prueba de Zivot-Andrews. Se empleó la prueba de Zivot y Andrews para realizar el contraste de raíz unitaria contra la alternativa de un único cambio estructural de la serie empleando la ecuación de regresión del modelo A el cual permite un único cambio en el nivel de la serie (pues no hay evidencia de tendencia determinística)

Ho: Hay evidencia de raíz unitaria en el proceso X_t

Ha: No existe raíz unitaria en el proceso X_t

El valor del estadístico de prueba obtenido fue -21.4193

Tabla 25. Valores críticos prueba de Zivot-Andrews

Valores Críticos			
	1pct	5pct	10pct
Tau 2	-5.34	-4.8	-4.58

Fuente: Elaboración Propia.

Como el estadístico de la prueba de Zivot y Andrews es menor que los valores críticos para todo α , se rechaza hipótesis nula y se concluye que no hay evidencia de raíz unitaria en el proceso bajo estudio. De igual manera se identifica un posible cambio de nivel en la observación 772. En la Tabla 26 se presentan los resultados del ajuste del modelo de regresión de la prueba realizada.

Tabla 26.. Regresión prueba de Zivot-Andrews-Modelo A

	Estimador	Error estándar	Estadístico t	P valor
Intercepto	1.9118	0.1014	18.854	< 2e-16 ***
y.l1	0.2412	0.0354	6.807	1.51e-11 ***
y.dl1	-0.0894	0.0275	-3.256	0.001159 **
Du	-0.2987	0.0804	-3.714	0.000213 ***
Signif. codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.' 0.1 ' ' 1

Fuente: Elaboración Propia.

A continuación se presenta la prueba de estabilidad estructural realizada sobre la serie transformada con el fin de confirmar lo identificado con la prueba de Zivot y Andrews.

PRUEBA DE NYBLOM - HANSEN

En esta prueba se establece el siguiente contraste de hipótesis:

Ho: Hay estabilidad en la estimación de los coeficientes las funciones identificadas.

Ha: No hay estabilidad en la estimación de los coeficientes las funciones identificadas.

En la Tabla 26 se presentan los resultados obtenidos de la prueba

Tabla 27. Prueba de Nyblom-Hansen

Estadístico de prueba	Valor p
0.8341	0.1085

Fuente: Elaboración Propia.

Como el valor p de la prueba de Nyblom -Hansen es mayor que los niveles de significancia nominal considerados del 5% no se rechaza hipótesis nula y se concluye que hay estabilidad en la estimación de los coeficientes bajo el modelo considerado. Lo anterior implica que no es necesario incluir un cambio de nivel en el modelo considerado para el análisis de proceso bajo estudio. A continuación se presentan las pruebas de raíz fraccional para verificar si en efecto o no hay evidencia de memoria larga en el proceso.

4.2.2. Pruebas de raíz fraccional estacionaria. En las pruebas que se presentan en esta sección se establece el siguiente contraste de hipótesis:

- Ho: $d = 0$. El proceso X_t corresponde a un proceso ARMA(p,q) estacionario e invertible.
- Ha: $d \neq 0$ con $-\frac{1}{2} < d < \frac{1}{2}$. El proceso X_t corresponde a un proceso ARFIMA(p,d,q) estacionario e invertible.

En la Tabla 28 se presentan los resultados de las pruebas de raíz fraccional estacionaria de Geweke y Porter –Hudak(GPH) y de Robinson(ROB)

Tabla 28. Prueba de raíz Fraccional

Prueba	d	Error estándar	Estadístico de prueba	Valor P
GPH	0.1434	0.0707	2.0286	0.0252
ROBINSON	0.1807	0.0975	1.8539	0.0348

Fuente: Elaboración Propia.

Como el valor p de las pruebas mencionadas es menor que el nivel de significancia nominal del 5% se rechaza hipótesis nula y se concluye que hay evidencia de raíz fraccional estacionaria en la serie transformada del número de horas de vuelo del equipo B-350.

4.2.3. Procedimiento de identificación y estimación - modelos ARFIMA (p,d,q). A continuación se ilustra la implementación del procedimiento propuesto en Castaño et. al. (2008) para la identificación y estimación de un modelo adecuado para la serie transformada del número de horas de vuelo del equipo SA-237 asumiendo que el término de error a_t en la ecuación 3.1 sigue una distribución normal estándar.

Primer paso: siguiendo la recomendación de Castaño (2008), aproxime la componente a corto plazo por medio de un proceso autorregresivo de orden $p^* = T^{1/4} = 1407^{1/4} \approx 6$. Los resultados del modelo aproximado ARFIMA(6,d,0) se presentan en la Tabla 29.

Tabla 29. Modelo aproximado ARFIMA (6,d,0)

Parámetro	Estimador	Error estándar	t value	Pr(> t)
phi(1)	-0.0263	0.1037	-0.2535	0.7998
phi(2)	0.0114	0.0597	0.1912	0.8483
phi(3)	-0.0438	0.0436	-1.0034	0.3157
phi(4)	-0.0238	0.0403	-0.5898	0.5553
phi(5)	-0.0065	0.0379	-0.1713	0.8639
phi(6)	0.0184	0.0362	0.5092	0.6106
d.f	0.1804	0.0994	1.8148	0.0696
Fitted mean	2.6778	0.0683	39.1926	< 2e-16 ***

Fuente: Elaboración Propia.

Los resultados del modelo aproximado ajustado presentados en la Tabla 30 presentaron coeficientes autoregresivos no significativos (valores p mayores al nivel de significancia nominal del 5%) que indican la sobre parametrización del mismo. Se reespecificó el modelo aproximado a un ARFIMA (0,d,0) obteniendo los resultados que se presentan en la Tabla 30:

Tabla 30. Modelo aproximado ARFIMA (0,d,0)

Parámetro	Estimador	Error estándar	t-value	Pr(> t)
df	0.1560	0.0216	7.2153	5.38e-13 ***
Fitted mean	2.6768	0.0610	43.9055	< 2e-16

Fuente: Elaboración Propia.

Tabla 31. Prueba de Ljung-Box para residuales modelo ARFIMA (0,d,0)

Estadístico de prueba	df	Valor p
14.7773	12	0.2538

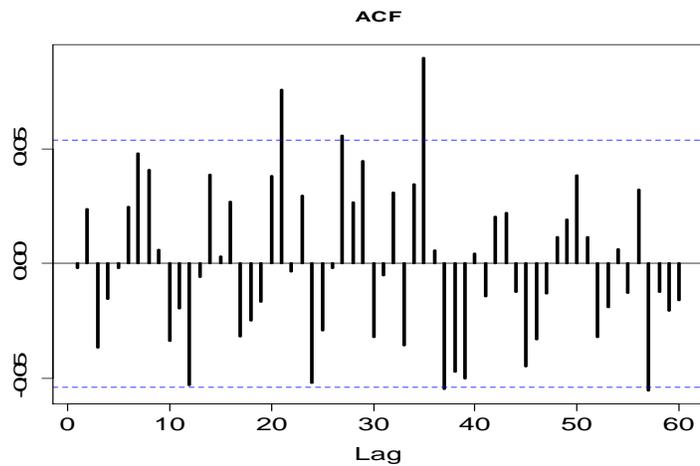
Fuente: Elaboración Propia.

En la Tabla 31 se presenta el resultado de la prueba de correlación serial de Ljung-Box sobre los residuales del modelo aproximado. Como el valor p de la prueba es mayor al nivel de significancia nominal del 5% se rechaza la hipótesis nula de correlación serial en los residuales del modelo aproximado anterior indica que el mencionado modelo captura la correlación serial presente en la serie transformada fraccionalmente diferenciada. Por lo enunciado anteriormente, se puede concluir al observar la Tabla 29 que la estimación del parámetro de diferenciación fraccional d en el modelo aproximado es significativa para un nivel de significancia nominal del 5%.

Segundo paso: Usar $\hat{d} = 0.1560$ para diferenciar fraccionalmente la serie transformada y sobre el proceso resultante utilizar las técnicas de identificación de un modelo ARMA(p,q) para el componente a corto plazo del proceso.

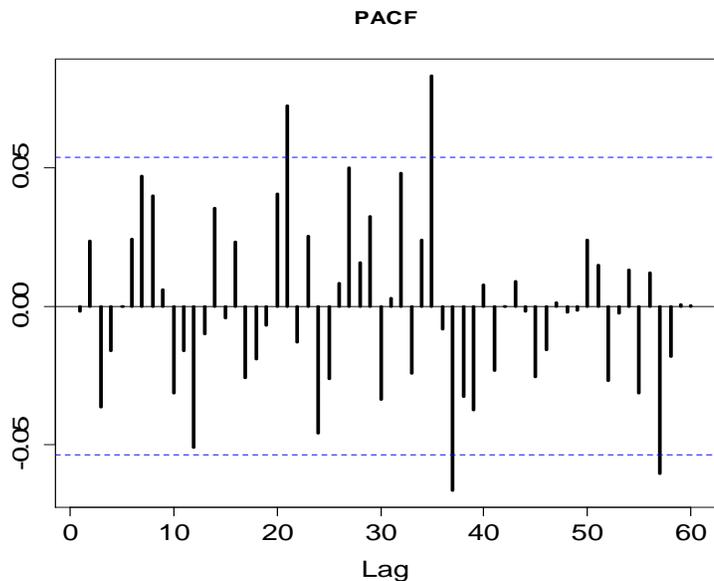
En el Gráfico 15, 16 y 17 se presentan la función de autocorrelación, la función de autocorrelación parcial y la función de autocorrelación extendida de los residuales del modelo aproximado.

Gráfico 15. ACF Serie Diferenciada Fraccionalmente



Fuente: Elaboración Propia

Gráfico 16. PACF Serie Diferenciada Fraccionalmente



Fuente: Elaboración Propia

Gráfico 17. EACF Serie diferenciada fraccionalmente

		AR/MA										
		0	1	2	3	4	5	6	7	8	9	10
0	0	0	0	0	0	0	0	0	0	0	0	0
1	X	0	0	0	0	0	0	0	0	0	0	0
2	X	X	0	0	0	0	0	0	0	0	0	0
3	X	X	X	0	0	0	0	0	0	0	0	0
4	X	X	0	0	0	0	0	0	0	0	0	0
5	X	X	0	X	X	0	0	0	0	0	0	0

Fuente: Elaboración Propia

Al observar los gráficos 15, 16 y 17, se puede concluir que el modelo ARMA(p,q) para el componente a corto plazo de la serie estudiada parece ser consistente con un modelo ARMA(0,0), lo anterior implica que no hay evidencia de componente a corto plazo en la serie transformada fraccionalmente diferenciada. En la Tabla 33 se presenta el resultado de la evaluación de diferentes modelos empleando criterio de información de Akaike (AIC) y el criterio de información bayesiano (BIC) empleando la función Autorima para la serie transformada diferenciada fraccionalmente.

Tabla 32. Resultado de los criterios AIC y BIC selección de modelos

Modelo	AIC	BIC
ARIMA(0,0,0)	2957.86	2963.05

Fuente: Elaboración Propia.

Los resultados obtenidos indican que el proceso ARFIMA(0,d,0) parece ser adecuado para modelar la serie transformada del número de horas de vuelo del equipo B-350.

Tercer paso: Estimar el modelo ARFIMA(0,d,0) identificado y validar los supuestos sobre sus residuales.

A continuación se presenta la verificación de los supuestos básicos implícitos en el modelo ARFIMA(0,d,0). De no cumplirse se determinarán las modificaciones necesarias y se repetirán las etapas anteriores hasta que la verificación indique los son resultados aceptables.

VALIDACIÓN SUPUESTOS DEL MODELO

Pruebas de incorrelación. Se consideraron las pruebas de Ljung-Box y Durbin-Watson en las cuales se considera el siguiente contraste de hipótesis para su implementación:

- Ho: No hay evidencia de correlación serial entre los residuales
- Ha: Si hay evidencia de correlación serial entre los residuales

Tabla 33. Prueba de Ljung-Box para residuales modelos ARFIMA (0,d,0)

Estadístico de prueba	df	Valor p
14.7773	12	0.2538

Fuente: Elaboración Propia.

Tabla 34. Prueba de Durbin-Watson para residuales modelo ARFIMA (0,d,0)

Estadístico de prueba	Valor p
2.0034	0.5245

Fuente: Elaboración Propia.

En las Tabla 33 y la Tabla 34 se puede observar que el valor p de ambas pruebas es mayor que el nivel de significancia considerado del 5% por lo que se puede concluir que los residuales del modelo ajustado no presentan correlación serial entre sí, lo anterior implica que el modelo ajustado captura el comportamiento del proceso adecuadamente, es decir, el modelo ajustado esta adecuadamente especificado.

PRUEBAS DE NORMALIDAD

Realizar estas pruebas tiene como finalidad el verificar que la secuencia de residuales del modelo ajustado se ajusta a la distribución normal. Para tal fin se consideraron las pruebas de normalidad de Shapiro-Wilk, Jarque – Bera, Kolmogorov - Smirnov y D'Agostino. A continuación se presentan los resultados obtenidos considerando el siguiente contraste de hipótesis:

- Ho: Hay normalidad en los residuales del modelo ajustado.
- Ha: No hay normalidad en los residuales del modelo ajustado.

Prueba de Shapiro-Wilk (S-W)

Tabla 35. Prueba de Shapiro-Wilk para residuales modelo ARFIMA (0,d,0)

Estadístico	Valor p
0.998	0.09235

Fuente: Elaboración Propia.

Según la prueba S-W, para un nivel de significancia nominal del 5% no se rechaza la hipótesis nula, lo cual implica que los residuales del modelo ARFIMA(0,d,0) ajustado siguen una distribución normal.

Prueba de Jarque-Bera (J-B)

Tabla 36. Prueba de Jarque-Bera para residuales modelo ARFIMA (0,d,0)

Estadístico	Valor p
4.2875	0.1172

Fuente: Elaboración Propia.

Según la prueba J-B, para un nivel de significancia nominal del 5% no se rechaza la hipótesis nula, lo cual implica que los residuales del modelo ARFIMA(0,d,0) ajustado siguen una distribución normal.

Prueba de Kolmogorov-Smirnov(K-S)

Tabla 37. Prueba de Kolmogorov-Smirnov para residuales modelo ARFIMA (0,d,0)

Estadístico	Valor p
0.0291	0.2138

Fuente: Elaboración Propia.

Según la prueba K-S, para un nivel de significancia nominal del 5% no se rechaza la hipótesis nula, lo cual implica que los residuales del modelo ARFIMA(0,d,0) ajustado siguen una distribución normal.

Prueba de D'Agostino

Tabla 38. Prueba de D`Agostino para residuales modelo ARFIMA (0,d,0)

	Estadístico	Valor p
Chi2 Omnibus	4.4046	0.1105
Z3 Skewness	-1.6047	0.1086
Z4 Kurtosis	-1.3527	0.1762

Fuente: Elaboración Propia.

Según la prueba de D'Agostino, para un nivel de significancia nominal del 5% no se rechaza la hipótesis nula, lo cual implica que los residuales del modelo ARFIMA(0,d,0) ajustado siguen una distribución normal.

PRUEBA CUSUM - NYBLOM HANSEN para residuales

En esta prueba se establece el siguiente contraste de hipótesis:

Ho: Hay estabilidad en la estimación de los coeficientes del modelo identificado.

Ha: No hay estabilidad en la estimación de los coeficientes del modelo identificado.

En la Tabla 39 se presentan los resultados obtenidos de la prueba

Tabla 39. Prueba de Nyblom-Hansen para residuales del modelo ajustado

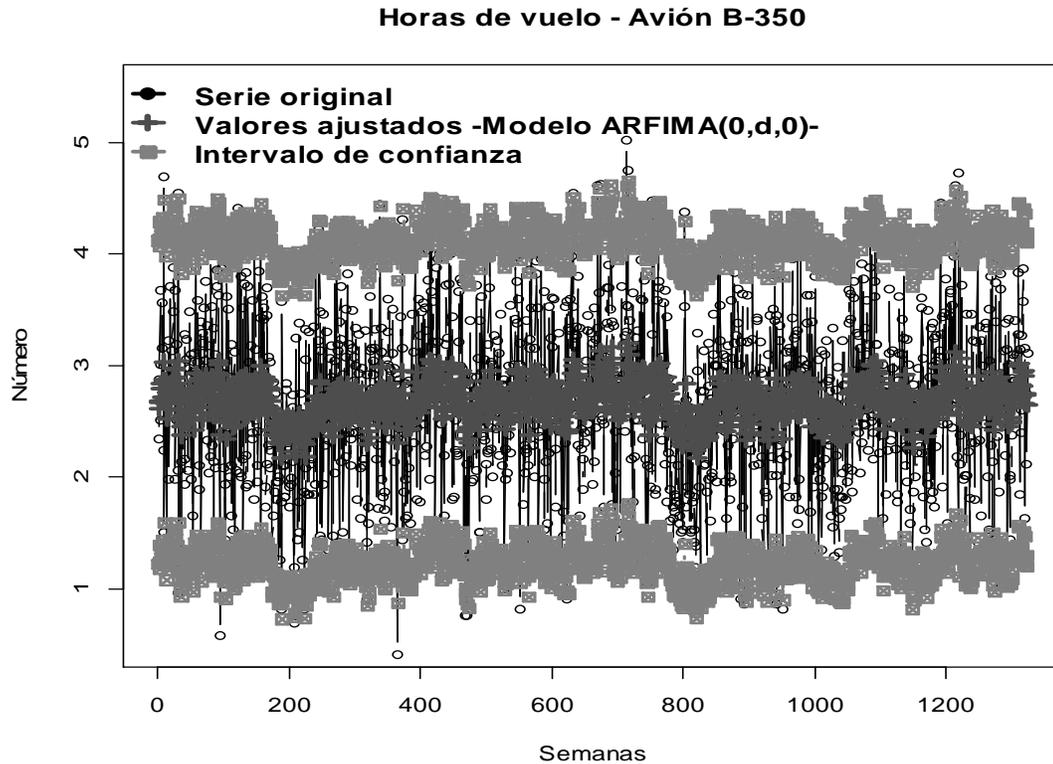
Estadístico de prueba	Valor p
0.587	0.4388

Fuente: Elaboración Propia.

Como el valor p de la prueba de Nyblom -Hansen es mayor que el nivel de significancia nominal considerado del 5%, no se rechaza hipótesis nula y se concluye que hay estabilidad en la estimación de los coeficientes del modelo identificado. Con lo cual se han verificado todos los supuestos del modelo y por lo tanto puede ser empleado para generar pronósticos del número de horas de vuelo del equipo B-350.

En el Gráfico 18 se presenta la serie transformada, los valores ajustados e intervalo de confianza del 95% del modelo ARFIMA(0,d,0) ajustado para el equipo B-350.

Gráfico 18. Serie transformada, valores ajustados e intervalo de confianza del 95% del modelo ajustado para el equipo B-350



Fuente: Elaboración propia

Cuarto paso: Uso del modelo identificado para pronosticar valores futuros del número de horas de vuelo del equipo B-350.

Teniendo el modelo ajustado es: ARFIMA (0,d,0) donde, $Z_t = \sqrt{X_t}$

$$(1 - B)^{0,156} Z_t = 2.6768 + a_t$$

$$\text{Donde, } a_t \sim RBG(0, \sigma_a^2)$$

Se presenta a continuación el pronóstico del número de horas de vuelo del equipo B-350 y su respectivo intervalo de confianza, calculados empleando dos métodos diferentes: el primero es el ajuste paramétrico o tradicional y el segundo es vía Bootstrap.

Tabla 40. Pronóstico e intervalo de confianza paramétrico

	1	2	3	4	5	6	7	8	9	10
Upper	4.978	4.988	4.991	4.993	4.993	4.992	4.992	4.992	4.991	4.990
95%	407	358	646	257	088	638	029	029	336	597
Forecast	2.759	2.742	2.736	2.733	2.730	2.728	2.726	2.725	2.723	2.722
	214	327	720	417	903	769	870	142	552	078
Lower	0.540	0.496	0.481	0.473	0.468	0.464	0.461	0.458	0.455	0.453
95%	021	296	794	910	549	450	102	255	768	559

Fuente: Elaboración Propia

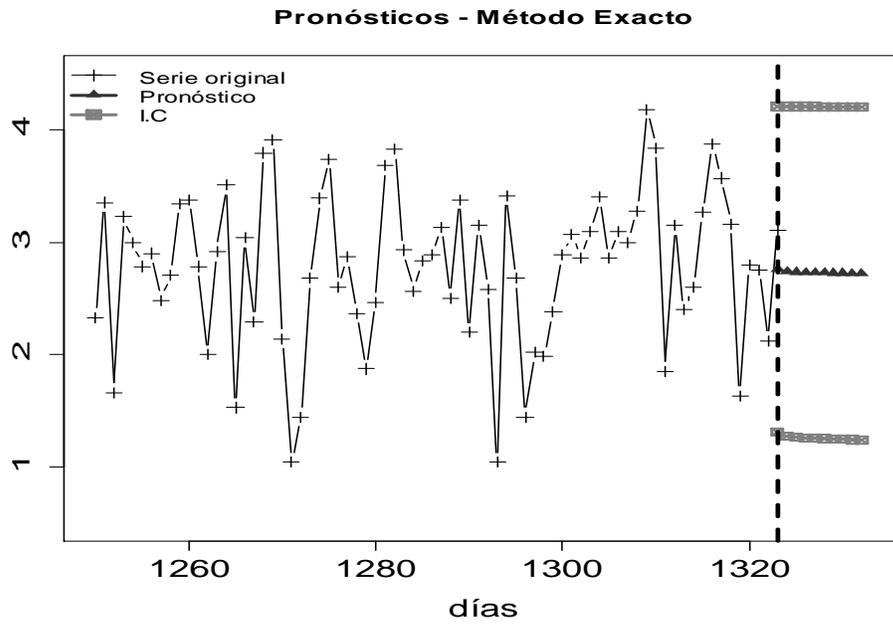
Tabla 41. Pronóstico e intervalo de confianza vía Bootstrap

	1	2	3	4	5	6	7	8	9	10
Upper	6.718	6.898	6.742	6.801	6.801	6.933	6.907	6.784	6.707	6.735
95%	77	20	13	96	96	11	68	36	76	46
Prediction	5.336	5.371	5.352	5.371	5.372	5.381	5.400	5.360	5.369	5.340
(Mean)	82	99	94	49	32	91	82	04	91	94
Lower	3.664	3.851	3.820	3.849	3.890	3.757	3.811	3.803	3.797	3.779
95%	44	30	37	81	20	90	64	40	86	36

Fuente: elaboración Propia

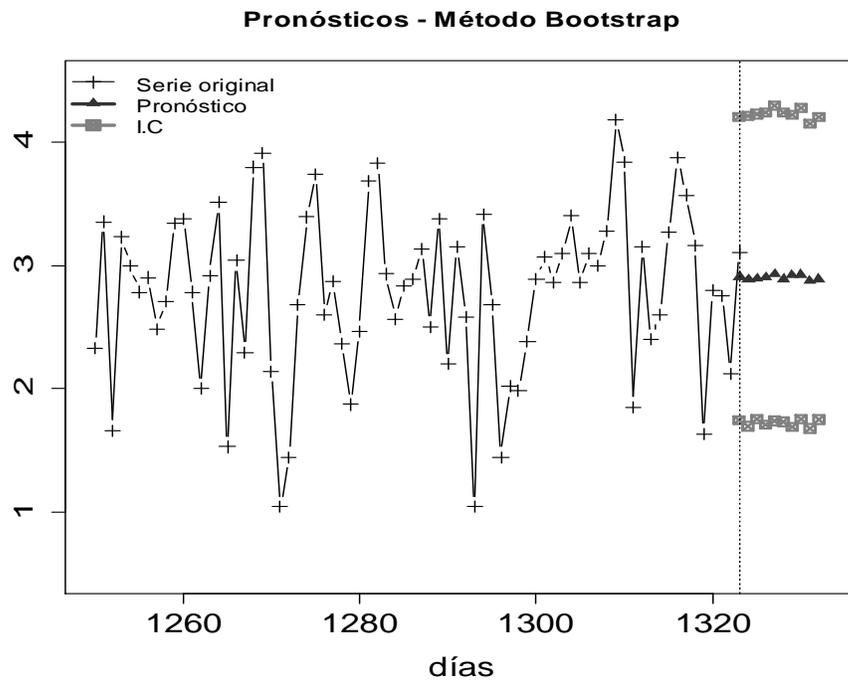
Estos son los pronósticos bajo la transformación Box Cox para el número de horas de vuelo para el equipo B-350. En los Gráficos 19 y 20 se presentan los pronósticos obtenidos y sus respectivos intervalos de confianza, empleando la verosimilitud del modelo ajustado y el método bootstrap.

Gráfico 19. Pronósticos obtenidos e intervalos de confianza método exacto



Fuente: Elaboración Propia

Gráfico 20. Pronósticos obtenidos e intervalos de confianza Bootstrap



Fuente: Elaboración Propia

5. CONCLUSIONES

Las siguientes son las conclusiones y recomendaciones obtenidas para el procedimiento de identificación del modelo ARFIMA (0,d,0), el cual permitirá pronosticar las horas a volar de los equipos aéreos SA-237 y B-350

- La elaboración de estos modelos de pronósticos servirán a la FAC para determinar la cantidad de horas de vuelo de los equipos; y de este modo ajustarse al presupuestos para así no solicitar adiciones presupuestales.
- Con los pronósticos de las horas de vuelo de los equipos B-350 y SA-237, permitirán calcular la compra de combustible necesaria para las operaciones a realizar.
- Con los pronósticos de las horas de vuelo de los equipos B-350 y SA-237, permitirá calcular la logística necesaria (repuestos y mantenimientos) para mantener listas las aeronaves.
- Tener el pronóstico de horas de vuelo de estos dos equipos permitirá al alto mando tomar decisiones acertadas en la planeación y estrategia militar.
- Es necesario elaborar modelos de pronósticos para la totalidad de los equipos de la FAC.
- Se realizaron pruebas de raíz unitaria al aeronave SA-237, obteniendo evidencia suficiente para rechazar la hipótesis nula que plantea la existencia de raíz unitaria en la serie. Empleando las pruebas de raíz fraccional se pudo establecer que la serie es estacionaria de memoria larga.
- No hay evidencia de tendencia determinística ni de componente estacional para el número de horas de vuelo de la aeronave SA-237
- Se identificó que el modelo adecuado para optimizar la serie de la aeronave SA-237 es un ARFIMA, los parámetros p y q , que mejor se ajustan fueron el modelo arma (0 0), por tanto el modelo completo resulto ser el ARFIMA (0, d, 0), siendo $d=0,1519$
- Se realizaron pruebas de raíz unitaria al aeronave B-350, obteniendo evidencia suficiente para rechazar la hipótesis nula que plantea la existencia de raíz unitaria en la serie. Empleando las pruebas de raíz fraccional se pudo establecer que la serie es estacionaria de memoria larga.

- No hay evidencia de tendencia determinística ni de componente estacional para el número de horas de vuelo de la aeronave B-350.
- Se realizaron pruebas de raíz fraccional a la serie obtenida de la aeronave B-350, obteniendo evidencia suficiente para estimar que la serie debe ser diferenciada fraccionalmente para poder identificar el modelo adecuado.
- Se identificó que el modelo adecuado para optimizar la serie de la aeronave B-350 es un ARFIMA, los parámetros p y q , que mejor se ajustan fueron el modelo arma $(0\ 0)$, por tanto el modelo completo resulto ser el ARFIMA $(0, d, 0)$, siendo $d=0,1560$
- Se efectuaron pruebas para validar supuestos de normalidad de los residuales del modelo, así mismo determinar la bondad de ajuste de la distribución de los residuales del modelo, indicando que el modelo es adecuado.

BIBLIOGRAFÍA

BERAN, Jan; PALMA, Maximum. Likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. Estados Unidos de América: Journal of the Royal Statistical Society, 1993. 57(4), p. 659–672

BOX, George; JENKINS Gwilym y REINSEL, Gregory. Time Series Analysis: Forecasting and Control, 4 ed. New Jersey.: John Wiley&Sons, 2008. 756 p.

BROCKWELL, Peter y DAVIS, Richard. Time Series: Theory and Methods. Second Edition, New York.: Springer-Verlag, 2006. P. 14

CASTAÑO, Elkin. Memorias del Curso Análisis en Series de Tiempo Lineales. Universidad Nacional de Colombia, Sede Medellín. 2012. 150 p.

CASTAÑO, GÓMEZ y GALLÓN. Una nueva prueba para el parámetro de diferenciación fraccional. Colombia: Revista Colombiana de **Estadística**, 2008. **p.70**

CENTRO DE INVESTIGACIONES HISTÓRICAS DE LA FUERZA AÉREA. *Historia de la Fuerza Aérea en Colombia* [en línea]. Colombia, CIHFA,SF. [Consulta: 27 de Mayo de 2014]. Disponible en: <https://www.fac.mil.co/?idcategoria=132>

DICKEY, David y FULLER Wayne. Distribution of the estimators for Autoregressive Time Series With a Unit Root. Journal of the American Statistical Association, Vol. 74, Issue 366 (Jun., 1979), p. 427-431.

DOLADO, Juan, GONZALO, Jesús y MAYORAL, Laura. Fraccional Dickey-Fuller Test for Unit Roots. *Econometrica*, vol. 70 N° 5 (septiembre, 2002), p. 1963-2006.

FLORES GIL, Javier. “Aplicación del método Bootstrap al contraste de hipótesis en la investigación educativa” *Revista de Educación* [en línea], 2003, [Consulta: 23 de Marzo de 2015]. Disponible en: <http://www.ince.mec.es/revistaeducacion/re>

FUERZA ARMADAS DE COLOMBIA. *Aeronaves de inteligencia en Colombia*. [en línea]. Colombia, FAC,SF. . [Consulta: 1 de Junio de 2014]. Disponible en: <http://www.webinfomil.com/2012/01/aeronaves-de-inteligencia-en-colombia.html>

GEWEKE, John y PORTER-HUDAK, Susan. The estimation and application of long-memory time series models USA: *Journal of Time Series Analysis* 4(4), 1983. p. 221–238.

GONZÁLEZ CASIMIRO, Pilar. Análisis de series temporales: modelos ARIMA, España: Universidad del País Vasco, Facultad de Ciencias Económicas, Departamento de Econometría y Estadística, 2009. 169 p.

GUERRERO, Víctor, Análisis estadístico de series de tiempo económicas, Colección CBI, México: Universidad Autónoma Metropolitana. 1991.150 p.

GRANGER, Clive y JOYEUX, Roselyne. An introduction to long-memory time series and fractional differencing, USA: *Journal of Time Series Analysis* 1, 1980. p. 15–29.

HOSKING, Jonathan. Fractional differencing, *Biometrika* USA: 1981.68(1), p. 165-176.

LEMUS, Diego y CASTAÑO, Elkin. “Prueba de hipótesis sobre la existencia de una raíz fraccional en una serie de tiempo no estacionaria”, *Lecturas De Economía* ISSN: 0120-2596 ; v.78 - p.153 - 184, 2013. Medellín (Colombia): Universidad de Antioquia.

LEÓN VARGAS, Karim. "Historia de la aviación en Colombia". *Revista Credencia* [en línea], 2011, [Consulta: 20 de Septiembre de 2014]. Disponible en: <http://www.revistacredencial.com/credencial/content/historia-de-la-aviacion-en-colombia-1911-1950>.

O-MAOPA. Manual de Operaciones Aéreas. Colombia:. Departamento de Planeación y Estadísticas, 2005. 120 p.

PALMA, Wilfredo. Long-Memory Time Series: Theory and Methods. Wiley-Interscience, New York: Hoboken, NJ, 2007

PHILIPS, Peter y PERRÓN, Pierre. Testing for a unit root in time series regression, USA: *Biometrika*, 1988. 75(2), p. 335–346.

ROBINSON, Peter. Efficient tests of nonstationary hypotheses. Newyork: *Journal of the American Statistical Association*, 1994 89(428), p. 1420–1437.

SAID, S. & DICKEY, D. Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order', USA: *Biometrika* 1984. 71, 599–607.p.

WEI, William. Time Series Analysis, Univariate and Multivariate Methods. México: Ed. Pearson, 2006. 614 p.

ZIVOT, Eric y ANDREWS, Donald. Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis, USA: *Journal of Business and Economic Statistics*, Vol. 10, No.3, p. 251-270, 1992