



UNIVERSIDAD DE SANTIAGO DE CHILE

FACULTAD DE CIENCIA

DEPARTAMENTO DE MATEMÁTICA Y CIENCIAS DE LA COMPUTACIÓN

INGENIERÍA ESTADÍSTICA

Revisión de Métodos para Estimar la Dependencia Temporal de Series de Tiempo Desigualmente Espaciadas

Jorge Quevedo Álvarez

Profesor Guía: Felipe Elorrieta López

Comisión: Claudia Matus, Abel Valdebenito

© Jorge Quevedo Álvarez.

Se autoriza la reproducción parcial o total de esta obra, con fines académicos, por cualquier forma, medio o procedimiento, siempre y cuando se incluya la cita bibliográfica del documento.

Resumen

La mayoría de los modelos de series de tiempo asumen que los datos provienen de observaciones que están igualmente espaciadas en el tiempo. Sin embargo, esta suposición no se sostiene en diversos campos científicos, como la astronomía, la economía y la paleoclimatología, entre otros.

Hay algunas técnicas que se ajustan a series de tiempo desigualmente espaciadas, como el proceso autorregresivo de tiempo continuo (CAR), el autorregresivo irregular discreto (IAR) o el autorregresivo irregular complejo discreto (CIAR).

Se muestra a través de simulaciones de Monte Carlo que el rendimiento de estos métodos es mucho mejor a diferencia de los modelos *ARMA* que asumen tiempos regulares, con una estimación del parámetro más precisa.

La metodología propuesta se aplica en tres diferentes series de tiempo desigualmente muestreadas a la magnitud de brillo de un objeto astronómico, al índice de capitalización de Colombia y la medición de la proporción de isótopos de las lluvias monzónicas del Océano Índico durante el Holoceno. En cada uno de estos ejemplos, se muestra que el ajuste de estos modelos es eficiente.

Por último, se muestra cómo el mejor modelo para cada situación, permite realizar predicciones.

Índice general

1. Introducción	10
1.1. Planteamiento del Problema	12
1.2. Objetivos	13
1.2.1. Objetivo General	13
1.2.2. Objetivos Específicos	13
1.3. Metodología	13
2. Preliminares	15
2.1. Series de Tiempo	15
2.1.1. Proceso Estocástico	15
2.1.2. Proceso Estacionario	15
2.1.3. Ruido Blanco a tiempo Discreto	16
2.1.4. Función de Autocovarianza	16
2.1.5. Función de Autocorrelación	16
2.1.6. Modelo AR	17
2.1.7. Modelo ARMA	17
2.1.8. Modelo de estado de espacio	18
2.1.9. Filtro de Kalman	18
3. Marco Teórico	20
3.1. Ruido Blanco a tiempo Continuo	20
3.2. Movimiento Browniano	20
3.3. Modelo CAR	21
3.3.1. CAR(1)	22
3.3.2. Reparametrización CAR	23
3.3.3. Estimación de parámetros por filtro de Kalman	23
3.3.4. Selección del modelo	26

3.4. Modelo IAR	26
3.4.1. Relación con CAR(1)	26
3.4.2. Relación con AR(1)	27
3.4.3. Estimación de Parámetros	27
3.5. Modelo CIAR	28
3.5.1. Estimación de Parámetros	29
4. Resultados	31
4.1. Simulación	32
4.1.1. Datos simulados a partir de un model <i>IAR</i>	33
4.1.2. Datos simulados a partir de un model <i>CIAR</i>	56
4.2. Datos reales	74
4.2.1. Base con datos Astronómicos	75
4.2.2. Índice de Capitalización	83
4.2.3. Proporción de Isótopos	92
4.3. Predicciones	100
5. Conclusión	103

Índice de figuras

4.1. Gráfico de datos simulados en base a un modelo $IAR(1)$ con datos irregularmente espaciados . . .	34
4.2. Gráfico de datos simulados a partir de un modelo $IAR(1)$ versus datos predichos por un modelo $AR(1)$ con datos irregularmente espaciados	36
4.3. Gráfico de datos simulados a partir de un modelo $IAR(1)$ versus datos predichos por un modelo $CAR(1)$ con datos irregularmente espaciados	39
4.4. Gráfico de datos simulados a partir de un modelo $IAR(1)$ versus datos predichos por un modelo $IAR(1)$ con estimación por filtro de Kalman, con datos irregularmente espaciados	42
4.5. Gráfico de datos simulados a partir de un modelo $IAR(1)$ versus datos predichos por un modelo $IAR(1)$ con estimación por máxima verosimilitud directa, con datos irregularmente espaciados . .	46
4.6. Gráfico de datos simulados a partir de un modelo $IAR(1)$ versus datos predichos por un modelo $IAR - T$ con estimación por máxima verosimilitud, con datos irregularmente espaciados	49
4.7. Gráfico de datos simulados a partir de un modelo $IAR(1)$ versus datos predichos por un modelo $CIAR$ con estimación por filtro de Kalman, con datos irregularmente espaciados	53
4.8. Gráfico de datos simulados en base a un modelo $CIAR$ con datos irregularmente espaciados	58
4.9. Gráfico de datos simulados a partir de un modelo $CIAR$ versus datos predichos por un modelo $AR(1)$ con datos irregularmente espaciados	59
4.10. Gráfico de datos simulados a partir de un modelo $CIAR$ versus datos predichos por un modelo $CAR(1)$ con datos irregularmente espaciados	62
4.11. Gráfico de datos simulados a partir de un modelo $CIAR$ versus datos predichos por un modelo $IAR(1)$ con estimación por filtro de Kalman, con datos irregularmente espaciados	64
4.12. Gráfico de datos simulados a partir de un modelo $CIAR$ versus datos predichos por un modelo $IAR(1)$ con estimación por máxima verosimilitud directa, con datos irregularmente espaciados . .	66
4.13. Gráfico de datos simulados a partir de un modelo $CIAR$ versus datos predichos por un modelo $IAR - T$ con estimación por máxima verosimilitud, con datos irregularmente espaciados	68
4.14. Gráfico de datos simulados a partir de un modelo $CIAR$ versus datos predichos por un modelo $CIAR$ con estimación por filtro de Kalman, con datos irregularmente espaciados	70

4.15. Gráfico de datos Blazar	75
4.16. Gráfico de datos Blazar versus datos predichos por un modelo $AR(1)$, con datos irregularmente espaciados	76
4.17. Gráfico de datos Blazar versus datos predichos por un modelo $CAR(1)$, con datos irregularmente espaciados	78
4.18. Gráfico de datos Blazar versus datos predichos por un modelo $IAR(1)$ con estimación por filtro de Kalman, con datos irregularmente espaciados	79
4.19. Gráfico de datos Blazar versus datos predichos por un modelo $IAR(1)$ con estimación por máxima verosimilitud directa, con datos irregularmente espaciados	80
4.20. Gráfico de datos Blazar versus datos predichos por un modelo $IAR - T$, con datos irregularmente espaciados	81
4.21. Gráfico de datos Blazar versus datos predichos por un modelo $CIAR$ con estimación por filtro de Kalman, con datos irregularmente espaciados	82
4.22. Grafico de datos de INDICE ICAP	84
4.23. Grafico de datos de INDICE ICAP	85
4.24. Gráfico de datos ICAP versus datos predichos por un modelo $AR(1)$, con datos irregularmente espaciados	86
4.25. Gráfico de datos ICAP versus datos predichos por un modelo $CAR(1)$, con datos irregularmente espaciados	87
4.26. Gráfico de datos ICAP versus datos predichos por un modelo $IAR(1)$ con estimación por filtro de Kalman, con datos irregularmente espaciados	88
4.27. Gráfico de datos ICAP versus datos predichos por un modelo $IAR(1)$ con estimación por máxima verosimilitud directa, con datos irregularmente espaciados	89
4.28. Gráfico de datos ICAP versus datos predichos por un modelo $IAR - T$, con datos irregularmente espaciados	90
4.29. Gráfico de datos ICAP versus datos predichos por un modelo $CIAR$ con estimación por filtro de Kalman, con datos irregularmente espaciados	91
4.30. Grafico de datos 4_7	93
4.31. Gráfico de datos 4_7 versus datos predichos por un modelo $AR(1)$, con datos irregularmente espaciados	94
4.32. Gráfico de datos 4_7 versus datos predichos por un modelo $CAR(1)$, con datos irregularmente espaciados	95
4.33. Gráfico de datos 4_7 versus datos predichos por un modelo $IAR(1)$ con estimación por filtro de Kalman, con datos irregularmente espaciados	96
4.34. Gráfico de datos 4_7 versus datos predichos por un modelo $IAR(1)$ con estimación por máxima verosimilitud directa, con datos irregularmente espaciados	97

4.35. Gráfico de datos 4_7 versus datos predichos por un modelo $IAR - T$, con datos irregularmente espaciados	98
4.36. Gráfico de datos 4_7 versus datos predichos por un modelo $CIAR$ con estimación por filtro de Kalman, con datos irregularmente espaciados	99
4.37. Gráfico de datos <i>Blazar</i> versus datos predichos por un modelo $CIAR$ con estimación por filtro de Kalman, con datos irregularmente espaciados	101
4.38. Gráfico de datos 4_7 versus datos predichos por un modelo $CIAR$ con estimación por filtro de Kalman, con datos irregularmente espaciados	101

Índice de cuadros

4.1. Descripción de las variables	34
4.2. Parámetro real vs Estimado; Tiempo de Computo	36
4.3. Bondad de Ajuste del Modelo $AR(1)$	37
4.4. Parámetro real vs Estimado; Tiempo de Computo	40
4.5. Bondad de Ajuste del Modelo $CAR(1)$	40
4.6. Parámetro real vs Estimado; Tiempo de Computo	43
4.7. Bondad de Ajuste del Modelo $IAR(1)$ filtro de Kalman	44
4.8. Parámetro real vs Estimado; Tiempo de Computo	46
4.9. Bondad de Ajuste del Modelo $IAR(1)$ máxima verosimilitud	47
4.10. Parámetro real vs Estimado; Tiempo de Computo	50
4.11. Bondad de Ajuste del Modelo $IAR - T$	50
4.12. Parámetro real vs Estimado; Tiempo de Computo	53
4.13. Bondad de Ajuste del Modelo $CIAR$ filtro de Kalman	54
4.14. Comparación de Modelos simulados por $IAR(1)$ y tiempos de naturaleza exponencial	55
4.15. Comparación de Modelos simulados por $IAR(1)$ y tiempos de naturaleza uniforme	56
4.16. Descripción de las variables	58
4.17. Parámetro real vs Estimado; Tiempo de Computo	60
4.18. Bondad de Ajuste del Modelo $AR(1)$	61
4.19. Parámetro real vs Estimado; Tiempo de Computo	62
4.20. Bondad de Ajuste del Modelo $CAR(1)$	63
4.21. Parámetro real vs Estimado; Tiempo de Computo	64
4.22. Bondad de Ajuste del Modelo $IAR(1)$ filtro de Kalman	65
4.23. Parámetro real vs Estimado; Tiempo de Computo	66
4.24. Bondad de Ajuste del Modelo $IAR(1)$ máxima verosimilitud	67
4.25. Parámetro real vs Estimado; Tiempo de Computo	69
4.26. Bondad de Ajuste del Modelo $IAR - T$	69

4.27. Parámetro real vs Estimado; Tiempo de Computo	71
4.28. Bondad de Ajuste del Modelo <i>CIAR</i> filtro de Kalman	71
4.29. Comparación de Modelos simulados por <i>CIAR</i> y tiempos de naturaleza exponencial	73
4.30. Comparación de Modelos simulados por <i>CIAR</i> y tiempos de naturaleza exponencial	74
4.31. Descripción de las variables	75
4.32. Análisis de Tendencia	76
4.33. Bondad de Ajuste del Modelo <i>AR</i> (1)	77
4.34. Bondad de Ajuste del Modelo <i>CAR</i> (1)	78
4.35. Bondad de Ajuste del Modelo <i>IAR</i> (1) filtro de Kalman	79
4.36. Bondad de Ajuste del Modelo <i>IAR</i> (1) máxima verosimilitud	81
4.37. Bondad de Ajuste del Modelo <i>IAR</i> – <i>T</i>	82
4.38. Bondad de Ajuste del Modelo <i>CIAR</i> filtro de Kalman	83
4.39. Análisis de Tendencia	84
4.40. Análisis de Tendencia	85
4.41. Descripción de las variables	85
4.42. Bondad de Ajuste del Modelo <i>AR</i> (1)	86
4.43. Bondad de Ajuste del Modelo <i>CAR</i> (1)	87
4.44. Bondad de Ajuste del Modelo <i>IAR</i> (1) filtro de Kalman	88
4.45. Bondad de Ajuste del Modelo <i>IAR</i> (1) máxima verosimilitud	90
4.46. Bondad de Ajuste del Modelo <i>IAR</i> – <i>T</i>	91
4.47. Bondad de Ajuste del Modelo <i>CIAR</i> filtro de Kalman	92
4.48. Descripción de las variables	93
4.49. Análisis de Tendencia	93
4.50. Bondad de Ajuste del Modelo <i>AR</i> (1)	94
4.51. Bondad de Ajuste del Modelo <i>CAR</i> (1)	95
4.52. Bondad de Ajuste del Modelo <i>IAR</i> (1) filtro de Kalman	96
4.53. Bondad de Ajuste del Modelo <i>IAR</i> (1) máxima verosimilitud	98
4.54. Bondad de Ajuste del Modelo <i>IAR</i> – <i>T</i>	99
4.55. Bondad de Ajuste del Modelo <i>CIAR</i> filtro de Kalman	100
4.56. Mejores modelos para cada Base de Datos Reales	101

Capítulo 1

Introducción

En la actualidad todo tipo de entidad de desarrollo laboral, ya sea una empresa, distintas áreas de salud, de estudios, etc., necesita generar proyecciones para tener una idea a priori del cómo se va a comportar la variable de interés medida en el tiempo, para así poder planificar, prever y/o prevenir ciertos fenómenos que puedan ocurrir.

ya sea con la meta de predecir ciertos cambios que puedan ocurrir en el futuro en una compañía y así poder planificar una respuesta concreta a estos, poder determinar el comportamiento de algún virus o

Para realizar estas proyecciones, es necesario tener información pasada del fenómeno y así poder generar modelos predictivos. Para cumplir con este objetivo, se propone el uso de métodos de series cronológicas o series de tiempo, la cual es una rama de la estadística enfocada en el análisis de datos coleccionados en el tiempo.

Si se quiere modelar datos que se recopilan, observan o registran en intervalos de tiempo regulares (diario, mensual, semestral, anual, etc.) se pueden utilizar modelos que asumen tiempos discretos como el modelo autorregresivo ($AR(p)$), modelo de medias móviles ($MA(q)$) o modelo autorregresivo de medias móviles ($ARMA(p, q)$), entre otros. Estos modelos podrían ser aplicados a cualquier estudio medible en el tiempo de forma regular que sea estacionario.

Tal como existen fenómenos que son medidos regularmente en el tiempo, es decir, poseen toda su información en períodos donde la distancia entre tiempos es constante. Están también los que se miden irregularmente en el tiempo, donde estos, a diferencia de los anteriormente nombrados, son observaciones y valores medidos en el tiempo donde la distancia entre tiempos consecutivos difiere [1, Eyheramendy et al. (2018)].

Las series de tiempo irregularmente espaciadas son observadas comúnmente en muchas disciplinas. Por ejemplo, en desastres naturales como terremotos, inundaciones o erupciones volcánicas con diferentes brechas de tiempo. En astronomía, las observaciones son obtenidas generalmente en intervalos de tiempo irregulares debido a, por

ejemplo, su dependencia de cielos despejados para poder obtener datos de observación.[1, Eyheramendy et al. (2018)]. Para este tipo de disciplinas, los modelos más comunes($AR(p)$, $MA(q)$ y $ARMA(p, q)$) tienden a fallar principalmente en la estimación de parámetros lo cual desencadena múltiples errores en la implementación del modelo, haciendo a este poco viable. Por esto será necesario recurrir a otro tipo de modelos de series de tiempo como el modelo autorregresivo de tiempo continuo ($CAR(p)$)[1, Eyheramendy et al. (2018)][2, Elorrieta et al. (2019)], modelo irregular autorregresivo (IAR)[1, Eyheramendy et al. (2018)][2, Elorrieta et al. (2019)] y el modelo irregular complejo autorregresivo ($CIAR$)[2, Elorrieta et al. (2019)], corrigiendo la estimación y brindando un mejor ajuste a la naturaleza de los datos . Además de métodos de comparación de autocorrelación[3, Rehfeld (2011)].

Con el objetivo de poder profundizar en estos modelos de series de tiempo irregularmente espaciadas, se van a comparar los resultados obtenidos por la librería AITS de R[5, Elorrieta et al. (2020)] que cuenta con las funciones necesarias para implementar tanto el modelo $CIAR$ como IAR con respecto a la librería CTS de R[4, Wang (2013)] que entrega las funciones requeridas para utilizar el modelo CAR .

1.1. Planteamiento del Problema

Para realizar cualquier tipo de análisis temporal, son necesarios los datos asociados al fenómeno. Estos datos pueden ser obtenidos principalmente en dos formatos, el primero es que la base tenga información regularmente espaciada. Por otra parte, el segundo formato es que se presenten los datos de forma irregular en el tiempo, de los cuales no existe mucho conocimiento sobre su modelamiento en la literatura actual.

En este estudio se dará prioridad a los modelos de series de tiempo irregularmente espaciados *CAR*, *IAR*, *CIAR* en conjunto con los métodos de comparación de autocorrelación. En base a éstos se quiere realizar comparaciones entre el paquete AITS y CTS de R.

El modelo *CAR* es para resolver una ecuación diferencial de orden 1 conducida por ruido blanco. Generalmente se utiliza al momento de integrar una caminata aleatoria continua (Comúnmente llamado movimiento Browniano o proceso Wiener). Una caminata aleatoria de tiempo continuo es el límite discreto del tiempo de una caminata aleatoria que va reduciendo su intervalo de tiempo[1, Eyheramendy et al. (2018)].

El modelo *IAR* es una extensión del modelo autorregresivo regular en el sentido de que bajo cierto supuesto paramétrico el proceso *IAR* se convierte en el modelo autorregresivo de orden 1 (*AR(1)*) para datos regularmente espaciados. Además, los dos primeros momentos de los procesos *CAR* e *IAR* son los mismos, por lo que bajo gaussianidad, ambos procesos son equivalentes. En otras palabras, se puede considerar que el proceso *IAR* es una representación discreta del proceso *CAR(1)*[6, Elorrieta (2018)].

Siguiendo el enfoque de [7, Sekita et al.(1991)] que construyen un modelo complejo autorregresivo para tiempos regulares. Se define el proceso complejo autorregresivo irregular (*CIAR*) de orden 1, el cual se asume está compuesto por una parte real observable y una imaginaria que será definida como un proceso latente. Este proceso al igual que el modelo *CARFIMA*[8, Tsai (2009)], pueden ser utilizados en series de tiempo irregularmente espaciadas con función de autocorrelación negativa. Una forma de diferenciarlos es que el proceso *CIAR* presenta un decaimiento exponencial mientras que el *CARFIMA* tiene un decaimiento más lento. El uso apropiado de estos modelos dependerá de la estructura de correlación de los datos.[6, Elorrieta (2018)].

Las tres implementaciones anteriormente expuestas serán utilizadas en datos astronómicos como paleoclimáticos, ya que comúnmente son datos irregularmente espaciados. Para así poder comparar los resultados que entregue el software R para las distintas librerías y determinar cual ejecuta un mejor ajuste.

1.2. Objetivos

1.2.1. Objetivo General

- Realizar una revisión de los modelos de series de tiempo para datos irregularmente espaciados.

1.2.2. Objetivos Específicos

- Hacer una revisión bibliográfica respecto a los métodos para estimar la dependencia de series de tiempo irregularmente espaciadas.
- Proponer y justificar modelos de series de tiempo para datos irregularmente espaciados (*CAR*, *IAR* y *CIAR*).
- Realizar un estudio de simulación que permita estudiar la precisión de los métodos de estimación de los parámetros de los modelos.
- Evaluar contextos en que el uso de cada modelo es apropiado.
- Implementar modelos en series de tiempo reales irregularmente muestreadas de distintas áreas de la investigación.
- Reevaluar contextos en que el uso de cada modelo es apropiado.

1.3. Metodología

Para dar inicio a este proyecto, se realizará un amplio estudio bibliográfico con respecto a los modelos de series de tiempo irregularmente muestreados, dándole principal énfasis al *CAR*, *IAR* y *CIAR*. La parte teórica cumplirá con un rol fundamental, ya que antes de poder implementar estos modelos será de gran importancia ver semejanzas y diferencias entre los métodos para conocer contextos en que debiesen funcionar mejor. Además se estudiarán otros métodos de estimación de dependencia temporal para datos irregularmente espaciados como la función de correlación discreta(*DCF*)[3, Rehfeld (2011)].

Posteriormente, se va a explorar en las librerías que posee el software R para la implementación de modelos de serie de tiempo para datos continuos, específicamente *CTS* y *AITTS*, las cuales tienen las funciones necesarias para implementar y estimar estos modelos.

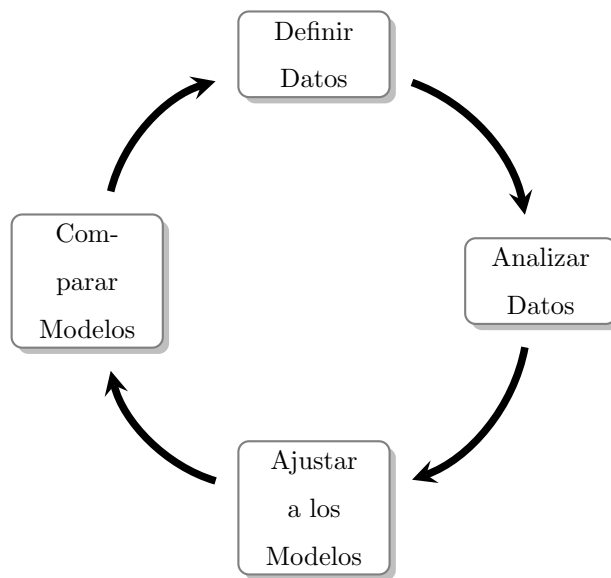
Una vez completo el estudio tanto teórico como práctico, específicamente la implementación con el software R. Se procede a testear estos modelos en datos simulados mediante el método de simulación de Monte Carlo con el objetivo de comprobar la eficiencia de cada uno de estos al momento de estimar los parámetros. Como fue mencionado con anterioridad, estas simulaciones serán ajustadas a distintos fenómenos por lo que dependiendo del

contexto, el ajuste de los modelos va a variar considerablemente.

Ya comprobada la eficiencia de estos modelos con datos simulados, se procede a implementar en datos reales los cuales constan de observaciones astronómicas y datos paleoclimáticos, donde es común observar que estos eventos sean de forma irregularmente espaciada. Al igual que con los datos simulados se va a evaluar el ajuste de cada modelo al fenómeno en estudio.

Finalmente, se recopilará la información entregada por el uso de estos modelos tanto en datos simulados como en los de carácter real, utilizando las distintas librerías proporcionadas por el software R, comparando resultados y analizando que combinación tanto de modelo como librería se ajusta mejor a los datos irregularmente espaciados.

A modo de resumen se plantea el siguiente esquema.



Capítulo 2

Preliminares

En este capítulo se definirán algunos conceptos necesarios para el entendimiento de este documento, los cuales serán secundados a partir de diversos textos.

2.1. Series de Tiempo

En muchos fenómenos sociales y naturales podemos encontrar variables medidas en periodos constantes de tiempo (días, meses, semestres, años, etc). Estas variables se pueden representar a partir de una serie de observaciones; a esta serie de observaciones la llamaremos serie cronológica o serie temporal. Uno de los usos más habituales de las series cronológicas es su análisis para predicción y pronóstico (como por ejemplo para series de datos demográficos, datos climáticos, etc). A continuación estudiaremos sus conceptos básicos.

2.1.1. Proceso Estocástico

Corresponde a un familia de variables aleatorias, donde existe un conjunto arbitrario de tiempo, denominado T , el cual es no aleatorio y es de la forma $T = \{1, 2, 3, \dots, n\}$, bajo esto, el proceso estocástico queda definido como: $Y = \{Y(t), t \in T\}$, o bien, $Y = \{Y_1, Y_2, Y_3, \dots, Y_n\}$.

2.1.2. Proceso Estacionario

Un proceso estocástico (serie temporal) es estacionaria cuando es estable a lo largo del tiempo, es decir, cuando la media y varianza son constantes en el tiempo.

2.1.2.1. Estacionaridad estricta o de primer orden

Un proceso estocástico $Y(t)$ se dice estrictamente estacionario, si la función de densidad conjunta, de cualesquiera de sus n v.a. medidas en instantes t_1, \dots, t_n , permanece constante cuando transcurre cualquier intervalo de tiempo, es decir.

$$F(y_1, y_2, \dots, y_n; t_1, t_2, \dots, t_n) = F(y_1, y_2, \dots, y_n; t_1 + k, t_2 + k, \dots, t_n + k)$$

2.1.2.2. Estacionaridad débil o de segundo orden

Un proceso estocástico $Y(t)$ se dice débilmente estacionario si:

1. $\mathbb{E}(Y_t) = \mu = cte < \infty, (\forall t ; t \in T)$
2. $\mathbb{V}(Y_t) = \sigma^2 = cte < \infty, (\forall t ; t \in T)$
3. $\mathbb{Cov}(Y_t, Y_s) = \gamma(t - s), (\forall t ; t \in T)(\forall s ; s \in T)$

Se puede demostrar que si un proceso estocástico es estrictamente estacionario y si $\mathbb{E}(y_1^2) < \infty$, también es débilmente estacionario.

2.1.3. Ruido Blanco a tiempo Discreto

Tanto en las series de tiempo discretas como continuas está presente el ruido blanco. En este apartado sólo se considerará el proceso de ruido blanco en tiempo discreto.

Un proceso estocástico de tiempo discreto $\{\epsilon_n, n \in N_0\}$ se denomina proceso de ruido blanco (RB) discreto si tiene las siguientes propiedades:

1. $\mathbb{E}(\epsilon_n) = \mu, n \in N_0$
2. $\mathbb{V}(\epsilon_n) = \sigma^2$
3. $\mathbb{Cov}(\epsilon_n, \epsilon_{n+k}) = 0$

Usualmente se considera el caso de $\mathbb{E}(\epsilon) = 0$, que se conoce como ruido blanco centrado. Se denota $\epsilon_n \sim RB(0, \sigma^2)$.

2.1.4. Función de Autocovarianza

Para un proceso débilmente estacionario se define la función de autocovarianza como:

$$\gamma(k) = \mathbb{Cov}(Y_t, Y_{t \pm k})$$

donde k es el retraso (lag) de la serie, es decir, es el período que separa dos observaciones.

2.1.5. Función de Autocorrelación

Para un proceso débilmente estacionario se define la función de autocorrelación como:

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)}$$

donde k es el retraso (lag) de la serie, es decir, es el período que separa dos observaciones.

2.1.6. Modelo AR

Un proceso estocástico $\{Y_t\} \in T, t \in \mathbb{Z}$ se dice autorregresivo de orden p si:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \epsilon_t \quad (2.1)$$

Donde $\epsilon_t \sim RB(0, \sigma^2)$, $Cov(\epsilon_t, Y_t) = 0$ para todo $j > 0$ y ϕ_1, \dots, ϕ_p son coeficientes fijos. Es posible reescribir la ecuación (2.1), a partir del operador de rezago B y el polinomio característico Φ como:

$$\Phi(B)Y_t = \epsilon_t$$

Donde $\Phi(y) = 1 - \phi_1 y - \phi_2 y^2 - \dots - \phi_p y^p$ y $BY_t = Y_{t-1}$.

Para que un proceso $AR(p)$ sea estacionario se debe comprobar que las raíces del polinomio autorregresivo estén fuera del círculo unitario, es decir $|\phi| < 1$. Ya cumpliéndose esto, las propiedades de este modelo pueden apreciarse para un modelo $AR(1)$ son.

1. $\mathbb{E}(Y_t) = 0$
2. $\mathbb{V}(Y_t) = \frac{\sigma_\epsilon^2}{1-\phi^2}$
3. $\gamma(k) = \phi^k \gamma(0)$, donde $\gamma(0) = \frac{\sigma_\epsilon^2}{1-\phi}$

2.1.7. Modelo ARMA

El Modelo autorregresivo de medias móviles de orden p, q , es una combinación entre el modelo autorregresivo de orden p y el de medias móviles de orden q .

Un proceso estocástico $\{Y_t\}, t \in \mathbb{N}$ sigue un proceso $ARMA(p, q)$ si:

$$Y_t - \phi_1 Y_{t-1} - \dots - \phi_p Y_{t-p} = \epsilon_t - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q} \quad (2.2)$$

Donde $\epsilon_t \sim RB(0, \sigma_\epsilon^2)$, $\phi_1, \dots, \phi_p; \theta_1, \dots, \theta_q$ son coeficientes fijos. Es posible reescribir la ecuación (2.2), a partir del operador de rezago B y los polinomios característicos Φ y Θ como:

$$\Phi(B)Y_t = \Theta(B)\epsilon_t$$

Donde $\Phi(y) = 1 - \phi_1 y - \dots - \phi_p y^p$ y $\Theta(y) = 1 - \theta_1 y - \dots - \theta_q y^q$.

Al igual que en el modelo $AR(1)$, la condición de estacionaridad es que $|\phi| < 1$. Ya cumpliéndose esto, las propiedades de este modelo pueden apreciarse para un modelo $ARMA(1, 1)$.

1. $\mathbb{E}(Y_t) = 0$
2. $\mathbb{V}(Y_t) = \frac{\sigma_\epsilon^2(1-2\phi\theta+\theta^2)}{1-\phi^2}$
3. $\gamma(k) = \phi^{k-1}\gamma(1)$, donde $\gamma(1) = \frac{\sigma_\epsilon^2(\phi+\theta)(1+\phi\theta)}{1-\phi^2}$
4. $\rho(k) = \phi^{k-1}\rho(1)$, donde $\rho(1) = \frac{(\phi+\theta)(1+\phi\theta)}{1+2\phi\theta+\theta^2}$

2.1.8. Modelo de estado de espacio

Un modelo de estado de espacio para una serie temporal $\{Y_t\}$ consiste en dos ecuaciones. La primera, conocida como la ecuación de observación, expresa la observación Y_t de w dimensiones como una función lineal de una variable de estado X_t de v dimensiones más ruido, es decir:

$$Y_t = G_t X_t + Z_t, \quad t = 1, 2, \dots \quad (2.3)$$

Donde $Z_t \sim RB(0, R_t)$ y G_t es la matriz de estado de $w \times v$ dimensiones, puede contener un estado o ser la matriz identidad. La segunda ecuación, llamada ecuación de estado, determina el estado X_{t+1} en términos del estado anterior X_t y un término de ruido [18, Brockwell et al (2002)]. La ecuación de estado es:

$$X_{t+1} = F_t X_t + V_t, \quad t = 1, 2, \dots \quad (2.4)$$

Donde $V_t \sim RB(0, Q_t)$ y está no correlacionado con Z_t ; y F_t es la matriz de transición de estado de $v \times v$ dimensiones. Es posible determinar el orden del sistema en base a esta matriz.

2.1.9. Filtro de Kalman

Al momento de estimar parámetros como de realizar predicciones en un modelo de estado de espacio es fundamental el filtro de Kalman ya que permite realizar solo un enfoque. En este caso será utilizado si se logra representar un modelo continuo a partir de uno de estado de espacio.

A partir de las ecuaciones (2.3) y (2.4) para poder realizar las predicciones en base a este algoritmo, es necesario definir un par de conceptos. Sea $S_t = \{Y_1, \dots, Y_t\}$ la información disponible en el tiempo t , donde el modelo es conocido. Se define $X_{t|j} = \mathbb{E}(X_t|S_j)$ y $\sum_{t|j} = \text{Var}(X_t|S_j)$ como la media condicional y varianza condicional de $X_t|S_j$, respectivamente. Además, sea $v_t = Y_t - Y_{t|t-1}$ y $V_t = \text{Var}(v_t|S_{t-1})$ el error de predicción a un paso y la varianza de este error de predicción dado S_{t-1} , respectivamente [19, Tsay (2005)].

Ahora, dado los valores iniciales $X_{1|0}$ y $\sum_{1|0}$ se tiene el filtro de Kalman por el modelo de estado de espacio para las ecuaciones (2.3) y (2.4), el cuál, es el siguiente:

1. $v_t = Y_t - G_t X_{t|t-1}$
2. $V_t = G_t \sum_{t|t-1} G_t' + R_t$
3. $K_t = F_t \sum_{t|t-1} G_t' V_t^{-1}$ donde K_t se conoce comúnmente como la ganancia de Kalman, que es el coeficiente de regresión de $X_{t|j}$ en v_t
4. $L_t = F_t - K_t G_t$
5. $X_{t+1|t} = F_t X_{t|t-1} + K_t v_t$
6. $\sum_{t+1|t} = F_t \sum_{t|t-1} L_t' + Q_t$

La idea de ocupar el filtro Kalman en este estudio, es que nos permite obtener la estimación de los parámetros de un modelo de estado de espacio por medio de la máxima verosimilitud.

Capítulo 3

Marco Teórico

3.1. Ruido Blanco a tiempo Continuo

Como se mencionó con anterioridad, las series de tiempo discretas y continuas difieren en el como se presenta el ruido blanco.

Un proceso estocástico de tiempo continuo $\{\epsilon_t, t \in [0, \infty)\}$ se denomina como proceso de ruido blanco continuo si tiene las siguientes propiedades:

1. $\mathbb{E}(\epsilon_t) = 0$
2. $\text{Cov}(\epsilon_t, \epsilon_s) = \delta(t - s)\sigma_\epsilon^2, \forall t, s \in \mathbb{R}^+$, donde δ es la función de masa de Dirac.[20, Pandit & Wu (1983)]

3.2. Movimiento Browniano

Este concepto es de alta relevancia, ya que los modelos a estudiar, están basados en este tipo de procesos. Además de que se establece un relación entre el movimiento Browniano y el ruido blanco continuo. Se utilizara el movimiento Browniano unidimensional de valor real.

Un proceso estocástico de tiempo continuo $\{W_t, t \in [0, \infty)\}$ se denomina movimiento Browniano si tiene las siguientes propiedades:

1. $W(0) = 0$
2. $\exists \sigma_W^2 \in [0, \infty)$, llamada varianza de $\{W(t), t \in [0, \infty)\}$, y así, $W(t) \sim N(0, t\sigma_W^2)$, es decir, $W(t)$ distribuye normal con media 0 y varianza $t\sigma_W^2$.
3. El proceso $\{W(t), t \in [0, \infty)\}$ aumenta de forma estacionaria, es decir, para todos $s > 0$ y $t > 0$, la distribución de $W(t + s) - W(t)$ es independiente del tiempo.

4. El proceso $\{W(t), t \in [0, \infty)\}$ aumenta de forma independiente, es decir, $W(t_2) - W(t_1)$ y $W(t_3) - W(t_2)$ son independientes para cualquier número real t_1, t_2 y t_3 . Los que además satisfacen: $0 \leq t_1 \leq t_2 \leq t_3$.

La existencia del movimiento browniano se ha demostrado de manera concluyente [10, Karatzas & Shreve (1988)] y [11, Rogers & Williams (1994)]. Además, está relacionado con el ruido blanco de tiempo continuo, proceso Gaussiano de una manera más específica.

Primero se define un proceso estocástico de tiempo continuo gaussiano como:

$$\int_0^t X(u)du \text{ es normalmente distribuida por todo } t \in (0, \infty)$$

Una vez definidos los conceptos anteriores se puede establecer la relación entre estos a partir del siguiente teorema.

Teorema 3.2.1.

Suponga que $\{\epsilon(t), t \in [0, \infty)\}$ es un proceso de ruido blanco continuo gaussiano. El proceso estocástico de tiempo continuo $\{W(t), t \in [0, \infty)\}$ esta definido por:

$$W(t) = \int_0^t \epsilon(u)du, t \in [0, \infty)$$

donde $\{W(t)\}$ es un movimiento Browniano[9, Roux (2002)].

Esta relación es de suma importancia al momento de definir modelos continuos. En este caso será aplicado al modelo continuo autoregresivo.

3.3. Modelo CAR

El modelo continuo autoregresivo *CAR*, es similar al proceso discreto *AR*, con el objetivo de resolver una ecuación diferencial estocástica, conducida por ruido blanco, esta está definida como:

$$Y^{(p)}(t) + \alpha_1 Y^{(p-1)}(t) + \dots + \alpha_{(p-1)} Y^{(1)}(t) + \alpha_p Y(t) = \epsilon_{(t)}, p \in \mathbb{Z}^+ \tag{3.1}$$

Donde $Y^{(i)}(t)$ es la *i*-ésima derivada de $Y(t)$ y $\epsilon(t)$ es la derivada formal de un proceso browniano $W(t)$.

Es posible reescribir la ecuación (3.1) como $\alpha(D)Y(t) = \epsilon(t)$, donde

$$\alpha(D) = D^p + \alpha_1 D^{p-1} + \dots + \alpha_{p-1} D + \alpha_p \tag{3.2}$$

D es el operador de derivada. La ecuación característica correspondiente viene dada por

$$\alpha(s) = s^p + \alpha_1 s^{p-1} + \dots + \alpha_{p-1} s + \alpha_s = 0 \quad (3.3)$$

Es importante detallar la ecuación característica ya que esta tiene relación con el punto de reparametrización más adelante.[12, Jones(1981)]

3.3.1. CAR(1)

A medida que el orden del modelo CAR va aumentando, también lo hace la dificultad de este, ya que repercute tanto en los cálculos como propiedades ya que se van haciendo más complicados. De manera ilustrativa se representará a partir del modelo continuo autorregresivo de orden 1 $CAR(1)$. Este está definido como:

$$\frac{d}{dt}Y(t) + \alpha Y(t) = \sigma \nu(t) + \beta \quad (3.4)$$

Donde $\nu(t)$ es el ruido blanco continuo, para que este modelo sea gaussiano se asume $\beta = 0$ por esto queda el modelo se la siguiente forma:

$$\frac{d}{dt}Y(t) + \alpha Y(t) = \sigma \epsilon(t) \quad (3.5)$$

Donde $\epsilon(t)$ es el ruido blanco continuo gaussiano. Así, la solución para (3.5) está dada por:

$$Y(t) = e^{-\alpha t} Y(0) + \sigma \int_0^t e^{-\alpha(t-u)} \epsilon(t)(u) \quad (3.6)$$

o en su forma Itô

$$Y(t) = e^{-\alpha t} Y(0) + e^{-\alpha t} I(t) \quad (3.7)$$

Donde $I(t) = \sigma^2 \int_0^t e^{\alpha u} \epsilon(t) du$ conocida como la integral de Itô [16, Brockwell & Davis (2016)]. Para que un proceso $CAR(1)$ sea estacionario se debe comprobar que el parámetro $\alpha > 0$. Ya cumpliéndose esto, las propiedades para un modelo $CAR(1)$ son.

1. $\mathbb{E}(Y(t)) = 0$
2. $\text{Var}(Y(t)) = \frac{\sigma^2}{2\alpha}$
3. $\gamma(k) = \frac{\sigma^2}{2\alpha} e^{-\alpha|k|}$
4. $\rho(k) = e^{-\alpha|k|}$

3.3.2. Reparametrización CAR

Con el propósito de asegurar la estabilidad del modelo, se construyó tanto una parametrización en los ceros de (3.3) r_1, \dots, r_p de $\alpha(s)$ como en la estructura del modelo (3.2) de [12, Jones (1981)], es decir.

- Parametrización de los ceros

$$\alpha(s) = \prod_{i=1}^p (s - r_i) \quad (3.8)$$

- Estructura del modelo

$$\alpha(D)Y(t) = \left(1 + \frac{D}{k}\right)^{p-1} \epsilon(t) \quad (3.9)$$

Con parámetro de escala $k > 0$. Esto introduce un operador de media móvil prescrito de orden $p - 1$ en el modelo, lo que hace que la selección del modelo sea conveniente junto con otros beneficios teóricos descritos en [13, Belcher et al.(1994)]. Es necesario explicar la parametrización ya que el paquete de R a utilizar la ejecuta.

3.3.3. Estimación de parámetros por filtro de Kalman

Es necesaria la aplicación del filtro de Kalman para estimar los parámetros del modelo (3.9), siguiendo a [12, Jones (1981)] & [13, Belcher et al. (1994)], para aplicar el filtro de Kalman, se requiere reescribir el modelo (3.9) en una forma de estado de espacio, que se puede encontrar en [17, Wiberg (1971)]. Sea el vector de estado $X(t) = (X(t), X'(t), X''(t), \dots, X^{(p-1)}(t))^T$ y $X'(t)$ la primera derivada de $X(t)$. La ecuación de estado viene dada por:

$$X'(t) = AX(t) + R\epsilon \quad (3.10)$$

Donde

$$A = \begin{pmatrix} 0 & 1 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ -\alpha_p & -\alpha_{p-1} & \cdots & -\alpha_1 \end{pmatrix} \quad (3.11)$$

y

$$R^T = \begin{pmatrix} 0 & 0 & \cdots & 1 \end{pmatrix} \quad (3.12)$$

En base a lo anteriormente planteado se puede definir la ecuación de observación como:

$$Y_k = HX(t_k) \quad (3.13)$$

Donde los elementos del vector $H_{1 \times p}$ vienen dados por:

$$H_i = \binom{p-1}{i-1} / k^{i-1} \quad i = 1, \dots, p \quad (3.14)$$

Suponga que la matriz A puede ser diagonalizado por $A = UDU^{-1}$, donde

$$U = \begin{pmatrix} 1 & 1 & \dots & 1 \\ r_1 & r_2 & \dots & r_p \\ r_1^2 & r_2^2 & \dots & r_p^2 \\ \vdots & \vdots & \ddots & \vdots \\ r_1^{p-1} & r_2^{p-1} & \dots & r_p^{p-1} \end{pmatrix} \quad (3.15)$$

r_1, r_2, \dots, r_p son las raíces de $\alpha(s)$, y D es una matriz diagonal, con estas raíces como su diagonal de elementos. En este caso, sea $X = U\psi$, y la ecuación de estado se convierte en

$$\psi' = D\psi + J\epsilon$$

Donde $J = U^{-1}R$, por consiguiente la ecuación de observación (3.13) se convierte en:

$$Y_k = G\psi(t_k) \quad (3.16)$$

donde $G = HU$. La condición necesaria y suficiente para la diagonalización de A es que A debe tener valores propios distintos.

Para iniciar las recursiones del filtro de Kalman hay que verificar si, la matriz de covarianza no condicional del vector de estado $X(t)$ es conocida [21, Doob, J. L. (1953)] y utilizando en [12, Jones, R. H. (1981)] y [22, Harvey, A. C. (1990)]. El estado inicial tanto para la versión estándar como para la versión diagonalizada se puede establecer como $X(0) = 0$ y $\psi_0 = 0$, respectivamente. La matriz de covarianza estacionaria Q satisface

$$Q = \sigma^2 \int_0^\infty e^{As} RR^T e^{A^T s} ds \quad (3.17)$$

Si A se puede diagonalizar, se puede demostrar que

$$Q_{\psi_i, j} = -\sigma^2 \frac{J_i \bar{J}_i}{r_i + \bar{r}_i} \quad (3.18)$$

Donde $J = U^{-1}$, \bar{J}_j y \bar{r}_j son las conjugados complejos de J_j y r_j , respectivamente.

A partir del parámetro de escala k definido como el recíproco del tiempo medio entre observaciones, el vector de estado estacionario inicial $\psi_0 = \psi(0|0) = 0$ y la matriz de covarianza estacionaria de estado inicial Q_ψ , los pasos para la recursión son los siguientes.

1. Predecir el estado. Sea

$$T_{\psi,k} = e^{D\delta_k}$$

una matriz diagonal, entonces

$$\psi(t_k|t_{k-1}) = T_{\psi,k}\psi(t_{k-1}|t_{k-1}).$$

2. Calcular la matriz de covarianzas de esta predicción:

$$P_{\psi}(t_k|t_{k-1}) = T_{\psi,k}[P_{\psi}(t_{k-1}|t_{k-1}) - Q_{\psi}]\bar{T}_{\psi,k} + Q_{\psi}.$$

3. Predecir la observación en el tiempo t_k :

$$Y_{\psi}(t_k|t_{k-1}) = G_{\psi}(t_k|t_{k-1})$$

4. Calcular la innovación:

$$v_{\psi}(t_k) = Y_{\psi}(t_k) - Y_{\psi}(t_k|t_{k-1})$$

y varianza

$$F_{\psi}(t_k) = GP_{\psi}(t_k|t_{k-1})\bar{G}^T + V$$

5. Actualizar la estimación del vector de estado:

$$\psi(t_k|t_k) = \psi(t_k|t_{k-1}) + P_{\psi}(t_k|t_{k-1})\bar{G}^T F_{\psi}^{-1}(t_k)v_{\psi}(t_k)$$

6. Actualizar la matriz de covarianzas:

$$P_{\psi}(t_k|t_k) = P_{\psi}(t_k|t_{k-1}) - P_{\psi}(t_k|t_{k-1})\bar{G}^T F_{\psi}^{-1}(t_k)G\bar{P}_{\psi}^T(t_k|t_{k-1})$$

7. El factor de escala desconocido σ^2 puede ser compensado dejando $\sigma^2 = 1$ temporalmente. $-2 \cdot \log$ de la verosimilitud es calculada por:

$$\log L_{\psi,k} = \sum_{t=1}^n \log F_{\psi}(t_k) + n \log \sum_{t=1}^n \frac{v_{\psi}^2(t_k)}{F_{\psi}(t_k)}$$

El factor de escala desconocido puede entonces estimarse mediante

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{t=1}^n \frac{v_{\psi}^2(t_k)}{F_{\psi}(t_k)} \quad (3.19)$$

Cuando una forma diagonal no es estable, se puede encontrar una recursión de filtro de Kalman de forma estándar

en [13, Belcher et al. (1994)] o [23, Wang (2004)]. Sin embargo, el costo computacional se reduce dramáticamente con la forma diagonal gracias a la matriz D .

Cuando la optimización no lineal se completa con éxito, además de la estimación por máxima verosimilitud de los parámetros y las variaciones de error, el filtro de Kalman devuelve la estimación óptima del estado y la matriz de covarianza en el último punto de tiempo.

3.3.4. Selección del modelo

Hay más de un criterio, pero la mejor práctica es calcular el AIC o BIC clásico [14, Akaike (1974)] & [15, Schwarz (1978)]. Ajustando los modelos de orden creciente p a la serie. El AIC se define como $n \cdot \log SS(p) + 2p$ y BIC se define como $n \cdot \log SS(p) + p \cdot \log(p)$ donde SS es la función de suma de cuadrados dada por [13, Belcher et al. (1994)].

3.4. Modelo IAR

El modelo autorregresivo irregular *IAR*, es una extensión tanto del modelo *AR*(1) como del modelo *CAR*(1), con el objetivo de adaptarse a estructuras con dependencia temporal de los errores y prevenir la mala estimación de parámetros lo que conlleva a poca explicación de variabilidad, esta está definida como:

$$Y_{t_j} = \phi^{t_j - t_{j-1}} Y_{t_{j-1}} + \sigma \sqrt{1 - \phi^{2(t_j - t_{j-1})}} \epsilon_{t_j} \quad (3.20)$$

Donde ϵ_{t_j} son variables aleatorias independientes con media cero y varianza unitaria.

Para el proceso *IAR* se cumplen las siguientes propiedades.

1. $\mathbb{E}(Y_{t_j}) = 0$
2. $\mathbb{V}(Y_{t_j}) = \sigma^2, \forall Y_{t_j}$
3. $\text{Cov}(Y_{t_k}, Y_{t_j}) = \sigma^2 \phi^{t_k - t_j}, \forall k \geq j$

Por lo tanto para dos observaciones en tiempos cualquier donde $s < t$

4. $\gamma(t - s) = \sigma^2 \phi^{t-s}$
5. $\rho(t - s) = \frac{\gamma(t-s)}{\gamma(0)}$

3.4.1. Relación con *CAR*(1)

Como se mencionó con anterioridad, el *IAR* es una extensión del modelo *CAR*, en este caso se demostrará el como a partir de ciertos supuestos el *IAR* es a la vez un *CAR* de orden 1.

A partir de la ecuación (3.7), al reemplazar $e^{-\alpha}$ con ϕ y definiendo $\sigma^2 = \frac{\sigma^2}{2\alpha}$. Se obtiene

$$Y(t) = e^{-\alpha t}Y(0) + \frac{\sigma^2}{2\alpha} \sqrt{1 - e^{-2\alpha t}} \epsilon(t)$$

Que corresponde al modelo autorregresivo continuo de orden 1 para datos irregularmente espaciados. Por lo tanto, el modelo *IAR* es efectivamente una extensión del modelo autorregresivo continuo.

3.4.2. Relación con AR(1)

Al igual que en el punto anterior, el *IAR* es una extensión del modelo *AR*, en este caso se demostrará bajo que condiciones el *IAR* es a la vez un *AR* de orden 1.

Se puede notar que si $t_j - t_{j-1} = 1, \forall j$, entonces de la ecuación (3.20), se obtiene

$$Y_{t_j} = \phi Y_{t_{j-1}} + \sigma \sqrt{1 - \phi^2} \epsilon_{t_j}, \text{ para } j = 2, \dots, n$$

Que corresponde al modelo autorregresivo de orden 1 para datos regularmente espaciados. Por lo tanto, el modelo *IAR* es efectivamente una extensión del modelo autorregresivo regular y así también es ergódico y estacionario.

3.4.3. Estimación de Parámetros

La verosimilitud de los datos $\{Y_{t_1}, \dots, Y_{t_n}\}$ se puede expresar como

$$f(Y_{t_1}, \dots, Y_{t_n}; \theta) = f(Y_{t_1}; \theta) f(Y_{t_2}|Y_{t_1}; \theta) \cdot \dots \cdot f(Y_{t_n}|Y_{t_{n-1}}; \theta)$$

donde $\theta = (\sigma^2, \phi)$ es el vector de parámetros del modelo. Para describir claramente el proceso de estimación, asumimos aquí que tanto la distribución marginal como la condicional de las series son gaussianas.

Asumiendo que

$$f(Y_{t_1}; \sigma^2, \phi) \sim N(0, \sigma^2) \text{ y además}$$

$$f(Y_{t_j}|Y_{t_{j-1}}; \sigma^2, \phi) \sim N(\phi^{t_j - t_{j-1}} Y_{t_{j-1}}, \sigma^2(1 - \phi^{2(t_j - t_{j-1})}))$$

Para $j = 2, \dots, n$. Basado en la ecuación (3.20), menos el logaritmo de la verosimilitud de este proceso, puede escribirse como

$$l(\theta) = \frac{n}{2} \log(2\pi) + \frac{1}{2} \sum_{j=1}^n \log v_{t_j} + \frac{1}{2} \sum_{j=1}^n \frac{e_{t_j}^2}{v_{t_j}} \quad (3.21)$$

Donde $e_{t_1} = y_{t_1}$, $e_{t_j} = y_{t_j} - \phi^{t_j - t_{j-1}} y_{t_{j-1}}$ para $j > 1$ y sus varianzas como $v_{t_j} = \text{Var}(e_{t_j})$.

Así el predictor con pasado finito del proceso en tiempo t_j , está dado por

$$\hat{y}_{t_1} = 0 \text{ y } \hat{y}_{t_j} = \phi^{t_j - t_{j-1}} y_{t_{j-1}} \text{ para } j = 2, \dots, n \quad (3.22)$$

Por lo tanto, $e_{t_j} = y_{t_j} - \hat{y}_{t_j}$ es el error de predicción con varianza

$$v_{t_1} = \text{Var}(e_{t_1}) = \sigma^2 \text{ y } v_{t_j} = P\text{Var}(e_{t_j}) = \sigma^2[1 - \phi^{2(t_j - t_{j-1})}], \text{ para } j = 2, \dots, n. \quad (3.23)$$

Por máxima verosimilitud directa aplicada a la ecuación (3.21) se obtiene

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^n \frac{(y_{t_j} - \hat{y}_{t_j})^2}{\tau_{t_j}}, \text{ donde } \tau_{t_j} = \frac{v_{t_j}}{\sigma^2} \quad (3.24)$$

Pero no es posible encontrar el estimador de ϕ por maxima verosimilitud directa, por esto se procede a utilizar distintos métodos iterativos.

3.5. Modelo CIAR

El modelo autorregresivo irregular complejo *CIAR*, es una extensión compleja del modelo *IAR*. A diferencia de otros modelos complejos, donde se asume que tanto la parte real como imaginaria son observadas, en el *CIAR* se asume que solo la parte real es observada mientras que la imaginaria es un proceso latente. A diferencia de los modelos anteriormente planteados en el capítulo 3, este cuenta con que puede detectar valores negativos en la ACF, corrigiendo problemas de parsimonia al utilizar otras alternativas que también las detectan.

Suponiendo que y_{t_j} es una secuencia de valores complejos, tal que $y_{t_j} = x_{t_j} + iz_{t_j}, \forall j = 1, \dots, n$ y asimismo $\phi = \phi^R + i\phi^I$ es el coeficiente complejo del modelo y $\epsilon_{t_j} = \epsilon_{t_j}^R + i\epsilon_{t_j}^I$ es el ruido blanco complejo. Este está definido como:

$$x_{t_j} + iz_{t_j} = (\phi^R + i\phi^I)^{t_j - t_{j-1}} (x_{t_{j-1}} + iz_{t_{j-1}}) + \sigma_{t_j} (\epsilon_{t_j}^R + i\epsilon_{t_j}^I) \quad (3.25)$$

Donde $\sigma_{t_j} = \sigma \sqrt{1 - |\phi^{t_j - t_{j-1}}|^2}$ y $|\cdot|$ es el módulo de un número complejo. Se asume que solo la parte real x_{t_j} es observada mientras que la parte imaginaria z_{t_j} es un proceso latente. Además se asume $\epsilon_{t_j}^R$ y $\epsilon_{t_j}^I$ con media 0 y varianzas $\text{Var}(\epsilon_{t_j}^R) = 1$ y $\text{Var}(\epsilon_{t_j}^I) = c$. Donde c es un parámetro fijo que toma valores en \mathbb{R}^+ .

Se define $\gamma(0) = \mathbb{E}(\bar{y}_{t_j} y_{t_j})$, $\gamma(k) = \mathbb{E}(\bar{y}_{t_{j+k}} y_{t_j})$ y $\rho(k)$, como la varianza, autocovarianza y autocorrelación del proceso y_{t_j} . respectivamente. Para el *CIAR* se cumplen las siguientes propiedades:

1. $\mathbb{E}(y_{t_j}) = 0$
2. $\text{Var}(y_{t_j}) = \gamma(0) = \mathbb{E}(\bar{y}_{t_j} y_{t_j}) = \sigma^2(1 + c)$
3. $\gamma(k) = \mathbb{E}(\bar{y}_{t_{j+k}} y_{t_j}) = \overline{\phi^{\Delta_k}} \sigma^2(1 + c)$
4. $\rho(k) = \overline{\phi^{\Delta_k}}$

Donde $\Delta_k = t_{j+k} - t_j$ que denota la diferencia temporal entre las observaciones de tiempo t_{j+k} y t_j . Además, \bar{y}_{t_j} es el conjugado complejo de y_{t_j} .

Si $|\phi| = |\phi^R + i\phi^I| < 1$ ocurre, es un proceso débilmente estacionario. Como se puede ver la ACF $\rho(k)$ del proceso *CIAR* decae a una velocidad $\phi^{t_{j+k}-t_j}$, es decir, tiene decaimiento exponencial.

3.5.1. Estimación de Parámetros

Es necesaria la aplicación del filtro de Kalman para estimar los parámetros del modelo (3.25), para aplicar el filtro de Kalman, se requiere reescribir el modelo (3.25) en una forma de espacio de estado. A partir de las ecuaciones (2.3) y (2.4) se puede escribir el modelo *CIAR* de la siguiente forma

$$x_{t_j} + iz_{t_j} = (\alpha_{t_j}^R + i\alpha_{t_j}^I)(x_{t_{j-1}} + iz_{t_{j-1}}) + \sigma_{t_j}(\epsilon_{t_j}^R + i\epsilon_{t_j}^I) \quad (3.26)$$

Donde $\alpha_{t_j}^R = |\phi|^{\delta_j} \cos(\delta_j \psi)$, $\alpha_{t_j}^I = |\phi|^{\delta_j} \sin(\delta_j \psi)$, $\delta_j = t_j - t_{j-1}$, $\psi = \arccos\left(\frac{\phi^R}{|\phi|}\right)$ y $\phi = \phi^R + i\phi^I$, según [2, Elorrieta et al.(2019)].

Así se puede separar tanto en su parte observada

$$x_{t_j} = \alpha_{t_j}^R x_{t_{j-1}} - \alpha_{t_j}^I z_{t_{j-1}} + \sigma_{t_j} \epsilon_{t_j}^R \quad (3.27)$$

y su parte latente

$$z_{t_j} = \alpha_{t_j}^I x_{t_{j-1}} + \alpha_{t_j}^R z_{t_{j-1}} + \sigma_{t_j} \epsilon_{t_j}^I \quad (3.28)$$

La ecuación (3.26) puede expresarse como un sistema de estado de espacio a partir de (2,4) – (2,3), asumiendo $t = t_j$ y $y_{t_j} = \begin{pmatrix} x_{t_j} \\ z_{t_j} \end{pmatrix}$ Dado que solo x_{t_j} se observa realmente, obtenemos $X_t = x_{t_j}$. Por lo tanto, $G_t = \begin{pmatrix} 1 & 0 \end{pmatrix}$ es la matriz de observación y se define la matriz de transición como

$$F_{t_j} = \begin{pmatrix} \alpha_{t_j}^R & -\alpha_{t_j}^I \\ \alpha_{t_j}^I & \alpha_{t_j}^R \end{pmatrix}, V_{t_j} = \sigma_{t_j} \begin{pmatrix} \epsilon_{t_j}^R \\ \epsilon_{t_j}^I \end{pmatrix} \text{ y } W_{t_j} = 0$$

(3.29)

Así, la observación y ecuaciones de estado de la representación de estado de espacio del modelo *CIAR* son

$$\begin{pmatrix} x_{t_j} \\ z_{t_j} \end{pmatrix} = \begin{pmatrix} \alpha_{t_j}^R & -\alpha_{t_j}^I \\ \alpha_{t_j}^I & \alpha_{t_j}^R \end{pmatrix} \begin{pmatrix} x_{t_{j-1}} \\ z_{t_{j-1}} \end{pmatrix} + \sigma_{t_j} \begin{pmatrix} \epsilon_{t_j}^R \\ \epsilon_{t_j}^I \end{pmatrix}, y_{t_j} = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} x_{t_j} \\ z_{t_j} \end{pmatrix} \quad (3.30)$$

Para este modelo de estado de espacio los predictores a un paso son $\hat{X}_{t_j} = P_{t_{j-1}}(X_{t_j})$ y sus matrices de error de covarianza $\Omega_{t_j} = \mathbb{E}[(X_{t_j} - \hat{X}_{t_j})(X_{t_j} - \hat{X}_{t_j})']$. Estos son determinados por

$$\hat{X}_{t_j} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Omega_{t_1} = \sigma^2 \begin{pmatrix} 1 & 0 \\ 0 & c \end{pmatrix} \quad (3.31)$$

Las recursiones de Kalman, para $j = 1, \dots, 8(n-1)$ se definen como

1. $\Lambda_{t_j} = G_{t_j} \Omega_{t_j} G_{t_j}'$
2. $\Theta_{t_j} = F_{t_j} \Omega_{t_j} G_{t_j}'$
3. $\Omega_{t_{j+1}} = F_{t_j} \Omega_{t_j} F_{t_j}' + Q_{t_j} - \Theta_{t_j} \Lambda_{t_j}^{-1} \Theta_{t_j}'$
4. $v_{t_j} = Y_{t_j} - G_{t_j} \hat{X}_{t_j}$
5. $\hat{X}_{t_{j+1}} = F_{t_j} \hat{X}_{t_j} + \Theta_{t_j} \Lambda_{t_j}^{-1} v_{t_j}$

Los parámetros del modelo *CIAR*, ϕ^R y ϕ^I se pueden obtener minimizando la función reducida de verosimilitud

$$l(\theta) \propto \frac{1}{n} \sum_{j=1}^n \left(\log(\Lambda_{t_j}) + \frac{v_{t_j}^2}{\Lambda_{t_j}} \right) \quad (3.32)$$

donde Λ_{t_j} y v_{t_j} provienen de la recursión de Kalman.

Capítulo 4

Resultados

En este capítulo serán revisados los resultados obtenidos de la estimación para cada modelo, y además, se evaluará el ajuste de cada uno de éstos.

Para probar tanto la eficiencia como eficacia de la estimación según los distintos métodos, en primer lugar se trabajó con simulaciones del proceso, con la finalidad de tener control sobre la naturaleza de los datos, y así, poder verificar que esta estimación sea certera, comprobando que se acerca al valor real ya establecido.

Para mostrar de una forma más específica las ventajas y defectos que tienen los métodos al momento de estimar, se trabajará de la siguiente manera:

1. En primer lugar, se simularán datos irregularmente espaciados a partir de un modelo *IAR* y se ajustarán estos datos para los modelos de series de tiempo *AR*, *IAR*, *CAR* y *CIAR*.
2. Luego se comparará cada modelo, esto con el fin de ver las virtudes y defectos que tiene cada modelo al estimar con tiempos irregularmente espaciados. Las medidas que se ocuparán para evaluar el desempeño de las estimaciones en cada modelo se definirán más adelante.
3. De la misma manera, se repetirá la metodología simulando datos irregularmente espaciados a partir de un modelo *CIAR* y se ajustarán estos datos para los modelos *AR*, *IAR*, *CAR* y *CIAR*.
4. Posteriormente, de igual forma se comparará cada modelo
5. - Finalmente, se ocupará la misma metodología de los puntos anteriores pero ahora con datos reales.

Como el objetivo de este análisis es la comparación entre modelos, se necesita tener una medida cuantitativa que ayude a determinar cual de los modelos contrastados produce estimaciones más certeras. Para esto, se utili-

zaron las siguientes medidas, definidas a continuación.

- Criterio de información de Akaike (*AIC*)

El criterio de información de Akaike esta basado en la teoría de la información que contiene la verosimilitud del modelo. Mide la calidad del ajuste de una serie y se calcula con la siguiente expresión:

$$AIC = 2k - 2\log(L)$$

Donde k es el número de parámetros en el modelo estadístico, y L es el máximo valor de la función de verosimilitud para el modelo estimado. Dado un conjunto de modelos candidatos para los datos, el modelo preferido es el que tiene el valor mínimo en el AIC

- Error Cuadrático Medio (*MSE*)

El error cuadrático medio es una forma de evaluar la diferencia entre un estimador y el valor real de la cantidad que se quiere calcular.

El MSE mide el promedio del cuadrado del error, siendo el error el valor en la que el estimador difiere de la cantidad a ser estimada.

Sea \hat{Y}_t un vector de n predicciones e Y_t el vector de los verdaderos valores para un tiempo $t = (t_1, \dots, t_n)$.

El error cuadrático medio viene dado por:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Un MSE de 0 significa que el estimador predice las observaciones con una precisión perfecta, por lo tanto, mientras más cercano sea este valor a cero, mejores serán las predicciones.

4.1. Simulación

Como se menciona con anterioridad, en primera instancia se simularán datos con el fin de ajustar los distintos tipos de modelos y así comparar su ajuste.

Una vez simulados, se procederá a ajustar los distintos modelos y además se utilizarán las medidas de desempeño $-\text{Loglik}$, *AIC* y *MSE* para ver la precisión de las estimaciones.

4.1.1. Datos simulados a partir de un model *IAR*

Para simular los datos con los que se trabajará en todos los modelos de esta sección, se utilizarán las funciones *gentime* e *IAR.sample*, ambas pertenecientes a la librería *AITS*.

A continuación el detalle de ambas funciones.

Función *gentime*

Descripción

Función para generar tiempos irregularmente espaciados a partir de una mezcla de distribuciones exponenciales.

Uso

```
gentime(n, lambda1 = 130, lambda2 = 6.5, p1 = 0.15, p2 = 0.85)
```

Argumentos

n	Un número entero positivo. Duración de los tiempos de observación.
lambda1	Media (1 / tasa) de la primera distribución exponencial.
lambda2	Media (1 / tasa) de la segunda distribución exponencial.
p1	Peso de la primera distribución exponencial.
p2	Peso de la segunda distribución exponencial.

Valor

Arreglo con tiempos de observación irregularmente espaciados.

Función *IAR.sample*

Descripción

Simula un modelo de serie temporal *IAR(1)*.

Uso

```
IAR.sample(phi, n = 300, sT)
```

Argumentos

phi	Un coeficiente del modelo <i>IAR(1)</i> . Un valor entre 0 y 1.
n	Duración de la serie temporal de salida. Un número entero estrictamente positivo.
sT	Matriz con tiempos de observación.

Valor

Arreglo con tiempos de observación irregularmente espaciados.

Se procede a realizar 100 simulaciones de datos en base a un modelo $IAR(1)$. Primero se generan tiempos irregularmente espaciados a partir de una mezcla de distribuciones exponenciales, definiendo su tamaño muestral, en este caso de 300 observaciones. Posterior a esto se asigna un valor arbitrario al coeficiente ϕ , en este caso de 0,9.

Es posible ver a través de las figuras como se comportan los datos, a continuación un ejemplo de 4 simulaciones.

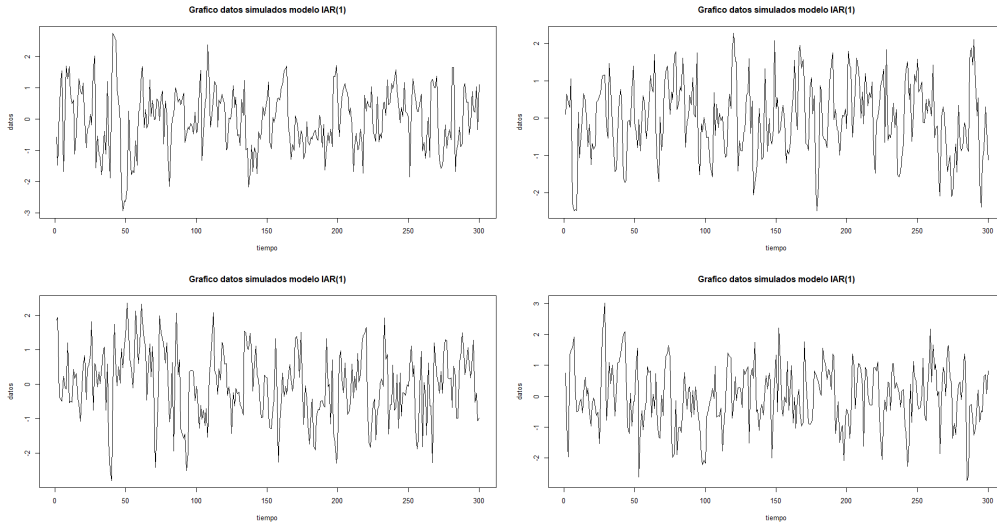


Figura 4.1: Gráfico de datos simulados en base a un modelo $IAR(1)$ con datos irregularmente espaciados

Simulación	Variable	Min	Q1	Mediana	Media	Q3	Max	Sd
1	y1	-2.9351	-0.7039	0.0338	0	0.7044	2.7456	1
1	diff(st)	0	2.2	5.4	21.5	12.5	337.6	49.09
2	y1	-2.4800	-0.6823	0.0392	0	0.6997	2.2702	1
2	diff(st)	0	2.3	5.9	35.1	14.9	621.3	90.83
3	y1	-2.791	-0.632	0.023	0	0.633	2.350	1
3	diff(st)	0	1.8	4.7	19.3	11.0	407.9	54.63
4	y1	-2.7349	-0.6343	-0.0002	0	0.7611	3.0179	1
4	diff(st)	0	2.4	6.8	29.8	14.5	470.3	72.58
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Media:	y1	-2.763	-0.68	-0.008275	0	0.6819	2.746	1
Media:	diff(st)	0.02192	2.254	5.651	24.79	13.16	546.7	65.26

Cuadro 4.1: Descripción de las variables

Observando tanto la figura (4.1) como el cuadro (4.1), a priori se puede apreciar tanto una media como una

dispersión de datos constante, esto quiere decir que la serie tiende a un comportamiento estacionario.

4.1.1.1. Ajuste por modelo *AR*

Se simularon valores desigualmente espaciados que se rigen por un modelo *IAR(1)*, posteriormente fueron ajustados a un modelo *AR(1)*, se utilizará la función *arma*, perteneciente a la librería *tseries*.

A continuación el detalle de la función.

Función *arma*

Descripción

Ajusta un modelo *ARMA* a una serie de tiempo univariante por mínimos cuadrados condicionales.

Uso

```
arma(x, order = c(1, 1), lag = NULL, coef = NULL, include.intercept = TRUE, series = NULL, qr.tol = 1e-07, ...)
```

Argumentos

<code>x</code>	Un vector numérico o una serie de tiempo.
<code>order</code>	Un vector entero bidimensional que da las órdenes del modelo a ajustar. El orden[1] corresponde a la parte AR y el orden[2] a la parte MA.
<code>lag</code>	Una lista con los componentes ar y ma. Cada componente es un vector entero, especificando los intervalos AR y MA que se incluyen en el modelo. Si se dan ambos, orden y retraso, sólo se utiliza la especificación de retraso.
<code>coef</code>	Si se da este vector numérico se utiliza como estimación inicial de los coeficientes.
<code>include.intercept</code>	¿Debería el modelo contener una intercepción
<code>serie</code>	Nombre para la serie.
<code>qr.tol</code>	El argumento tol para qr cuando se calculan los errores estándar asintóticos de coef.

Valor

<code>fitted.values</code>	La serie ajustada.
----------------------------	--------------------

Con el objetivo de evaluar el ajuste del modelo propuesto a los datos simulados, se generó un ciclo, donde se muestra de forma gráfica el ajuste del modelo para cuatro simulaciones.

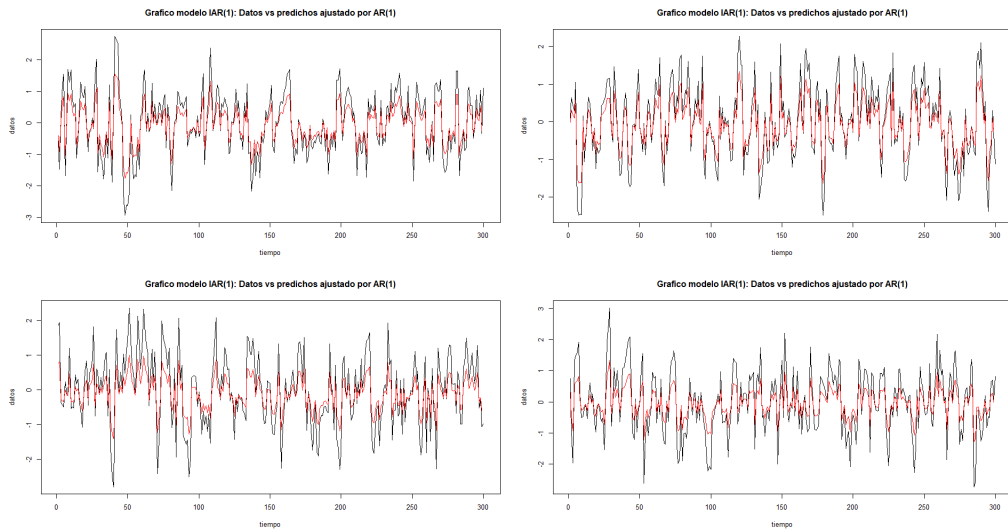


Figura 4.2: Gráfico de datos simulados a partir de un modelo $IAR(1)$ versus datos predichos por un modelo $AR(1)$ con datos irregularmente espaciados

De la figura (4.2) se aprecia en color negro los datos simulados por el modelo $IAR(1)$ y en rojo los datos estimados por el modelo $AR(1)$. A priori se puede observar que el modelo propuesto tiene un ajuste regular, exceptuando los tiempos en que los datos aumentan o disminuyen considerablemente su valor.

En el siguiente cuadro se evidencia parte de los resultados obtenidos sobre el parámetro ϕ y el tiempo de computo.

Modelo $AR(1)$			
Parámetro Real	Parámetro Estimado	Diferencia	Tiempo de Computo
ϕ	$\hat{\phi}$	$ \phi - \hat{\phi} $	Segundos
0.9	0.5415	0.3585	0
0.9	0.5368	0.3632	0
0.9	0.5049	0.3951	0
0.9	0.4579	0.4421	0
0.9	0.3953	0.5047	0.01
0.9	0.5658	0.3342	0
\vdots	\vdots	\vdots	\vdots
Medias:	$\bar{\hat{\phi}}=0.5111$	$ \bar{\phi} - \bar{\hat{\phi}} =0.3889$	0.0033
Desv Estándar:	sd=0.0528		0.006675

Cuadro 4.2: Parámetro real vs Estimado; Tiempo de Computo

Como se puede observar en el cuadro (4.2), el modelo $AR(1)$ tuvo un tiempo de cómputo muy pequeño lo que

sería muy útil al momento de trabajar con un gran volumen de datos, sin embargo, la estimación del parámetro esta lejos de la real. Tal como se pudo apreciar en la figura (4.2), el modelo propuesto no se ajusta de buena manera a la serie simulada.

Con el objetivo de terminar de confirmar lo anteriormente planteado, se analiza la bondad del ajuste a partir de los indicadores $-Loglik$ que es el logaritmo de la verosimilitud negativa de los datos, AIC que es el criterio de información de Akaike, el MSE que es el error cuadrado medio del modelo y el $MSE2$ que es el error cuadrado medio de la predicción (10% de los valores reales menos los predichos por el 90%).

	Modelo $AR(1)$			
	-Loglik	AIC	MSE	MSE2
	394.8	793.7	0.711	1.142
	419	842	0.7243	0.8731
	357.5	718.9	0.7494	0.403
	385.7	775.4	0.7899	0.8613
	369.4	742.8	0.8486	0.5487
	370.5	745	0.6769	1.244
	⋮	⋮	⋮	⋮
Medias:	379.5	762.9	0.7431	0.9599
Desv Estándar:	16.24	32.48	0.05509	0.6758

Cuadro 4.3: Bondad de Ajuste del Modelo $AR(1)$

Confirmando lo planteado en los puntos anteriores, a partir del cuadro (4.3), se puede notar que el modelo $AR(1)$ tuvo tanto una media de $-Loglik$ como AIC de las simulaciones muy alta. Por esto no sería recomendable el uso de un modelo $AR(1)$ para tiempo desigualmente espaciados.

4.1.1.2. Ajuste por modelo CAR

Se simularon valores desigualmente espaciados que se rigen por un modelo $IAR(1)$, posteriormente fueron ajustados a un modelo $CAR(1)$, se utilizará la función *carma*, perteneciente a la librería *growth*.

A continuación el detalle de la función.

Función *carma*

Descripción

Está diseñado para manejar un polinomio dentro de una matriz de diseño de sujetos con observaciones desigual-

mente espaciadas que pueden ser en diferentes momentos para diferentes sujetos. Se toma el origen del tiempo como el tiempo medio de todos los sujetos. Se supone que los errores internos del sujeto son gaussianos independientes o tienen una estructura gaussiana $ARMA(p, q)$ de tiempo continuo con la opción de incluir el error de medición. Se supone que los coeficientes aleatorios entre sujetos tienen una matriz de covarianza arbitraria. La matriz de diseño de efectos fijos es un polinomio de igual o mayor orden que la matriz de diseño interno. Esta matriz se puede aumentar mediante covariables multiplicadas por matrices de diseño polinomial de cualquier orden hasta el orden de la primera partición de la matriz de diseño. El método se basa en la probabilidad máxima exacta utilizando el filtro de Kalman para calcular la probabilidad.

Uso

```
carma(response = NULL, ccov = NULL, times = NULL, torder = 0, interaction, arma = c(0, 0, 0), parma = NULL, pre = NULL, position = NULL, iopt = TRUE, resid = TRUE, transform = "identity", delta = NULL, envir = parent.frame(), print.level = 0, typsize = abs(p), ndigit = 10, gradtol = 1e-05, steptol = 1e-05, iterlim = 100, fscale = 1, stepmax = 10 * sqrt(p %% p))
```

Argumentos

response	Una lista de matrices de dos columnas con valores de respuesta y tiempos para cada individuo, una matriz o marco de datos de valores de respuesta, o un objeto de cualquier clase.
ccov	Una matriz con columnas de covariables, con una fila por individuo, una fórmula modelo que usa vectores del mismo tamaño o un objeto de clase
times	Cuando la respuesta es una matriz, un vector de tiempos posiblemente desigualmente espaciados cuando son iguales para todos los individuos o una matriz de tiempos. No es necesario si está igualmente espaciado. Ignorado si la respuesta tiene clase.
torder	Orden del polinomio en el tiempo para ser ajustado.
arma	Vector de tres valores: orden de AR, orden de MA, indicador binario de presencia de error de medición. No se requiere para un AR (1) si se proporciona una estimación inicial. Si solo se proporciona un valor, se supone que es el orden del AR.
parma	Estimaciones iniciales de los parámetros ARMA. Por ejemplo, con arma = c(1,0,0), un AR (1), el parámetro es parma [1] = log(theta), donde theta es el coeficiente autorregresivo de tiempo continuo positivo. El coeficiente de autorregresión de paso finito para un paso de longitud delta es entonces alfa = exp(-delta * theta).

Valor

Se devuelve una lista de la clase carma que contiene toda la información relevante calculada, incluidos los códigos de error.

coef	Coefficientes.
deviance	Devianza.

residuals

Residuos.

mprofile

Perfiles marginales especiales con SE.

Con el objetivo de evaluar el ajuste del modelo propuesto a los datos simulados, se generó un ciclo, donde se muestra de forma gráfica el ajuste del modelo para cuatro simulaciones.

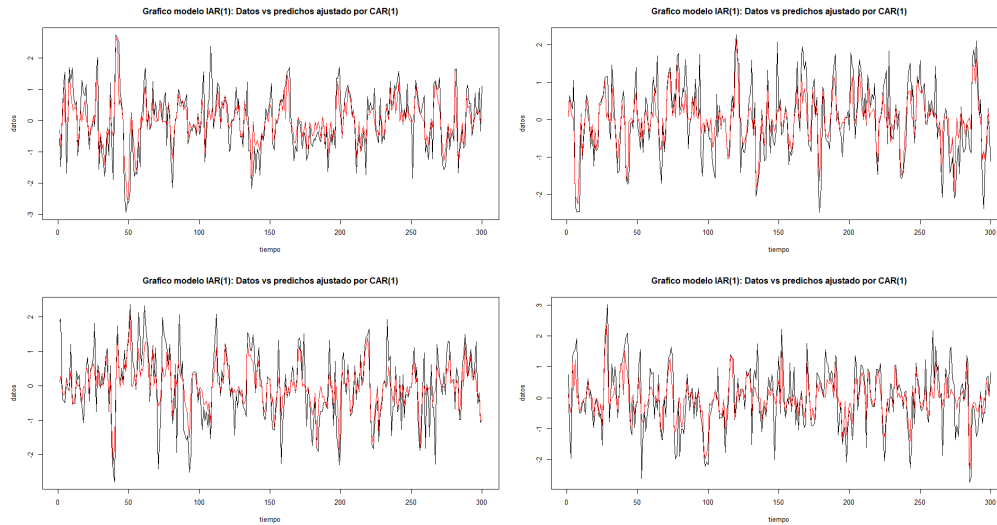


Figura 4.3: Gráfico de datos simulados a partir de un modelo $IAR(1)$ versus datos predichos por un modelo $CAR(1)$ con datos irregularmente espaciados

De la figura (4.3) se aprecia en color negro los datos simulados por el modelo $IAR(1)$ y en rojo los datos estimados por el modelo $CAR(1)$. A priori se puede observar que el modelo propuesto tiene un buen ajuste, exceptuando los tiempos en que los datos aumentan o disminuyen considerablemente su valor.

En el siguiente cuadro se evidencia parte de los resultados obtenidos sobre el parámetro ϕ y el tiempo de computo.

Modelo $CAR(1)$			
Parámetro Real	Parámetro Estimado	Diferencia	Tiempo de Computo
ϕ	$\hat{\phi}$	$ \phi - \hat{\phi} $	Segundos
0.9	0.9153	0.0153	0
0.9	0.909	0.009	0
0.9	0.8885	0.0115	0
0.9	0.8905	0.0095	0.01
0.9	0.8951	0.0049	0
0.9	0.9071	0.0071	0
\vdots	\vdots	\vdots	\vdots
Medias:	$\bar{\hat{\phi}}=0.9139$	$ \bar{\phi} - \bar{\hat{\phi}} = 0.0139$	0.0034
Desv Estándar:	0.03995		0.007138

Cuadro 4.4: Parámetro real vs Estimado; Tiempo de Computo

Como se puede observar en el cuadro (4.4), el modelo $CAR(1)$ tuvo un tiempo de cómputo muy pequeño lo que sería muy útil al momento de trabajar con un gran volumen de datos, además la estimación del parámetro es bastante cercana a la real. Tal como se pudo apreciar en la figura (4.3), el modelo propuesto se ajusta de buena manera a la serie simulada.

Con el objetivo de corroborar lo anteriormente planteado, se analiza nuevamente la bondad del ajuste a partir de los indicadores $-Loglik$, AIC , MSE y $MSE2$.

Modelo $CAR(1)$				
	-Loglik	AIC	MSE	MSE2
	148.5	301.1	0.3017	0.9681
	149.3	302.5	0.3002	2.154
	162.2	328.4	0.3449	0.6401
	160	323.9	0.3603	1.433
	169.3	342.6	0.383	2.186
	156.5	317.1	0.3064	1.741
	\vdots	\vdots	\vdots	\vdots
Medias:	165.1	334.1	0.2836	1.227
Desv Estándar:	14.05	28.1	0.1301	0.5011

Cuadro 4.5: Bondad de Ajuste del Modelo $CAR(1)$

Confirmando lo planteado en los puntos anteriores, a partir del cuadro (4.5), se puede notar que el modelo $CAR(1)$ tuvo tanto una media de $-Loglik$ como AIC de las simulaciones baja. Por esto sería recomendable el uso de un modelo $CAR(1)$ para tiempo desigualmente espaciados.

4.1.1.3. Ajuste por modelo IAR Kalman

Se simularon valores desigualmente espaciados que se rigen por un modelo $IAR(1)$, posteriormente fueron ajustados a un modelo IAR con estimación de parámetros por filtro de Kalman, se utilizarán las funciones $IAR.phi.kalman$ e $IAR.kalman$, pertenecientes a la librería $AITS$.

A continuación el detalle de las funciones.

Función $IAR.phi.kalman$

Descripción

Esta función devuelve el logaritmo de la verosimilitud negativa del proceso IAR dado un valor específico de ϕ

Uso

$IAR.phi.kalman(x, y, yerr, t, zero.mean = "TRUE", standardized = "TRUE")$

Argumentos

y	Matriz con las observaciones de series de tiempo.
$yerr$	Matriz con las desviaciones estándar del error de medición.
t	Matriz con los tiempos de observación irregulares.
$zero.mean$	Lógico; si es verdadero, la matriz y tiene media cero; si es falso, y tiene una media diferente de cero.
$standardized$	Lógico; si es verdadero, la matriz y está estandarizada; si es falso, y contiene la serie de tiempo sin procesar

Valor

Valor de el logaritmo de la verosimilitud negativa evaluada en ϕ .

Función $IAR.kalman$

Descripción

Estimación de máxima verosimilitud del parámetro ϕ del modelo IAR . El procedimiento de estimación utiliza el filtro de Kalman para encontrar la máxima verosimilitud.

Uso

IAR.kalman(y, sT, delta = 0, zero.mean = "TRUE", standardized = "TRUE")

Argumentos

y	Matriz con las observaciones de series de tiempo.
sT	Matriz con los tiempos de observación irregulares.
delta	Matriz con las desviaciones estándar del error de medición.
zero.mean	Lógico; si es verdadero, la matriz y tiene media cero; si es falso, y tiene una media diferente de cero.
standardized	Lógico; si es verdadero, la matriz y está estandarizada; si es falso, y contiene la serie de tiempo sin procesar

Valor

Una lista con los siguientes componentes:

phi	EMV del coeficiente phi del modelo IAR.
ll	Valor de la probabilidad logarítmica negativa evaluada en phi.

Con el objetivo de evaluar el ajuste del modelo propuesto a los datos simulados, se generó un ciclo, donde se muestra de forma gráfica el ajuste del modelo para cuatro simulaciones.

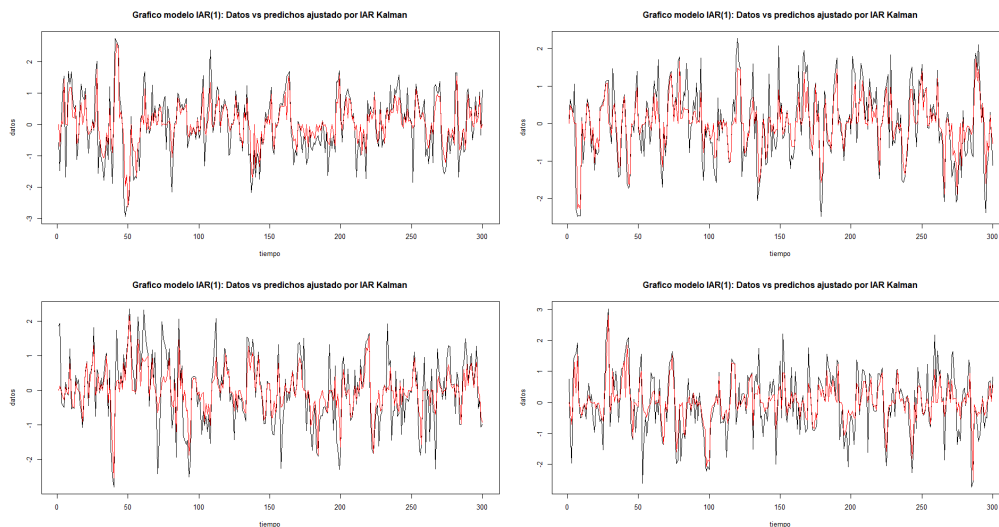


Figura 4.4: Gráfico de datos simulados a partir de un modelo $IAR(1)$ versus datos predichos por un modelo $IAR(1)$ con estimación por filtro de Kalman, con datos irregularmente espaciados

De la figura (4.4) se aprecia en color negro los datos simulados por el modelo $IAR(1)$ y en rojo los datos estimados por el modelo $IAR(1)$ con estimación por filtro de Kalman. A priori se puede observar que el modelo propuesto tiene un buen ajuste, esto era de esperarse considerando que los datos provienen del mismo modelo.

En el siguiente cuadro se evidencia parte de los resultados obtenidos sobre el parámetro ϕ y el tiempo de computo.

Modelo $IAR(1)$ filtro de Kalman			
Parámetro Real	Parámetro Estimado	Diferencia	Tiempo de Computo
ϕ	$\hat{\phi}$	$ \phi - \hat{\phi} $	Segundos
0.9	0.9184	0.0184	0.21
0.9	0.9164	0.0164	0.21
0.9	0.8844	0.0156	0.17
0.9	0.8988	0.0012	0.17
0.9	0.8908	0.0092	0.17
0.9	0.9097	0.0097	0.16
\vdots	\vdots	\vdots	\vdots
Medias:	$\bar{\hat{\phi}}=0.8994$	$ \bar{\phi} - \bar{\hat{\phi}} =0.0006$	0.1686
Desv Estándar:	0.01511		0.01583

Cuadro 4.6: Parámetro real vs Estimado; Tiempo de Computo

Como se puede observar en el cuadro (4.6), el modelo $IAR(1)$ con estimación por filtro de Kalman, tuvo un tiempo de cómputo muy pequeño y además la diferencia entre la estimación del parámetro versus el parámetro real es casi un cero numérico. Tal como se pudo apreciar en la figura (4.4), el modelo propuesto se ajusta de buena manera a la serie simulada.

Con el objetivo de corroborar lo anteriormente planteado, se analiza nuevamente la bondad del ajuste a partir de los indicadores $-Loglik$, AIC , MSE y $MSE2$

Modelo $IAR(1)$ filtro de Kalman			
-Loglik	AIC	MSE	MSE2
0.1458	4.292	0.2722	1.116
0.1592	4.318	0.2839	1.049
0.3386	4.677	0.3235	0.4792
0.3025	4.605	0.3385	1.222
0.4178	4.836	0.3914	1.219
0.256	4.512	0.3198	1.417
⋮	⋮	⋮	⋮
Medias:	0.2953	4.591	0.3331
Desv Estándar:	0.1026	0.2052	0.05049

Cuadro 4.7: Bondad de Ajuste del Modelo $IAR(1)$ filtro de Kalman

Confirmando lo planteado en los puntos anteriores, a partir del cuadro (4.7), se puede notar que el modelo $IAR(1)$ con estimación por filtro de Kalman tuvo tanto una media de $-Loglik$, AIC y MSE de las simulaciones bastante baja. Esto era de esperarse considerando que los datos provienen del mismo modelo, por todo esto, es recomendable el uso del modelo $IAR(1)$ por filtro de Kalman en series desigualmente espaciadas con ϕ positivo.

4.1.1.4. Ajuste por modelo IAR $loglik$

Se simularon valores desigualmente espaciados que se rigen por un modelo $IAR(1)$, posteriormente fueron ajustados a un modelo IAR con estimación de parámetros por máxima verosimilitud directa, se utilizarán las funciones $IAR.phi.loglik$ e $IAR.loglik$, pertenecientes a la librería $AITS$.

A continuación el detalle de las funciones.

Función $IAR.phi.loglik$

Descripción

Esta función devuelve el logaritmo de la verosimilitud negativa del modelo IAR para un valor específico de ϕ

Uso

$IAR.phi.loglik(x, y, sT, delta = 0, include.mean = "FALSE", standardized = "TRUE")$

Argumentos

- y Matriz con las observaciones de series de tiempo.
- sT Matriz con los tiempos de observación irregulares.
- $delta$ Matriz con las desviaciones estándar del error de medición.

include.mean	Lógico; si es verdadero, la matriz y tiene media cero; si es falso, y tiene una media diferente de cero.
standarized	Lógico; si es verdadero, la matriz y está estandarizada; si es falso, y contiene la serie de tiempo sin procesar

Valor

Valor de el logaritmo de la verosimilitud negativa evaluada en phi.

Función *IAR.loglik*

Descripción

Estimación de máxima verosimilitud del modelo IAR

Uso

`IAR.loglik(y, sT, delta = 0, include.mean = "FALSE", standarized = "TRUE")`

Argumentos

y	Matriz con las observaciones de series de tiempo.
sT	Matriz con los tiempos de observación irregulares.
delta	Matriz con las desviaciones estándar del error de medición.
include.mean	Lógico; si es verdadero, la matriz y tiene media cero; si es falso, y tiene una media diferente de cero.
standarized	Lógico; si es verdadero, la matriz y está estandarizada; si es falso, y contiene la serie de tiempo sin procesar

Valor

Una lista con los siguientes componentes:

phi	EMV del coeficiente phi del modelo IAR.
ll	Valor de la probabilidad logarítmica negativa evaluada en phi.

Con el objetivo de evaluar el ajuste del modelo propuesto a los datos simulados, se generó un ciclo, donde se muestra de forma gráfica el ajuste del modelo para cuatro simulaciones.

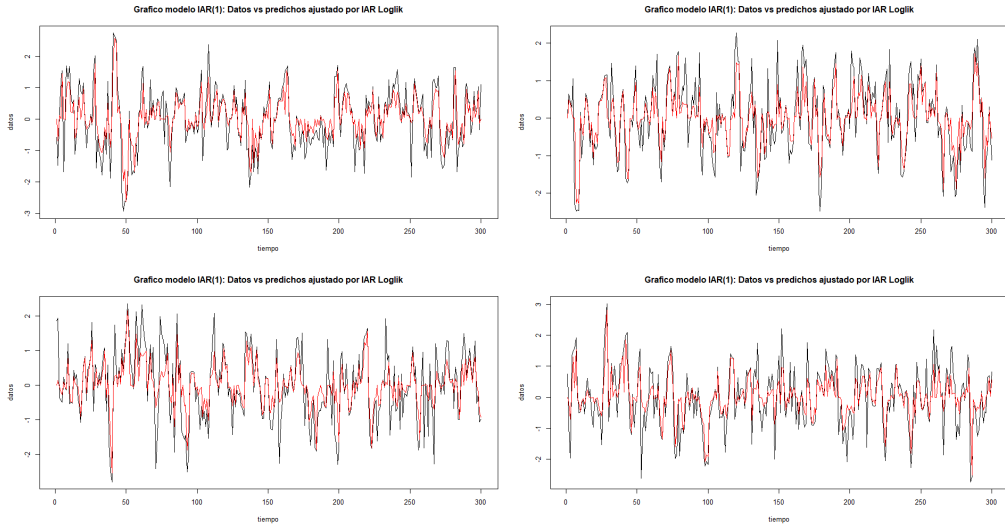


Figura 4.5: Gráfico de datos simulados a partir de un modelo $IAR(1)$ versus datos predichos por un modelo $IAR(1)$ con estimación por máxima verosimilitud directa, con datos irregularmente espaciados

De la figura (4.5) se aprecia en color negro los datos simulados por el modelo $IAR(1)$ y en rojo los datos estimados por el modelo $IAR(1)$ con estimación por verosimilitud directa. A priori se puede observar que el modelo propuesto tiene un buen ajuste.

En el siguiente cuadro se evidencia parte de los resultados obtenidos sobre el parámetro ϕ y el tiempo de computo.

Modelo $IAR(1)$ máxima verosimilitud			
Parámetro Real	Parámetro Estimado	Diferencia	Tiempo de Computo
ϕ	$\hat{\phi}$	$ \phi - \hat{\phi} $	Segundos
0.9	0.9182	0.0182	0.03
0.9	0.9165	0.0165	0
0.9	0.885	0.015	0
0.9	0.8988	0.0012	0
0.9	0.8908	0.0092	0.01
0.9	0.9097	0.0097	0
\vdots	\vdots	\vdots	\vdots
Medias:	$\bar{\hat{\phi}}=0.8995$	$ \bar{\phi} - \bar{\hat{\phi}} =0.0005$	0.001
Desv Estándar:	0.01504		0.004144

Cuadro 4.8: Parámetro real vs Estimado; Tiempo de Computo

Como se puede observar en el cuadro (4.8), el modelo $IAR(1)$ con estimación por verosimilitud directa, tuvo un

tiempo de cómputo muy pequeño incluso menor que el del modelo $IAR(1)$ por estimación de Kalman, esto se debe a que el uso de estas recursiones para disminuir el costo computacional se destacan más en modelos con tiempos de compilación mucho mayores. Además la diferencia entre la estimación del parámetro versus el parámetro real es casi un cero numérico. Tal como se pudo apreciar en la figura (4.5), el modelo propuesto se ajusta de buena manera a la serie simulada.

Con el objetivo de corroborar lo anteriormente planteado, se analiza nuevamente la bondad del ajuste a partir de los indicadores $-Loglik$, AIC , MSE y $MSE2$.

Modelo $IAR(1)$ máxima verosimilitud				
	-Loglik	AIC	MSE	MSE2
	298	600	0.2722	1.115
	300	604.1	0.2837	1.049
	323.7	651.4	0.3222	0.4792
	321	646.1	0.3285	1.222
	339.6	683.2	0.3914	1.219
	313.7	631.4	0.3197	1.417
	⋮	⋮	⋮	⋮
Medias:	319.5	643.1	0.3329	0.8448
Desv Estándar:	15.42	30.84	0.0504	0.3155

Cuadro 4.9: Bondad de Ajuste del Modelo $IAR(1)$ máxima verosimilitud

De forma contraria a lo planteado en los puntos anteriores, a partir del cuadro (4.9), se puede notar que el modelo $IAR(1)$ con estimación por máxima verosimilitud directa, tuvo tanto una media de $-Loglik$ como AIC de las simulaciones muy alta. Por esto no sería recomendable el uso de un modelo $IAR(1)$ con estimación por máxima verosimilitud directa.

4.1.1.5. Ajuste por modelo $IAR - T$

Se simularon valores desigualmente espaciados que se rigen por un modelo $IAR(1)$, posteriormente fueron ajustados a un modelo $IAR - T$ con estimación de parámetros por máxima verosimilitud, se utilizarán las funciones $IAR.phi.t$ e $IAR.t$, pertenecientes a la librería $AITS$.

A continuación el detalle de las funciones.

Descripción

Esta función devuelve el logaritmo de la verosimilitud negativa del IAR-T dados valores específicos de phi y sigma

Uso

```
IAR.phi.t(x, y, sT, nu = 3)
```

Argumentos

x	Una matriz con los parámetros del modelo IAR-T. El primer elemento de la matriz corresponde al parámetro phi y el segundo elemento al parámetro de escala sigma.
y	Matriz con las observaciones de series de tiempo.
sT	Matriz con los tiempos de observación irregulares.
nu	Grados de libertad.

Valor

Valor de el logaritmo de la verosimilitud negativa evaluada en phi, mu y sigma.

Función *IAR.t*

Descripción

Estimación de máxima verosimilitud del modelo IAR-T.

Uso

```
IAR.t(y, sT, nu = 3)
```

Argumentos

y	Matriz con las observaciones de series de tiempo.
sT	Matriz con los tiempos de observación irregulares.
nu	Grados de libertad.

Valor

Una lista con los siguientes componentes:

phi	EMV del coeficiente phi del modelo IAR-T.
ll	Valor de la probabilidad logarítmica negativa evaluada en phi.
sigma	EMV del coeficiente sigma del modelo IAR-T.

Con el objetivo de evaluar el ajuste del modelo propuesto a los datos simulados, se generó un ciclo, donde se muestra de forma gráfica el ajuste del modelo para cuatro simulaciones.

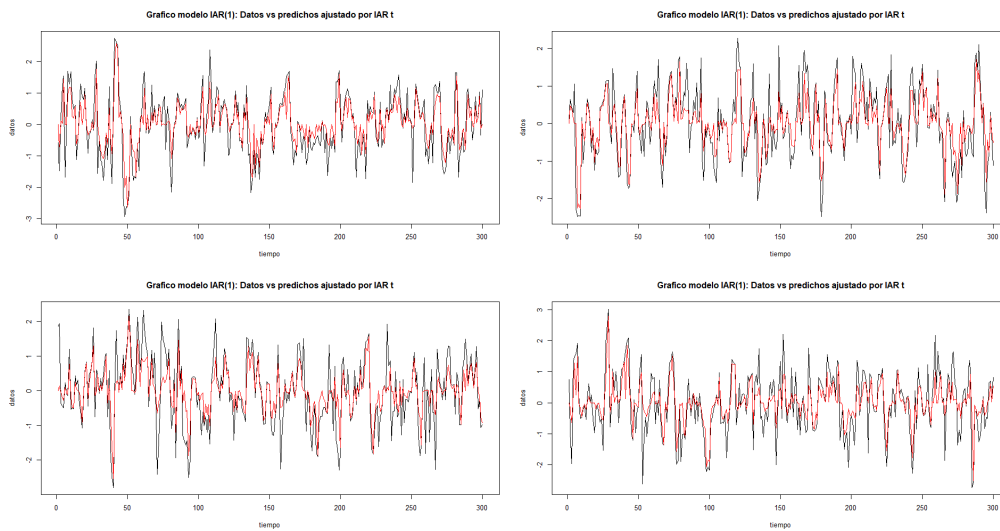


Figura 4.6: Gráfico de datos simulados a partir de un modelo $IAR(1)$ versus datos predichos por un modelo $IAR - T$ con estimación por máxima verosimilitud, con datos irregularmente espaciados

De la figura (4.6) se aprecia en color negro los datos simulados por el modelo $IAR(1)$ y en rojo los datos estimados por el modelo $IAR - T$, donde a diferencia de los modelos IAR planteados anteriormente este no asume gaussianidad de datos, sino que estos provienen de una distribución t-student. A priori se puede observar que el modelo propuesto tiene un buen ajuste, sin embargo no tan bueno como el de sus análogos normalmente distribuidos.

En el siguiente cuadro se evidencia parte de los resultados obtenidos sobre el parámetro ϕ y el tiempo de computo.

Modelo $IAR - T$			
Parámetro Real	Parámetro Estimado	Diferencia	Tiempo de Computo
ϕ	$\hat{\phi}$	$ \phi - \hat{\phi} $	Segundos
0.9	0.9207	0.0207	0.11
0.9	0.9112	0.0112	0.08
0.9	0.8847	0.0153	0.06
0.9	0.8903	0.0097	0.08
0.9	0.8977	0.0023	0.11
0.9	0.8977	0.0023	0.09
\vdots	\vdots	\vdots	\vdots
Medias:	$\bar{\hat{\phi}}=0.8992$	$ \bar{\phi} - \bar{\hat{\phi}} =0.0008$	0.0783
Desv Estándar:	0.0171		0.01164

Cuadro 4.10: Parámetro real vs Estimado; Tiempo de Computo

Como se puede observar en el cuadro (4.10), el modelo $IAR - T$, tuvo un tiempo de cómputo muy pequeño y además la diferencia entre la estimación del parámetro versus el parámetro real es casi un cero numérico. Tal como se pudo apreciar en la figura (4.6), el modelo propuesto se ajusta de buena manera a la serie simulada.

Con el objetivo de corroborar lo anteriormente planteado, se analiza nuevamente la bondad del ajuste a partir de los indicadores $-Loglik$, AIC , MSE y $MSE2$.

Modelo $IAR - T$				
	-Loglik	AIC	MSE	MSE2
	315.6	635.1	0.2663	1.114
	316.2	636.4	0.2947	1.067
	341.6	687.2	0.323	0.4781
	330.8	665.6	0.3476	1.24
	351.4	706.7	0.3764	1.218
	330.3	664.7	0.3473	1.416
	\vdots	\vdots	\vdots	\vdots
Medias:	332.9	669.8	0.3331	0.8467
Desv Estándar:	16.09	32.19	0.05439	0.3175

Cuadro 4.11: Bondad de Ajuste del Modelo $IAR - T$

Confirmando lo planteado en los puntos anteriores, a partir del cuadro (4.11), se puede notar que el modelo $IAR - T$, tuvo tanto una media de $-Loglik$, AIC bastante altas. Sin embargo también tiene un MSE de las simulaciones bastante baja. Por todo esto, es recomendable el uso del modelo $IAR - T$ en series desigualmente espaciadas con ϕ positivo.

4.1.1.6. Ajuste por modelo $CIAR$ Kalman

Se simularon valores desigualmente espaciados que se rigen por un modelo $IAR(1)$, posteriormente fueron ajustados a un modelo $CIAR$ con estimación de parámetros por filtro de Kalman, se utilizarán las funciones $CIAR.phi.kalman$ y $CIAR.kalman$, pertenecientes a la librería $AITS$.

A continuación el detalle de la función.

Función $CIAR.phi.kalman$

Descripción

Esta función devuelve el logaritmo de la verosimilitud negativa del proceso $CIAR$ dados valores específicos de $phi.R$ y phi .

Uso

$CIAR.phi.kalman(x, y, t, yerr, zero.mean = "TRUE", standarized = "TRUE", c = 1)$

Argumentos

x	Una matriz con los parámetros del modelo $CIAR$. Los elementos de la matriz son, en orden, la parte real ($phi.R$) y la imaginaria ($phi.I$) del coeficiente de Modelo $CIAR$.
y	Matriz con las observaciones de series de tiempo.
t	Matriz con los tiempos de observación irregulares.
$yerr$	Matriz con las desviaciones estándar del error de medición.
$zero.mean$	Lógico; si es verdadero, la matriz y tiene media cero; si es falso, y tiene una media diferente de cero.
$standarized$	Lógico; si es verdadero, la matriz y está estandarizada; si es falso, y contiene la serie de tiempo sin procesar
c	Parámetro de molestia correspondiente a la varianza de la parte imaginaria.

Valor

Valor de el logaritmo de la verosimilitud negativa evaluada en $phiR$ y $phiI$.

Función $CIAR.kalman$

Descripción

Estimación de máxima verosimilitud de los parámetros del modelo CIAR $\phi.R$ y $\phi.I$. El procedimiento utiliza el filtro de Kalman para encontrar la máxima verosimilitud.

Uso

`CIAR.kalman(y, t, delta = 0, zero.mean = "TRUE", standardized = "TRUE", c = 1, niter = 10, seed = 1234)`

Argumentos

<code>y</code>	Matriz con las observaciones de series de tiempo.
<code>t</code>	Matriz con los tiempos de observación irregulares.
<code>delta</code>	Matriz con las desviaciones estándar del error de medición.
<code>zero.mean</code>	Lógico; si es verdadero, la matriz <code>y</code> tiene media cero; si es falso, <code>y</code> tiene una media diferente de cero.
<code>standardized</code>	Lógico; si es verdadero, la matriz <code>y</code> está estandarizada; si es falso, <code>y</code> contiene la serie de tiempo sin procesar.
<code>c</code>	Parámetro de molestia correspondiente a la varianza de la parte imaginaria.
<code>niter</code>	Número de iteraciones en las que se repetirá la función <code>nlminb</code> .
<code>seed</code>	Un solo valor, interpretado como la semilla del proceso aleatorio.

Valor

Una lista con los siguientes componentes:

<code>phi.R</code>	EMV de la parte Real del coeficiente del modelo CIAR ($\phi.R$).
<code>ll</code>	Valor de la probabilidad logarítmica negativa evaluada en <code>phi</code> .
<code>phi.I</code>	EMV de la parte imaginaria del coeficiente del modelo CIAR ($\phi.I$).

Con el objetivo de evaluar el ajuste del modelo propuesto a los datos simulados, se generó un ciclo, donde se muestra de forma gráfica el ajuste del modelo para cuatro simulaciones.

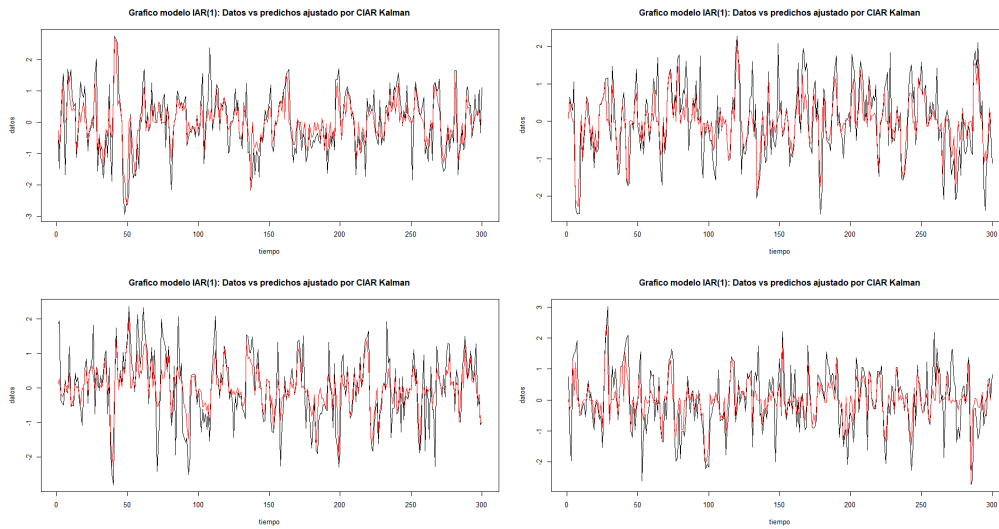


Figura 4.7: Gráfico de datos simulados a partir de un modelo $IAR(1)$ versus datos predichos por un modelo $CIAR$ con estimación por filtro de Kalman, con datos irregularmente espaciados

De la figura (4.7) se aprecia en color negro los datos simulados por el modelo $IAR(1)$ y en rojo los datos estimados por el modelo $CIAR$ con estimación por filtro de Kalman. A priori se puede observar que el modelo propuesto tiene un buen ajuste, a pesar de que los datos no provienen del mismo modelo.

En el siguiente cuadro se evidencia parte de los resultados obtenidos sobre el parámetro ϕ y el tiempo de computo.

Modelo $CIAR$ filtro de Kalman			
Parámetro Real	Parámetro Estimado	Diferencia	Tiempo de Computo
ϕ	$\hat{\phi}$	$ \phi - \hat{\phi} $	Segundos
0.9	0.9183	0.0184	5.44
0.9	0.9215	0.0216	3.66
0.9	0.8844	0.0146	4.19
0.9	0.9079	0.0077	4.94
0.9	0.9	0.0001	5.1
0.9	0.9097	0.0097	4.83
\vdots	\vdots	\vdots	\vdots
Medias:	$\bar{\hat{\phi}}=0.9024$	$ \bar{\phi} - \bar{\hat{\phi}} =0.0024$	5.472
Desv Estándar:	0.01549		1.67

Cuadro 4.12: Parámetro real vs Estimado; Tiempo de Computo

Como se puede observar en el cuadro (4.12), el modelo $CIAR$ con estimación por filtro de Kalman, a diferencia

de los modelos anteriores tuvo un tiempo de cómputo bastante mayor debido a que tiene un parámetro más que estimar y el cálculo de su verosimilitud es más complejo, obviando esta desventaja con respecto a los otros modelos, la diferencia entre la estimación del parámetro y el parámetro real es casi un cero numérico. Tal como se pudo apreciar en la figura (4.7), el modelo propuesto se ajusta de buena manera a la serie simulada.

Con el objetivo de corroborar lo anteriormente planteado, se analiza nuevamente la bondad del ajuste a partir de los indicadores $-Loglik$, AIC , MSE y $MSE2$.

	Modelo $CIAR$ filtro de Kalman			
	-Loglik	AIC	MSE	MSE2
	0.1458	6.292	0.5673	0.8179
	0.1548	6.31	0.5511	1.041
	0.3386	6.677	0.6422	0.4243
	0.291	6.582	0.6572	0.939
	0.4127	6.825	0.7051	0.9404
	0.256	6.512	0.6038	1.123
	⋮	⋮	⋮	⋮
Medias:	0.2932	6.586	0.632	0.7668
Desv Estándar:	0.1029	0.2057	0.05195	0.2507

Cuadro 4.13: Bondad de Ajuste del Modelo $CIAR$ filtro de Kalman

Confirmando lo planteado en los puntos anteriores, a partir del cuadro (4.13), se puede notar que el modelo $CIAR$ con estimación por filtro de Kalman, tuvo tanto una media de $-Loglik$ y MSE de las simulaciones bastante baja. Además su AIC también es bastante pequeño considerando que este modelo tiene un parámetro más de castigo en el criterio de Akaike. Por todo esto, es recomendable el uso del modelo $CIAR$ con estimación por filtro de Kalman en series desigualmente espaciadas con ϕ positivo.

4.1.1.7. Comparación de Resultados para Datos Simulados por Modelo $IAR(1)$ y Tiempos de Naturaleza Exponencial

A continuación una tabla resumen donde se comparan los datos entregados por los puntos anteriores:

	Media					Desv Estándar			
	Dif	TdC	-LL	AIC	MSE	TdC	-LL	AIC	MSE
Modelo $AR(1)$	0.3889	0.0033	379.5	762.9	0.7431	0.006675	16.24	32.48	0.05509
Modelo CAR	0.0139	0.0034	165.1	334.1	0.2836	0.007138	14.05	28.1	0.1301
Modelo $IAR.k$	0.0006	0.1686	0.2953	4.591	0.3331	0.01583	0.1026	0.2052	0.05049
Modelo $IAR.l$	0.0005	0.001	319.5	643.1	0.3329	0.004144	15.42	30.84	0.0504
Modelo $IAR.t$	0.0008	0.0783	332.9	669.8	0.3331	0.01164	16.09	32.19	0.05439
Modelo CIAR	0.0024	5.472	0.2932	6.586	0.632	1.67	0.1029	0.2057	0.05195

Cuadro 4.14: Comparación de Modelos simulados por $IAR(1)$ y tiempos de naturaleza exponencial

Como se puede notar en el cuadro (4.14), los modelos propuestos con peor desempeño para el ajuste fueron el modelo $AR(1)$ e $IAR.l$. Se puede notar que en la estimación de parámetros, el modelo $AR(1)$ no tuvo el mismo desempeño que los demás, repercutiendo en el ajuste como se puede ver de forma más notoria tanto en el $-LL$ como en el AIC , esto se debe a que no tiene la capacidad de adaptarse a series irregularmente espaciadas.

Por otra parte tanto el ajuste del modelo $IAR(1)$ con estimación por filtro de Kalman como el modelo $CIAR$ con estimación por filtro de Kalman tuvieron el mejor desempeño. Ambos con poca diferencia entre el valor real del parámetro y el estimado, provocando una incidencia directa en los indicadores de bondad de ajuste. Es importante destacar que si bien, ambos tuvieron una buena estimación, el modelo IAR tuvo un tiempo de computo ampliamente menor, es por esto que para esta primera simulación el modelo más óptimo a utilizar es el $IAR(1)$ con estimación por filtro de Kalman.

4.1.1.8. Comparación de Resultados para Datos Simulados por Modelo $IAR(1)$ y Tiempos de Naturaleza Uniforme

Análogamente se realiza el mismo procedimiento antes descrito, a partir de tiempos con naturaleza uniforme, con el objetivo de encontrar nuevas condiciones para el uso de los modelos. A continuación una tabla resumen donde se comparan los datos entregados según los puntos anteriores:

	Media					Desv Estándar			
	Dif	TdC	-LL	AIC	MSE	TdC	-LL	AIC	MSE
Modelo AR(1)	0.014	0.0031	176.5	356.9	0.2265	0.007063	12.62	25.25	0.05867
Modelo CAR	0.0049	0.0041	92.79	189.6	0.01215	0.007797	16.94	33.88	0.008398
Modelo IAR.k	0.0141	0.1711	-0.5934	2.813	0.01907	0.02197	0.2246	0.4492	0.01037
Modelo IAR.l	0.0141	0.0041	185.4	374.9	0.01906	0.007667	33.78	67.57	0.01032
Modelo IAR.t	0.0125	0.1668	1.915	7.829	0.01859	0.01797	0.1223	0.2446	0.01024
Modelo CIAR	0.014	4.832	-0.5963	4.807	0.2135	1.228	0.2253	0.4505	0.05029

Cuadro 4.15: Comparación de Modelos simulados por $IAR(1)$ y tiempos de naturaleza uniforme

De igual forma que en la comparación anterior, se puede notar en el cuadro (4.15) que los modelos propuestos con peor desempeño para el ajuste fueron el modelo $AR(1)$ e $IAR.l$. Una vez más el modelo $AR(1)$ no tuvo un buen desempeño para estimar los parámetros, repercutiendo en el ajuste como se puede ver de forma más notoria tanto en el $-LL$ como en el AIC .

Se repiten los resultados del modelo $IAR(1)$ con estimación por filtro de Kalman como el del modelo $CIAR$ con estimación por filtro de Kalman con el mejor desempeño. Ambos con poca diferencia entre el valor real del parámetro y el estimado, mostrando un mejor ajuste según los indicadores $-LL$ como AIC . Con esto se confirma que bajo el caso de estar en presencia de series desigualmente espaciadas con ϕ positivo estos modelos se ajustan con gran precisión.

4.1.2. Datos simulados a partir de un model $CIAR$

Para simular la base de datos con la que se trabajará en todos los modelos de esta sección, se utilizarán las funciones $gentime$ y $CIAR.sample$, ambas pertenecientes a la librería $AITs$.

A continuación el detalle de la función $CIAR.sample$.

Función $CIAR.sample$

Descripción

Simula un modelo de serie temporal $CIAR$.

Uso

$CIAR.sample(n, sT, phi.R, phi.I, rho = 0, c = 1)$

Argumentos

n Duración de la serie temporal de salida. Un número entero estrictamente positivo.

sT	Matriz con tiempos de observación.
phi.R	Parte real del coeficiente del modelo CIAR. Un valor entre -1 y 1.
phi.I	Parte imaginaria del coeficiente del modelo CIAR. Un valor entre -1 y 1.
rho	Correlación entre la parte real e imaginaria del proceso. Un valor entre -1 y 1.
c	Parámetro de molestia correspondiente a la varianza de la parte imaginaria.

Valor

Lista con los siguientes componentes:

y	Matriz con la parte real simulada del proceso CIAR.
t	Matriz con tiempos de observación.
sigma	Matriz de covarianza del proceso.

Se procede a realizar 100 simulaciones de datos en base a un modelo *CIAR*. Primero se generan tiempos irregularmente espaciados a partir de una mezcla de distribuciones exponenciales, definiendo su tamaño muestral, en este caso de 300 observaciones. Posterior a esto se asigna un valor al coeficiente ϕ , en este caso de $-0,9$, con el objetivo de que sólo el modelo *CIAR* tenga una buena estimación de parámetros por su capacidad de detectar valores negativos en la ACF.

Es posible ver a través de las figuras como se comportan los datos, a continuación un ejemplo de 4 simulaciones.

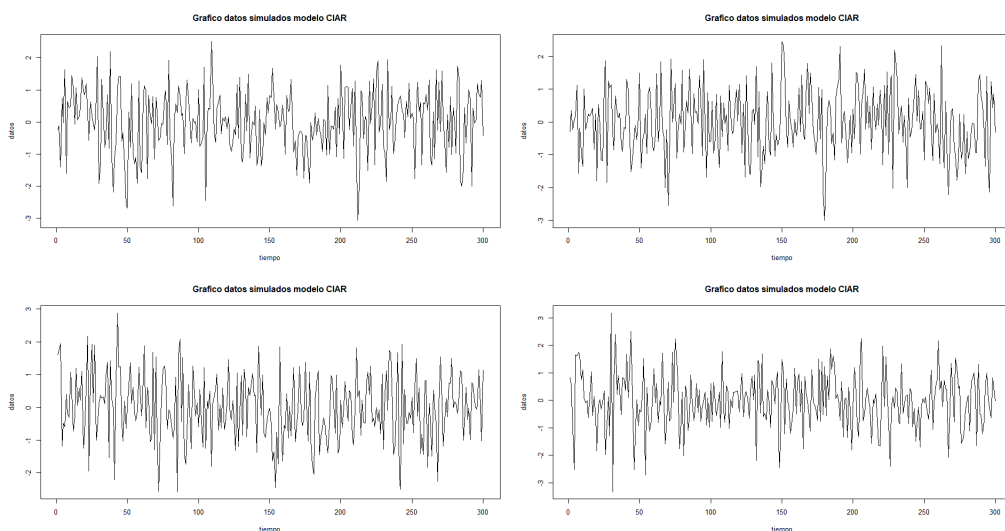


Figura 4.8: Gráfico de datos simulados en base a un modelo *CIAR* con datos irregularmente espaciados

Simulación	Variable	Min	Q1	Mediana	Media	Q3	Max	Sd
1	y1	-3.0735	-0.6402	0.0432	0	0.7546	2.5022	1
1	diff(st)	0	2.2	5.4	21.5	12.5	337.6	49.09
2	y1	-3.0035	-0.7051	-0.0528	0	0.7801	2.4502	1
2	diff(st)	0	2.3	5.9	35.1	14.9	621.3	90.83
3	y1	-2.5774	-0.6548	0.0047	0	0.6799	2.8806	1
3	diff(st)	0	1.8	4.7	19.3	11.0	407.9	54.63
4	y1	-3.326	-0.637	0.007	0	0.621	3.172	1
4	diff(st)	0	2.4	6.8	29.8	14.5	470.3	72.58
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Media:	y1	-2.882	-0.679	-0.002538	0	0.679	2.796	1
Media:	diff(st)	0.02192	2.254	5.651	24.79	13.16	546.7	65.26

Cuadro 4.16: Descripción de las variables

Observando tanto la figura (4.8) como el cuadro (4.16), a priori se puede apreciar tanto una media como una dispersión de datos constante, esto quiere decir que la serie tiende a un comportamiento estacionario, esto es de esperarse considerando que anteriormente se mencionó que estos modelos tienden a ser estacionarios de segundo orden.

4.1.2.1. Ajuste por modelo AR

Se simularon valores desigualmente espaciados que se rigen por un modelo $CIAR$, posteriormente fueron ajustados a un modelo $AR(1)$, se utilizará la función *arma*, perteneciente a la librería *tseries*.

Con el objetivo de evaluar el ajuste del modelo propuesto a los datos simulados, se generó un ciclo, donde se muestra de forma gráfica el ajuste del modelo para cuatro simulaciones.

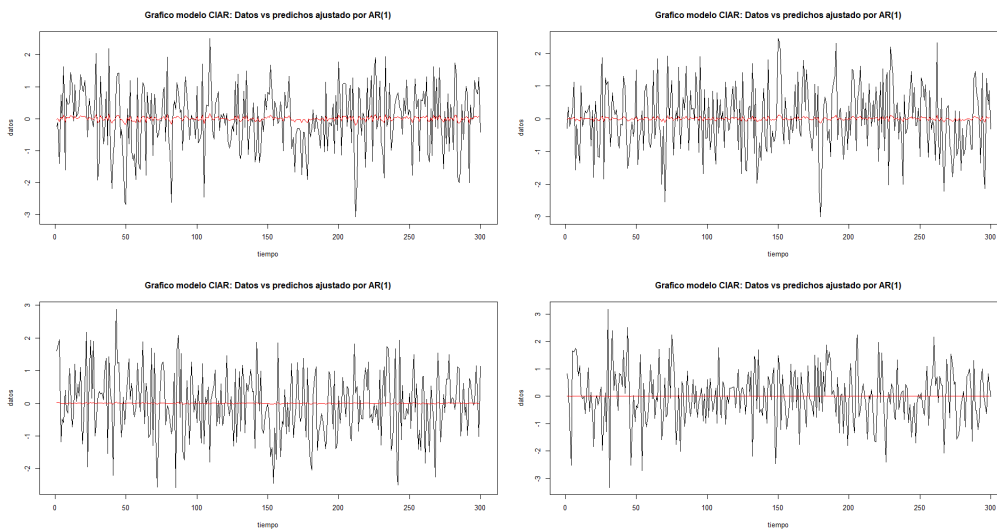


Figura 4.9: Gráfico de datos simulados a partir de un modelo $CIAR$ versus datos predichos por un modelo $AR(1)$ con datos irregularmente espaciados

De la figura (4.9) se aprecia en color negro los datos simulados por el modelo $CIAR$ y en rojo los datos estimados por el modelo $AR(1)$. A priori se puede observar que el modelo propuesto tiene un mal ajuste, se asumió previamente que este ocurriría sabiendo que el modelo AR no se ajusta bien a series de tiempo desigualmente espaciadas, además de que no sería capaz de adaptarse a un ϕ negativo.

En el siguiente cuadro se evidencia parte de los resultados obtenidos sobre el parámetro ϕ y el tiempo de computo.

Modelo $AR(1)$			
Parámetro Real	Parámetro Estimado	Diferencia	Tiempo de Computo
ϕ	$\hat{\phi}$	$ \phi - \hat{\phi} $	Segundos
-0.9	0.06594	0.96594	0.02
-0.9	0.04509	0.94509	0
-0.9	0.01206	0.91206	0
-0.9	0.00167	0.90167	0.01
-0.9	-0.1023	0.7977	0
-0.9	0.02094	0.92094	0.02
\vdots	\vdots	\vdots	\vdots
Medias:	$\bar{\hat{\phi}}=0.006095$	$ \bar{\phi} - \bar{\hat{\phi}} =0.906095$	0.0035
Desv Estándar:	0.06461		0.007017

Cuadro 4.17: Parámetro real vs Estimado; Tiempo de Computo

Como se puede observar en el cuadro (4.17), el modelo $AR(1)$ tuvo un tiempo de cómputo muy pequeño lo que sería muy útil al momento de trabajar con un gran volumen de datos, sin embargo, la estimación del parámetro esta completamente errónea. Tal como se pudo apreciar en la figura (4.9), el modelo propuesto no se ajusta de buena manera a la serie simulada.

Con el objetivo de terminar de confirmar lo anteriormente planteado, se analiza la bondad del ajuste a partir de los indicadores $-Loglik$ que es el logaritmo de la verosimilitud negativa de los datos, AIC que es el criterio de información de Akaike y el MSE que es el error cuadrado medio del modelo y el $MSE2$ que es el error cuadrado medio de la predicción (10 % de los valores reales menos los predichos por el 90 %).

	Modelo $AR(1)$			
	-Loglik	AIC	MSE	MSE2
	445.6	853.1	0.9954	1.144
	431.4	853.7	0.9977	1.125
	417.7	854.3	0.9913	0.5085
	411.5	854.4	0.9977	0.7985
	418.8	851.2	0.9895	1.437
	413.9	854.2	0.9961	1.387
	⋮	⋮	⋮	⋮
Medias:	424.6	853.1	0.9928	1.014
Desv Estándar:	0.9492	1.898	0.007277	0.3055

Cuadro 4.18: Bondad de Ajuste del Modelo $AR(1)$

Confirmando lo planteado en los puntos anteriores, a partir del cuadro (4.18), se puede notar que el modelo $AR(1)$ tuvo tanto una media de $-Loglik$ como AIC de las simulaciones muy alta. Por esto no sería recomendable el uso de un modelo $AR(1)$ para tiempo desigualmente espaciados.

4.1.2.2. Ajuste por modelo CAR

Se simularon valores desigualmente espaciados que se rigen por un modelo $CIAR$, posteriormente fueron ajustados a un modelo $CAR(1)$, se utilizará la función *carma*, perteneciente a la librería *growth*.

Con el objetivo de evaluar el ajuste del modelo propuesto a los datos simulados, se generó un ciclo, donde se muestra de forma gráfica el ajuste del modelo para cuatro simulaciones.

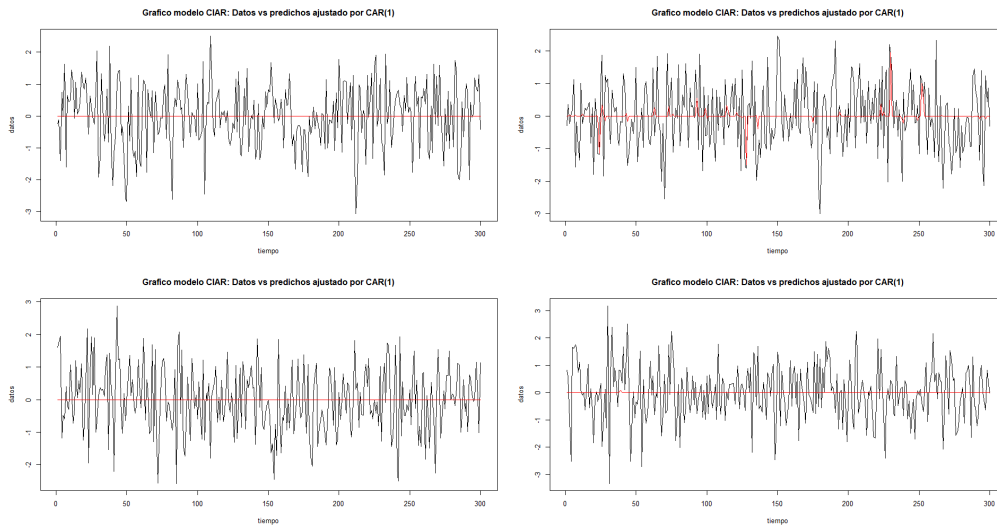


Figura 4.10: Gráfico de datos simulados a partir de un modelo *CIAR* versus datos predichos por un modelo *CAR(1)* con datos irregularmente espaciados

De la figura (4.10) se aprecia en color negro los datos simulados por el modelo *CIAR* y en rojo los datos estimados por el modelo *CAR(1)*. A priori se puede observar que el modelo propuesto tiene un mal ajuste, se asumía previamente que este ocurriría sabiendo que el modelo *CAR* no es capaz de adaptarse a un ϕ negativo.

En el siguiente cuadro se evidencia parte de los resultados obtenidos sobre el parámetro ϕ y el tiempo de computo.

Modelo <i>CAR(1)</i>			
Parámetro Real	Parámetro Estimado	Diferencia	Tiempo de Computo
ϕ	$\hat{\phi}$	$ \phi - \hat{\phi} $	Segundos
-0.9	0	0.9	0
-0.9	0.02986	0.92986	0
-0.9	0	0.9	0
-0.9	$8,152e^{-235}$	0.9	0.02
-0.9	0	0.9	0
-0.9	0	0.9	0
\vdots	\vdots	\vdots	\vdots
Medias:	$\bar{\hat{\phi}}=0.005721$	$ \bar{\phi} - \bar{\hat{\phi}} = 0.905721$	0.0045
Desv Estándar:	0.03005		0.007703

Cuadro 4.19: Parámetro real vs Estimado; Tiempo de Computo

Como se puede observar en el cuadro (4.19), el modelo *CAR(1)* tuvo un tiempo de cómputo muy pequeño lo

que sería muy útil al momento de trabajar con un gran volumen de datos, sin embargo, la estimación del parámetro esta completamente errónea. Tal como se pudo apreciar en la figura (4.10), el modelo propuesto no se ajusta de buena manera a la serie simulada.

Con el objetivo de corroborar lo anteriormente planteado, se analiza nuevamente la bondad del ajuste a partir de los indicadores $-Loglik$, AIC , MSE y $MSE2$.

	Modelo $CAR(1)$			
	-Loglik	AIC	MSE	MSE2
	212.1	428.2	0.9967	1.142
	207.1	418.2	0.9669	1.119
	212.1	428.2	0.9967	0.5023
	212.1	428.1	0.9962	0.7981
	212.1	428.2	0.9967	1.445
	212.1	428.2	0.9967	1.371
	⋮	⋮	⋮	⋮
Medias:	211.6	427.3	0.9945	1.007
Desv Estándar:	1.327	2.655	0.006493	0.3002

Cuadro 4.20: Bondad de Ajuste del Modelo $CAR(1)$

Confirmando lo planteado en los puntos anteriores, a partir del cuadro (4.20), se puede notar que el modelo $CAR(1)$ tuvo tanto una media de $-Loglik$ como AIC de las simulaciones muy alta. Esto era de esperarse considerando que no tiene la capacidad de adaptarse a un ϕ negativo.

4.1.2.3. Ajuste por modelo IAR Kalman

Se simularon valores desigualmente espaciados que se rigen por un modelo $CIAR$, posteriormente fueron ajustados a un modelo IAR con estimación de parámetros por filtro de Kalman, se utilizará la función $IAR.kalman$, perteneciente a la librería $AITS$.

Con el objetivo de evaluar el ajuste del modelo propuesto a los datos simulados, se generó un ciclo, donde se muestra de forma gráfica el ajuste del modelo para cuatro simulaciones.

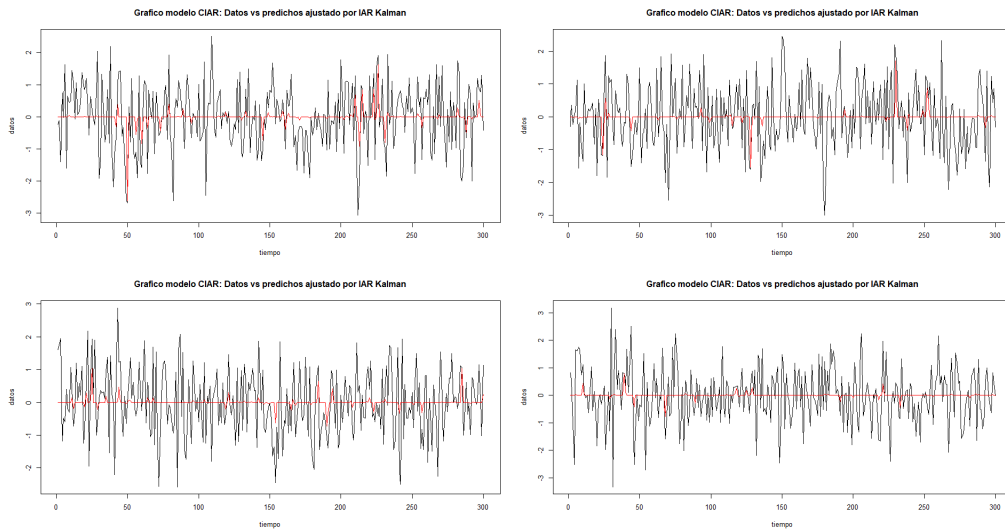


Figura 4.11: Gráfico de datos simulados a partir de un modelo *CIAR* versus datos predichos por un modelo *IAR(1)* con estimación por filtro de Kalman, con datos irregularmente espaciados

De la figura (4.11) se aprecia en color negro los datos simulados por el modelo *CIAR* y en rojo los datos estimados por el modelo *IAR(1)* con estimación por filtro de Kalman. A priori se puede observar que el modelo propuesto tiene un mal ajuste, esto era de esperarse considerando que no cuenta con la capacidad de adaptarse a ACF con valores negativos.

En el siguiente cuadro se evidencia parte de los resultados obtenidos sobre el parámetro ϕ y el tiempo de computo.

Modelo <i>IAR(1)</i> filtro de Kalman			
Parámetro Real	Parámetro Estimado	Diferencia	Tiempo de Computo
ϕ	$\hat{\phi}$	$ \phi - \hat{\phi} $	Segundos
-0.9	0.1145	1.0145	0.18
-0.9	0.03132	0.93132	0.19
-0.9	0.05456	0.95456	0.21
-0.9	0.03025	0.93025	0.16
-0.9	0.01825	0.91825	0.2
-0.9	0.09983	1.09983	0.16
\vdots	\vdots	\vdots	\vdots
Medias:	$\bar{\hat{\phi}}=0.06396$	$ \bar{\phi} - \bar{\hat{\phi}} =0.96396$	0.1753
Desv Estándar:	0.05582		0.03433

Cuadro 4.21: Parámetro real vs Estimado; Tiempo de Computo

Como se puede observar en el cuadro (4.21), el modelo $IAR(1)$ con estimación por filtro de Kalman, tuvo un tiempo de cómputo muy pequeño, sin embargo, la estimación del parámetro esta completamente errónea. Tal como se pudo apreciar en la figura (4.11), el modelo propuesto no se ajusta de buena manera a la serie simulada.

Con el objetivo de corroborar lo anteriormente planteado, se analiza nuevamente la bondad del ajuste a partir de los indicadores $-Loglik$, AIC , MSE y $MSE2$.

	Modelo $IAR(1)$ filtro de Kalman			
	-Loglik	AIC	MSE	MSE2
	0.9015	5.803	0.8892	1.116
	0.9304	5.861	0.9386	1.129
	0.9491	5.898	0.9253	0.6096
	0.9728	5.946	0.9628	0.8102
	0.9699	5.94	0.9752	1.441
	0.9438	5.888	0.936	1.422
	⋮	⋮	⋮	⋮
Medias:	0.9453	5.891	0.9319	1.029
Desv Estándar:	0.02317	0.04635	0.02948	0.3195

Cuadro 4.22: Bondad de Ajuste del Modelo $IAR(1)$ filtro de Kalman

Confirmando lo planteado en los puntos anteriores, a partir del cuadro (4.22), se puede notar que el modelo $IAR(1)$ con estimación por filtro de Kalman tuvo tanto una media de $-Loglik$ y AIC de las simulaciones bastante baja, sin embargo el MSE , muestra su deficiencia al momento de estimar el parámetro. Esto era de esperarse considerando que no tiene la capacidad de adaptarse a un ϕ negativo.

4.1.2.4. Ajuste por modelo IAR $loglik$

Se simularon valores desigualmente espaciados que se rigen por un modelo $CIAR$, posteriormente fueron ajustados a un modelo IAR con estimación de parámetros por máxima verosimilitud directa, se utilizará la función $IAR.loglik$, perteneciente a la librería $AITS$.

Con el objetivo de evaluar el ajuste del modelo propuesto a los datos simulados, se generó un ciclo, donde se muestra de forma gráfica el ajuste del modelo para cuatro simulaciones.

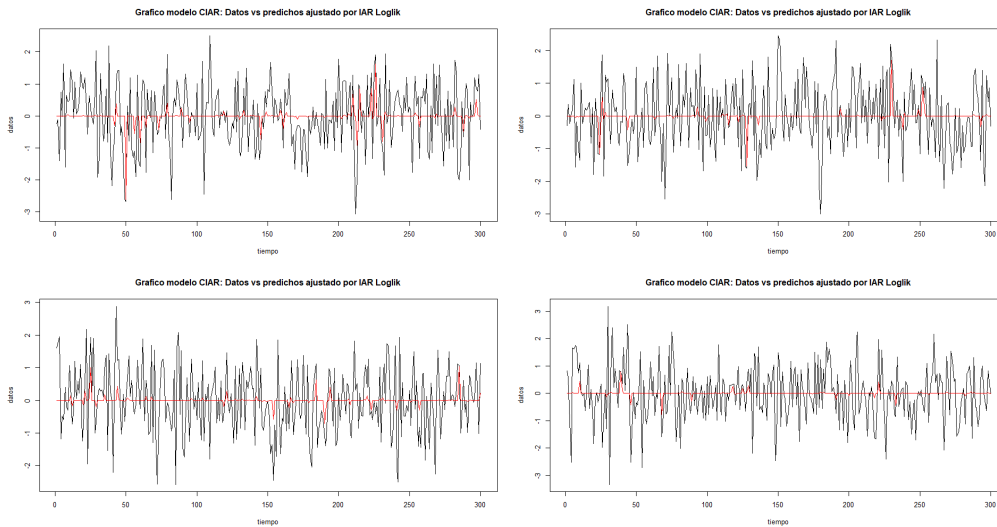


Figura 4.12: Gráfico de datos simulados a partir de un modelo *CIAR* versus datos predichos por un modelo *IAR(1)* con estimación por máxima verosimilitud directa, con datos irregularmente espaciados

De la figura (4.12) se aprecia en color negro los datos simulados por el modelo *CIAR* y en rojo los datos estimados por el modelo *IAR(1)* con estimación por filtro verosimilitud directa. A priori se puede observar que el modelo propuesto tiene un mal ajuste, esto era de esperarse considerando que no cuenta con la capacidad de adaptarse a ACF con valores negativos.

En el siguiente cuadro se evidencia parte de los resultados obtenidos sobre el parámetro ϕ y el tiempo de computo.

Modelo <i>IAR(1)</i> máxima verosimilitud			
Parámetro Real	Parámetro Estimado	Diferencia	Tiempo de Computo
ϕ	$\hat{\phi}$	$ \phi - \hat{\phi} $	Segundos
-0.9	0.1145	1.0145	0.01
-0.9	0.03132	0.93132	0
-0.9	0.0461	0.9461	0
-0.9	0.03025	0.93025	0
-0.9	0.01825	0.91825	0
-0.9	0.09983	0.99983	0
\vdots	\vdots	\vdots	\vdots
Medias:	$\bar{\hat{\phi}}=0.06374$	$ \bar{\phi} - \bar{\hat{\phi}} =0.96374$	0.0015
Desv Estándar:	0.05579		0.005198

Cuadro 4.23: Parámetro real vs Estimado; Tiempo de Computo

Como se puede observar en el cuadro (4.23), el modelo $IAR(1)$ con estimación por verosimilitud directa, tuvo un tiempo de cómputo muy pequeño, sin embargo, la estimación del parámetro esta completamente errónea. Tal como se pudo apreciar en la figura (4.12), el modelo propuesto no se ajusta de buena manera a la serie simulada.

Con el objetivo de corroborar lo anteriormente planteado, se analiza nuevamente la bondad del ajuste a partir de los indicadores $-Loglik$, AIC , MSE y $MSE2$.

Modelo $IAR(1)$ máxima verosimilitud				
	-Loglik	AIC	MSE	MSE2
	411	825.9	0.8892	1.116
	415.2	834.5	0.9386	1.129
	417.7	839.4	0.9295	0.6096
	421.3	846.5	0.9628	0.8102
	421.2	846.3	0.9752	1.441
	417.8	839.5	0.936	1.422
	⋮	⋮	⋮	⋮
Medias:	417.3	838.7	0.932	1.029
Desv Estándar:	3.448	6.895	0.02929	0.3194

Cuadro 4.24: Bondad de Ajuste del Modelo $IAR(1)$ máxima verosimilitud

Confirmando lo planteado en los puntos anteriores, a partir del cuadro (4.24), se puede notar que el modelo $IAR(1)$ con estimación por máxima verosimilitud directa tuvo tanto una media de $-Loglik$ y AIC de las simulaciones bastante alta, además el MSE muestra su deficiencia al momento de estimar el parámetro. Esto era de esperarse considerando que no tiene la capacidad de adaptarse a un ϕ negativo.

4.1.2.5. Ajuste por modelo $IAR - T$

Se simularon valores desigualmente espaciados que se rigen por un modelo $CIAR$, posteriormente fueron ajustados a un modelo $IAR - T$ con estimación de parámetros por máxima verosimilitud, se utilizará la función $IAR.t$, perteneciente a la librería $AITS$.

Con el objetivo de evaluar el ajuste del modelo propuesto a los datos simulados, se generó un ciclo, donde se muestra de forma gráfica el ajuste del modelo para cuatro simulaciones.

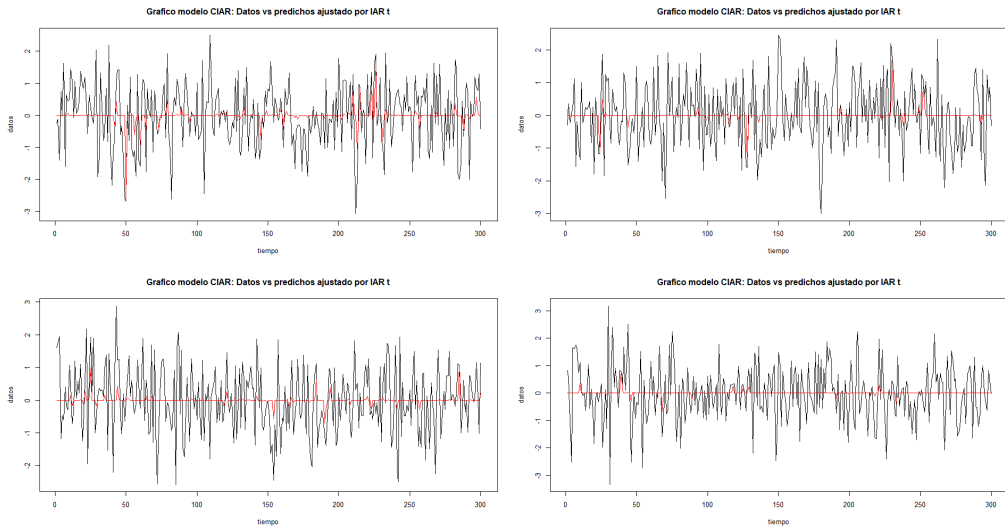


Figura 4.13: Gráfico de datos simulados a partir de un modelo *CIAR* versus datos predichos por un modelo *IAR – T* con estimación por máxima verosimilitud, con datos irregularmente espaciados

De la figura (4.13) se aprecia en color negro los datos simulados por el modelo *CIAR* y en rojo los datos estimados por el modelo *IAR – T*, donde a diferencia de los modelos *IAR* planteados anteriormente este no asume gaussianidad de datos, sino que estos provienen de una distribución t-student. A priori se puede observar que el modelo propuesto tiene un mal ajuste, esto era de esperarse dado el ejemplo de los modelos anteriores, considerando que no cuenta con la capacidad de adaptarse a ACF con valores negativos.

En el siguiente cuadro se evidencia parte de los resultados obtenidos sobre el parámetro ϕ y el tiempo de computo.

Modelo $IAR - T$			
Parámetro Real	Parámetro Estimado	Diferencia	Tiempo de Computo
ϕ	$\hat{\phi}$	$ \phi - \hat{\phi} $	Segundos
-0.9	0.1537	1.0537	0.11
-0.9	0.01934	0.91934	0.09
-0.9	0.04328	0.94328	0.08
-0.9	0.01506	0.91506	0.09
-0.9	0.01849	0.91849	0.11
-0.9	0.05226	0.95226	0.07
\vdots	\vdots	\vdots	\vdots
Medias:	$\bar{\hat{\phi}}=0.06794$	$ \bar{\phi} - \bar{\hat{\phi}} =0.96794$	0.0789
Desv Estándar:	0.06322		0.01814

Cuadro 4.25: Parámetro real vs Estimado; Tiempo de Computo

Como se puede observar en el cuadro (4.25), el modelo $IAR - T$, tuvo un tiempo de cómputo muy pequeño, sin embargo, la estimación del parámetro esta completamente errónea. Tal como se pudo apreciar en la figura (4.13), el modelo propuesto no se ajusta de buena manera a la serie simulada.

Con el objetivo de corroborar lo anteriormente planteado, se analiza nuevamente la bondad del ajuste a partir de los indicadores $-Loglik$, AIC , MSE y $MSE2$.

Modelo $IAR - T$				
	-Loglik	AIC	MSE	MSE2
	424.9	853.8	0.8776	1.111
	430.8	865.5	0.9448	1.123
	433.8	871.7	0.931	0.6096
	426.4	856.7	0.9703	0.8076
	433.1	870.1	0.9751	1.442
	434	872	0.948	1.411
	\vdots	\vdots	\vdots	\vdots
Medias:	430.2	864.5	0.9299	1.03
Desv Estándar:	5.31	10.62	0.03342	0.3206

Cuadro 4.26: Bondad de Ajuste del Modelo $IAR - T$

Confirmando lo planteado en los puntos anteriores, a partir del cuadro (4.26), se puede notar que el modelo

$IAR - T$, tuvo tanto una media de $-Loglik$ y AIC de las simulaciones muy alta, además el MSE , muestra su deficiencia al momento de estimar el parámetro. Esto era de esperarse considerando que no tiene la capacidad de adaptarse a un ϕ negativo.

4.1.2.6. Ajuste por modelo *CIAR Kalman*

Se simularon valores desigualmente espaciados que se rigen por un modelo *CIAR*, posteriormente fueron ajustados a un modelo *CIAR* con estimación de parámetros por filtro de Kalman, se utilizará la función *CIAR.kalman*, perteneciente a la librería *AITIS*.

Con el objetivo de evaluar el ajuste del modelo propuesto a los datos simulados, se generó un ciclo, donde se muestra de forma gráfica el ajuste del modelo para cuatro simulaciones.

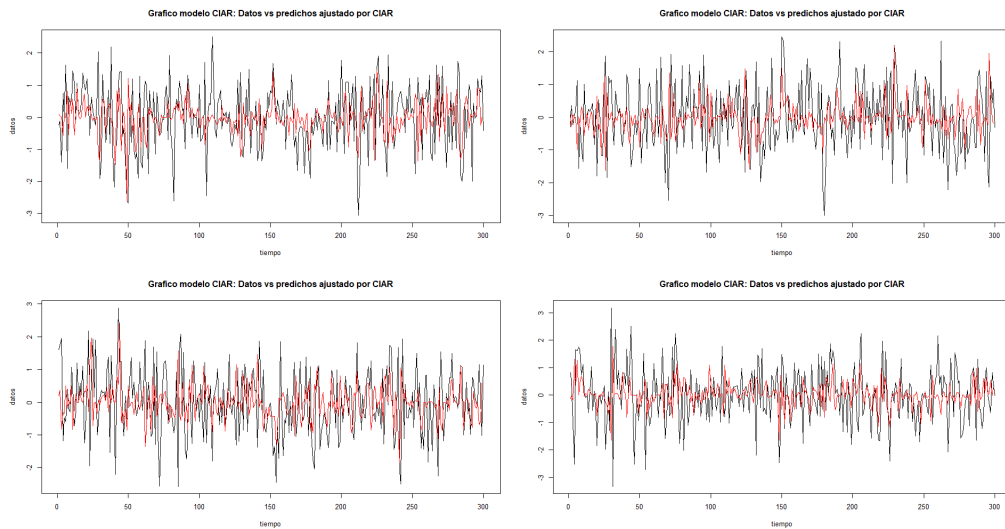


Figura 4.14: Gráfico de datos simulados a partir de un modelo *CIAR* versus datos predichos por un modelo *CIAR* con estimación por filtro de Kalman, con datos irregularmente espaciados

De la figura (4.14) se aprecia en color negro los datos simulados por el modelo *CIAR* y en rojo los datos estimados por el modelo *CIAR* con estimación por filtro de Kalman. El objetivo inicial era que sólo este modelo tuviese un buen ajuste y a priori se puede observar que este objetivo se cumple.

En el siguiente cuadro se evidencia parte de los resultados obtenidos sobre el parámetro ϕ y el tiempo de computo.

Modelo <i>CIAR</i> filtro de Kalman			
Parámetro Real	Parámetro Estimado	Diferencia	Tiempo de Computo
ϕ	$\hat{\phi}$	$ \phi - \hat{\phi} $	Segundos
-0.9	-0.9089	0.0089	4.83
-0.9	-0.8868	0.0132	5.19
-0.9	-0.9117	0.0117	5.12
-0.9	-0.8939	0.061	4.11
-0.9	-0.911	0.011	7.55
-0.9	-0.8899	0.0101	7.97
\vdots	\vdots	\vdots	\vdots
Medias:	$\bar{\hat{\phi}}=-0.8964$	$ \bar{\phi} - \bar{\hat{\phi}} =0.036$	6.598
Desv Estándar:	0.01536		1.996

Cuadro 4.27: Parámetro real vs Estimado; Tiempo de Computo

Como se puede observar en el cuadro (4.27), el modelo *CIAR* con estimación por filtro de Kalman, a diferencia de los modelos anteriores tuvo un tiempo de cómputo bastante mayor debido a su parámetro extra y que el cálculo de su verosimilitud es más complicada y extensa, sin embargo, la diferencia entre la estimación del parámetro y el parámetro real es muy poca. Tal como se pudo apreciar en la figura (4.14), el modelo propuesto se ajusta de buena manera a la serie simulada.

Con el objetivo de corroborar lo anteriormente planteado, se analiza nuevamente la bondad del ajuste a partir de los indicadores $-Loglik$, AIC , MSE y $MSE2$.

Modelo <i>CIAR</i> filtro de Kalman				
	-Loglik	AIC	MSE	MSE2
	0.5731	7.146	0.7199	1.127
	0.6201	7.24	0.7576	1.522
	0.565	7.13	0.7122	0.737
	0.7355	7.471	0.8339	0.7449
	0.5895	7.179	0.7663	1.598
	0.6809	7.362	0.795	1.334
	\vdots	\vdots	\vdots	\vdots
Medias:	0.6328	7.266	0.7687	1.067
Desv Estándar:	0.07109	0.1422	0.0474	0.3456

Cuadro 4.28: Bondad de Ajuste del Modelo *CIAR* filtro de Kalman

Confirmando lo planteado en los puntos anteriores, a partir del cuadro (4.28), se puede notar que el modelo *CIAR* con estimación por filtro de Kalman, tuvo tanto una media de $-Loglik$ y AIC de las simulaciones bastante baja. Además su MSE a diferencia de los demás modelos también es bastante pequeño. Por todo esto, es recomendable el uso del modelo *CIAR* con estimación por filtro de Kalman en series desigualmente espaciadas con ϕ negativo.

4.1.2.7. Comparación de Resultados para Datos Simulados por Modelo *CIAR* y Tiempos de Naturaleza Exponencial

A continuación una tabla resumen donde se comparan los datos entregados por los puntos anteriores:

	Media					Desv Estándar			
	Dif	TdC	-LL	AIC	SCE	TdC	-LL	AIC	SCE
Modelo AR(1)	0.906085	0.0057	424.6	853.1	0.9928	0.005881	0.9492	1.898	0.007277
Modelo CAR	0.905721	0.0045	211.6	427.3	0.9945	0.007703	1.327	2.655	0.006493
Modelo IAR.k	0.96402	0.1763	0.9479	5.896	0.9346	0.02769	0.02382	0.04763	0.0295
Modelo IAR.l	0.9638	0.0012	417.7	839.5	0.9347	0.00433	3.525	7.05	0.0293
Modelo IAR.t	0.96793	0.0777	1.873	7.747	0.9326	0.01523	0.08406	0.1681	0.0332
Modelo CIAR	0.0028	7.03	0.6322	7.264	0.77	2.992	0.07072	0.1414	0.04751

Cuadro 4.29: Comparación de Modelos simulados por *CIAR* y tiempos de naturaleza exponencial

Como se puede notar en el cuadro (4.29), todos los modelos propuestos exceptuando al modelo *CIAR* tuvieron un mal desempeño. Se puede notar principalmente en el indicador de bondad de ajuste *MSE*, esto se debe a que no tienen la capacidad de detectar valores negativos en la ACF.

Por otra parte y como se esperaba, el modelo *CIAR* con estimación por filtro de Kalman fue el único con un buen desempeño. Con poca diferencia entre el valor real del parámetro y el estimado. Es por esto que para esta simulación el modelo más óptimo a utilizar es el *CIAR* con estimación por filtro de Kalman.

4.1.2.8. Comparación de Resultados para Datos Simulados por Modelo *CIAR* y Tiempos de Naturaleza Uniforme

Análogamente se realiza el mismo procedimiento antes descrito, a partir de tiempos con naturaleza uniforme, con el objetivo de encontrar nuevas condiciones para el uso de los modelos. A continuación una tabla resumen donde se comparan los datos entregados según los puntos anteriores:

	Media					Desv Estándar			
	Dif	TdC	-LL	AIC	SCE	TdC	-LL	AIC	SCE
Modelo AR(1)	0.334	0.0035	364.2	732.4	0.6769	0.0073	24.93	49.86	0.05605
Modelo CAR	0.9	0.0014	212.1	428.2	0.9972	0.005322	$4,12e^{-14}$	$8,239e^{-14}$	0.003982
Modelo IAR.k	0.9001	0.275	0.9945	5.989	0.9929	0.0309	0.004341	0.008682	0.006639
Modelo IAR.l	0.9001	0.0058	424.9	853.8	0.9929	0.009447	0.9953	1.991	0.006639
Modelo IAR.t	0.9	0.1153	1.884	7.767	0.9944	0.08257	0.09493	0.1899	0.006631
Modelo CIAR	0.004714	6.524	-0.2212	5.558	0.3112	2.812	0.1491	0.2982	0.04718

Cuadro 4.30: Comparación de Modelos simulados por *CIAR* y tiempos de naturaleza exponencial

De que igual forma que en la comparación anterior, se puede notar en el cuadro (4.30) que todos los modelos propuestos tienen un mal desempeño con respecto a la estimación de parámetros exceptuando el modelo *CIAR*. Esto se ve reflejo en el *MSE* de cada uno.

Se repiten los resultados del cuadro anterior con el del modelo *CIAR* con estimación por filtro de Kalman como el mejor el mejor modelo para series desigualmente espaciadas con ϕ negativo.

4.2. Datos reales

Todo el análisis anterior se hizo en base a datos simulados, obteniendo fundamentos de como se comportan y el poder de estimación de cada modelo de acuerdo al contexto. A continuación, se trabajará con datos reales relacionados con la astronomía, la economía y la paleoclimatología.

La idea es ver qué modelo produce mejores resultados en cuanto a la estimación y predicción.

Una vez revisada y estudiada la base de datos, se procederá a ajustar los distintos modelos y además se utilizarán las medidas de desempeño $-Loglik$, *AIC*, *MSE* y *MSE2* para ver la precisión de las estimaciones y predicciones.

4.2.1. Base con datos Astronómicos

Esta base de datos tiene 105 observaciones y dos variables. La primera variable *newt*, tiene los registros temporales por año, medidos de forma irregular. La variable *newm* tiene la medición de la magnitud de brillo de un Blazar. Un Blazar es una fuente de energía muy compacta y altamente variable, asociada a un agujero negro situado en el centro de una galaxia. Los blazares están entre los fenómenos más violentos del universo y son un tema importante en la astronomía extragaláctica.

A continuación un gráfico de su comportamiento.

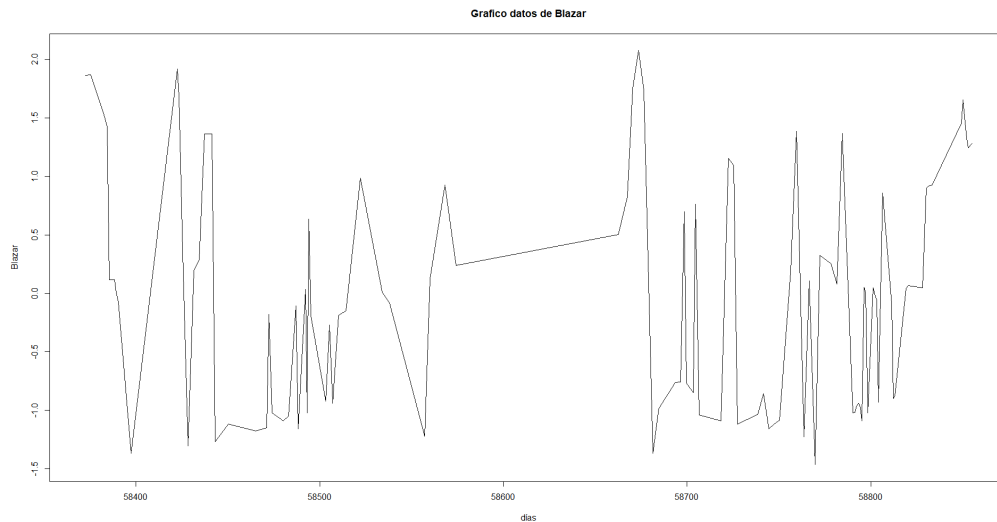


Figura 4.15: Gráfico de datos Blazar

En la siguiente tabla se puede observar la descripción de las variables donde *Blazar* es la variable *newm* estandarizada y *diff(st)* es la diferencia entre el valor de la variable *newt* en un tiempo i versus un tiempo $i + 1$.

Variable	Min	Q1	Mediana	Media	Q3	Max	Sd
Blazar	-1.4642	-0.9839	0.0077	0	0.8592	2.0750	1
diff(st)	0.87	1.04	2.97	4.64	4.26	88.29	9.179

Cuadro 4.31: Descripción de las variables

Al ver tanto la figura (4.15) como el cuadro (4.31), a priori se puede apreciar tanto una media como una dispersión de datos constante, esto quiere decir que la serie tiende a un comportamiento estacionario.

Con el objetivo de corroborar esta información se realiza el test de Cox-Stuart, se basa en variables aleatorias binomiales y permite estudiar la presencia de tendencias. Se tiene: $\{X_1, \dots, X_n\}$, conjunto de observaciones.

Se forman pares entre las observaciones:

- Si n es par: $\{X_1, X_{c+1}\}, \dots, \{X_c, X_n\}$, donde $c = \frac{n}{2}$
- Si n es impar: $\{X_1, X_{c+1}\}, \dots, \{X_{c-1}, X_n\}$, donde $c = \frac{n-1}{2}$

Estadístico: $T =$ número de signos positivos en la diferencia de $(X_{i+c} - X_i)$

Región Crítica: $\{m.a|T > \omega_{1-\alpha}\}$ donde $\alpha = 0,1$.

H_0 : La serie no presenta tendencia creciente vs H_1 : La serie presenta tendencia creciente

Los resultados fueron los siguientes:

	P-valor	Decisión
Cox-Stuart	0.398	No se Rechaza H_0

Cuadro 4.32: Análisis de Tendencia

Al ver que la serie no presenta tendencia, se procede a ajustar los modelos propuestos.

4.2.1.1. Ajuste por modelo AR

Tras comprobar que la series es apta para el estudio se procede a ajustar los datos a un modelo $AR(1)$, se utilizará la función *arma*, perteneciente a la librería *tseries*.

A continuación se muestra de forma gráfica el ajuste del modelo.

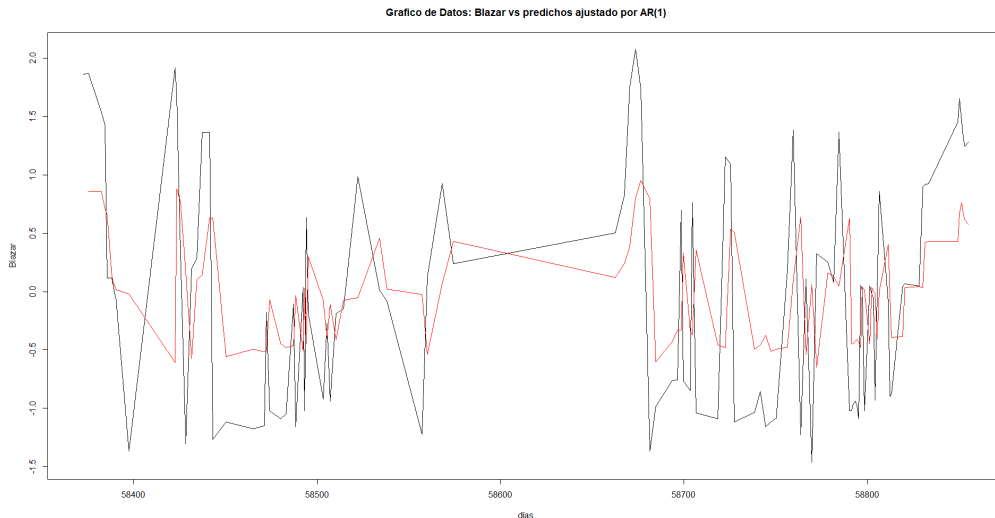


Figura 4.16: Gráfico de datos Blazar versus datos predichos por un modelo $AR(1)$, con datos irregularmente espaciados

De la figura (4.16) se aprecia en color negro los datos pertenecientes a la base Blazar y en rojo los datos estimados por el modelo $AR(1)$. A priori se puede observar que el modelo propuesto no tiene un gran ajuste, viendose reflejado principalmente donde los datos presentan una mayor amplitud, se asumía previamente que esto ocurriría sabiendo que el modelo AR no se ajusta bien a series de tiempo desigualmente espaciadas.

Con el objetivo de analizar el desempeño del modelo de forma más concreta se analiza tanto el tiempo de cómputo (TdC) como los indicadores de bondad de ajuste $-Loglik$, AIC , MSE y $MSE2$. Además para comprobar el supuesto de no autocorrelación en los residuos, se utiliza el test de Ljung-Box que tiene por hipótesis nula que los residuos no estan correlacionados.

Modelo $AR(1)$						
ϕ	TdC	-Loglik	AIC	MSE	MSE2	LB.p-value
0.4525	0	136.9893	277.9787	0.7757668	1.537	0.822

Cuadro 4.33: Bondad de Ajuste del Modelo $AR(1)$

Confirmando lo planteado a priori por la figura (4.16), a partir del cuadro (4.33), se puede notar que el modelo $AR(1)$ tuvo tanto un $-Loglik$ como AIC muy alta.

Por esto no sería recomendable el uso de un modelo $AR(1)$ para tiempo desigualmente espaciados.

4.2.1.2. Ajuste por modelo CAR

De igual forma se procede a ajustar los datos a un modelo $CAR(1)$, se utilizará la función *carma*, perteneciente a la librería *growth*.

A continuación se muestra de forma gráfica el ajuste del modelo.

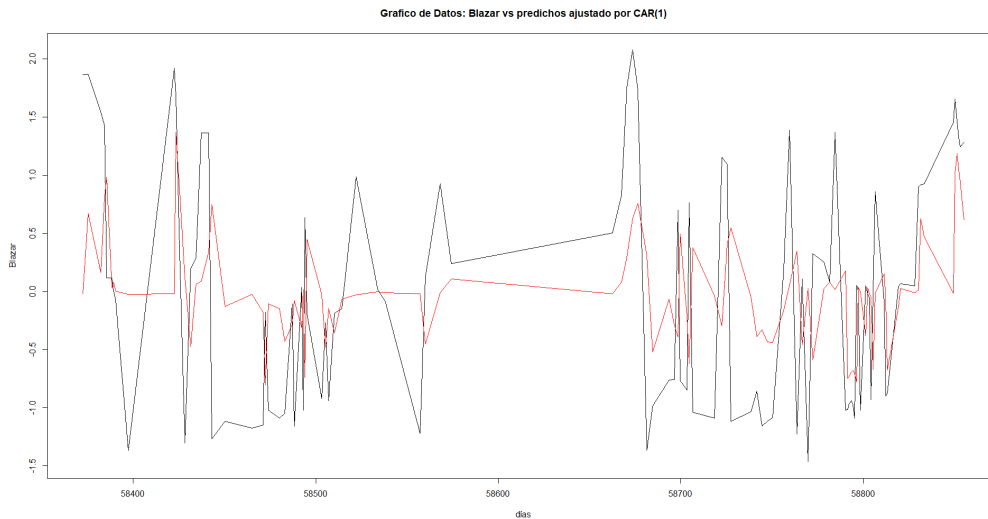


Figura 4.17: Gráfico de datos Blazar versus datos predichos por un modelo $CAR(1)$, con datos irregularmente espaciados

De la figura (4.17) se aprecia en color negro los datos pertenecientes a la base Blazar y en rojo los datos estimados por el modelo $CAR(1)$. A priori se puede observar que el modelo propuesto no tiene un gran ajuste, viendose reflejado principalmente donde los datos presentan una mayor amplitud.

Con el objetivo de analizar el desempeño del modelo de forma más concreta se analiza tanto el tiempo de cómputo (TdC) como los indicadores de bondad de ajuste $-Loglik$, AIC , MSE y $MSE2$. Además para comprobar el supuesto de no autocorrelación en los residuos, se utiliza el test de Ljung-Box.

Modelo $CAR(1)$						
ϕ	TdC	-Loglik	AIC	MSE	MSE2	LB.p-value
0.7155	0	67.4	138.8	0.8128	0.4679	0.7

Cuadro 4.34: Bondad de Ajuste del Modelo $CAR(1)$

Si bien a priori por la figura (4.17) no se ve un buen ajuste, a partir del cuadro (4.34), se puede notar que el modelo $CAR(1)$, tuvo tanto un $-Loglik$ como AIC bajo, lo que podría dar a entender que hubo un buen ajuste. Además fué de los pocos modelos que cumple con el supuesto de no autocorrelación de los residuos del modelo.

Por esto sería recomendable el uso de un modelo $CAR(1)$ para la serie *Blazar*.

4.2.1.3. Ajuste por modelo *IAR Kalman*

De igual forma se procede a ajustar los datos a un modelo *IAR(1)* con estimación por filtro de Kalman, se utilizará la función *IAR.kalman*, perteneciente a la librería *AITS*.

A continuación se muestra de forma gráfica el ajuste del modelo.

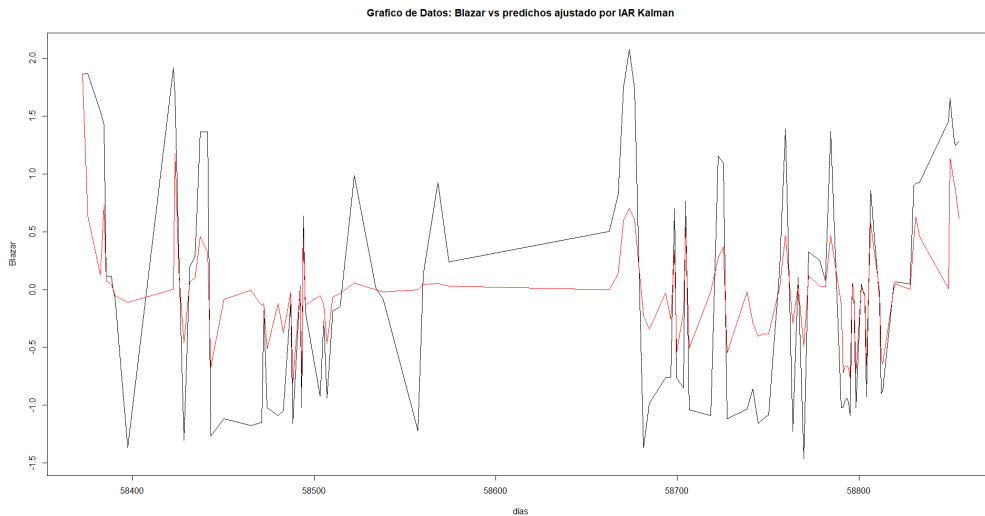


Figura 4.18: Gráfico de datos Blazar versus datos predichos por un modelo *IAR(1)* con estimación por filtro de Kalman, con datos irregularmente espaciados

De la figura (4.18) se aprecia en color negro los datos pertenecientes a la base Blazar y en rojo los datos estimados por el modelo *IAR(1)* con estimación por filtro de Kalman. A priori se puede observar que el modelo propuesto no tiene un gran ajuste, viéndose reflejado principalmente donde los datos presentan una mayor amplitud.

Con el objetivo de analizar el desempeño del modelo de forma más concreta se analiza tanto el tiempo de cómputo (*TdC*) como los indicadores de bondad de ajuste $-Loglik$, *AIC*, *MSE* y *MSE2*. Además para comprobar el supuesto de no autocorrelación en los residuos, se utiliza el test de Ljung-Box.

Modelo <i>IAR(1)</i> filtro de Kalman						
ϕ	TdC	-Loglik	AIC	MSE	MSE2	LB.p-value
0.6994	0.07	0.7465811	5.493162	0.4332851	0.4803	0.001

Cuadro 4.35: Bondad de Ajuste del Modelo *IAR(1)* filtro de Kalman

Si bien a priori por la figura (4.18) no se ve un buen ajuste, a partir del cuadro (4.35), se puede notar que el modelo *IAR(1)* con estimación por filtro de Kalman, tuvo un $-Loglik$, *AIC* y *MSE* bajos, lo que podría dar a entender que hubo un buen ajuste. Sin embargo, no cumple con el supuesto de no autocorrelación de los residuos

del modelo.

Por esto no sería recomendable el uso de un modelo $IAR(1)$ con estimación por filtro de Kalman para la serie *Blazar*.

4.2.1.4. Ajuste por modelo IAR loglik

De igual forma se procede a ajustar los datos a un modelo $IAR(1)$ con estimación por máxima verosimilitud directa, se utilizará la función $IAR.loglik$, perteneciente a la librería *AITS*.

A continuación se muestra de forma gráfica el ajuste del modelo.

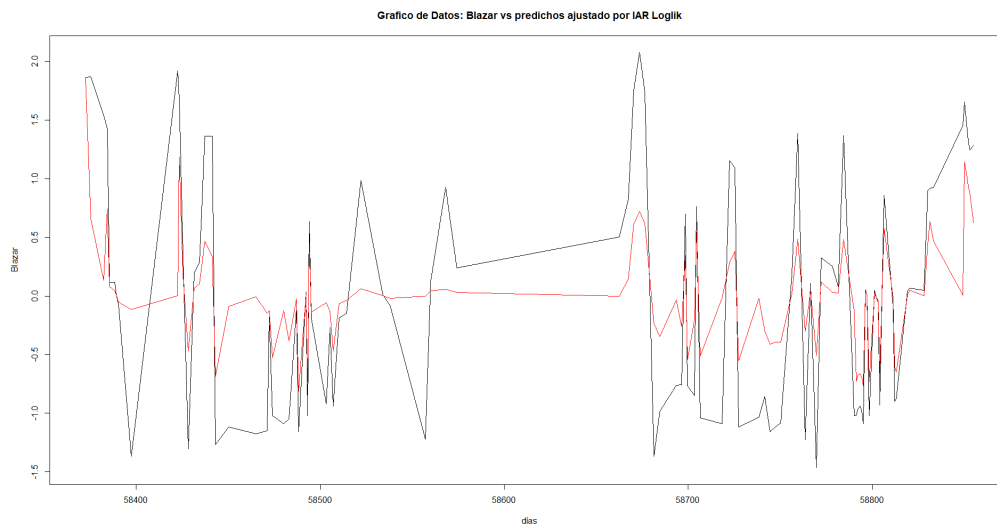


Figura 4.19: Gráfico de datos Blazar versus datos predichos por un modelo $IAR(1)$ con estimación por máxima verosimilitud directa, con datos irregularmente espaciados

De la figura (4.19) se aprecia en color negro los datos pertenecientes a la base Blazar y en rojo los datos estimados por el modelo $IAR(1)$ con estimación por máxima verosimilitud directa. A priori se puede observar que el modelo propuesto no tiene un gran ajuste, viendose reflejado principalmente donde los datos presentan una mayor amplitud.

Con el objetivo de analizar el desempeño del modelo de forma más concreta se analiza tanto el tiempo de cómputo (TdC) como los indicadores de bondad de ajuste $-Loglik$, AIC , MSE y $MSE2$. Además para comprobar el supuesto de no autocorrelación en los residuos, se utiliza el test de Ljung-Box.

Modelo $IAR(1)$ máxima verosimilitud						
ϕ	TdC	-Loglik	AIC	MSE	MSE2	LB.p-value
0.7051	0.02	134.1174	272.2349	0.4261956	0.4759	0.001

Cuadro 4.36: Bondad de Ajuste del Modelo $IAR(1)$ máxima verosimilitud

Confirmando lo planteado por la figura (4.19), a partir del cuadro (4.36), se puede notar que el modelo $IAR(1)$ con estimación por máxima verosimilitud directa, tuvo tanto un $-Loglik$ como AIC altos. Además no cumple con el supuesto de no autocorrelación de los residuos del modelo.

Por esto no sería recomendable el uso de un modelo $IAR(1)$ con estimación por máxima verosimilitud directa para la serie *Blazar*.

4.2.1.5. Ajuste por modelo $IAR - T$

De igual forma se procede a ajustar los datos a un modelo $IAR - T$, se utilizará la función $IAR.t$, perteneciente a la librería *AITS*.

A continuación se muestra de forma gráfica el ajuste del modelo.

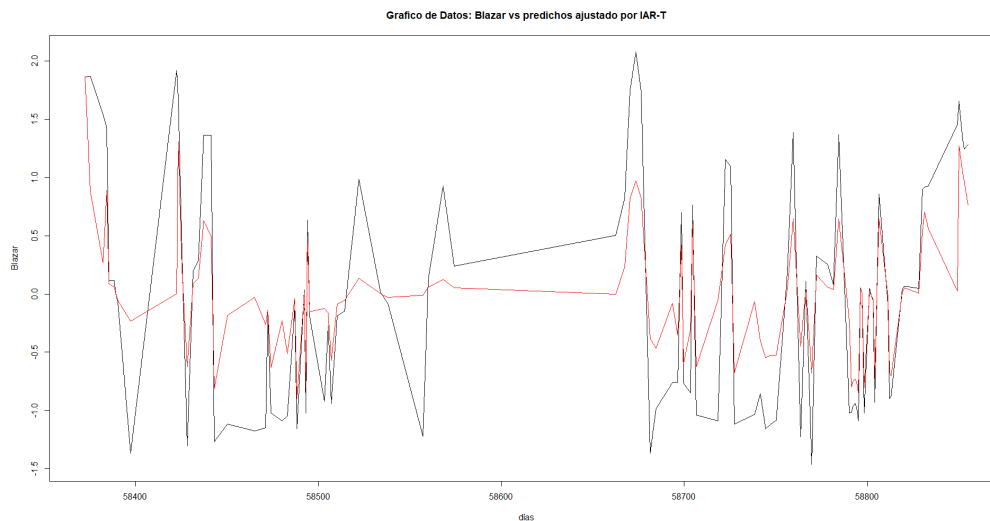


Figura 4.20: Gráfico de datos Blazar versus datos predichos por un modelo $IAR - T$, con datos irregularmente espaciados

De la figura (4.20) se aprecia en color negro los datos pertenecientes a la base Blazar y en rojo los datos estimados por el modelo $IAR - T$. A priori se puede observar que el modelo propuesto no tiene un gran ajuste, viéndose reflejado principalmente donde los datos presentan una mayor amplitud al igual que en los modelos IAR anteriores.

Con el objetivo de analizar el desempeño del modelo de forma más concreta se analiza tanto el tiempo de cómputo (TdC) como los indicadores de bondad de ajuste $-Loglik$, AIC , MSE y $MSE2$. Además para comprobar el supuesto de no autocorrelación en los residuos, se utiliza el test de Ljung-Box.

Modelo $IAR - T$						
ϕ	TdC	-Loglik	AIC	MSE	MSE2	LB.p-value
0.7777	0.05	143.8408	291.6816	0.3306454	0.4219	0.003091

Cuadro 4.37: Bondad de Ajuste del Modelo $IAR - T$

Confirmando lo planteado por la figura (4.20), a partir del cuadro (4.37), se puede notar que el modelo $IAR - T$, tuvo tanto un $-Loglik$ como AIC altos. Además no cumple con el supuesto de no autocorrelación de los residuos del modelo.

Por esto no sería recomendable el uso de un modelo $IAR - T$ para la serie *Blazar*.

4.2.1.6. Ajuste por modelo $CIAR$ Kalman

De igual forma se procede a ajustar los datos a un modelo $CIAR$ con estimación por filtro de Kalman, se utilizará la función $CIAR.kalman$, perteneciente a la librería *AIMS*.

A continuación se muestra de forma gráfica el ajuste del modelo.

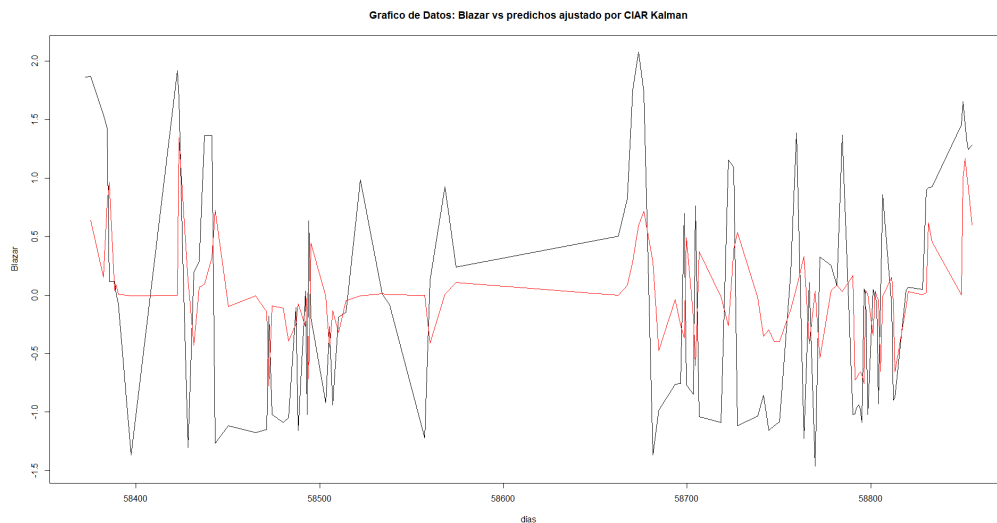


Figura 4.21: Gráfico de datos Blazar versus datos predichos por un modelo $CIAR$ con estimación por filtro de Kalman, con datos irregularmente espaciados

De la figura (4.21) se aprecia en color negro los datos pertenecientes a la base Blazar y en rojo los datos estima-

dos por el modelo *CIAR* con estimación por filtro de Kalman. A priori se puede observar que el modelo propuesto no tiene un gran ajuste, viendose reflejado principalmente donde los datos presentan una mayor amplitud al igual que en los modelos anteriores.

Con el objetivo de analizar el desempeño del modelo de forma más concreta se analiza tanto el tiempo de cómputo (*TdC*) como los indicadores de bondad de ajuste $-Loglik$, *AIC*, *MSE* y *MSE2*. Además para comprobar el supuesto de no autocorrelación en los residuos, se utiliza el test de Ljung-Box.

Modelo <i>CIAR</i> filtro de Kalman						
ϕ	TdC	-Loglik	AIC	MSE	MSE2	LB.p-value
0.6994	1.10	0.7465811	7.493162	0.788314	0.5832	0.8

Cuadro 4.38: Bondad de Ajuste del Modelo *CIAR* filtro de Kalman

Si bien por la figura (4.21) no se ve un buen ajuste, a partir del cuadro (4.38), se puede notar que el modelo *CIAR* con estimación por filtro de Kalman, tuvo tanto un $-Loglik$ como *AIC* bajos, lo que podría dar a entender que hubo un buen ajuste. Además fué de los pocos modelos que cumple con el supuesto de no autocorrelación de los residuos del modelo.

Por esto sería recomendable el uso de un modelo *CIAR* con estimación por filtro de Kalman para la serie *Blazar*.

4.2.2. Índice de Capitalización

La segunda base de datos llamada *INDICE_ICAP* tiene 770 observaciones y dos variables. La primera variable *FECHA*, tiene los registros temporales por días medidos de forma continua e irregular. La variable *VALOR* tiene la medición del índice de capitalización de Colombia. Este índice se refiere al porcentaje de capital neto que mantiene cada banco.

A continuación un gráfico de su comportamiento.

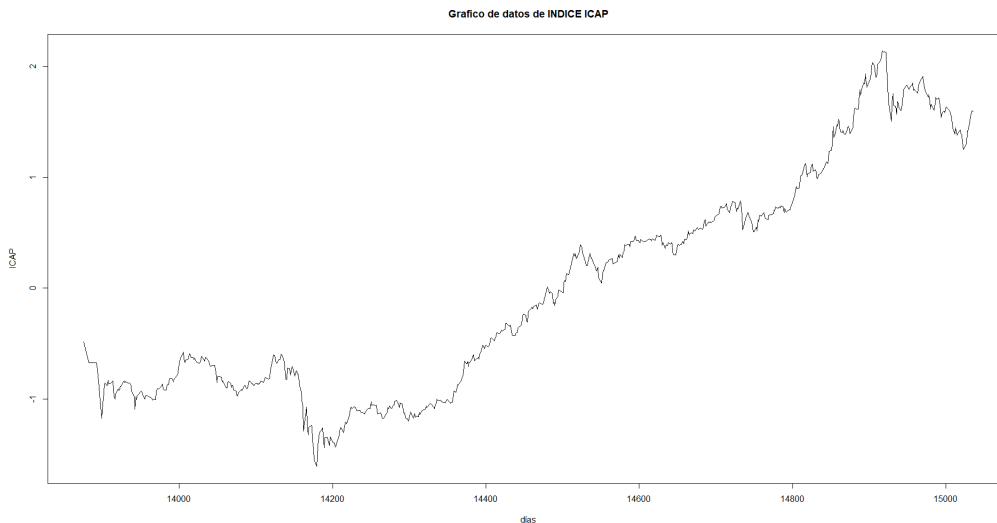


Figura 4.22: Grafico de datos de INDICE ICAP

Al ver la figura (4.22), a priori se puede apreciar una tendencia positiva, esto quiere decir que la serie tiende a un comportamiento no estacionario. Con el objetivo de corroborar esta información se realiza el test de Cox-Stuart para tendencia.

Como se mencionó anteriormente, el test de Cox-Stuart se basa en variables aleatorias binomiales y permite estudiar la presencia de tendencias.

Los resultados fueron los siguientes:

	P-valor	Decisión
Cox-Stuart	$< 2e^{-16}$	Se Rechaza H_0

Cuadro 4.39: Análisis de Tendencia

Al ver que la serie presenta tendencia, se procede a quitar la tendencia de los datos a partir del suavizado *lowess*[24, Cleveland (1981)], se implementó antes de realizar el ajuste a cada modelo con un valor $f = 0,1$ (Esto da la proporción de puntos en la gráfica que influyen en la suavidad en cada valor. Los valores más altos dan más suavidad), quedando de la siguiente forma:

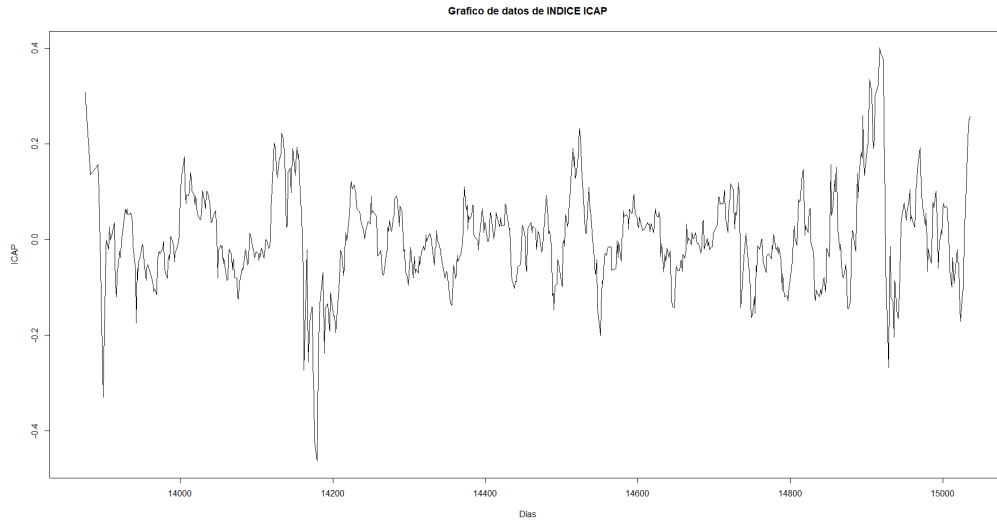


Figura 4.23: Grafico de datos de INDICE ICAP

Al ver la figura (4.23), a priori se puede apreciar que desaparece la tendencia positiva. Con el objetivo de corroborar esta información se realiza nuevamente el test de Cox-Stuart para tendencia. Los resultados fueron los siguientes:

	P-valor	Decisión
Cox-Stuart	0.1	No se Rechaza H_0

Cuadro 4.40: Análisis de Tendencia

En la siguiente tabla se puede observar la descripción de las variables donde $y1$ es la variable *VALOR* estandarizada y $diff(st)$ es la diferencia entre el valor de la variable *FECHA* en un tiempo i versus un tiempo $i + 1$ posterior al retiro de tendencia.

Variable	Min	Q1	Mediana	Media	Q3	Max	Sd
y1	-0.4634	-0.0519	-0.0015	0.0035	0.0551	0.4003	0.09981
diff(st)	1	1	1	1.51	1	7	1.001

Cuadro 4.41: Descripción de las variables

Tanto la media como la dispersión de los datos es constante, esto quiere decir que la serie tiende a un comportamiento estacionario y al ver que la serie efectivamente no presenta tendencia, se procede a ajustar los modelos propuestos.

4.2.2.1. Ajuste por modelo *AR*

Tras comprobar que la series es apta para el estudio se procede a ajustar los datos a un modelo *AR(1)*, se utilizará la función *arma*, perteneciente a la librería *tseries*.

A continuación se muestra de forma gráfica el ajuste del modelo.

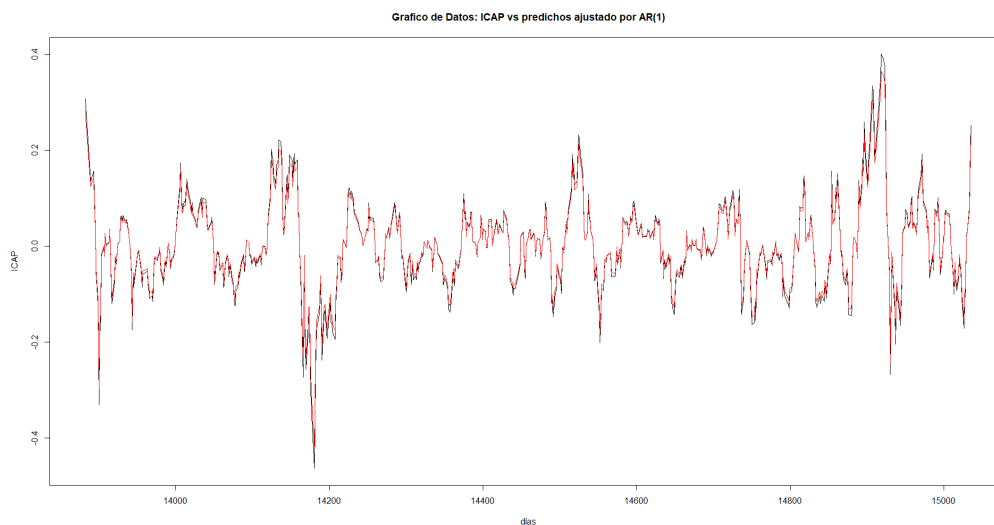


Figura 4.24: Gráfico de datos ICAP versus datos predichos por un modelo $AR(1)$, con datos irregularmente espaciados

De la figura (4.24) se aprecia en color negro los datos pertenecientes a la base *INDICE_ICAP* y en rojo los datos estimados por el modelo $AR(1)$. A priori se puede observar que el modelo propuesto tiene un buen ajuste.

Con el objetivo de analizar el desempeño del modelo de forma más concreta se analiza tanto el tiempo de cómputo (TdC) como los indicadores de bondad de ajuste $-Loglik$, AIC , MSE y $MSE2$. Además para comprobar el supuesto de no autocorrelación en los residuos, se utiliza el test de Ljung-Box que tiene por hipótesis nula que los residuos no están correlacionados.

Modelo $AR(1)$					
ϕ	TdC	$-Loglik$	AIC	MSE	LB.p-value
0.9093	0	-1325	-2646	$8,173e^{-05}$	0.02

Cuadro 4.42: Bondad de Ajuste del Modelo $AR(1)$

A partir del cuadro (4.42), se puede notar que el modelo $AR(1)$ tuvo un $-Loglik$, AIC y MSE muy bajo. Sin embargo, no cumple con el supuesto de no autocorrelación de los residuos del modelo.

Por esto no sería recomendable el uso de un modelo $AR(1)$ para la serie *INDICE_ICAP*.

4.2.2.2. Ajuste por modelo CAR

Tras comprobar que la series es apta para el estudio se procede a ajustar los datos a un modelo $CAR(1)$, se utilizará la función *carma*, perteneciente a la librería *growth*.

A continuación se muestra de forma gráfica el ajuste del modelo.

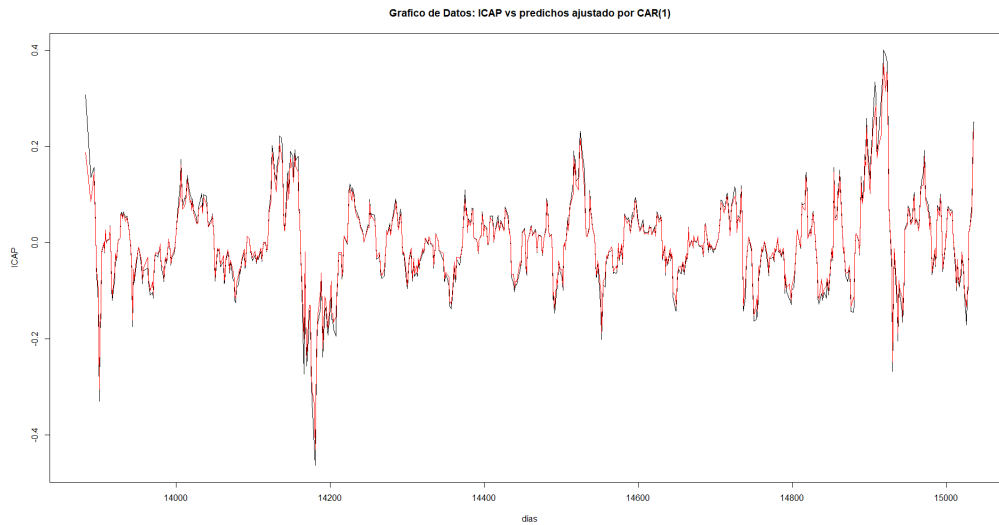


Figura 4.25: Gráfico de datos ICAP versus datos predichos por un modelo $CAR(1)$, con datos irregularmente espaciados

De la figura (4.25) se aprecia en color negro los datos pertenecientes a la base *INDICE_ICAP* y en rojo los datos estimados por el modelo $CAR(1)$. A priori se puede observar que el modelo propuesto tiene un buen ajuste.

Con el objetivo de analizar el desempeño del modelo de forma más concreta se analiza tanto el tiempo de cómputo (TdC) como los indicadores de bondad de ajuste $-Loglik$, AIC , MSE y $MSE2$. Además para comprobar el supuesto de no autocorrelación en los residuos, se utiliza el test de Ljung-Box que tiene por hipótesis nula que los residuos no estan correlacionados.

Modelo $CAR(1)$					
ϕ	TdC	-Loglik	AIC	MSE	LB.p-value
0.9285	0.02	-656.2	-1308	0.0001579	$< 2e^{-16}$

Cuadro 4.43: Bondad de Ajuste del Modelo $CAR(1)$

A partir del cuadro (4.43), se puede notar que el modelo $CAR(1)$ tuvo un $-Loglik$, AIC y MSE muy bajo. Sin embargo, no cumple con el supuesto de no autocorrelación de los residuos del modelo.

Por esto no sería recomendable el uso de un modelo $CAR(1)$ para la serie $INDICE_ICAP$.

4.2.2.3. Ajuste por modelo IAR Kalman

De igual forma se procede a ajustar los datos a un modelo $IAR(1)$ con estimación por filtro de Kalman, se utilizará la función $IAR.kalman$, perteneciente a la librería $AITS$.

A continuación se muestra de forma gráfica el ajuste del modelo.

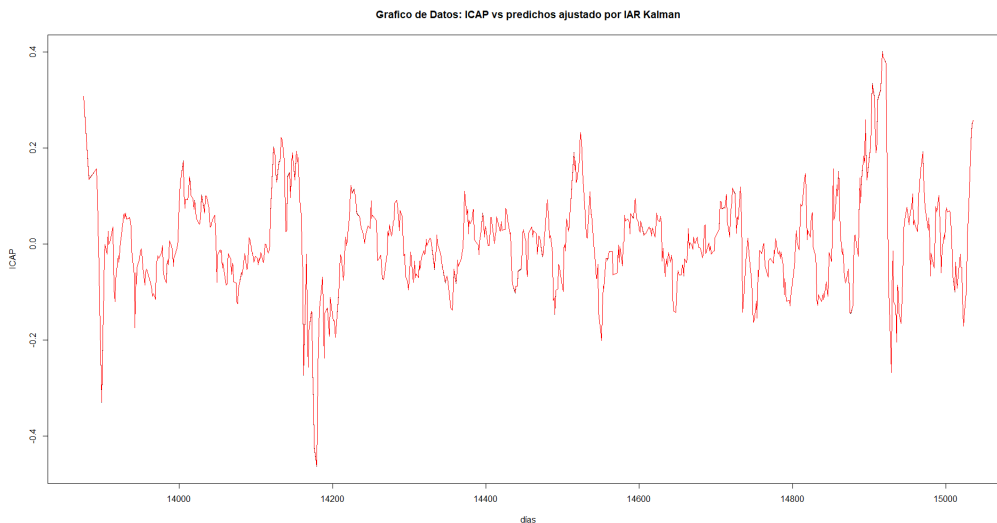


Figura 4.26: Gráfico de datos $ICAP$ versus datos predichos por un modelo $IAR(1)$ con estimación por filtro de Kalman, con datos irregularmente espaciados

De la figura (4.26) se aprecia en color negro los datos pertenecientes a la base $INDICE_ICAP$ y en rojo los datos estimados por el modelo $IAR(1)$ con estimación por filtro de Kalman. A priori se puede observar que el modelo propuesto tiene un excelente ajuste.

Con el objetivo de analizar el desempeño del modelo de forma más concreta se analiza tanto el tiempo de cómputo (TdC) como los indicadores de bondad de ajuste $-Loglik$, AIC , MSE y $MSE2$. Además para comprobar el supuesto de no autocorrelación en los residuos, se utiliza el test de Ljung-Box.

Modelo $IAR(1)$ filtro de Kalman					
ϕ	TdC	-Loglik	AIC	MSE	LB.p-value
0.9992	0.69	-5.182	-6.364	$2,141e^{-08}$	$< 2e^{-16}$

Cuadro 4.44: Bondad de Ajuste del Modelo $IAR(1)$ filtro de Kalman

Confirmando lo planteado a priori por la figura (4.26), a partir del cuadro (4.44), se puede notar que el modelo

$IAR(1)$ con estimación por filtro de Kalman, tuvo un $-Loglik$, AIC y MSE muy bajo. Sin embargo no cumple con el supuesto de no autocorrelación de los residuos del modelo.

Por esto no sería recomendable el uso de un modelo $IAR(1)$ con estimación por filtro de Kalman para la serie $INDICE_ICAP$.

4.2.2.4. Ajuste por modelo IAR $loglik$

De igual forma se procede a ajustar los datos a un modelo $IAR(1)$ con estimación por máxima verosimilitud directa, se utilizará la función $IAR.loglik$, perteneciente a la librería $AITS$.

A continuación se muestra de forma gráfica el ajuste del modelo.

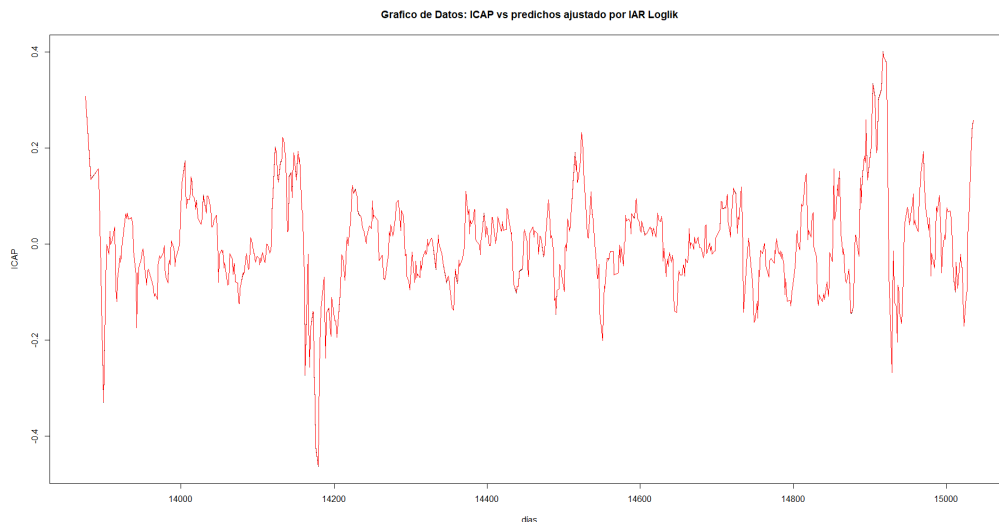


Figura 4.27: Gráfico de datos ICAP versus datos predichos por un modelo $IAR(1)$ con estimación por máxima verosimilitud directa, con datos irregularmente espaciados

De la figura (4.27) se aprecia en color negro los datos pertenecientes a la base $INDICE_ICAP$ y en rojo los datos estimados por el modelo $IAR(1)$ con estimación por máxima verosimilitud directa. A priori se puede observar que el modelo propuesto tiene un excelente ajuste.

Con el objetivo de analizar el desempeño del modelo de forma más concreta se analiza tanto el tiempo de cómputo (TdC) como los indicadores de bondad de ajuste $-Loglik$, AIC , MSE y $MSE2$. Además para comprobar el supuesto de no autocorrelación en los residuos, se utiliza el test de Ljung-Box.

Modelo $IAR(1)$ máxima verosimilitud					
ϕ	TdC	-Loglik	AIC	MSE	LB.p-value
0.9992	0.01	-1291	-2577	$2,141e^{-08}$	$< 2e^{-16}$

Cuadro 4.45: Bondad de Ajuste del Modelo $IAR(1)$ máxima verosimilitud

Confirmando lo planteado a priori por la figura (4.27), a partir del cuadro (4.45), se puede notar que el modelo $IAR(1)$ con estimación por máxima verosimilitud directa, tuvo un $-Loglik$, AIC y MSE muy bajo. Sin embargo no cumple con el supuesto de no autocorrelación de los residuos del modelo.

Por esto no sería recomendable el uso de un modelo $IAR(1)$ con estimación por máxima verosimilitud directa para la serie $INDICE_ICAP$.

4.2.2.5. Ajuste por modelo $IAR - T$

De igual forma se procede a ajustar los datos a un modelo $IAR - T$, se utilizará la función $IAR.t$, perteneciente a la librería $AITS$.

A continuación se muestra de forma gráfica el ajuste del modelo.

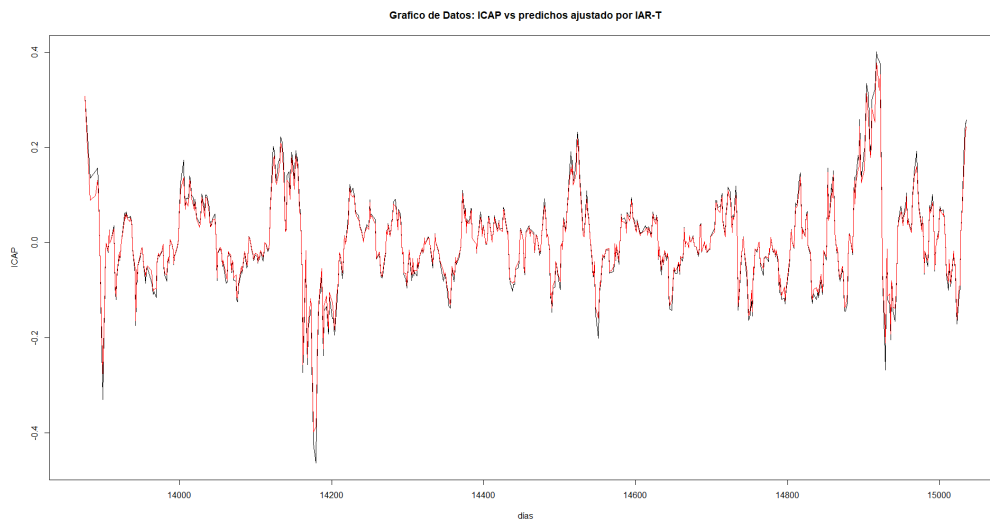


Figura 4.28: Gráfico de datos ICAP versus datos predichos por un modelo $IAR - T$, con datos irregularmente espaciados

De la figura (4.28) se aprecia en color negro los datos pertenecientes a la base $INDICE_ICAP$ y en rojo los datos estimados por el modelo $IAR - T$. A priori se puede observar que el modelo propuesto tiene un buen ajuste.

Con el objetivo de analizar el desempeño del modelo de forma más concreta se analiza tanto el tiempo de

cómputo (TdC) como los indicadores de bondad de ajuste $-Loglik$, AIC , MSE y $MSE2$. Además para comprobar el supuesto de no autocorrelación en los residuos, se utiliza el test de Ljung-Box.

Modelo $IAR - T$					
ϕ	TdC	-Loglik	AIC	MSE	LB.p-value
0.9418	0.2	-1366	-2728	0.0001076	$< 2e^{-16}$

Cuadro 4.46: Bondad de Ajuste del Modelo $IAR - T$

Confirmando lo planteado a priori por la figura (4.28), a partir del cuadro (4.46), se puede notar que el modelo $IAR - T$, tuvo un $-Loglik$, AIC y MSE muy bajo. Sin embargo no cumple con el supuesto de no autocorrelación de los residuos del modelo.

Por esto no sería recomendable el uso de un modelo $IAR - T$ para la serie $INDICE_{ICAP}$.

4.2.2.6. Ajuste por modelo $CIAR$ Kalman

De igual forma se procede a ajustar los datos a un modelo $CIAR$ con estimación por filtro de Kalman, se utilizará la función $CIAR.kalman$, perteneciente a la librería $AITS$.

A continuación se muestra de forma gráfica el ajuste del modelo.

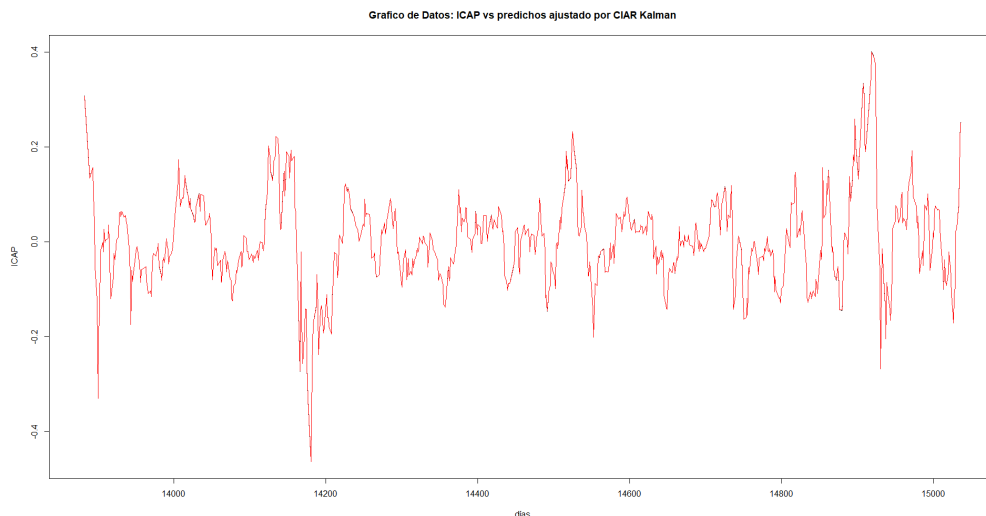


Figura 4.29: Gráfico de datos ICAP versus datos predichos por un modelo $CIAR$ con estimación por filtro de Kalman, con datos irregularmente espaciados

De la figura (4.29) se aprecia en color negro los datos pertenecientes a la base $INDICE_{ICAP}$ y en rojo los datos estimados por el modelo $CIAR$ con estimación por filtro de Kalman. A priori se puede observar que el

modelo propuesto tiene un buen ajuste.

Con el objetivo de analizar el desempeño del modelo de forma más concreta se analiza tanto el tiempo de cómputo (TdC) como los indicadores de bondad de ajuste $-Loglik$, AIC , MSE y $MSE2$. Además para comprobar el supuesto de no autocorrelación en los residuos, se utiliza el test de Ljung-Box.

Modelo $CIAR$ filtro de Kalman					
ϕ	TdC	-Loglik	AIC	MSE	LB.p-value
0.9992	37.14	-5.182	-4.364	$2,277e^{-08}$	$< 2e^{-16}$

Cuadro 4.47: Bondad de Ajuste del Modelo $CIAR$ filtro de Kalman

Confirmando lo planteado a priori por la figura (4.29), a partir del cuadro (4.47), se puede notar que el modelo $CIAR$ con estimación por filtro de Kalman, tuvo un $-Loglik$, AIC y MSE muy bajo. Sin embargo no cumple con el supuesto de no autocorrelación de los residuos del modelo.

Por esto no sería recomendable el uso de un modelo $CIAR$ con estimación por filtro de Kalman para la serie $INDICEICAP$.

4.2.3. Proporción de Isótopos

La tercera base de datos llamada 4.7 tiene 1345 observaciones y dos variables. La primera variable $X1$, tiene los registros temporales por años medidos de forma continua e irregular. La variable $X2$ tiene la medición de la proporción del isótopo pesado frente al ligero ($\delta^{18}O$), de las lluvias monzónicas del Océano Índico durante el Holoceno

A continuación un gráfico de su comportamiento.

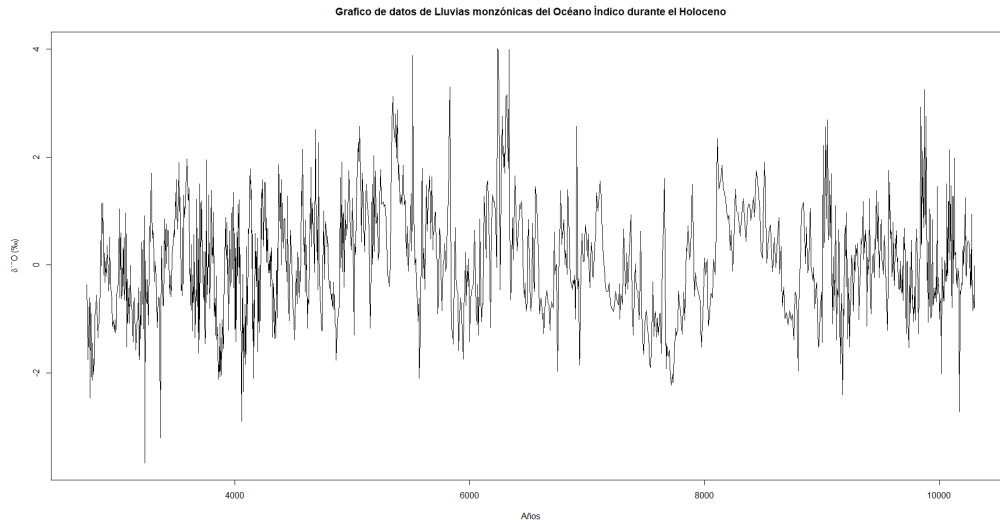


Figura 4.30: Grafico de datos 4.7

Al ver la figura (4.30), a priori se puede apreciar una tendencia positiva, esto quiere decir que la serie tiende a un comportamiento no estacionario.

En la siguiente tabla se puede observar la descripción de las variables donde $y1$ es la variable $X2$ estandarizada y $diff(st)$ es la diferencia entre el valor de la variable $X1$ en un tiempo i versus un tiempo $i + 1$.

Variable	Min	Q1	Mediana	Media	Q3	Max	Sd
y1	-3.675	-0.689	-0.051	0	0.638	4.006	1
diff(st)	0.262	3.550	5.133	5.625	7.186	19.599	2.732

Cuadro 4.48: Descripción de las variables

Con el objetivo de corroborar esta información se realiza el test de Cox-Stuart para tendencia, se basa en variables aleatorias binomiales y permite estudiar la presencia de tendencias.

Los resultados fueron los siguientes:

	P-valor	Decisión
Cox-Stuart	0.04	No se Rechaza H_0

Cuadro 4.49: Análisis de Tendencia

Al ver que la serie con un 1 % de significancia no presenta tendencia, se procede a ajustar los modelos propuestos.

4.2.3.1. Ajuste por modelo AR

Tras comprobar que la series es apta para el estudio se procede a ajustar los datos a un modelo $AR(1)$, se utilizará la función *arma*, perteneciente a la librería *tseries*.

A continuación se muestra de forma gráfica el ajuste del modelo.

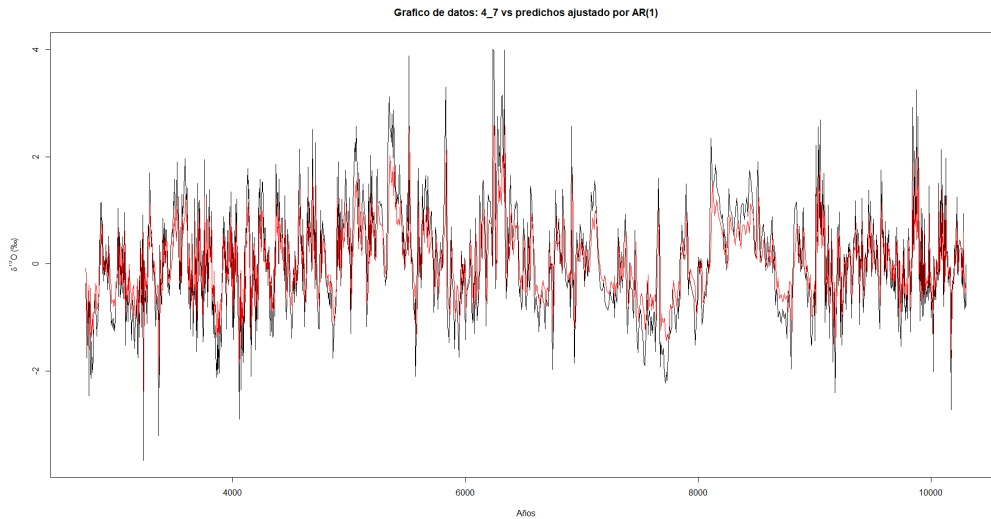


Figura 4.31: Gráfico de datos 4_7 versus datos predichos por un modelo $AR(1)$, con datos irregularmente espaciados

De la figura (4.31) se aprecia en color negro los datos pertenecientes a la base 4_7 y en rojo los datos estimados por el modelo $AR(1)$. A priori se puede observar que el modelo propuesto tiene un buen ajuste, viéndose afectado principalmente donde los datos presentan una mayor amplitud, se asumía previamente que esto ocurriría sabiendo que el modelo AR no se ajusta bien a series de tiempo desigualmente espaciadas.

Con el objetivo de analizar el desempeño del modelo de forma más concreta se analiza tanto el tiempo de cómputo (TdC) como los indicadores de bondad de ajuste $-Loglik$, AIC , MSE y $MSE2$. Además para comprobar el supuesto de no autocorrelación en los residuos, se utiliza el test de Ljung-Box que tiene por hipótesis nula que los residuos no están correlacionados.

Modelo $AR(1)$						
ϕ	TdC	-Loglik	AIC	MSE	MSE2	LB.p-value
0.6458	0.01	1545	3093	0.5819	0.7951	$4e^{-06}$

Cuadro 4.50: Bondad de Ajuste del Modelo $AR(1)$

Confirmando lo planteado a priori por la figura (4.31), a partir del cuadro (4.50), se puede notar que el modelo $AR(1)$ tuvo tanto un $-Loglik$ como AIC muy altos, además de no cumplir el supuesto de no autocorrelación.

Por esto no sería recomendable el uso de un modelo $AR(1)$ para la serie 4_7.

4.2.3.2. Ajuste por modelo CAR

De igual forma se procede a ajustar los datos a un modelo $CAR(1)$, se utilizará la función *carma*, perteneciente a la librería *growth*.

A continuación se muestra de forma gráfica el ajuste del modelo.

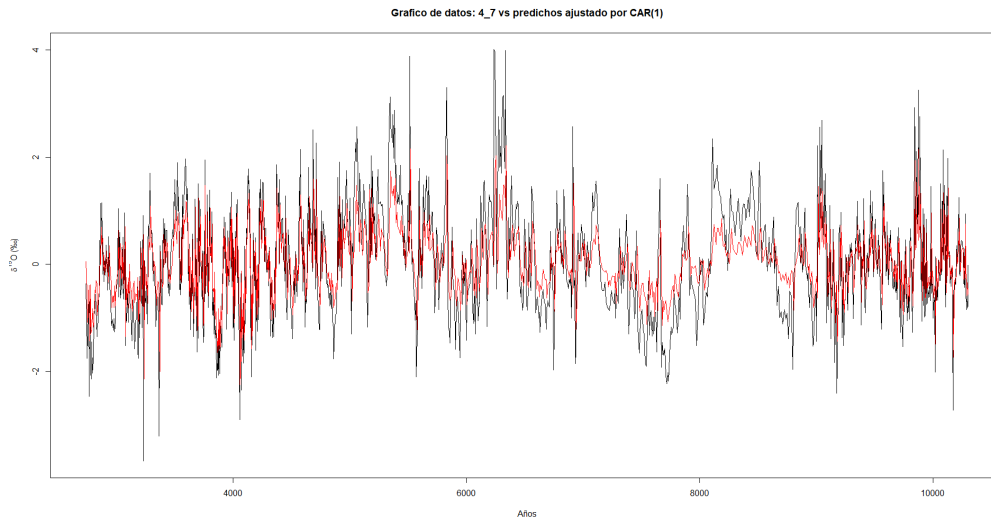


Figura 4.32: Gráfico de datos 4.7 versus datos predichos por un modelo $CAR(1)$, con datos irregularmente espaciados

De la figura (4.32) se aprecia en color negro los datos pertenecientes a la base 4.7 y en rojo los datos estimados por el modelo $CAR(1)$. A priori se puede observar que el modelo propuesto tiene un buen ajuste.

Con el objetivo de analizar el desempeño del modelo de forma más concreta se analiza tanto el tiempo de cómputo (TdC) como los indicadores de bondad de ajuste $-Loglik$, AIC , MSE y $MSE2$. Además para comprobar el supuesto de no autocorrelación en los residuos, se utiliza el test de Ljung-Box.

Modelo $CAR(1)$						
ϕ	TdC	-Loglik	AIC	MSE	MSE2	LB.p-value
0.9091	0	802.2	1608	0.6095	1.015	0.09

Cuadro 4.51: Bondad de Ajuste del Modelo $CAR(1)$

Confirmando lo planteado a priori por la figura (4.32), a partir del cuadro (4.51), se puede notar que el modelo $CAR(1)$, tuvo un $-Loglik$, AIC y MSE altos. Sin embargo es de los pocos modelos que cumple con el supuesto de no autocorrelación de los residuos del modelo.

Por esto es recomendable el uso de un modelo $CAR(1)$ para la serie 4.7.

4.2.3.3. Ajuste por modelo IAR Kalman

De igual forma se procede a ajustar los datos a un modelo $IAR(1)$ con estimación por filtro de Kalman, se utilizará la función $IAR.kalman$, perteneciente a la librería $AITS$.

A continuación se muestra de forma gráfica el ajuste del modelo.

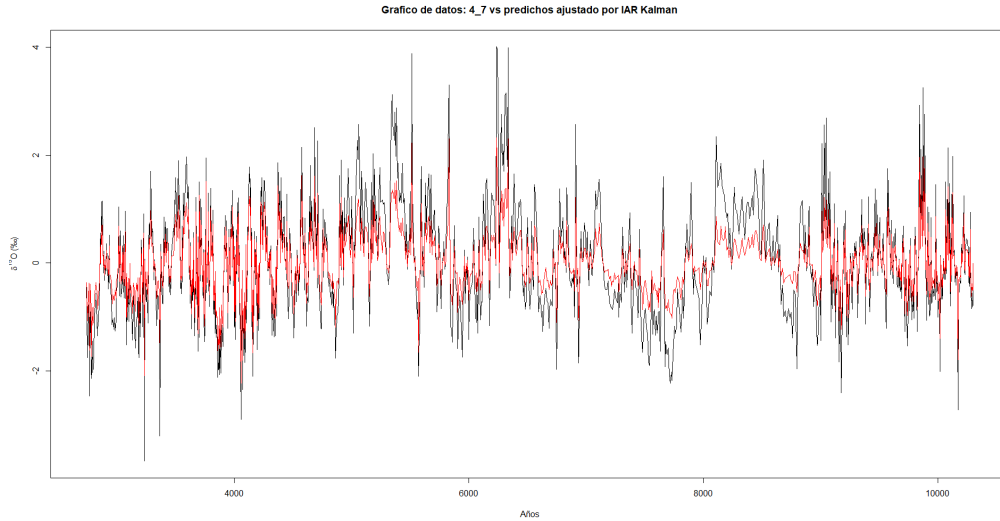


Figura 4.33: Gráfico de datos 4.7 versus datos predichos por un modelo $IAR(1)$ con estimación por filtro de Kalman, con datos irregularmente espaciados

De la figura (4.33) se aprecia en color negro los datos pertenecientes a la base 4.7 y en rojo los datos estimados por el modelo $IAR(1)$ con estimación por filtro de Kalman. A priori se puede observar que el modelo propuesto tiene un buen ajuste.

Con el objetivo de analizar el desempeño del modelo de forma más concreta se analiza tanto el tiempo de cómputo (TdC) como los indicadores de bondad de ajuste $-Loglik$, AIC , MSE y $MSE2$. Además para comprobar el supuesto de no autocorrelación en los residuos, se utiliza el test de Ljung-Box.

Modelo $IAR(1)$ filtro de Kalman						
ϕ	TdC	-Loglik	AIC	MSE	MSE2	LB.p-value
0.9034	0.9	0.5514	5.103	0.2014	0.7652	$< 2e^{-16}$

Cuadro 4.52: Bondad de Ajuste del Modelo $IAR(1)$ filtro de Kalman

Confirmando lo planteado a priori por la figura (4.33), a partir del cuadro (4.52), se puede notar que el modelo

$IAR(1)$ con estimación por filtro de Kalman, tuvo un $-Loglik$, AIC y MSE muy bajos. Sin embargo no cumple con el supuesto de no autocorrelación de los residuos del modelo

Por esto no sería recomendable el uso de un modelo $IAR(1)$ con estimación por filtro de Kalman para la serie 4.7.

4.2.3.4. Ajuste por modelo IAR $loglik$

De igual forma se procede a ajustar los datos a un modelo $IAR(1)$ con estimación por máxima verosimilitud directa, se utilizará la función $IAR.loglik$, perteneciente a la librería $AITS$.

A continuación se muestra de forma gráfica el ajuste del modelo.

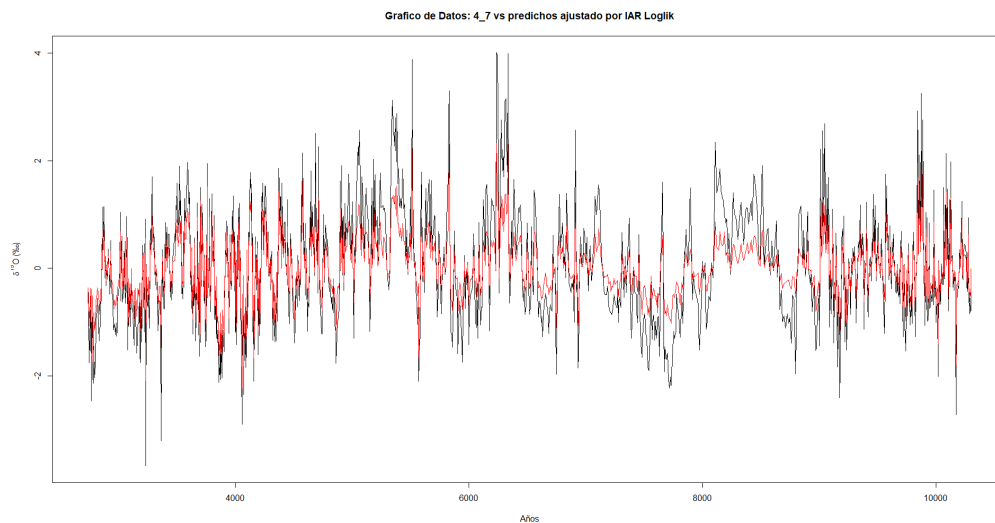


Figura 4.34: Gráfico de datos 4.7 versus datos predichos por un modelo $IAR(1)$ con estimación por máxima verosimilitud directa, con datos irregularmente espaciados

De la figura (4.34) se aprecia en color negro los datos pertenecientes a la base 4.7 y en rojo los datos estimados por el modelo $IAR(1)$ con estimación por máxima verosimilitud directa. A priori se puede observar que el modelo propuesto tiene un buen ajuste.

Con el objetivo de analizar el desempeño del modelo de forma más concreta se analiza tanto el tiempo de cómputo (TdC) como los indicadores de bondad de ajuste $-Loglik$, AIC , MSE y $MSE2$. Además para comprobar el supuesto de no autocorrelación en los residuos, se utiliza el test de Ljung-Box.

Modelo $IAR(1)$ máxima verosimilitud						
ϕ	TdC	-Loglik	AIC	MSE	MSE2	LB.p-value
0.9034	0.02	1607	3217	0.2015	0.7653	$< 2e^{-16}$

Cuadro 4.53: Bondad de Ajuste del Modelo $IAR(1)$ máxima verosimilitud

Confirmando lo planteado a priori por la figura (4.34), a partir del cuadro (4.53), se puede notar que el modelo $IAR(1)$ con estimación por máxima verosimilitud directa, tuvo tanto un $-Loglik$ como AIC altos. Además no cumple con el supuesto de no autocorrelación de los residuos del modelo.

Por esto no sería recomendable el uso de un modelo $IAR(1)$ con estimación por máxima verosimilitud directa para la serie 4_7.

4.2.3.5. Ajuste por modelo $IAR - T$

De igual forma se procede a ajustar los datos a un modelo $IAR - T$, se utilizará la función $IAR.t$, perteneciente a la librería $AITS$.

A continuación se muestra de forma gráfica el ajuste del modelo.

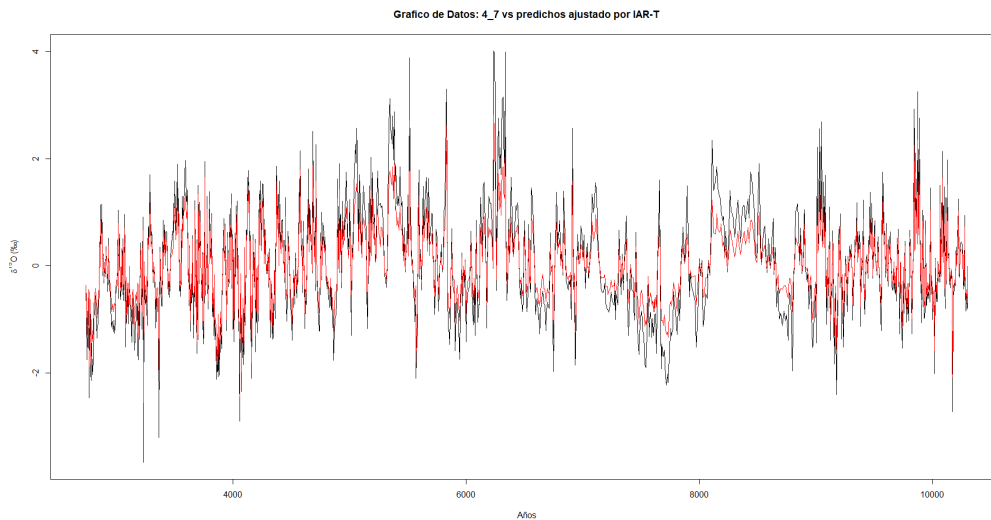


Figura 4.35: Gráfico de datos 4_7 versus datos predichos por un modelo $IAR - T$, con datos irregularmente espaciados

De la figura (4.35) se aprecia en color negro los datos pertenecientes a la base 4_7 y en rojo los datos estimados por el modelo $IAR - T$. A priori se puede observar que el modelo propuesto tiene un ajuste bueno aunque inferior al de los demás IAR .

Con el objetivo de analizar el desempeño del modelo de forma más concreta se analiza tanto el tiempo de cómputo (TdC) como los indicadores de bondad de ajuste $-Loglik$, AIC , MSE y $MSE2$. Además para comprobar el supuesto de no autocorrelación en los residuos, se utiliza el test de Ljung-Box.

Modelo $IAR - T$						
ϕ	TdC	-Loglik	AIC	MSE	MSE2	LB.p-value
0.9365	0.52	1559	3122	0.1068	0.8255	$< 2e^{-16}$

Cuadro 4.54: Bondad de Ajuste del Modelo $IAR - T$

Confirmando lo planteado a priori por la figura (4.35), a partir del cuadro (4.54), se puede notar que el modelo $IAR - T$, tuvo un $-Loglik$, AIC y MSE más alto que sus símiles IAR . Además no cumple con el supuesto de no autocorrelación de los residuos del modelo.

Por esto no sería recomendable el uso de un modelo $IAR - T$ para la serie 4.7.

4.2.3.6. Ajuste por modelo $CIAR$ Kalman

De igual forma se procede a ajustar los datos a un modelo $CIAR$ con estimación por filtro de Kalman, se utilizará la función $CIAR.kalman$, perteneciente a la librería $AITS$.

A continuación se muestra de forma gráfica el ajuste del modelo.

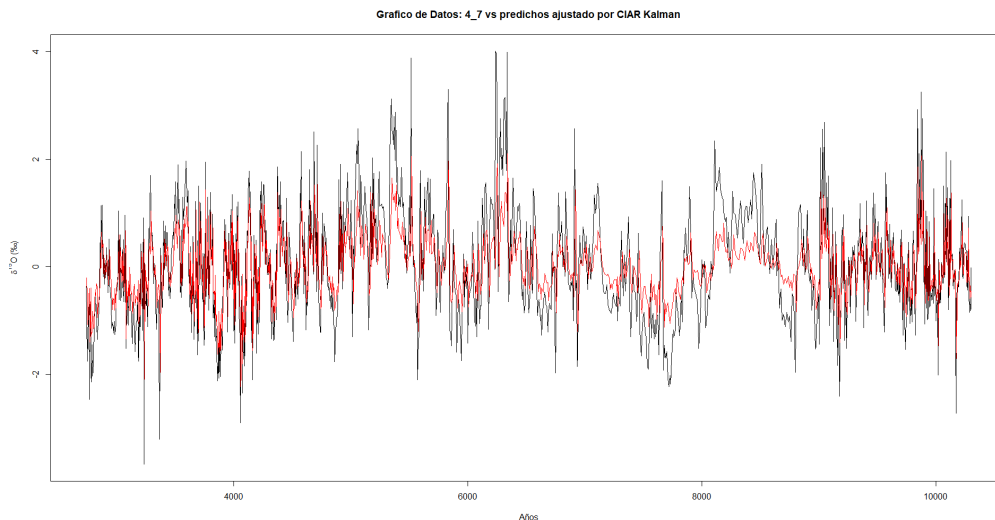


Figura 4.36: Gráfico de datos 4.7 versus datos predichos por un modelo $CIAR$ con estimación por filtro de Kalman, con datos irregularmente espaciados

De la figura (4.36) se aprecia en color negro los datos pertenecientes a la base 4.7 y en rojo los datos estimados

por el modelo *CIAR* con estimación por filtro de Kalman. A priori se puede observar que el modelo propuesto tiene un buen ajuste, viendose afectado principalmente donde los datos presentan una mayor amplitud.

Con el objetivo de analizar el desempeño del modelo de forma más concreta se analiza tanto el tiempo de cómputo (*TdC*) como los indicadores de bondad de ajuste $-Loglik$, *AIC*, *MSE* y *MSE2*. Además para comprobar el supuesto de no autocorrelación en los residuos, se utiliza el test de Ljung-Box.

Modelo <i>CIAR</i> filtro de Kalman						
ϕ	TdC	-Loglik	AIC	MSE	MSE2	LB.p-value
0.9034	22.36	0.5514	7.103	0.6143	1.103	0.5

Cuadro 4.55: Bondad de Ajuste del Modelo *CIAR* filtro de Kalman

Confirmando lo planteado a priori por la figura (4.36), a partir del cuadro (4.55), se puede notar que el modelo *CIAR* con estimación por filtro de Kalman, tuvo un $-Loglik$, *AIC* y *MSE* bajos. Además cumple con el supuesto de no autocorrelación de los residuos del modelo.

Por esto es recomendable el uso de un modelo *CIAR* con estimación por filtro de Kalman para la serie 4.7.

4.3. Predicciones

En este último apartado se profundiza en la predicción para las tres bases con datos reales antes mencionadas. El procedimiento fue separar los datos tanto en un parte de entrenamiento (90%) como una de predicción (10% restante), posterior a esto se ajustaron los datos de las bases *Blazar* y 4.7 a los modelos *AR*, *CAR*, *IAR* y sus variantes y el modelo *CIAR*, mas no para la base *ICAP* ya que ningún modelo propuesto se ajusto bien a este tipo de datos viendose reflejado en el test de blancura aplicado.

A partir de los valores ajustados y los parámetros estimados se realizó la predicción a un paso de cada base con cada modelo para el 10% restante, posterior a esto se compararon las predicciones por medio de los indicadores antes mencionados, utilizando además el indicador *MSE2*, el cual es el error cuadrático medio entre los valores predichos y la base de predicción para determinar que tan bueno es el poder predictivo del modelo seleccionado.

A continuación los mejores resultados para cada base.

Base	Modelo	TdC	AIC	-Loglik	MSE	MSE2	LB.p-value
Blazar	CIAR	1.10	0.7465811	7.493162	0.788314	1.294	0.8
4.7	CIAR	22.36	0.5514	7.103	0.6143	0.9103	0.5

Cuadro 4.56: Mejores modelos para cada Base de Datos Reales

Por último se graficó la predicción para cada base de datos utilizando los modelos con mejores resultados.

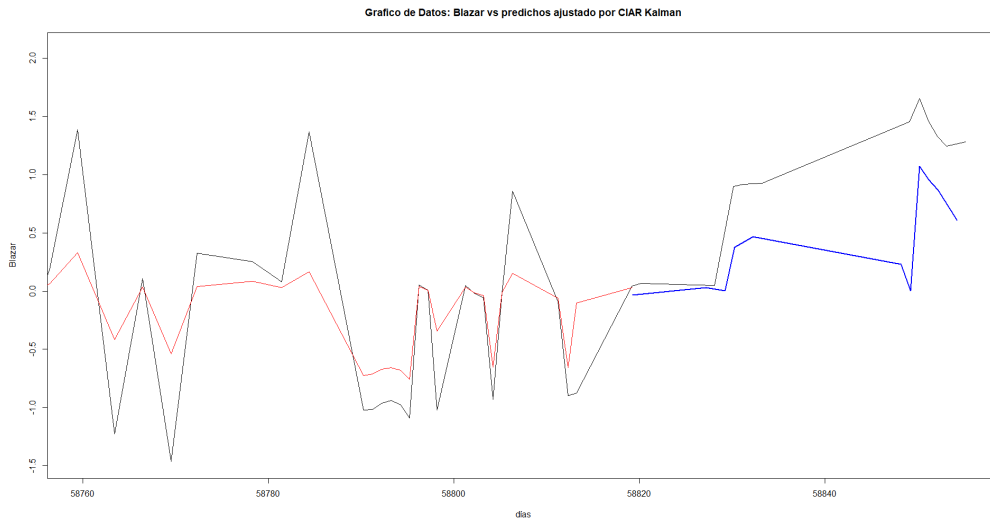


Figura 4.37: Gráfico de datos *Blazar* versus datos predichos por un modelo *CIAR* con estimación por filtro de Kalman, con datos irregularmente espaciados

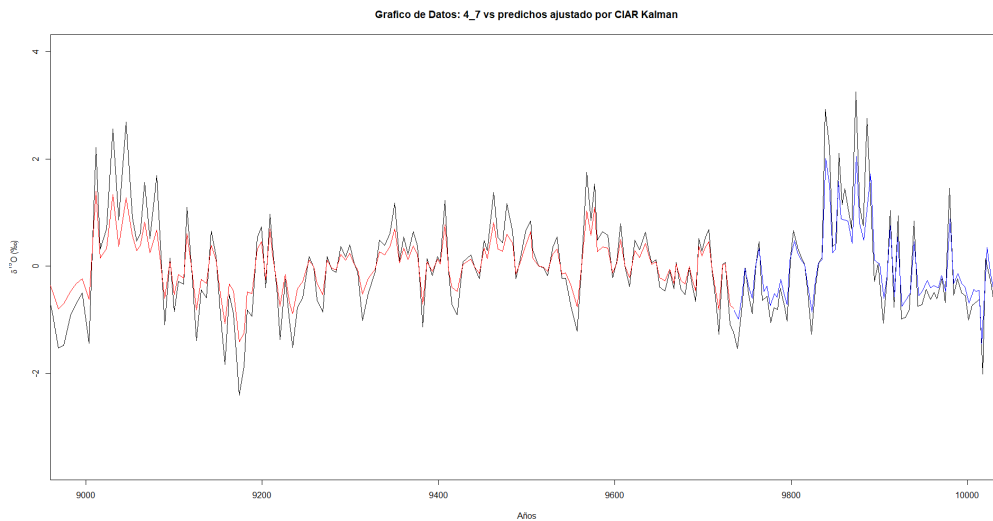


Figura 4.38: Gráfico de datos 4.7 versus datos predichos por un modelo *CIAR* con estimación por filtro de Kalman, con datos irregularmente espaciados

De las figuras (4.37, 4.38) se aprecia en color negro los datos pertenecientes a cada base, en rojo los datos estimados por el mejor modelo y en azul la predicción realizada por cada modelo.

En el primer gráfico se puede notar que el ajuste no fue tan bueno, sin embargo esto es común tomando en cuenta que las bases con datos astronómicos tienen saltos considerablemente grandes y poca información. Además se observa una predicción coherente con respecto al ajuste con problemas principalmente para los aumentos repentinos en la magnitud de la variable Blazar.

Por último el gráfico perteneciente a la serie 4_7 presentó un buen ajuste, esto se debe a que los saltos temporales entre cada observación eran más constantes y tenía un número mayor de observaciones que la base anteriormente nombrada. Además se observa una predicción de las últimas 135 observaciones bastante exacta respaldada por un error cuadrático medio de 0.9103.

Capítulo 5

Conclusión

En la literatura hay escasa información para el desarrollo de los modelos irregularmente espaciados (*CAR*, *IAR* y *CIAR*), sin embargo esto refleja que aún se puede seguir mejorando y desarrollando nuevas herramientas para el uso y la implementación de estas técnicas, en este estudio cada modelo tuvo tanto ventajas como desventajas frente a los demás.

En primero lugar para los datos simulados por un modelo *IAR* se ajustaron todos los modelos antes mencionados, efectivamente todos tuvieron buen ajuste exceptuando el modelo autorregresivo discreto común, ya que no tenía la capacidad de adaptarse a datos irregularmente espaciados, sin embargo su análogo continuo que sí cuenta con la adaptabilidad a este tipo de datos se ajustó de buena manera, teniendo diferencias pequeñas entre el valor del parámetro real y estimado, además de indicadores como el $-Loglik$ o el *AIC* pequeños.

Por otra parte los modelos discretos diseñados para series irregularmente muestreadas tuvieron un gran desempeño, destacando principalmente los *IAR* por sobre el *CIAR*, con esto se pudo apreciar la gran velocidad de cómputo de estos modelos en comparación a su extensión compleja.

Sin embargo para los datos simulados por un *CIAR* con parámetro negativo ningún modelo fue eficiente, esto no fue realmente un problema ya que se esperaba que en este tipo de datos los modelo no tuviera la capacidad de realizar un buen ajuste, exceptuando al autorregresivo irregular complejo, donde se pudo resaltar la capacidad de adaptación de este modelo con respecto a los demás.

En el caso de los datos con información real (no simulada), dependía mucho de la naturaleza de los datos recolectados, ya que como se pudo ver en la sección de resultados en algunos casos hubo modelos que se ajustaron mejor, mientras que en otros fue nula.

Para la primera base con datos astronómicos, como se mencionó anteriormente depende de varios factores el poder recolectar la información. Uno de los más importantes son los cielos despejados, ya que de no presentarse esta situación comienzan a variar los intervalos de observación.

En primera instancia para la base *Blazar* se notó que la variabilidad de los saltos temporales era bastante alta en comparación a lo que se había observado previamente, posterior a eso y siendo la primera base con datos reales que se manipuló, de inmediato se podía notar que ningún modelo tenía un gran ajuste de forma gráfica, además pocos modelos cumplían con el supuesto de no autocorrelación. Sin embargo, todos entregaban un valor del parámetro ϕ bastante similar.

De los tres modelos que cumplieron con el test de *Ljung – Box* para no autocorrelación de los residuos el que presentó mejores resultados con respecto a los indicadores ya mencionados fue el modelo *CIAR*. A partir del gráfico de predicciones, da la idea de que su poder predictivo no es el más óptimo para este tipo de datos, y ahí es donde se corrobora que aun existe tanto escasa información como herramientas para el desarrollo de estos fenómenos, dejando abierta la posibilidad de seguir avanzando en este campo.

La segunda base correspondiente al índice de capitalización de Colombia presentó varios problemas para el total funcionamiento de los modelos. En primera instancia fue la clara presencia de tendencia en los datos. Posterior a eso los modelos discretos presentaron un ajuste muy bueno, mas no cumplieron con el test de blancura, esto es un segundo fundamento a la falta de herramientas para el desarrollo de las series irregularmente espaciadas.

Finalmente para la serie con datos paleoclimáticos, se evidenció un mejor ajuste de todos los modelos en relación a los resultados arrojados por los datos astronómicos. Obviando la comparación ya antes realizada entre todos los modelos para esta serie, se pudo observar en el cuadro (4.56) que los indicadores fueron considerablemente menores en todos los casos. Además de que la predicción fue mucho mas exacta como se pudo apreciar en la figura (4.38). Por estos fundamentos, se plantea que en lo que respecta a datos paleoclimáticos dentro de lo que es el estudio de series desigualmente espaciadas ya existe un modelo predictivo suficientemente efectivo.

Por último, espero que en futuros estudios se puedan crear más paquetes para modelar este tipo de procesos, crear funciones donde se permita hacer predicciones más precisas de estos modelos y seguir enriqueciendo la teoría de cada uno. Además se propone el realizar este mismo análisis a datos financieros, los cuales no pudieron ser predichos por el incumplimiento del supuesto de no autocorrelación, como otros de distinta naturaleza, las que pueden estar ya reconocidas como irregularmente espaciadas o agregar nuevos casos a la literatura y así analizar el que tan bueno puede ser tanto el ajuste como las predicciones para ese tipo de datos, sin olvidar la posible incompatibilidad entre los paquetes ya existentes y el supuesto de no autocorrelación.

Apéndice

Función generadora de tiempo irregulares a partir de mezcla de distribuciones exponenciales.

```
gentime <- function(n,lambda1=130,lambda2=6.5,p1=0.15,p2=0.85)
{
dT<-rep(0,n)
a<-sample(c(lambda1,lambda2),size=n,prob=c(p1,p2),replace=TRUE)
dT[which(a==lambda1)]=rexp(length(which(a==lambda1)),rate=1/lambda1)
dT[which(a==lambda2)]=rexp(length(which(a==lambda2)),rate=1/lambda2)
sT=cumsum(dT)
return(sT)
}
```

Función para simular un modelo de serie temporal CIAR

```
CIAR.sample <- function(n,sT,phi.R,phi.I,rho=0,c=1)
{
delta<-diff(sT)
x=matrix(0,nrow=2,ncol=n)
F=matrix(0,nrow=2,ncol=2)
phi=complex(1,real=phi.R,imaginary=phi.I)
if(Mod(phi)>=1)
stop("Mod of Phi must be less than one")
Phi=Mod(phi)
psi<-acos(phi.R/Phi)
e.R=rnorm(n)
e.I=rnorm(n)
state.error=rbind(e.R,e.I)
Sigma=matrix(1,nrow=2,ncol=2)
Sigma[1,1]=1
Sigma[2,2]=c
Sigma[1,2]=rho*sqrt(Sigma[1,1])*sqrt(Sigma[2,2])
Sigma[2,1]=Sigma[1,2]
B=svd(Sigma)
```

```

A=matrix(0,nrow=2,ncol=2)
diag(A)=sqrt(B$d)
Sigma.root=(B$u)%*%A%*%B$u
state.error=Sigma.root%*%state.error
G=matrix(0,nrow=1,ncol=2)
G[1,1]=1
y=numeric()
x[,1]=state.error[,1]
for(i in 1:(n-1))
{
phi2.R<-(Phi**delta[i])*cos(delta[i]*psi)
phi2.I<-(Phi**delta[i])*sin(delta[i]*psi)
phi2<-1-Mod(phi**delta[i])**2
F[1,1]=phi2.R
F[1,2]=-phi2.I
F[2,1]=phi2.I
F[2,2]=phi2.R
x[,i+1]=F%*%x[,i]+sqrt(phi2)*state.error[,i]
y[i]=G%*%x[,i]
}
y[n]=G%*%x[,n]
return(list(t=sT,y=y,Sigma=Sigma))
}

```

Función que entrega el valor de logaritmo de la verosimilitud negativa de CIAR.kalman

```

CIAR.phi.kalman <- function(x,y,t,yerr,zero.mean='TRUE',standarized='TRUE',c=1)
{
sigmay<-1
if(standarized=='FALSE')
sigmay<-var(y)
if(zero.mean=='FALSE')
y=y-mean(y)
n=length(y)
Sighat=sigmay*matrix(c(1,0,0,c),2,2)

```

```

xhat=matrix(0,nrow=2,ncol=n)
delta<-diff(t)
Q=Sighat
phi.R=x[1]
phi.I=x[2]
#DEFINITION OF F
F=matrix(0,nrow=2,ncol=2)
G=matrix(0,nrow=1,ncol=2)
G[1,1]=1
#MOD PHI MUST BE LESS THAN ONE
phi=complex(1,real=phi.R,imaginary=phi.I)
Phi=Mod(phi)
psi<-acos(phi.R/Phi)
sum.Lambda=0
sum.error=0
phi=ifelse(is.na(phi)==TRUE,1.1,phi)
if(Mod(phi)<1){
for(i in 1:(n-1))
{
Lambda=G%*%Sighat%*%t(G) + yerr[i+1]**2
if(Lambda<=0 | is.na(Lambda)==TRUE)
{
sum.Lambda<-n*1e10
break;
}
phi2.R<-(Phi**delta[i])*cos(delta[i]*psi)
phi2.I<-(Phi**delta[i])*sin(delta[i]*psi)
F[1,1]=phi2.R
F[1,2]=-phi2.I
F[2,1]=phi2.I
F[2,2]=phi2.R
phi2<-1-Mod(phi**delta[i])**2
Qt<-phi2*Q
sum.Lambda=sum.Lambda+log(Lambda)
Theta=F%*%Sighat%*%t(G)

```

```

sum.error= sum.error+ ( y[i]-G**xhat[,i] )**2/Lambda
xhat[,i+1]=F**xhat[,i]+Theta**solve(Lambda)**(y[i]-G**xhat[,i])
Sighat=F**Sighat**t(F)+ Qt - Theta**solve(Lambda)**t(Theta)
}
yhat=G**xhat
out<-ifelse(is.na(sum.Lambda)==TRUE,1e10,(sum.Lambda + sum.error)/n)
}
else out=1e10
return(out)
}

```

Función de estimación de máxima verosimilitud de los parámetros del modelo CIAR: ϕ^R y ϕ^I , por filtro de kalman

```

CIAR.kalman <- function(y,t,delta=0,zero.mean='TRUE',standarized='TRUE',c=1,niter=10,seed=1234)
{
set.seed(seed)
aux<-1e10
value<-1e10
br<-0
if(sum(delta)==0){
delta=rep(0,length(y))}
for(i in 1:niter)
{
phi.R=2*runif(1)-1
phi.I=2*runif(1)-1
#print(complex(1,real=phi.R,imaginary=phi.I))
if(Mod(complex(1,real=phi.R,imaginary=phi.I))<1)
{
optim<-nlminb(start=c(phi.R,phi.I),objective=CIAR.phi.kalman,y=y,t=t,yerr=delta,zero.mean=zero.mean)
value<-optim$objective
}
if(aux>value)
{
par<-optim$par

```

```

aux<-value
br<-br+1
}
if(aux<=value & br>1 & i>trunc(niter/2))
break;
}
if(aux==1e10)
par<-c(0,0)
return(list(phiR=par[1],phiI=par[2],ll=aux))
}

```

Función para simular un modelo de serie temporal IAR (1)

```

IAR.sample <- function(phi,n=100,sT)
{
Sigma=matrix(0,ncol=n,nrow=n)
for( i in 1:n){
d<-sT[i]-sT[i:n]
Sigma[i,i:n]=phi**abs(d)
Sigma[i:n,i]=Sigma[i,i:n]
}
b=eigen(Sigma)
V <- b$vector
A=V **% diag(sqrt(b$values)) **% t(V)
e=rnorm(n)
y=as.vector(A**%e)
out=list(series=y,times=sT)
return(out)
}

```

Función que entrega el valor de logaritmo de la verosimilitud negativa de IAR.kalman

```

IAR.phi.kalman <- function(x,y,yerr,t,zero.mean='TRUE',standarized='TRUE')
{
sigmay<-1

```

```

if(standardized=='FALSE')
sigmay<-var(y)
if(zero.mean=='FALSE')
y=y-mean(y)
n=length(y)
Sighat=sigmay*diag(1)
xhat=matrix(0,nrow=1,ncol=n)
delta<-diff(t)
Q=Sighat
phi=x
#DEFINITION OF F
F=matrix(0,nrow=1,ncol=1)
G=matrix(0,nrow=1,ncol=1)
G[1,1]=1
#MOD PHI MUST BE LESS THAN ONE
sum.Lambda=0
sum.error=0
phi=ifelse(is.na(phi)==TRUE,1.1,phi)
if(abs(phi)<1){
for(i in 1:(n-1))
{
Lambda=G%%Sighat%%t(G) + yerr[i+1]**2
if(Lambda<=0 | is.na(Lambda)==TRUE)
{
sum.Lambda<-n*1e10
break;
}
phi2<-phi**delta[i]
F[1,1]=phi2
phi2<-1-phi**(2*delta[i])
Qt<-phi2*Q
sum.Lambda=sum.Lambda+log(Lambda)
Theta=F%%Sighat%%t(G)
sum.error= sum.error+ ( y[i]-G%%xhat[,i] )**2/Lambda
xhat[,i+1]=F%%xhat[,i]+Theta%%solve(Lambda)%(y[i]-G%%xhat[,i])

```

```

Sighat=F%%Sighat%%t(F)+ Qt - Theta%%solve(Lambda)%%t(Theta)
}
yhat=G%%xhat
out<-ifelse(is.na(sum.Lambda)==TRUE,1e10,(sum.Lambda + sum.error)/n)
}
else out=1e10
return(out)
}

```

Función de estimación de máxima verosimilitud del parámetro del modelo IAR: ϕ , por filtro de kalman

```

IAR.kalman <- function (y, sT, delta=0,zero.mean='TRUE',standarized='TRUE')
{
if(sum(delta)==0){
delta=rep(0,length(y))}
out = optimize(IAR.phi.kalman, interval = c(0, 1), y = y, t = sT, yerr=delta,zero.mean=zero.mean,standarized=standarized)
phi = out$minimum
ll = out$objective
return(list(phi = phi, kalman = ll))
}

```

Función que entrega el valor de logaritmo de la verosimilitud negativa de IAR.loglik

```

IAR.phi.loglik <- function(x,y,sT,delta=0,include.mean='FALSE',standarized='TRUE')
{
sigma=1
mu=0
if(sum(delta)==0){
delta=rep(0,length(y))}
if(standarized=='FALSE')
sigma=var(y)
if(include.mean=='TRUE')
mu=mean(y)
n=length(y)

```

```

delta<-delta[-1]
d<-diff(sT)
phi=x**d
yhat=mu+phi*(y[-n]-mu)
cte=(n/2)*log(2*pi)
s1=cte+0.5*sum(log(sigma*(1-phi**2)+delta**2)+(y[-1]-yhat)**2/(sigma*(1-phi**2)+delta**2))
return(s1)
}

```

Función de estimación de máxima verosimilitud directa del parámetro del modelo IAR: ϕ

```

IAR.loglik <- function(y,sT,delta=0,include.mean='FALSE',standarized='TRUE')
{
if(sum(delta)==0){
delta=rep(0,length(y))}
out=optimize(IAR.phi.loglik,interval=c(0,1),y=y,sT=sT,delta=delta,include.mean=include.mean,standar
phi=out$minimum
ll=out$objective
return(list(phi=phi,loglik=ll))
}

```

Función que entrega el valor de logaritmo de la verosimilitud negativa de IAR.t

```

IAR.phi.t <- function (x, y, sT, nu=3) #Minus Log Full Likelihood Function
{
sigma=x[2]
x=x[1]
n = length(y)
d <- diff(sT)
xd=x**d
yhat = xd * y[-n] #Mean of conditional distribution
gL=sigma*(1-xd**(2))*((nu-2)/nu) #Variance of conditional distribution
cte = (n-1)*log((gamma((nu+1)/2)/(gamma(nu/2)*sqrt(nu*pi))))
stand=((y[-1]-yhat)/sqrt(gL))**2
s1=sum(0.5*log(gL))

```

```

s2=sum(log(1 + (1/nu)*stand))
out= cte - s1 - ((nu+1)/2)*s2 -0.5*(log(2*pi) + y[1]**2)
out=-out #-Log Likelihood (We want to minimize it)
return(out)
}

```

Función de estimación de máxima verosimilitud directa del parámetro del modelo IAR-T: ϕ

```

IAR.t <- function (y, sT,nu=3) #Find minimum of IAR.phi.gamma
{
aux<-1e10
value<-1e10
br<-0
for(i in 1:20)
{
phi=runif(1)
sigma=var(y)*runif(1)
optim<-nlminb(start=c(phi,sigma),objective=IAR.phi.t,y=y,sT=sT,nu=nu,lower=c(0,0.0001),upper=c(0.99
value<-optim$objective
#print(c(optim$objective,optim$par,aux>value))
if(aux>value)
{
par<-optim$par
aux<-value
br<-br+1
}
if(aux<=value & br>10 & i>15)
break;
}
if(aux==1e10)
par<-c(0,0)
return(list(phi=par[1],sigma=par[2],ll=aux))
}

```

Para cada base real se utilizo una arquitectura similar y así obtener: Tiempo de Computo,

Parámetro, Valores Ajustados, Residuos del modelo, Logaritmo de la Verosimilitud Negativa, AIC, Error Cuadrático Medio, Gráfico de datos vs los datos ajustados y el test de Ljung-Box para no Autocorrelación de los Residuos. Para cada modelo

```
st=df[[1]]
y=df[[2]]
y1=(y-mean(y))/sd(y)
tiempo_computo.1=system.time(CIAR.kalman(y=y1,t=st))
data_temporal1=CIAR.kalman(y=y1,t=st)

modelo1=CIAR.fit(x=c(data_temporal1$phiR,data_temporal1$phiI),y=y1,t=st)
numericol=as.numeric(modelo1$yhat[-1])
ree1=y1[-1]-numericol

ECM1=mean(ree1**2)
phi1 = data_temporal1[[1]]
loglik1=data_temporal1[[3]]
AIC.1=6+(2*loglik1)

plot(st,y1, type="l", xlab="días", ylab="Blazar")
title(main="Grafico de Datos: Blazar vs predichos ajustado por CIAR Kalman")
lines(st[-1],numericol, col="red")

Box.test(ree1,lag = 1, type = c("Ljung-Box"))

tiempo_computo.2=system.time(IAR.kalman(y=y1,sT=st))
data_temporal2=IAR.kalman(y=y1,sT=st)
n=length(y1)
d=c(0,diff(st))
phi.k=data_temporal2$phi**d
yhat.k=phi.k*as.vector(c(y1[1:(n)]))
ree.k=y1-yhat.k

ECM.k=mean(ree.k**2)
phi.k2= data_temporal2[[1]]
```

```

loglik.k=data_temporal2[[2]]
AIC.k=4+(2*loglik.k)

plot(st,y1, type="l", xlab="días", ylab="Blazar")
title(main="Grafico de Datos: Blazar vs predichos ajustado por IAR Kalman")
lines(st,yhat.k, col="red")

Box.test(ree.k,lag = 1, type = c("Ljung-Box"))

tiempo_computo.3=system.time(IAR.loglik(y=y1,sT=st))
data_temporal3=IAR.loglik(y=y1,sT=st)
n=length(y1)
d=c(0,diff(st))
phi.l=data_temporal3$phi**d
yhat.l=phi.l*as.vector(c(y1[1:(n)]))
ree.l=y1-yhat.l

ECM.l=mean(ree.l**2)
phi.l2= data_temporal3[[1]]
loglik.l=data_temporal3[[2]]
AIC.l=4+(2*loglik.l)

plot(st,y1, type="l", xlab="días", ylab="Blazar")
title(main="Grafico de Datos: Blazar vs predichos ajustado por IAR Loglik")
lines(st,yhat.l, col="red")

Box.test(ree.l,lag = 1, type = c("Ljung-Box"))

tiempo_computo.4=system.time(IAR.t(y=y1,sT=st))
data_temporal4=IAR.t(y=y1,sT=st)
n=length(y1)
d=c(0,diff(st))
phi.t=data_temporal4$phi**d
yhat.t=phi.t*as.vector(c(y1[1:(n)]))
ree.t=y1-yhat.t

```

```

ECM.t=mean(ree.t**2)
phi.t2= data_temporal4[[1]]
loglik.t=data_temporal4[[3]]
AIC.t=4+(2*loglik.t)

plot(st,y1, type="l", xlab="días", ylab="Blazar")
title(main="Grafico de Datos: Blazar vs predichos ajustado por IAR-T")
lines(st,yhat.t, col="red")

Box.test(ree.t,lag = 1, type = c("Ljung-Box"))

```

```

tiempo_computo.5=system.time(arima(y1,order=c(1,0,0)))
data_temporal5=arima(y1,order=c(1,0,0))
param.ar=as.numeric(data_temporal5[[1]])
parametro.ar=param.ar[[1]]
numerico.ar=fitted(data_temporal5)[-1]
ree.ar=y1[-1]-numerico.ar
ECM.ar=mean(ree.ar**2)
phi.ar = parametro.ar[[1]]
AIC.ar=data_temporal5[[6]]
ll.ar=(AIC.ar-4)/2

```

```

plot(st,y1, type="l", xlab="días", ylab="Blazar")
title(main="Grafico de Datos: Blazar vs predichos ajustado por AR(1)")
lines(st[-1],numerico.ar, col="red")

```

```

Box.test(data_temporal5$residuals,lag = 1, type = c("Ljung-Box"))

```

```

tiempo_computo.6=system.time(carma(y1,times=st,arma = c(1, 0,0),parma=c(0.9)))
data_temporal6=carma(y1,times=st,arma = c(1, 0,0),parma=c(0.9))
param.car=exp(-exp(data_temporal6$coef))
numerico.car=data_temporal6$rpred
ree.car=y1-numerico.car

```

```
ECM.car=mean(ree.car**2)
phi.car = parametro.car[[1]]
AIC.car=data_temporal6$aic
ll.car=(AIC.car-4)/2
st2=st[-1]
plot(st, y1, type="l", xlab="días", ylab="Blazar")
title(main="Grafico de Datos: Blazar vs predichos ajustado por CAR(1)")
lines(st,numerico.car, col="red")
```

```
Box.test(ree.car, type = c("Ljung-Box"))
```

Bibliografía

- [1] EYHERAMENDY, S. & ELORRIETA, F. & PALMA, W. (2018). *An irregular discrete time series model to identify residuals with autocorrelation in astronomical light curves.*
- [2] ELORRIETA, F. & EYHERAMENDY, S. & PALMA, W. (2019). *Discrete-time autoregressive model for unequally spaced time-series observations.*
- [3] REHFELD, K. & MARWAN, N. & HEITZIG, J. & KURTHS, J. (2011). *Comparison of correlation analysis techniques for irregularly sampled time series.*
- [4] WANG, Z. (2013). *CTS: An R Package for Continuous Time Autoregressive Models via Kalman Filter.*
- [5] ELORRIETA, F. & EYHERAMENDY, S. & PALMA, W. (2020). *AIMS: Analysis of Irregular Time Series.*
- [6] ELORRIETA, F. (2018). *Classification and Modeling of time series of astronomical data.*
- [7] SEKITA, I. & KURITA, T. & OTSU, N. (1991). *Complex autoregressive model and its properties.*
- [8] TSAI, H. (2009). *On continuous-time autoregressive fractionally integrated moving average processes.*
- [9] ROUX, A. (2002). *Some properties of continuous-time autoregressive moving average time series models .*
- [10] SHREVE, S.E. & KARATZAS, I. (1991). *Brownian motion and stochastic calculus.* Graduate Texts in Mathematics, 113.
- [11] ROGERS, L. C. G. & WILLIAMS, D. (1994), *Diffusions, Markov processes and martingales.*
- [12] JONES, R. H. (1981). *Fitting a continuous time autoregression to discrete data. In Applied time series analysis II*
- [13] BELCHER J & HAMPTON JS & TUNNICLIFFE W (1994). *Parameterization of Continuous Time Autoregressive Models for Irregularly Sampled Time Series Data. Journal of the Royal Statistical Society. B, 56, 141–155.*
- [14] AKAIKE, H. (1974). *A new look at the statistical model identification. IEEE transactions on automatic control. 19(6), 716-723.*

- [15] SCHWARZ, G. (1978). *Estimating the dimension of a model. The annals of statistics*.6(2), 461-464.
- [16] BROCKWELL P. & DAVIS R. (2016). *Introduction to Time Series and Forecasting*, 3rd edn.
- [17] WIBERG, D. M. (1971). *Schaum's outline of theory and problems of state space and linear systems (p. 117)*.
- [18] BROCKWELL, P. J. & DAVIS, R. A. & CALDER, M. V. (2002). *Introduction to time series and forecasting (Vol. 2)*.
- [19] TSAY, R. S. (2005). *Analysis of financial time series (Vol. 543)*.
- [20] PANDIT, S. M. & WU, S. M. (1983). *Time series and system analysis with applications (Vol. 3)*.
- [21] DOOB, J. L. (1953). *Stochastic processes (Vol. 7, No. 2)*.
- [22] HARVEY, A. C. (1990). *Forecasting, structural time series models and the Kalman filter*.
- [23] WANG, Z (2004). *The Application of the Kalman Filter to Nonstationary Time Series through. Time Deformation*.
- [24] CLEVELAND, W. S. (1981). *LOWESS: A program for smoothing scatterplots by robust locally weighted regression. The American Statistician, 35, 54*.