

UNIVERSIDAD DE CONCEPCIÓN
DIRECCIÓN DE POSTGRADO
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
PROGRAMA MAGÍSTER EN ESTADÍSTICA



**Intervalos de Predicción en Procesos Localmente
Estacionarios mediante una técnica de Bootstrap: Una
Aplicación a los índices de COVID-19 en Chile**

Tesis presentada a la Facultad de Ciencias Físicas y Matemáticas de la
Universidad de Concepción para optar al grado académico de Magíster
en Estadística

Por: Oscar Gabriel Rubilar Parra

Profesor Guía: Dr. Guillermo Patricio Ferreira Cabezas

2023
Concepción - Chile

© 2023, Intervalos de Predicción en Procesos Localmente Estacionarios mediante una técnica de Bootstrap: Una Aplicación a los índices de COVID-19 en Chile.

Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo la cita bibliográfica del documento.

Dedicatoria

Dedicado a Dios y a mi esposa Cindy

Agradecimientos

En estos párrafos, quiero expresar mi más sincero agradecimiento y cariño a quienes me han acompañado en el transcurso de esta etapa, siendo de una enorme ayuda, a los cuales dedico esta obra.

En primer lugar, agradecer a Dios por ser mi amigo en todo momento, dándome apoyo y ánimo en las horas más difíciles y por sobre todo el entendimiento necesario para efectuar este trabajo.

Agradezco a mi amada esposa Cindy por su motivación en este proceso, por entender las largas horas de estudio durante estos 2 años y por ser mi apoyo en cada momento.

A mi profesor guía de tesis, Guillermo Ferreira, por darme la oportunidad de sumarme a esta investigación y en entregarme todas las herramientas necesarias para desarrollarme de la mejor manera en esta disciplina, a través de su buena disposición y excelente trato que me manifestó desde el primer momento.

Un reconocimiento también a los profesores Bernardo Lagos, Joel Muñoz y Jorge Matteu, quienes brindaron toda su experiencia para solucionar dudas y confeccionar esta investigación con la mayor rigurosidad posible.

También a mi compañero de estudios, Jorge Muñoz, por su motivación en el transcurso

de estos 2 años para seguir estudiando y avanzando en este campo del saber.

Finalmente, a la institución educativa en la cual trabajo, CEALA, por entenderme desde el primero momento en este desafío personal y profesional que tenía y en otorgar todas las facilidades en relación a mi horario y permisos para participar de esta gran experiencia.

Oscar Gabriel Rubilar Parra

Resumen

La predicción es uno de los objetivos principales en el análisis de series de tiempo y hay una creciente literatura dedicada a obtener resultados deseables en la construcción de intervalos. La representación de Modelos Espacio Estado ha sido ampliamente utilizada para este propósito. Permite obtener intervalos de predicción asumiendo innovaciones gaussianas y utilizar las ecuaciones de predicción del filtro de Kalman con un sistema de matrices invariante en el tiempo. Sin embargo, esta metodología de construir intervalos de predicción es muy restrictiva e imprecisa para modelar series temporales no estacionarias.

Este documento propone un procedimiento «bootstrap» para construir intervalos de predicción que tienen en cuenta la incertidumbre de la estimación de los parámetros, al momento que permite un sistema variable en el tiempo de matrices en la representación del Espacio Estado del modelo.

Finalmente, este procedimiento puede manejar procesos localmente estacionarios y no gaussianos de futuras innovaciones, presentando algunos estudios de simulación y un análisis práctico relacionado al número de casos diarios de COVID-19 en Chile.

Palabras claves: *Bootstrap - Filtro de Kalman - Localmente Estacionario - Representación Espacio Estado - Modelos variables en el tiempo.*

Abstract

Forecasting is one of the main goals in time series analysis and there is an increasing literature devoted to obtain optimal predictions and desirable coverage prediction intervals. The state space representation has been widely used for this purpose. It allows to obtain prediction intervals by assuming Gaussian innovations and uses the prediction equations of the Kalman filter with a time-invariant system of matrices. However, this methodology of constructing prediction intervals is very restrictive and inaccurate for modeling non-stationary time series.

This paper proposes a bootstrap procedure for constructing prediction intervals that takes into account the parameter estimation uncertainty, while allowing for a time-varying system of matrices in the state space representation of the model.

Our procedure can also handle locally stationary and non-Gaussian processes of future innovations. We present some simulation studies and a practical analysis related to the number of daily cases of COVID-19 in Chile.

Keywords: *Bootstrapping · Kalman Filter · Local Stationarity · State Space System · Time-Varying Models.*

Índice general

Dedicatoria	3
Agradecimientos	4
Resumen	6
Abstract	7
Índice de figuras	9
Índice de cuadros	11
1. Introducción	13
2. Marco Teórico	17
2.1. Series de Tiempo	17
2.2. Series de Tiempo Estacionarias	17
2.3. Series de Tiempo No Estacionarias	18
2.4. Procesos localmente estacionarios (LS)	19
2.5. Modelo de regresión con errores LS	21
2.6. Representación Modelos Espacio Estado (SS)	24
2.7. Filtro de Kalman	26
2.8. Intervalos de predicción	30
2.8.1. Filtro de Kalman con valores perdidos en procesos LS	30
2.8.2. Intervalos de predicción bootstrap en procesos LS	32

ÍNDICE GENERAL	9
<hr/>	
3. Estudio de Simulación	36
3.1. Métodos de predicción ST y SSB	37
3.2. Paralelización SSB	42
4. Análisis de una serie temporal relacionada a los casos de COVID-19 en Chile	44
5. Conclusiones	52
Bibliografía	54

Índice de figuras

2.1. El tiempo reescalado para un procedimiento de predicción en una serie LSFN	29
3.1. Estimaciones de las densidades de Y_{n+k} para $k = 15$ cuando las innovaciones son normales.	41
3.2. Estimaciones de las densidades de Y_{n+k} para $k = 15$ cuando las innovaciones son chi-cuadrado con 1 grado de libertad.	42
4.1. Casos diarios COVID-19 en Chile, desde el 01 de abril del 2022 hasta el 14 de enero del 2023.	45
4.2. Casos confirmados de COVID-19 en Chile: ACF Muestral: (a) Observaciones de 1 a 72, (b) Observaciones de 73 a 145, (c) Observaciones de 146 a 218, (d) Observaciones de 219 a 289.	46
4.3. Casos confirmados de COVID-19 en Chile. (a) Estimaciones del parámetro de larga memoria (b) Estimaciones de la varianza del ruido. En todos los paneles la línea continua representa el LSFN estacionario, la línea discontinua horizontal indica el modelo FN estacionario y los puntos representan el enfoque heurístico.	47
4.4. ACF Muestral: (a) Autocorrelaciones muestrales y (b) Autocorrelaciones parciales del error estandarizado de un paso adelante.	48
4.5. Densidad empírica e histograma del error estandarizado de un paso adelante.	50

4.6. Pronóstico de varios pasos fuera de la muestra de las últimas 50 observaciones y bandas de predicción del 95 % considerando la data total disponible.	50
4.7. Pronóstico de varios pasos fuera de muestra de las últimas 50 observaciones y bandas de predicción del 95 %.	51

Índice de cuadros

3.1. Coberturas promedio de Monte Carlo, longitud y porcentaje de observaciones restantes a la derecha y a la izquierda de los intervalos de predicción para Y_{n+k} los que son construidos usando las metodologías ST y SSB cuando z_t es Normal y la cobertura nominal es del 95%. . .	39
3.2. Coberturas promedio de Monte Carlo, longitud y porcentaje de observaciones restantes a la derecha y a la izquierda de los intervalos de predicción para Y_{n+k} los que son construidos usando las metodologías ST y SSB cuando z_t es $\chi^2_{(1)}$ y la cobertura nominal es del 95%.	40
3.3. Resultados de simulación de bootstrap.	43
4.1. Estimación de los parámetros del modelo: Casos diarios de la pandemia COVID-19 en Chile	49

Capítulo 1

Introducción

En la vasta literatura relacionada con las series de tiempo, se encuentra una estructura estacionaria en la mayoría de las metodologías utilizadas para el análisis de estos procesos temporales. Sin embargo, la realidad indica que hay una gran cantidad de contextos que no presentan dicha estructura. Esta problemática, afecta considerablemente las inferencias y/o predicciones del proceso. Debido a esto, muchos autores han propuesto técnicas para hacer frente al fenómeno de la no estacionariedad. Por ejemplo, [28], [27] y [34] proporcionaron el concepto de espectro evolutivo. En la década de los noventa, R. Dahlhaus, propuso en [4] y [5] una definición formal de una familia de procesos localmente estacionarios (LS en adelante, por su sigla en inglés de Local Stationarity) que generó una colección relevante de obras, véase: [6], [16], [8], [9], [3], [23], [24] y [36] entre otros. Cabe señalar que estos procesos tienen como finalidad, realizar una especie de bloques en series no estacionarias, suponiendo que aquellos si siguen una estructura estacionaria.

Uno de los principales objetivos de las series de tiempo estacionarias es proporcionar herramientas para predecir el fenómeno en estudio. En este contexto, existe una literatura sustancial sobre la predicción de procesos estacionarios utilizando el filtro de Kalman (KF en adelante, por su sigla en inglés de Kalman Filter). Este algoritmo, elaborado en la década de los sesenta, proporciona un procedimiento útil para estimar

y predecir series de tiempo. Para poder realizar esto, se requiere saber los verdaderos parámetros que impulsan esos modelos y que en la práctica, los parámetros verdaderos son reemplazados por estimadores consistentes, que producen incertidumbre en el análisis estadístico que el KF por sí solo no es capaz de manejar. Para superar estos inconvenientes, varios autores han propuesto procedimientos para incorporar esta incertidumbre en los intervalos de predicción de los Modelos Espacio Estado (en adelante, SS, por su sigla en inglés de Space State). En primer lugar, varias propuestas se basan en la distribución asintótica del parámetro estimador, ver: [17], [14], [29] y [13]. Estos procedimientos son inadecuados con tamaños de muestra pequeños, porque la distribución asintótica podría ser una pobre aproximación de la distribución muestral finita del estimador de parámetros. Segundo, también se han propuesto enfoques totalmente bayesianos; sin embargo, los cálculos pueden volverse muy pesados y llevar mucho tiempo, ver [10]. Finalmente, [26] y [30] utilizaron procedimientos de arranque, conocidos como «bootstrap», que tienen la ventaja de ser computacionalmente simples incluso en modelos relativamente complicados. Todos estos trabajos incorporan el parámetro incertidumbre en el error cuadrático medio de predicción, pero su enfoque no está en la predicción.

[35] proponen un procedimiento de arranque para incorporar la incertidumbre del parámetro en el intervalo de predicción. Sin embargo, este procedimiento se basa en una representación hacia atrás del modelo que no siempre existe. Posteriormente, [31] proponen un procedimiento bootstrap alternativo para construir intervalos de predicción en series de tiempo adaptados en una representación SS que no depende de la necesidad de una representación hacia atrás del modelo. Este procedimiento se basa en el formulario de innovación (en adelante, IF por su sigla en inglés de Innovation Form), que es una representación alternativa del modelo SS basado en un solo término de innovación. El principal inconveniente de este procedimiento, es que los parámetros son constantes en el tiempo, y en consecuencia su aplicabilidad es limitada.

Por otra parte, la literatura sobre la predicción de modelos variantes en el tiempo es

menos abundante. Recientemente, [24] y [12] propusieron predicciones a través de pronósticos de uno y de varios pasos, tratando el futuro como observaciones faltantes y luego usando un KF. Tal procedimiento de pronóstico explota una característica elegante de los métodos SS, que permiten obtener predicciones de valores futuros de la serie junto con sus correspondientes errores cuadráticos medios (en adelante, MSE, por su sigla en inglés de Mean Square Error); ver [11] para más detalles. La principal ventaja de usar modelos SS se asocia con la posibilidad de trabajar con componentes no observados. Estos componentes incluyen cualquier valor faltante en la serie y también valores futuros. Por lo tanto, la aplicación de tales modelos no necesita ninguna modificación o transformación de los datos para realizar predicciones con sus correspondientes intervalos de predicción.

En línea con el párrafo anterior, la contribución de este trabajo es desarrollar y validar un procedimiento de arranque «bootstrap» para obtener intervalos de predicción de observaciones futuras en una representación de modelos SS, específicamente de series de tiempo LS. Es decir, procesos caracterizados por parámetros que describan el nivel y la varianza que pueden ser manejados por funciones de tiempo. Nos desviamos sustancialmente de los enfoques propuestos por [24] y [12] porque nuestro procedimiento de arranque se basa en el IF de la representación SS sin asumir la existencia de valores faltantes en los datos. Nuestra principal contribución es incorporar la incertidumbre del parámetro al ejecutar el KF para construir intervalos de predicción permitiendo parámetros variables en el tiempo.

Los resultados de la predicción en los casos confirmados de COVID-19 en Chile, mediante la aplicación bootstrap se presentan en los siguientes capítulos:

- **Capítulo 2:** Marco Teórico - Se presentan los fundamentos teóricos de series de tiempo estacionarias, no estacionarias, localmente estacionarias, modelos espacio estado y filtro de Kalman. En relación a las metodologías que se pondrán en comparación en la presente tesis, se describirán los intervalos de predicción

usando Filtro de Kalman con valores perdidos y Bootstrap.

- **Capítulo 3:** Estudio de Simulación - Se describen los resultados del desempeño de ambas metodologías.
- **Capítulo 4:** Aplicación a datos reales - Se presenta el contexto del país en relación al número de casos confirmados de COVID-19 y posteriormente se efectúa la estimación de los parámetros de acuerdo al modelo que se propuso para esa serie, como así también, la interpretación de las diferentes predicciones que ambas metodologías arrojaron.
- **Capítulo 5:** Conclusiones - Se presentan las conclusiones, análisis desprendidos de la realización de la tesis y futuros trabajos que se podrían desarrollar a partir de este trabajo.

Capítulo 2

Marco Teórico

2.1. Series de Tiempo

En muchos fenómenos naturales o sociales se pueden medir ciertas variables y asignar un valor numérico a cada observación. Cuando se hace esto, se puede hablar de una serie de tiempo que de acuerdo a [2] es:

Definition 1 *Un conjunto de observaciones $\{X_t\}$, cada una de las cuales se registra en un tiempo específico t .*

Ahora, una **serie de tiempo discreta** es una en el que el conjunto T_0 de tiempos en los que se realizan las observaciones, resulta ser un conjunto discreto, como lo es el caso, por ejemplo, cuando las observaciones se realizan en intervalos de tiempo fijos. Mientras que las **series de tiempo de tiempo continuas** se obtienen cuando las observaciones se registran continuamente durante algún intervalo de tiempo, por ejemplo, cuando $T_0 = [0, 1]$.

2.2. Series de Tiempo Estacionarias

La propiedad de estacionariedad de una serie temporal, según [2], es importante en la teoría de la estimación y asintótica de los modelos de series temporales.

Definition 2 Una serie temporal discreta $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$, se dice que es estrictamente estacionaria si para cualquier t_1, t_2, \dots, t_n y para cualquier $k > 0$, la distribución de probabilidad conjunta de $\{X_{t_1}, X_{t_2}, \dots, X_{t_n}\}$ es idéntica a la distribución de probabilidad conjunta de $\{X_{t_1+k}, X_{t_2+k}, \dots, X_{t_n+k}\}$

Restringiendo la atención a aquellas propiedades que dependen sólo de la momentos de primer y segundo orden de $\{X_t\}$, podemos precisar esta idea con la siguientes definiciones:

Definition 3 Sea $\{X_t\}$ una serie de temporal con $\mathbb{E}(X_t^2) < \infty$. La media de $\{X_t\}$ es:

$$\mu_X = \mathbb{E}(X_t)$$

Definition 4 La función de covarianza de $\{X_t\}$ es:

$$\gamma_X = \text{Cov}(X_r, X_s) = \mathbb{E}[(X_r - \mu_X(r))(X_s - \mu_X(s))]$$

para todo número entero r y s

Definition 5 $\{X_t\}$ es una serie débilmente estacionaria o de segundo orden si:

- I. μ_X es independiente de t
- II. $\gamma_X(t+h, t)$ es independiente de t para cada h

Observación: Cabe señalar que si $\{X_t\}$ es estrictamente estacionaria entonces es estacionaria de segundo orden.

2.3. Series de Tiempo No Estacionarias

Definition 6 Una serie temporal $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$, como considera [19], se dice que no es estacionaria, cuando las propiedades estadísticas de al menos una secuencia

finita $\{X_{t_1}, X_{t_2}, \dots, X_{t_n}\}$ con $n \geq 1$ son diferentes para al menos un número entero $k > 0$ en la secuencia $\{X_{t_1+k}, X_{t_2+k}, \dots, X_{t_n+k}\}$

Esto, lo complementa también [22], mencionando que también existen series, que no cumplen los requisitos para decir que el proceso sea estacionario, los cuales presentan parámetros que no permanecen constantes en el tiempo. Este tipo de procesos motivan el estudio de métodos para abordar series en las que se presenta variabilidad en sus parámetros. Particularmente, se pondrá atención a Procesos Localmente Estacionarios, concepto que será definido a continuación.

2.4. Procesos localmente estacionarios (LS)

[4], presenta una de las primeras aproximaciones a este contexto. Primero, menciona que uno de los problemas más difíciles de resolver cuando se trata de procesos no estacionarios es como establecer una teoría asintótica adecuada. Las consideraciones asintóticas son necesarias en el análisis de series temporales para simplificar la situación, ya que es inútil hacer cálculos para un tamaño de muestra finito.

Sin embargo, si $\{X_1, X_2, \dots, X_n\}$ son observaciones de un proceso arbitrario no estacionario, dejar que n tienda al infinito, es decir, extender el proceso hacia el futuro, no dará ninguna información sobre el comportamiento del proceso al principio del intervalo de tiempo que se ve reflejado con una función desconocida $a(t)$. Por lo tanto, necesitamos un concepto asintótico.

Supongamos para eso, que observamos:

$$X_t = a(t)X_{t-1} + \varepsilon_t \quad (2.1)$$

con ε_t se distribuye iid $N(0, \sigma^2)$ para $t = 1, 2, \dots, n$. La inferencia en este caso, signifi-

ca la inferencia para lo desconocido de la función $a(t)$ en el intervalo $[1, n]$, teniendo información sobre $a(t)$ en la cuadrícula $\{1, 2, 3, \dots, n\}$.

Para analizar las propiedades asintóticas de los procesos que satisfacen, [4] estableció un reescalamiento del tiempo mediante el cambio de variable $u = \frac{t}{n}$, con esto, se puede analizar de mejor manera las propiedades asintóticas de los estimadores. Se tiene además que si n es lo suficientemente grande, se tendrá mayor información disponible de la estructura local de la serie, obteniendo una grilla más fina del intervalo $[0, 1]$, por lo cual, la representación de esta regresión no paramétrica, queda:

$$X_{t,n} = a\left(\frac{t}{n}\right)X_{t-1,n} + \varepsilon_t \quad (2.2)$$

para $t = 1, 2, \dots, n$ (donde a se reescala en el intervalo $[0, 1]$)

Ahora, para definir una clase general de procesos no estacionarios que incluya la expresión anterior, se representa a continuación, la definición de procesos localmente estacionarios (LS en adelante):

Definition 7 *Un proceso localmente estacionario (LS), puede ser definido en la siguiente representación espectral variable en el tiempo, la cual es:*

$$Y_{t,n} = \mu\left(\frac{t}{n}\right) + \int_{-\pi}^{\pi} A\left(\frac{t}{n}, \lambda\right) e^{i\lambda t} d\xi(\lambda) \quad (2.3)$$

para $t = 1, 2, \dots, n$ con función de transferencia A^0 y de tendencia μ . Y además:

- $\xi(\lambda)$ sea un movimiento Browniano en $[-\pi, \pi]$
- y exista una función A , 2π - periódica con preimagen en $A : (0, 1] \times \mathbb{R}$ en \mathbb{C} , con $A(u, -\lambda) = \overline{A(u, \lambda)}$ y

$$\sup_{t,\lambda} |A_{t,n}^0(\lambda) - A(\frac{t}{n}, \lambda)| \leq \frac{K}{n} < \infty, \quad (2.4)$$

para todo n con K una constante positiva.

Cabe señalar, que la suavidad de $A(\lambda)$ en $\frac{t}{n}$ garantiza que este tipo de procesos tenga un comportamiento localmente estacionario.

2.5. Modelo de regresión con errores LS

En esta sección, se presenta un modelo de regresión gaussiana con errores LS, a través de la siguiente representación espectral:

$$Y_{t,n} = X(\frac{t}{n})\beta(\frac{t}{n}) + \varepsilon_{t,n} \quad (2.5)$$

con

$$\varepsilon_{t,n} = \begin{cases} e^{i\lambda t} A_{t,T}^0(\lambda) d\xi(\lambda) & \lambda \in [-\pi, \pi] \\ 0 & \lambda \notin [-\pi, \pi] \end{cases}$$

para $t = 1, \dots, n$, donde $X(\frac{t}{n}) = (X_1(\frac{t}{n}), \dots, X_k(\frac{t}{n}))^\top$ es un k -vector de regresores no estocásticos, $\beta(\frac{t}{n}) = (\beta_1(\frac{t}{n}), \dots, \beta_k(\frac{t}{n}))^\top$, el cual tiene k -parámetros de regresión desconocidos que varían en el tiempo, $A_{t,T}^0$ es una función de transferencia y \top es el operador de transposición. Para más detalles, ver sobre los procesos de LS en el artículo presentado en [7].

Un caso especial del modelo de regresión definido a través de la ecuación 2.5, se obtiene considerando los errores $\{\varepsilon_{t,n}\}$ a partir de una representación de expansión de media móvil infinita (MA(∞)), es decir:

$$\varepsilon_{t,n} = \sigma(\frac{t}{n}) \sum_{j=0}^{\infty} \psi_j(\frac{t}{n}) z_{t-j} \quad (2.6)$$

donde $\sigma(\frac{t}{n})$ es la varianza variable en el tiempo, $\{\psi_j(\cdot)\}$ son coeficientes que satisfacen: $\psi_0(u) = 1$ y $\sum_{j=0}^{\infty} \psi_j(u)^2 < \infty$ para todo $u \in [0, 1]$, mientras que $\{z_t\}$ es una secuencia de variables aleatorias independientes y distribuidas de forma idéntica con media cero y varianza 1.

A continuación, se llamará a la siguiente secuencia de procesos $\{Y_{t,n}, t = 1, \dots, n\}$ que satisfacen las ecuaciones 2.5 y 2.6 un modelo de regresión con errores variables en el tiempo, siendo este proceso aproximado localmente por procesos estacionarios. Por lo tanto, podemos construir el proceso LS a través de:

$$Y_{t,n} = X\left(\frac{t}{n}\right)\beta\left(\frac{t}{n}\right) + \sigma\left(\frac{t}{n}\right) \sum_{j=0}^{\infty} \psi_j\left(\frac{t}{n}\right) z_{t-j} \quad (2.7)$$

En este caso, la función de transferencia (A^0) de este proceso en la ecuación 2.7 se encuentra dada por: $A_{t,n}^0(\lambda) = \sigma\left(\frac{t}{n}\right) \sum_{j=0}^{\infty} \psi_j\left(\frac{t}{n}\right) e^{-i\lambda j} = A\left(\frac{t}{n}, \lambda\right)$, por lo cual este modelo será denotado en lo sucesivo como LSMA(∞).

Es importante señalar que en este tipo de modelos, la descomposición de $\{\varepsilon_{t,n}\}$ proporciona dos tipos de dependencias en la clase de procesos LS: **de corta y larga memoria**, quienes son definidos a continuación.

Definition 8 *Un proceso $\{\varepsilon_{t,n}\}$ con representación infinita de medias móviles definidas por la ecuación 2.7, se dice **Localmente Estacionario de Corta Memoria**, si existen constantes positivas a y K tales que los coeficientes de la expansión de medias móviles cumplen lo siguiente:*

$$\forall j \geq 1, \forall u \in [0, 1], |\psi_j(u)| \leq K e^{-aj} \quad (2.8)$$

Para ilustrar, el caso de un proceso de corta memoria, se considerará una extensión del modelo LSMA, con uno en donde los errores tienen promedio móvil autorregresivos

LSARMA(p, q), siendo definido de la siguiente manera por [12]:

$$\varepsilon_{t,n} = \sum_{j=1}^p \phi_j\left(\frac{t}{n}\right) \varepsilon_{t-j,n} + \sum_{j=1}^q \theta_j\left(\frac{t}{n}\right) z_{t-j,n} + z_t \quad (2.9)$$

para $t = 1, 2, \dots, T$. Supongamos que $p = q = 1$, entonces $\varepsilon_{t,n} = \phi_j\left(\frac{t}{n}\right) \varepsilon_{t-j,n} + \theta_j\left(\frac{t}{n}\right) z_{t-j,n} + z_t$, siendo este proceso un caso particular del modelo de regresión expresada en 2.7 con coeficientes $\psi_j(u) = (\phi(u) - \theta(u)) = \phi^{j-1}$ para $j \geq 1$ y $\psi_j = 1$ para $j = 0$ en un LSMA(∞).

Cuando $q = 0$, se tiene un LS autorregresivo LSAR(p) definido en Dahlhaus (1997), donde $\varepsilon_{t,n} = \sum_{j=1}^p \phi_j\left(\frac{t}{n}\right) \varepsilon_{t-j,n} + z_t$, para $t = 1, 2, \dots, T$. Si $p = 1$, luego $\varepsilon_{t,n} = \phi_j\left(\frac{t}{n}\right) \varepsilon_{t-j,n} + z_t$, este proceso se transforma en un caso particular del modelo de regresión en 2.7 con $\psi_j(u) = \psi_j = \phi^j$ en un proceso LSMA(∞), siendo utilizado más adelante en esta investigación.

Definition 9 Un proceso $\{\varepsilon_{t,n}\}$ con representación infinita de medias móviles definidas por la ecuación 2.7, se dice **Localmente Estacionario de Larga Memoria**, si existe una constante positiva K tal que los coeficientes de la expansión de medias móviles cumplen lo siguiente:

$$\forall u \in [0, 1], \exists d \in \left(0, \frac{1}{2}\right), |\psi_j(u)| \leq K j^{d-1} \quad (2.10)$$

Cabe señalar, que los modelos ARIMA son uno de los más utilizados para capturar la dependencia del término de error en aplicaciones empíricas; esta es la razón por la que hemos considerado una generalización no estacionaria con parámetros variables en el tiempo, esto es:

$$\Phi\left(\frac{t}{n}, B\right) \varepsilon_{t,n} = (1 - B)^{-d\left(\frac{t}{T}\right)} \Theta\left(\frac{t}{n}, B\right) \sigma_z\left(\frac{t}{n}\right) z_t, \quad t = 0, 1, \dots, n, \quad (2.11)$$

donde $\frac{t}{n} \in [0, 1]$, $\Phi(u, B) = 1 - \phi_1(u)B - \dots - \phi_p(u)B^p$ es una componente autorregresiva, $\Theta(u, B) = 1 + \theta_1(u) + \dots + \theta_q(u)B^q$ es una componente de medias móviles, $d(u)$ es un parámetro de larga memoria, $\sigma(u)$ es un factor que determina la varianza para $u \in [0, 1]$ y $\{z_t\}$ es una secuencia de ruido blanco Gaussiana de media cero y varianza unitaria. Esta clase de modelos es una extensión del proceso autorregresivo fraccional integrado de medias móviles (ARFIMA), donde $\Phi(u, B)$, $\Theta(u, B)$, $d(u)$ y $\sigma_z(u)$ no dependen de u .

En lo sucesivo, el modelo escrito en 2.11, con el proceso autorregresivo ($\Theta(\cdot, \cdot) = 1$) y el ruido fraccional ($\Phi(\cdot, \cdot) = \Theta(\cdot, \cdot) = 1$), será conocido como un proceso localmente estacionario de ruido fraccionado (LSFN en adelante, por su sigla en inglés de Locally stationary fractional noise process).

A continuación, se describirá lo referido a modelos SS, para enlazar lo de KF con lo de LS.

2.6. Representación Modelos Espacio Estado (SS)

Definition 10 De acuerdo a [24], un modelo de espacio estado para una serie temporal (posiblemente multivariable y en este caso no estacionaria) $\{Y_t, t = 1, 2, \dots, n\}$ consta de dos ecuaciones. Estas son:

$$\begin{aligned} Y_{t,n} &= G_{t,n}X_{t,n} + W_{t,n} \\ X_{t+1,n} &= F_{t,n}X_{t,n} + V_{t,n} \end{aligned} \quad (2.12)$$

para $t = 1, 2, \dots, n$ La primera ecuación es conocida como la **ecuación de observación**, donde $Y_{t,n}$ es un vector de estado, $G_{t,n}$ es el operador de observación y $W_{t,n}$ es un ruido de observación con varianza $R_{t,n}$. La segunda ecuación, se llama **ecuación de estado**, la cual determina el estado $X_{t+1,n}$ en el tiempo $t + 1$ en términos del estado anterior

$X_{t,n}$ y un término de ruido, donde $F_{t,n}$ es un operador de transición de estado y $V_{t,n}$ es un ruido de estado con varianza $Q_{t,n}$.

2.7. Filtro de Kalman

En esta sección se explicarán las ecuaciones que definen este filtro y para ello, es importante representar la serie temporal mediante un sistema de ecuaciones SS, en función de la cual serán definidas las ecuaciones que conforman las recursiones de Kalman, basándose en las definiciones dadas por [2].

Definition 11 Para un conjunto de observaciones $\{Y_t\}$, en el contexto de una formulación SS, interesa obtener, en términos del error cuadrático medio, la mejor estimación lineal del vector de estado X_t en términos de las observaciones Y_1, Y_2, \dots y un vector aleatorio Y_0 ortogonal a V_t y W_t para todo $t \geq 1$. La estimación de X_t en términos de:

- Y_0, \dots, Y_{t-1} define el problema de predicción.
- Y_0, \dots, Y_t define el problema de filtrado.
- $Y_0, \dots, Y_n, n > t$ define el problema de suavizamiento.

Dado un vector aleatorio $X = (X_1, \dots, X_v)^\top$, se define:

$$P_t(X) = (P_t(X_1), \dots, P_t(X_v))^\top, \quad (2.13)$$

donde $P_t(X_i) = P(X_i | Y_0, \dots, Y_t)$ es el mejor predictor lineal de X_i en términos de Y_0, \dots, Y_t . El predictor a un paso $\hat{X} = P_{t-1}(X_t)$ y la correspondiente matriz de covarianza del error $\Omega_t = \mathbb{E} \left[(X_t - \hat{X}_t) (X_t - \hat{X}_t)^\top \right]$ están únicamente determinadas por las condiciones iniciales:

$$\hat{X}_1 = P(X_1 | Y_0), \Omega_1 = \mathbb{E} \left[(X_1 - \hat{X}_1) (X_1 - \hat{X}_1)^\top \right] \text{ y las recursiones}$$

$$\begin{aligned} \hat{X}_{t+1} &= F_t \hat{X}_t + \Theta_t \Delta_t^{-1} (Y_t - G_t \hat{X}_t) \\ \Omega_{t+1} &= F_t \Omega_t F_t^\top + Q_t - \Theta_t \Delta_t^{-1} \Theta_t^\top \end{aligned} \quad (2.14)$$

donde:

$$\Delta_t = G_t \Omega_t G_t^\top$$

$$\Theta_t = F_t \Omega_t G_t^\top$$

y Δ_t^{-1} es la inversa generalizada de Δ_t .

De acuerdo a lo trabajado por [12], a continuación se presenta, el sistema SS y el algoritmo KF para estimar y predecir el modelo de regresión en 2.5, donde el término del error sigue la estructura dada por la ecuación 2.6. Consideremos el sistema SS no estacionario:

$$\begin{aligned} Y_{t,n} &= \begin{bmatrix} G_{t,n} & X_{t,n} \end{bmatrix} \begin{bmatrix} \xi_{t,n} \\ \beta_{t,n} \end{bmatrix} + W_{t,n} \\ \begin{bmatrix} \xi_{t+1,n} \\ \beta_{t+1,n} \end{bmatrix} &= \begin{bmatrix} F_{t,n} & 0 \\ 0 & I_k \end{bmatrix} \begin{bmatrix} \xi_{t,n} \\ \beta_{t,n} \end{bmatrix} + \begin{bmatrix} H \\ 0 \end{bmatrix} V_{t,n} \end{aligned} \quad (2.15)$$

donde $\{Y_{t,n}\}$ son las observaciones, $G_{t,n}$ es el operador de observación, $X_{t,n}$ es un vector de variables predeterminadas, $\begin{bmatrix} \xi_{t,n} & \beta_{t,n} \end{bmatrix}^\top$ es el vector de estado y $W_{t,n}$ es un ruido de observación con varianza $R_{t,n}$. En la otra ecuación, $F_{t,n}$ es un operador de transición de estado, I_k es la matriz identidad de orden $k \times k$ y $V_{t,n}$ es un ruido de estado con varianza $Q_{t,n}$.

Por razones de eficiencia computacional, se truncará la expansión en la ecuación 2.7 en $m \in \mathbb{N}$ componentes, de modo que una aproximación para $\{Y_{t,n}\}$ se puede escribir como:

$$Y_{t,n} = X\left(\frac{t}{n}\right)\beta\left(\frac{t}{n}\right) + \sigma\left(\frac{t}{n}\right) \sum_{j=0}^m \psi_j\left(\frac{t}{n}\right) z_{t-j} \quad (2.16)$$

para $t = 1, 2, \dots, n$ y algún entero positivo m . Así, la representación SS del modelo está dada por:

$$\begin{aligned}
Y_{t,n} &= \begin{bmatrix} G_{t,n} & X_{t,n} \end{bmatrix} \begin{bmatrix} \xi_{t,n} \\ \beta_{t,n} \end{bmatrix} \\
\begin{bmatrix} \xi_{t+1,n} \\ \beta_{t+1,n} \end{bmatrix} &= \begin{bmatrix} F_{t,n} & 0 \\ 0 & I_k \end{bmatrix} \begin{bmatrix} \xi_{t,n} \\ \beta_{t,n} \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ & & & & 0 \end{bmatrix}^\top z_{t+1},
\end{aligned} \tag{2.17}$$

para $t = 1, 2, \dots, n$, donde $G_{t,n} = \sigma(\frac{t}{n}) [1 \ \psi_1(\frac{t}{n}) \ \psi_2(\frac{t}{n}) \ \psi_3(\frac{t}{n}) \ \dots \ \psi_m(\frac{t}{n})]$, $F_{t,n} = \begin{bmatrix} 0 & 0 \\ I_m & 0 \end{bmatrix}$ y $W_{t,n} = 0$. En esta ecuación, se puede utilizar las ecuaciones de KF para estimar los parámetros del modelo en la representación SS dada en la última ecuación. Cabe señalar, que este filtro permite obtener recursivamente la predicción de valores futuros de la serie. Con este fin, se considera los vectores siguientes: $G_{t,n}^\dagger = \begin{bmatrix} G_{t,n} & X_{t,n} \end{bmatrix}$, $\Psi_{t,n} = \begin{bmatrix} \xi_{t,n} & \beta_{t,n} \end{bmatrix}^\top$, $F_{t,n}^\dagger = \begin{bmatrix} F_{t,n} & 0 \\ 0 & I_k \end{bmatrix}$ y $P = \begin{bmatrix} H \\ 0 \end{bmatrix}$.

Luego, se puede reescribir el sistema SS original como:

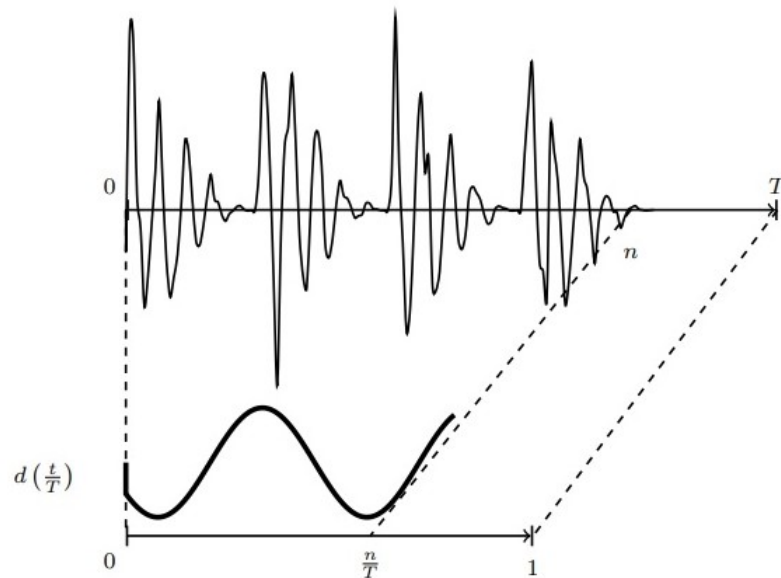
$$\begin{aligned}
Y_{t,n} &= G_{t,n}^\dagger \Psi_{t,n} \\
\Psi_{t+1,n} &= F_{t,n}^\dagger \Psi_{t,n} + PV_{t,n}
\end{aligned} \tag{2.18}$$

donde $H = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \end{bmatrix}^\top$ y $V_{t,n} = z_{t+1}$. Ahora, el siguiente sistema, resume la ecuación del algoritmo que describe el KF, considerando toda la información anterior, esto es:

$$\begin{aligned}
\hat{\Psi}_{t+1,n} &= F_{t,n}^\dagger \hat{\Psi}_{t,n} + \Theta_{t,n} \Delta_{t,n}^{-1} \Upsilon_{t,n} \\
\Omega_{t+1,n} &= F_{t,n}^\dagger \Omega_{t,n} \left(F_{t,n}^\dagger \right)^\top + Q_{t,n} - \Theta_{t,n} \Delta_{t,n}^{-1} \Theta_{t,n}^\top
\end{aligned} \tag{2.19}$$

donde $\Upsilon_{t,n} = Y_{t,n} - G_{t,n}^\dagger \hat{\Psi}_{t,n}$ es la secuencia de innovación, $\Delta_{t,n}$ es su varianza dada por $\Delta_{t,n} = G_{t,n}^\dagger \Omega_{t,n} \left(G_{t,n}^\dagger \right)^\top$ y $\Theta_{t,n} = F_{t,n}^\dagger \Omega_{t,n} \left(G_{t,n}^\dagger \right)^\top$. Es notable que los parámetros que varían en el tiempo cambian cuando llega una nueva observación. Por lo tanto, para obtener predicciones de varios pasos Y_{n+K} para $K > 0$, se redefine el tamaño de muestra $T = n + K$ para que los valores futuros del dominio de tiempo $[0, T]$ pertenezcan al dominio de tiempo reescalado en el intervalo $[0, 1]$, lo que implica que el vector

Figura 2.1: El tiempo reescalado para un procedimiento de predicción en una serie LSFN



Fuente: Elaboración propia.

paramétrico variable en el tiempo $\theta(u)$ se estima en el intervalo $I = [0, \frac{n}{T}]$. Si consideramos un proceso LSFN, entonces el valor de $d(u)$ fuera de I es el valor predicho del parámetro de memoria larga variable en el tiempo en el tiempo reescalado (ver Figura 2.1 y [21]).

2.8. Intervalos de predicción

En esta sección, se describen los procedimientos para obtener predicciones de uno y varios pasos, junto con sus correspondientes intervalos de predicción mediante una representación SS del modelo. La sección se divide en dos partes. El primero se basa en la propuesta presentada en [24], dedicado a la predicción de procesos LS. En esta propuesta, los pronósticos se obtienen de las ecuaciones recursivas de Kalman tratando las observaciones Y_{n+k} como datos faltantes. El último, analiza un procedimiento de arranque, denominado bootstrap, que complementa los aportes mencionados por [26], [30], [35], [31], [24] y [12].

2.8.1. Filtro de Kalman con valores perdidos en procesos LS

El método SS y su algoritmo KF asociado, proporciona una metodología simple para el manejo de valores perdidos. En este caso, los valores futuros de las series de tiempo observadas son tratados como si estuvieran perdiendo observaciones. Supongamos un conjunto de observaciones faltantes $Y_{n+k,T}$, con $k = 1, \dots, K$ para cualquier $K > 0$, donde $T = n + K$. Sin pérdida de generalidad, se define como 0 el vector $\Upsilon_{t,T}$ y la matriz $\Theta_{t,T}$ del KF, de modo que las actualizaciones de Kalman se convierten en:

$$\begin{aligned}\hat{\Psi}_{t+1,T} &= F_{t,T}^\dagger \Psi_{t,T} \\ \Omega_{t+1,T} &= F_{t,n}^\dagger \Omega_{t,n} \left(F_{t,n}^\dagger \right)^\top + Q_{t,T}\end{aligned}\tag{2.20}$$

para $t = n + 1, \dots, T$. Con el fin de predecir $Y_{n+k,T}$ junto con su varianza del error de pronóstico, se puede obtener tratando $Y_{t,T}$ para $t > n$ como observaciones faltantes y continuando el KF más allá de $t = n$ con $\Upsilon_{t,T} = 0$ y $\Theta_{t,T} = 0$ para $t > n$. Este procedimiento para el pronóstico es una característica elegante de los métodos SS para el análisis de series de tiempo.

Lo propuesto a continuación por [24], consiste en el pronóstico de varios pasos, los cuales se pueden calcular de la siguiente manera:

Sea $\tilde{Y}_{n+k,T} = \mathbb{E}(Y_{n+k,T} | Y_{n,T}, Y_{n-1,T}, \dots, Y_{1,T})$ el predictor de k pasos basado en el pasado finito para $1 \leq n+k \leq T$. Estos pronósticos y el MSE se obtienen de las ecuaciones recursivas del KF en 2.20. La propuesta hecha por [24] asume los errores de predicción futuros $e_{n+k,T} = Y_{n+k,T} - \tilde{Y}_{n+k,T}$ gaussianos. Por lo tanto, el intervalo de predicción de k para Y_{n+k} viene dado por:

$$\left[\tilde{Y}_{n+k,T} - z_{1-\frac{\alpha}{2}} \sqrt{\Delta_{n+k,T}}, \tilde{Y}_{n+k,T} + z_{1-\frac{\alpha}{2}} \sqrt{\Delta_{n+k,T}} \right] \quad (2.21)$$

donde $z_{1-\frac{\alpha}{2}}$ es el percentil $(1 - \frac{\alpha}{2})$ de la distribución normal estándar e $\tilde{Y}_{n+k,T}$ es:

$$\begin{aligned} \tilde{Y}_{n+k,T} &= G_{n+k,T}^\dagger \hat{\Psi}_{n+k,T} \\ &= X_{n+k,T} \beta_{n+k,T} + \hat{G}_{n+k,T} \hat{\xi}_{n+k,T} \end{aligned} \quad (2.22)$$

para $k = 1, \dots, K$, donde $G_{n+k,T} = [1 \quad \psi_1(\frac{n+k}{T}) \quad \psi_2(\frac{n+k}{T}) \quad \psi_3(\frac{n+k}{T}) \quad \dots \quad \psi_m(\frac{n+k}{T})]$.

Finalmente, la predicción varianza del error $\Delta_{n+k,T}$ de esta propuesta satisface:

$$\Delta_{n+k,T} = G_{n+k,T}^\dagger (F^\dagger)^k \Omega_{n,T} ((F^\dagger)^\top)^k (G_{n+k,T}^\dagger)^\top + G_{n+k,T}^\dagger \sum_{j=0}^{k-1} (F^\dagger)^j Q ((F^\dagger)^\top)^j (G_{n+k,T}^\dagger)^\top \quad (2.23)$$

En la práctica, los parámetros desconocidos involucrados en 2.22 son reemplazados por estimadores consistente, generalmente obtenidos mediante Cuasi-Máxima Verosimilitud (QML en adelante, por su sigla en inglés de Quasi-Maximum Likelihood); ver [15] y [11]. Las expresiones descritas en 2.21 son la forma más fácil de obtener intervalos de predicción, ya que asumen normalidad. **Esta metodología se denominará estándar** (en adelante, ST, por las dos primeras letras de la palabra en inglés: Standard). Nótese que los intervalos de predicción construidos como en 2.21 generalmente subestiman la variabilidad del error de pronóstico porque no toman en cuenta la incertidumbre debido a la estimación de los parámetros, ver [25]. Lo anterior, provoca que

estos intervalos podrían tener una cobertura inexacta en la predicción de los errores.

2.8.2. Intervalos de predicción bootstrap en procesos LS

De acuerdo a [18] el bootstrap, como método fue conceptualizado y descrito sistemáticamente por Efron y Tibshirani en la década de los setenta y noventa respectivamente.

Es un método que trata una muestra observada como una población finita y generan muestras aleatorias a partir de ella para estimar características poblacionales y hacer inferencia sobre la población muestral, incluyendo la construcción de intervalos de confianza o el test de significación estadística.

Ahora, como menciona [32], existen diversas estrategias para elaborar un intervalo de confianza, entre las cuales cabe destacar el método percentil, percentil corregido para el sesgo y percentil corregido para el sesgo acelerado. De hecho, esta enumeración recoge las estrategias desarrolladas por Bradley Efron, a las cuales deben añadirse otras como el método percentil-t, percentil automático, iteración pivotal, etc.

Considerado lo anterior, se trabajará en este estudio con el método percentil, ya que de acuerdo a [32] pueden considerarse como una estrategia adecuada cuando el estimador carece de sesgo.

A continuación, se propone un procedimiento de arranque para calcular la contabilidad de intervalos de predicción tanto para los errores futuros desconocidos como para la incertidumbre de la muestra asociada con el uso de los parámetros estimados en lugar de los verdaderos para ejecutar el KF. Como en [35], esta propuesta se basa en el IF, que se define en términos de un conjunto único de perturbaciones.

$$\begin{aligned} Y_{t,T} &= G_{t,T}^\dagger \Psi_{t,T} + \Upsilon_{t,T} \\ \hat{\Psi}_{t+1,T} &= F_{t,T}^\dagger \Psi_{t,T} + \Theta_{t,T} \Delta_{t,T}^{-1} \Upsilon_{t,T} \end{aligned} \quad (2.24)$$

Este procedimiento fue propuesto por [30], quien asumió que las matrices de 2.24 son invariantes en el tiempo. Por lo tanto, nuestro enfoque generaliza el trabajo previo al permitir que las matrices del sistema en la representación SS dada por 2.15 varíen con el tiempo.

Para describir el procedimiento, que ayude a construir intervalos de predicción bootstrap, se aproxima directamente la distribución de $Y_{n+k,T} : k = 1, \dots, K$ condicionalmente en los datos disponibles por la distribución de réplicas de bootstrap que incorporan simultáneamente la variabilidad debida a la estimación de parámetros y la incertidumbre como resultado de futuras innovaciones desconocidas.

Nuestro procedimiento se basa en la convergencia asintótica del método empírico bootstrap de las innovaciones a la distribución empírica de las innovaciones, es decir:

$$F_{\hat{Y}_{t,T}^{s*}} \rightarrow F_{\hat{Y}_{t,T}^s}$$

Para más detalles de esta convergencia ver [33], los cuales muestran la validez asintótica de la QML bajo la representación IF.

El procedimiento consta de los siguientes pasos:

Paso 1: Estimar los parámetros del modelo en 2.17 mediante QML (se denotará como $\hat{\theta}$) y obtener las innovaciones estandarizadas $\{\hat{Y}_{t,T}^s; 1 \leq t \leq n\}$

Paso 2: Obtener una secuencia de innovaciones estandarizadas bootstrap $\{\hat{Y}_{t,T}^s; 1 \leq t \leq T\}$ mediante sorteos aleatorios con reemplazo de las innovaciones estandarizadas, $\hat{Y}_{t,T}^s$

Paso 3: Usando las innovaciones de bootstrap para $1 \leq t \leq n$ y el IF en 2.24, calcular una réplica de bootstrap $\{\Upsilon_{t,T}^*; 1 \leq t \leq n\}$ que imiten la estructura de la serie original. Luego, se deben estimar los parámetros de arranque correspondientes, $\hat{\theta}^*$

Paso 4: Ejecutar el KF con $\hat{\theta}^*$ y las observaciones originales. Obtener una réplica de bootstrap del vector de estado en el tiempo n que incorpora la incertidumbre debida a la estimación de parámetros, $\hat{\Psi}_{n,T}^*$

Paso 5: Obtener predicciones de k -pasos adelante de bootstrap para out-sample, $\{\hat{Y}_{n+k,T}^*; 1 \leq k \leq K\}$ para las siguientes expresiones:

$$\begin{aligned}\hat{\Psi}_{n+k,T}^* &= (F^{\dagger*})^k \hat{\Psi}_{n,T}^* + \sum_{j=0}^{k-1} (F^{\dagger*})^{k-1-j} \hat{\Theta}_{n+j,T}^* \left(\hat{\Delta}_{n+j,T}^*\right)^{-1} \hat{Y}_{n+j,T}^* \\ \hat{Y}_{n+k,T}^* &= G_{n+k,T}^{\dagger*} (F^{\dagger*})^k \hat{\Psi}_{n,T}^* + G_{n+k,T}^{\dagger*} \sum_{j=0}^{k-1} (F^{\dagger*})^{k-1-j} \hat{\Theta}_{n+j,T}^* \hat{\Delta}_{n+j,T}^{*-1} \hat{Y}_{n+j,T}^* \hat{Y}_{n+k,T}^*\end{aligned}\quad (2.25)$$

para $k = 1, \dots, K$ y $\hat{Y}_{n,T}^* = Y_{n,T} - G_{n,T}^{\dagger*} \hat{\Psi}_{n,T}^*$, donde su varianza del error de predicción es dado por:

$$\hat{\Delta}_{T+k,T}^* = G_{n+k,T}^{\dagger*} \left(F^{\dagger*}\right)^k \Omega_{T,T}^* \left(\left(F^{\dagger*}\right)^{\top}\right)^k \left(G_{n+k,T}^{\dagger*}\right)^{\top} + G_{n+k,T}^{\dagger*} \sum_{j=0}^{k-1} \left(\left(F^{\dagger*}\right)^{\top}\right)^j \left(G_{n+k,T}^{\dagger*}\right)^{\top}\quad (2.26)$$

donde $G_{n+k,T}^{\dagger*}$ viene dado por 2.22. Observe que el sombrero en la parte superior de las matrices significa que se obtienen sustituyendo los parámetros por su correspondiente bootstrap estimados.

Los pasos 2 a 5 se repiten B veces obteniendo B réplicas de arranque $\hat{Y}_{n+k,T}^*$. Algunos comentarios están en orden: primero, el sistema de matrices en las ecuaciones 2.25 y 2.26 varían con el tiempo; segundo, la distribución empírica de $\hat{Y}_{n+k,T}^*$ incorpora la variabilidad debida a innovaciones futuras desconocidas y la variabilidad debida a la estimación de parámetros en un solo paso.

El procedimiento anterior se conocerá como **Bootstrap de Espacio Estado** (SSB en adelante, por su sigla en inglés de State Space Bootstrap). Los intervalos de predicción de SSB finalmente se construyen a través de:

$$\left[Q_{\frac{\alpha}{2}}^*, \hat{Y}_{n+k,T}^*, Q_{1-\frac{\alpha}{2}}^*, \hat{Y}_{n+k,T}^* \right] \quad (2.27)$$

donde $Q_{\frac{\alpha}{2}}^*, \hat{Y}_{n+k,T}^*$ es el percentil $\frac{\alpha}{2}$ de la distribución empírica bootstrap $\hat{Y}_{n+k,T}^*$

Capítulo 3

Estudio de Simulación

En este apartado, se presentan los resultados de experimentos MonteCarlo, los cuales tienen como propósito analizar el comportamiento de muestras finitas de predictores de uno y de varios pasos en procesos LS de corta y larga memoria. Se estudió el comportamiento de las bandas de predicción obtenidas a través de las metodologías ST y SSB, definidas respectivamente en las expresiones 2.21 y 2.27. Se consideró también **la cobertura, longitud y el porcentaje de observaciones omitidas tanto a la derecha como a la izquierda de los límites de los intervalos de predicción** y luego se compararon estas medidas con la cobertura nominal $1 - \alpha$ igual a 0,95. Se consideraron simulaciones de una serie particular generada por el proceso LSMA(∞) discutido en la sección 2.5, generando $R = 1000$ réplicas de series de tamaños $n = 64, 128, 256$ y 512.

Para cada serie simulada, $\{y_{1,T}^{(r)}, \dots, y_{n,T}^{(r)}\}$, $r = 1, 2, \dots, R$, se generaron R observaciones de $y_{n+k,T}^{(r)}$ para horizontes de predicción k . Luego, para esa serie en particular, se obtuvo un $100(1 - \alpha)\%$ intervalo de predicción denotado por (L^*, U^*) (basado en $B = 1000$ intervalos de réplica bootstrap) y se estimó la cobertura a través de:

$$\widehat{1 - \alpha^*} = \#(L^* \leq y_{n+k,T}^{(r)} \leq U^*)/R,$$

donde $y_{n+k,T}^{(r)}$ ($r = 1, 2, \dots, R$) son los valores generados anteriormente. Una vez realizado esto, se llevaron a cabo 1000 experimentos de MonteCarlo, informando en la siguiente sección la cobertura y longitud promedio y proporción de cobertura de las observaciones a la izquierda y a la derecha para cada intervalo de predicción. Cabe señalar que se denotarán aquellas medidas con las siguientes palabras en inglés, esto es: Coverage, Leng, Below y Above.

En lo que sigue, se considerará el parámetro de regresión que es variable en el tiempo, conocido como: β , de la siguiente forma: $\beta\left(\frac{t}{T}\right) = 0$. Esta consideración se efectúa por meras razones de eficiencia computacional en la formulación de modelos SS. Para obtener más detalles sobre los parámetros de regresión variables en el tiempo con diferentes procesos dinámicos, se remite al lector a [20] y [12].

3.1. Métodos de predicción ST y SSB

Los mejores predictores lineales de muestra detallados en las ecuaciones 2.22 y 2.25 fueron calculados para los horizontes de predicción $k = 1, 5$ y 15 . Con el fin de comprobar la idoneidad de los valores de predicción para $\tilde{Y}_{n+k,T}$, se usó $Y_{n+k,T}$ y \hat{Y}_T^* para los métodos ST y SSB respectivamente. Para todo lo anterior, se consideraron 2 modelos:

- I. Un proceso de corta memoria, esto es: LSAR(1)
- II. Un proceso de larga memoria, esto es: LSFN

Para el primer modelo, se consideró el tiempo variable $\phi(u)$ y el factor de escala $\sigma(u)$, quienes son definidos por las siguientes funciones:

$$\phi(u) = a_0 + a_1 u, \quad \sigma(u) = b_0 + b_1 u, \quad (3.1)$$

para $u \in [0, 1]$ con $(a, b) = (a_0, a_1, b_0, b_1) = (-0,4, 0,8, 0,5, 0,5)$. Para el segundo modelo, los parámetros $d(u)$ y $\sigma(u)$ siguieron la misma estructura en relación a las funciones del modelo anterior, eso si con parámetros, $(a_0, a_1, b_0, b_1) = (0,2, 0,25, 0,5, 0,5)$.

Las estimaciones QML de los parámetros se utilizaron con el filtro de Kalman con truncamiento $m = 30$ y observaciones $t = 1, \dots, n$ para obtener las estimaciones de las innovaciones y sus varianzas. Los valores restantes, es decir, $t = n + 1, \dots, n + K$ se dejaron para el pronóstico fuera de la muestra. Finalmente, se asumieron dos distribuciones de probabilidad para las innovaciones, estas son: la **Gaussiana y una centrada re-escala Chi-Cuadrado con 1 grado de libertad**.

El Cuadro 3.1, muestra las coberturas promedio de Monte Carlo de cada una de las bandas (coverage), longitud promedio (length) y proporción de cobertura de las observaciones a la izquierda y a la derecha para cada intervalo de predicción (below y above) cuando las innovaciones son gaussianas.

Los resultados de estas simulaciones muestran que a medida que aumentan los horizontes de predicción, el SSB tiene una cobertura y longitud promedio mayor que la cobertura ST en el proceso LSAR(1). Cabe señalar, que en ambos procedimientos disminuye el porcentaje de observaciones ubicadas tanto a la izquierda como a la derecha, aunque en SSB dicha disminución es más brusca, concentrando una mayor cobertura y siendo más próxima a la nominal. Mientras que para el proceso LSFN, la situación es diferente solo para los resultados originados por SSB, ya que a mayor horizonte de predicción, la cobertura disminuye y alejándose de la nominal. Por otro lado, la proporción a los costados crece a medida que aumentan los horizontes de predicción. La Figura 3.1, muestra las estimaciones del núcleo ST representadas por puntos delgados, mientras que la línea discontinua gruesa indica las densidades SSB para predicción de 15 pasos por delante en una determinada serie generada por un proceso LSFN, junto con su densidad empírica, representada por la línea continua. Además, cada panel presenta las estimaciones del Kernel para diferentes tamaños de muestra, panel (a) $n = 64$, panel (b) $n = 128$, panel (c) $n = 256$ y panel (d) $n = 512$. En todos los gráficos, se puede comprobar que la densidad del método SSB se encuentra más cercana al método empírico en relación a la densidad del kernel ST.

Cuadro 3.1: Coberturas promedio de Monte Carlo, longitud y porcentaje de observaciones restantes a la derecha y a la izquierda de los intervalos de predicción para Y_{n+k} los que son construidos usando las metodologías ST y SSB cuando z_t es Normal y la cobertura nominal es del 95 %.

Model	k	Mean coverage		Mean coverage in tails		Mean length	
		ST	SSB	ST Below / Above	SSB Below / Above	ST	SSB
<i>n</i> = 64							
LSAR(1)	1	0.8786	0.9041	0.0611 / 0.0601	0.0463 / 0.0495	3.8411	4.0271
	5	0.9341	0.9613	0.0327 / 0.0330	0.0198 / 0.0187	4.3733	5.0254
	15	0.9353	0.9808	0.0325 / 0.0321	0.0095 / 0.0095	4.7691	6.9980
LSFN	1	0.5986	0.9281	0.1966 / 0.2047	0.0351 / 0.0367	3.9263	3.9336
	5	0.7244	0.9074	0.1403 / 0.1352	0.0443 / 0.0482	4.5820	4.5811
	15	0.7784	0.8627	0.1121 / 0.1094	0.0589 / 0.0783	4.8653	4.9239
<i>n</i> = 128							
LSAR(1)	1	0.8921	0.9041	0.0534 / 0.0544	0.0477 / 0.0480	3.8887	3.9509
	5	0.9438	0.9560	0.0281 / 0.0280	0.0221 / 0.0218	4.3459	4.5829
	15	0.9450	0.9659	0.0275 / 0.0274	0.0167 / 0.0173	4.4069	5.2039
LSFN	1	0.6006	0.9350	0.2025 / 0.1968	0.0316 / 0.0333	3.9469	3.9245
	5	0.7330	0.9162	0.1389 / 0.1280	0.0384 / 0.0452	4.7057	4.6299
	15	0.7989	0.8881	0.0992 / 0.1018	0.0526 / 0.0592	5.0370	4.9934
<i>n</i> = 256							
LSAR(1)	1	0.8933	0.8862	0.0484 / 0.0583	0.0542 / 0.0596	3.9068	3.8315
	5	0.9424	0.9395	0.0294 / 0.0282	0.0268 / 0.0337	4.2672	4.2610
	15	0.9524	0.9480	0.0224 / 0.0252	0.0246 / 0.0274	4.3454	4.4267
LSFN	1	0.6185	0.9303	0.1937 / 0.1878	0.0372 / 0.0325	3.9520	3.8208
	5	0.7380	0.9054	0.1312 / 0.1308	0.0475 / 0.0471	4.7133	4.4790
	15	0.8079	0.8619	0.0941 / 0.0980	0.0659 / 0.0722	5.0933	4.8039
<i>n</i> = 512							
LSAR(1)	1	0.9026	0.9030	0.0481 / 0.0492	0.0486 / 0.0482	3.9168	3.9171
	5	0.9486	0.9497	0.0258 / 0.0255	0.0252 / 0.0250	4.2811	4.3087
	15	0.9484	0.9517	0.0255 / 0.0260	0.0238 / 0.0243	4.2860	4.3496
LSFN	1	0.5836	0.9400	0.2215 / 0.1949	0.0306 / 0.0293	3.9874	3.9024
	5	0.7419	0.8938	0.1291 / 0.1290	0.0572 / 0.0490	4.7986	4.4804
	15	0.8178	0.8779	0.0927 / 0.0895	0.0492 / 0.0729	5.1929	4.8874

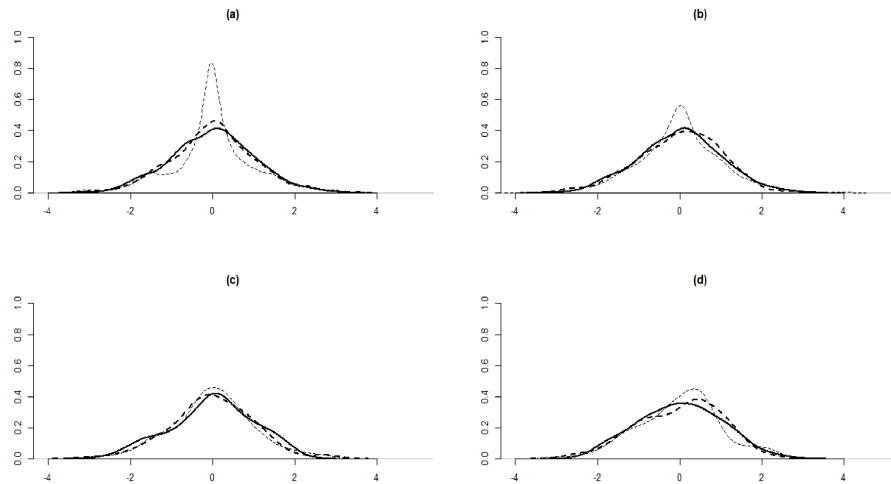
Fuente: Elaboración propia.

Cuadro 3.2: Coberturas promedio de Monte Carlo, longitud y porcentaje de observaciones restantes a la derecha y a la izquierda de los intervalos de predicción para Y_{n+k} los que son construidos usando las metodologías ST y SSB cuando z_t es $\chi^2_{(1)}$ y la cobertura nominal es del 95 %.

Model	k	Mean coverage		Mean coverage in tails		Mean length	
		ST	SSB	ST Below / Above	SSB Below / Above	ST	SSB
<i>n</i> = 64							
LSAR(1)	1	0.8313	0.8342	0.0701 / 0.0985	0.0738 / 0.0919	5.9623	4.6696
	5	0.9206	0.9414	0.0095 / 0.0698	0.0002 / 0.0583	9.2539	7.3743
	15	0.9237	0.9681	0.0000 / 0.0762	0.0000 / 0.0319	12.071	10.6799
LSFN	1	0.9586	0.7831	0.0011 / 0.0402	0.2164 / 0.0003	4.8947	9.9977
	5	0.9572	0.9904	0.0000 / 0.0427	0.0084 / 0.0011	6.4373	8.7788
	15	0.8843	0.7829	0.0000 / 0.1156	0.2166 / 0.0003	5.7832	9.9977
<i>n</i> = 128							
LSAR(1)	1	0.8921	0.9282	0.0534 / 0.0544	0.0073 / 0.0644	3.8887	6.9141
	5	0.9431	0.9530	0.0051 / 0.0517	0.0221 / 0.0218	9.4151	6.9406
	15	0.9471	0.9544	0.0000 / 0.0528	0.0000 / 0.0455	10.7574	9.3096
LSFN	1	0.9572	0.9826	0.0028 / 0.0399	0.0141 / 0.0031	5.5781	7.4412
	5	0.9627	0.9928	0.0000 / 0.0372	0.0065 / 0.0006	6.1277	9.5868
	15	0.9261	0.7182	0.0000 / 0.0738	0.2814 / 0.0002	5.8143	10.5981
<i>n</i> = 256							
LSAR(1)	1	0.8659	0.7737	0.0579 / 0.0762	0.1077 / 0.1186	6.2207	3.9986
	5	0.9446	0.9001	0.0097 / 0.0457	0.0093 / 0.0906	9.1192	5.9769
	15	0.9487	0.9203	0.0000 / 0.0513	0.0000 / 0.0797	9.8272	7.2387
LSFN	1	0.9624	0.9884	0.0016 / 0.0359	0.0099 / 0.0016	4.8132	8.1908
	5	0.9631	0.9885	0.0000 / 0.0368	0.0110 / 0.0003	5.2956	10.1046
	15	0.9357	0.6468	0.0000 / 0.0642	0.3530 / 0.0001	6.1347	11.4364
<i>n</i> = 512							
LSAR(1)	1	0.8921	0.7800	0.0314 / 0.0763	0.1260 / 0.0938	6.1986	4.0862
	5	0.9546	0.9202	0.0001 / 0.0451	0.0013 / 0.0783	9.1621	5.9984
	15	0.9504	0.9517	0.0000 / 0.0495	0.0000 / 0.0789	9.5875	6.7111
LSFN	1	0.9534	0.9886	0.0020 / 0.0445	0.0100 / 0.0014	4.9377	8.8523
	5	0.9621	0.9685	0.0000 / 0.0378	0.0305 / 0.0010	5.4335	10.8483
	15	0.9299	0.4788	0.0000 / 0.0700	0.5212 / 0.0000	6.0726	12.1469

Fuente: Elaboración propia.

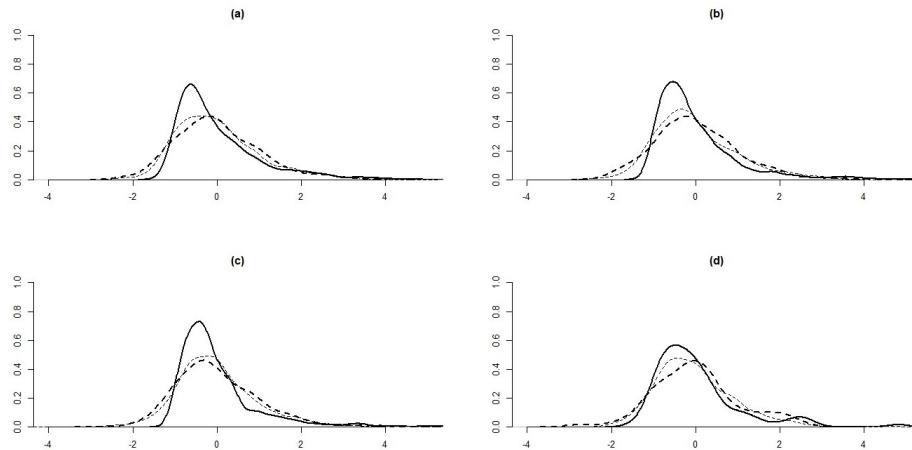
Figura 3.1: Estimaciones de las densidades de Y_{n+k} para $k = 15$ cuando las innovaciones son normales.



Fuente: Elaboración propia.

En relación con el Cuadro 3.2, en ambos procedimientos, la longitud promedio aumenta a medida que se crece en los horizontes de predicción. En los resultados originados en una serie LSAR(1) para los dos procedimientos, el comportamiento es el mismo que ocurre cuando las innovaciones son gaussianas, es decir, aumenta la cobertura y disminuyen los porcentaje de observaciones ubicadas tanto a la izquierda como a la derecha, aunque la cobertura para el procedimiento SSB, resulta ser más cercana a la nominal, lo que indica que este método puede capturar de mejor forma la asimetría de las innovaciones, en comparación a ST. Situación distinta ocurre en la serie LSFN, donde se observa una disminución para un horizonte de $k=15$ de la cobertura en ambos procedimientos, siendo más pronunciada en SSB, donde esta se aleja de la nominal, aumentando los porcentajes de observaciones en la izquierda, mientras que en ST, crecen los porcentajes de observaciones a la derecha. La Figura 3.2, presenta las estimaciones de ambos métodos con la misma simbología y tamaños de muestra que la figura anterior. En estos gráficos, se puede comprobar que no existe diferencia visual entre las densidades de los métodos SSB y SST en comparación con la empírica.

Figura 3.2: Estimaciones de las densidades de Y_{n+k} para $k = 15$ cuando las innovaciones son chi-cuadrado con 1 grado de libertad.



Fuente: Elaboración propia.

En conclusión, de los Cuadros 3.1 y 3.2, podemos concluir que en la mayoría de los casos simulados, la cobertura promedio de los intervalos ST es menor a que la cobertura promedio de los intervalos SSB, situación que se puede explicar por el hecho de que los intervalos ST no incorporan la incertidumbre que es ocasionada en la estimación de los parámetros.

3.2. Paralelización SSB

Al trabajar en el lenguaje de programación estadístico R, existen códigos que pueden demorar mucho tiempo, haciendo los procesos muy costosos computacionalmente y demorando las conclusiones de los análisis. Un ejemplo de lo anterior, es lo que sucede con el procedimiento bootstrap. Normalmente, al trabajar en los computadores actuales, se tiende a realizar un procesamiento serial, esto es cuando un problema se divide en una serie de instrucciones, que se ejecutan secuencialmente una tras otra, en un solo procesador. Sin embargo, ese no es el único camino que existe para trabajar computacionalmente. En este nuevo camino, de acuerdo a lo que describe [1], el problema se divide en partes discretas a las que pueden aplicarse una serie de instrucciones

simultáneamente en diferentes procesadores, lo que sus autores lo denominan procesamiento en paralelo o paralelización, técnica utilizada en el transcurso de este informe para encontrar los resultados de los Cuadros 3.1 y 3.2 a través de la metodología SSB.

En lo sucesivo, se presentan 2 ejemplos a través del Cuadro 3.3, el cual entrega los tiempos de comparación del código bootstrap sin o con procesamiento en paralelo, para series LSAR(1) con tamaño muestral $n = 64$ y 128 , asumiendo innovaciones gaussianas, luego de haber aplicado 1000 repeticiones de Monte Carlo y con horizonte de predicción $k = 1$ y 5 respectivamente. Cabe señalar, que se comparará la información llamada «elapsed» del programa R, la cual indica el tiempo real de la ejecución de los códigos en segundos. Para la paralelización se utilizaron 10 núcleos y con el fin de observar alguna tendencia en los tiempos se llevaron a cabo 3 veces estas simulaciones para cada tipo de experimento.

Cuadro 3.3: Resultados de simulación de bootstrap.

n	k	Elapsed Sin paralelizar	Elapsed Con paralelizar
64	1	195947.920	55616.634
64	1	209986.330	55404.797
64	1	191287.800	55712.988
128	5	408432.400	118606.877
128	5	405962.500	118457.016
128	5	402470.540	117953.625

Fuente: Elaboración propia.

Al analizar los tiempos del cuadro anterior, se puede constatar que el procesamiento en paralelo disminuye en promedio para ambos tamaños muestrales y horizontes de predicción un 71 % los tiempos de ejecución, siendo mucho menos intensivo en su trabajo y entregando resultados más certeros que los de la metodología ST.

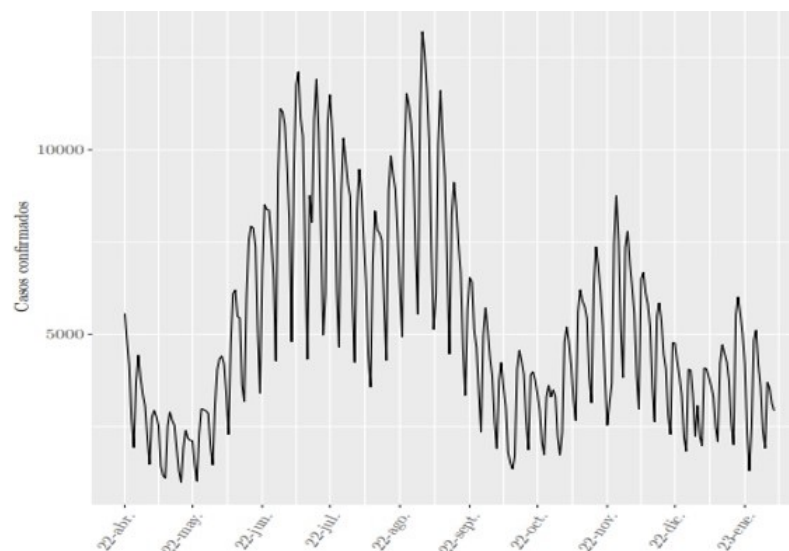
Capítulo 4

Análisis de una serie temporal relacionada a los casos de COVID-19 en Chile

En esta sección se describe la aplicación de la predicción de un modelo de regresión en el contexto de errores de larga memoria localmente estacionarios (LSFN). Para dicho efecto, se estimó el vector de regresores y errores a través del filtro de Kalman usando el sistema SS dado por 2.17. Para esta ilustración del modelo de regresión con errores localmente estacionarios, se aplicó la metodología bootstrap, específicamente, a los datos relativos al número de casos confirmados de la pandemia de COVID-19 en Chile, los cuales se encuentran disponibles en la página web del Ministerio de Salud de dicho país.

Esta pandemia originada por el virus SARS-CoV-2 ha estado presente en el país, desde el 02 de marzo del 2020, día en que se detectó el primer caso confirmado en la ciudad de Talca. Esta enfermedad durante el año 2020, llevó a la aplicación de distintas medidas para contener su avance, como por ejemplo: cuarentenas, cierres de lugares con alta afluencia de público, suspensión de eventos deportivos y musicales, entre otros. También en materia económica, causó inestabilidad en los principales mercados y un

Figura 4.1: Casos diarios COVID-19 en Chile, desde el 01 de abril del 2022 hasta el 14 de enero del 2023.



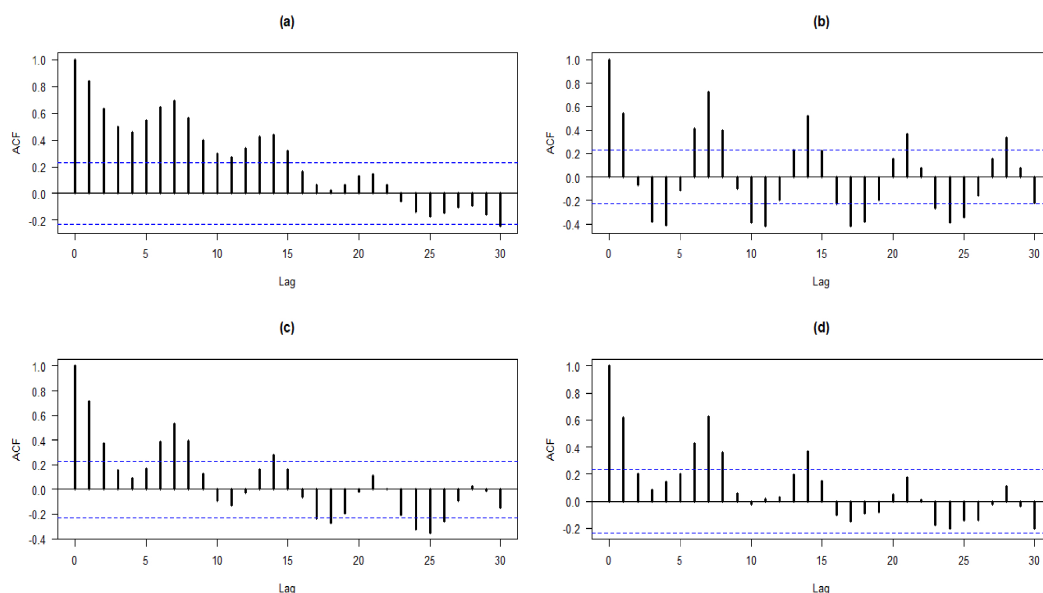
Fuente: Elaboración propia.

aumento en los niveles de la cesantía, lo que llevó al gobierno a desarrollar el mayor paquete fiscal de su historia, el cual consistió principalmente en transferencias monetarias directas a la población, siendo reconocido por la CEPAL (Comisión Económica para América Latina y el Caribe), como el país de América Latina que mayor ayuda económica entregó a sus habitantes. Dicho de esa forma, el objetivo de este estudio es analizar los casos diarios desde el 01 de abril del 2022 hasta el 14 de enero 2023 (ver Figura 4.1), época en donde el país ya se encontraba con una relativa normalidad, la gran parte de la población se encontraba vacunada con 3 dosis y el contexto sanitario marcaba una baja considerable en los casos diarios, después del aumento significativo que produjo la entrada de la variante Omicrón, proveniente de Sudáfrica, a principios de ese año.

De lo anterior y observando la Figura 4.1, se propone el siguiente modelo de regresión:

$$Y_{t,T} = \mu_1 + \mu_2 \cos\left(2\pi\left(\frac{t}{T}\right) - 210\right) + \varepsilon_t \quad (4.1)$$

Figura 4.2: Casos confirmados de COVID-19 en Chile: ACF Muestral: (a) Observaciones de 1 a 72, (b) Observaciones de 73 a 145, (c) Observaciones de 146 a 218, (d) Observaciones de 219 a 289.



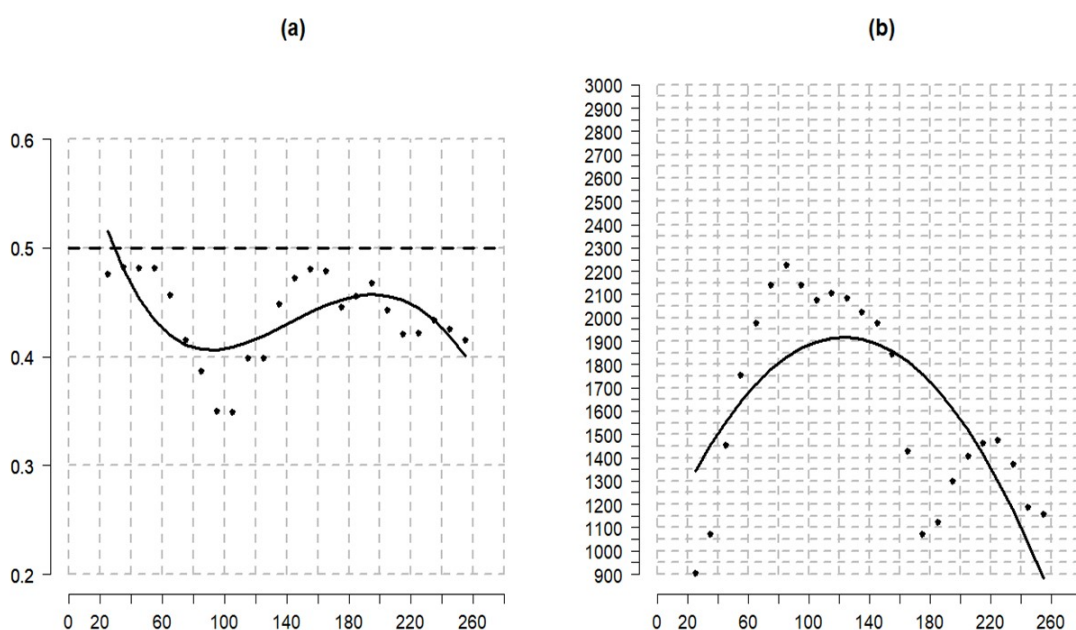
Fuente: Elaboración propia.

donde ε_t se distribuye *iid* $N(0, \sigma_\varepsilon^2)$.

Analizando el ACF empírico de los datos considerandos, ver Figura 4.2, revela que el grado de larga memoria parece ser constante a lo largo del tiempo en todos los bloques. De hecho, los valores de la muestra ACF de las primeras 72 observaciones, que se encuentran en el panel (a), son más altas que las valores muestrales correspondientes de ACF para los otros tres períodos considerados, ver paneles (b)-(c)-(d) y va disminuyendo a medida que se avanza en los datos.

Dicho lo anterior, resulta apropiado mencionar que las innovaciones ε_t en el modelo de regresión lineal dado en 4.1 tienen una estructura de memoria larga variable en el tiempo. Entonces, para conocer la estructura dinámica de los parámetros se propone un análisis exploratorio de los datos de la siguiente manera: la serie cuenta con un dominio de observación $\{1, \dots, T\}$, se divide en M bloques superpuestos de longitud N y

Figura 4.3: Casos confirmados de COVID-19 en Chile. (a) Estimaciones del parámetro de larga memoria (b) Estimaciones de la varianza del ruido. En todos los paneles la línea continua representa el LSFN estacionario, la línea discontinua horizontal indica el modelo FN estacionario y los puntos representan el enfoque heurístico.

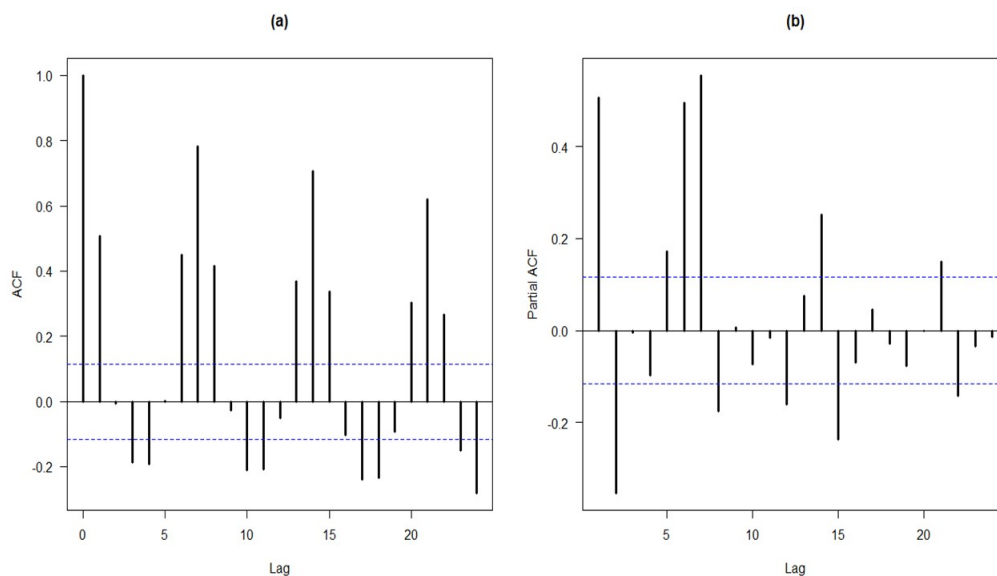


Fuente: Elaboración propia.

con un desplazamiento S de modo que $T = S(M - 1) + N$ y el punto medio del j -ésimo bloque, $j = 1, \dots, M$, es $[t_j]$ con $t_j = S(j - 1) + N/2$.

Para nuestro análisis empírico, la muestra posee $T = 289$ observaciones, la que se subdivide en $M = 24$ bloques de longitud $N = 50$ c/u, cambiando $S = 10$ posiciones hacia adelante. Es por eso, que el primer bloque o ventana es $(Y_{1,T}, \dots, Y_{50,T})$, $(Y_{11,T}, \dots, Y_{60,T})$ y así sucesivamente. Dentro de cada ventana se estimó los parámetros de un proceso estacionario FN(d) (FN por su sigla en inglés de Fractal Noise). La Figura 4.3 muestra las estimaciones heurísticas, las que se representan en la imagen a través de puntos. Por otro lado, la línea punteada muestra las estimaciones de los parámetros (d, σ) para el caso estacionario considerando la data disponible; mientras que la línea continua

Figura 4.4: ACF Muestral: (a) Autocorrelaciones muestrales y (b) Autocorrelaciones parciales del error estandarizado de un paso adelante.



Fuente: Elaboración propia.

indica las estimaciones del filtro de Kalman de las variables temporales $(d(u), \sigma(u))$. De esta figura, se puede sugerir una función polinómica de cuarto grado y cuadrática, para $(d(u), \sigma(u))$, respectivamente; es decir:

$$d(u) = a_0 + a_1u + a_2u^2 + a_3u^3 + a_4u^4 \quad \sigma(u) = b_0 + b_1u + b_2u^2 \quad (4.2)$$

donde $u \in [0, 1]$.

Las estimaciones de cuasi máxima verosimilitud de $a = (a_0, a_1, a_2, a_3, a_4)$ y $b = (b_0, b_1, b_2)$ fueron obtenidas mediante el estimador del filtro de Kalman con nivel de truncamiento $m = 30$. Las estimaciones referidas a cada parámetro se presentan en la tabla de a continuación. Finalmente, las desviaciones estándar y la prueba t fueron obtenidas en el Cuadro 4.1 a través de simulaciones Monte Carlo, con 1000 repeticiones para series de igual longitud que los datos empíricos del COVID-19 en Chile. Cabe señalar, que el estadístico t para todos los parámetros son altamente significativos al 5% de confianza.

Cuadro 4.1: Estimación de los parámetros del modelo: Casos diarios de la pandemia COVID-19 en Chile

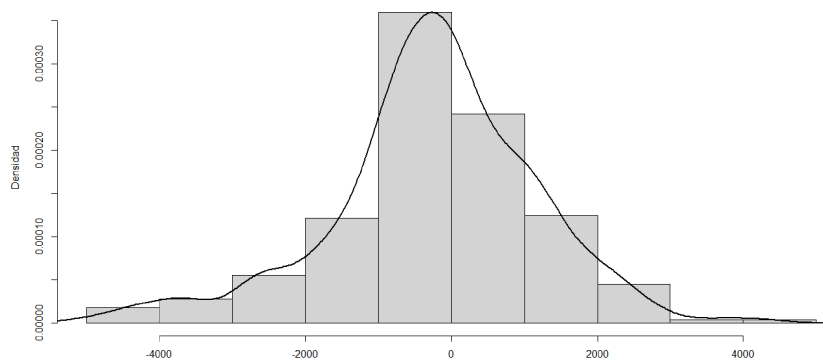
Parámetros	Estimadores	Desviación Estándar	$ t - \text{calculado} $
a_0	0.28647	0.00027	1050.38400
a_1	-0.28701	0.00005	5709.22400
a_2	0.28120	0.00008	3340.51600
a_3	0.46189	0.00008	5719.11200
a_4	0.14887	0.00007	2125.79800
b_0	1.03689	0.00022	4647.90300
b_1	0.07482	0.00024	303.950200
b_2	0.59132	0.00025	2292.68100
μ_1	5995.00000	0.00001	$5.530 \cdot 10^8$
μ_2	5499.99998	0.00001	$4.640 \cdot 10^8$

Fuente: Elaboración propia.

La Figura 4.4 muestra el correlograma, mientras que la Figura 4.5 presenta la estimación kernel de la densidad de los errores estandarizados de un paso adelante. Las correlaciones y las correlaciones parciales son significativas, mientras que por otro lado, la densidad de los errores muestra que están lejos de presentar la distribución normal. Por lo tanto, aunque el modelo de regresión parece apropiado, para representar las dependencias de los casos confirmados de COVID-19 en Chile, resulta conveniente implementar un procedimiento de predicción que tenga en cuenta la no normalidad de los errores.

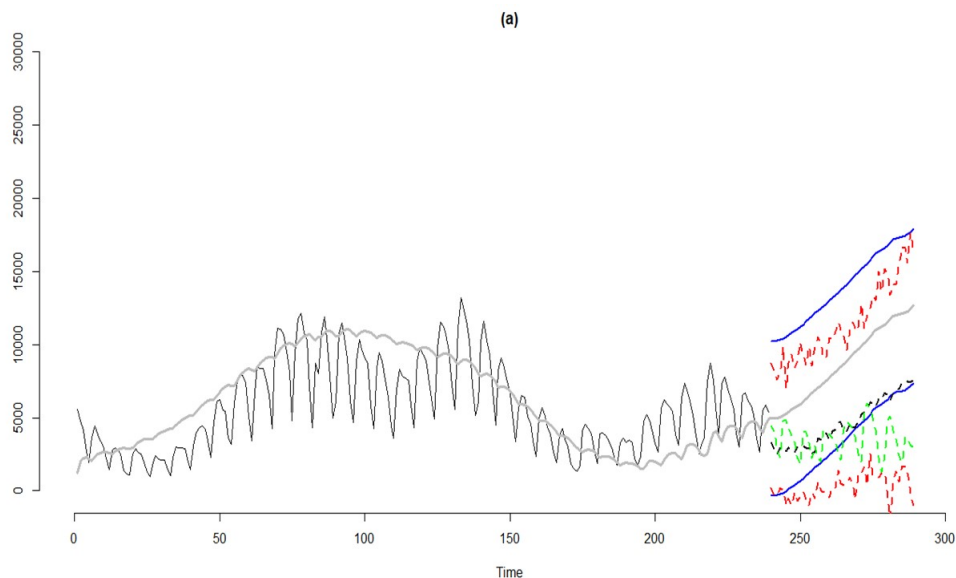
Las Figuras 4.6 y 4.7 muestran la predicción del bloque $\hat{Y}_{240:289}$ y sus respectivos intervalos de predicción 50 pasos adelante usando los procedimientos ST y SSB. En primer lugar, de ambas figuras, se logra evidenciar que el procedimiento SSB (líneas rojas) genera intervalos que son capaces de capturar de forma más precisa la predicción de los

Figura 4.5: Densidad empírica e histograma del error estandarizado de un paso adelante.



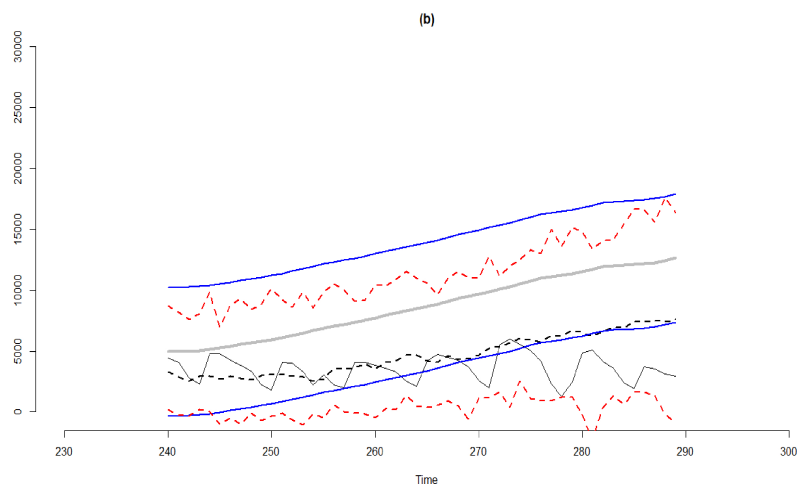
Fuente: Elaboración propia.

Figura 4.6: Pronóstico de varios pasos fuera de la muestra de las últimas 50 observaciones y bandas de predicción del 95 % considerando la data total disponible.



Fuente: Elaboración propia.

Figura 4.7: Pronóstico de varios pasos fuera de muestra de las últimas 50 observaciones y bandas de predicción del 95 %.



Fuente: Elaboración propia.

casos diarios (línea discontinua negra), el ajuste del modelo (línea gris) y la totalidad de los valores empíricos de la data disponible (línea verde), mientras que los intervalos ST (líneas azules) no logran contener la data empírica de de la serie en estudio 50 pasos adelante. Lo anterior, era de esperar dado que incorporan la incertidumbre debida a la estimación de los parámetros. En segundo lugar, tenga en cuenta que las observaciones correspondientes a la última medición del período diciembre 2022 a enero 2023 de los datos de casos diarios de COVID-19 en Chile, caen fuera de los intervalos de predicción de ST. Sin embargo, los intervalos de arranque aún contienen estos valores.

Capítulo 5

Conclusiones

Este estudio propone un nuevo procedimiento para obtener intervalos de predicción bootstrap en el contexto de procesos localmente estacionarios utilizando modelos de espacio estado. Este procedimiento se basa en obtener la densidad de futuras observaciones en un solo paso, incorporando simultáneamente las incertidumbres, las cuales son originadas por la estimación de parámetros y la distribución de errores.

En este trabajo implementamos los modelos de espacio estado donde los sistemas de las matrices varían en el tiempo. Otro aspecto importante es que los intervalos bootstrap son más adecuados que los intervalos estándar, principalmente en presencia de errores no normales. Por otro lado, nuestro procedimiento de arranque propuesto se implementa para obtener intervalos en valores futuros de una serie de datos sobre los casos diarios de COVID-19 en Chile. En este caso, los intervalos bootstrap considerados en este artículo son más precisos que los intervalos ST y logran capturar la predicción de los casos. Cabe señalar que esta metodología no se restringe únicamente al ámbito de las series temporales sobre la salud, como fue la aplicación en este estudio, ya que puede utilizarse en varias disciplinas cuyos datos contienen cambios suaves en su estructura de segunda orden. Por ejemplo, numerosos trabajos que tratan con parámetros variables en el tiempo han incorporado este procedimiento en procesamiento de señales y explorado en ciencias como la: geofísica, econometría neuro-

ciencia y finanzas. Estos son solo algunos ejemplos de una gran cantidad de series de tiempo no estacionarias que existen.

Finalmente, se debe resaltar la importancia de contar con una metodología que incorpore la incertidumbre en la estimación de los parámetros, la no estacionariedad y la no gaussianidad de los datos.

Bibliografía

- [1] Barney, B. et al. (2010). Introduction to parallel computing. *Lawrence Livermore National Laboratory*, 6(13):10.
 - [2] Brockwell, P. J. and Davis, R. A. (2002). *Introduction to time series and forecasting*. Springer.
 - [3] Chandler, G. and Polonik, W. (2006). Discrimination of locally stationary time series based on the excess mass functional. *Journal of the American Statistical Association*, 101(473):240–253.
 - [4] Dahlhaus, R. (1996). Asymptotic statistical inference for nonstationary processes with evolutionary spectra. In *Athens conference on applied probability and time series analysis*, pages 145–159. Springer.
 - [5] Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *The Annals of Statistics*, 25(1):1–37.
 - [6] Dahlhaus, R. (2000). A likelihood approximation for locally stationary processes. *The Annals of Statistics*, 28(6):1762–1794.
 - [7] Dahlhaus, R. (2012). Locally stationary processes. In *Handbook of statistics*, volume 30, pages 351–413. Elsevier.
 - [8] Dahlhaus, R. and Polonik, W. (2006). Nonparametric quasi-maximum likelihood estimation for gaussian locally stationary processes.
 - [9] Dahlhaus, R. and Polonik, W. (2009). Empirical spectral processes for locally stationary time series. *Bernoulli*, 15(1):1–39.
-

-
- [10] Durbin, J. and Koopman, S. J. (2000). Time series analysis of non-gaussian observations based on state space models from both classical and bayesian perspectives. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(1):3–56.
- [11] Durbin, J. and Koopman, S. J. (2012). *Time series analysis by state space methods*, volume 38. OUP Oxford.
- [12] Ferreira, G., Rodríguez, A., and Lagos, B. (2013). Kalman filter estimation for a regression model with locally stationary errors. *Computational Statistics & Data Analysis*, 62:52–69.
- [13] Gonçalves, A. M. and Costa, M. (2013). Predicting seasonal and hydro-meteorological impact in environmental variables modelling via kalman filtering. *Stochastic Environmental Research and Risk Assessment*, 27(5):1021–1038.
- [14] Hamilton, J. D. (1986). A standard error for the estimated state vector of a state-space model. *Journal of Econometrics*, 33(3):387–397.
- [15] Harvey, A. C. (1990). Forecasting, structural time series models and the kalman filter.
- [16] Jensen, M. J. and Whitcher, B. (2000). Time-varying long-memory in volatility: detection and estimation with wavelets. *preprint, University of Missouri and Eurandom*.
- [17] Kohn, R. and Ansley, C. F. (1986). Prediction mean squared error for state space models with estimated parameters. *Biometrika*, 73(2):467–473.
- [18] Ledesma, R. (2008). Introducción al bootstrap. desarrollo de un ejemplo acompañado de software de aplicación. *Tutorials in quantitative methods for psychology*, 4(2):51–60.
- [19] Mauricio, J. A. (2007). Análisis de series temporales. *Universidad Complutence de Madrid*.
- [20] Mergner, S. (2009). *Applications of state space models in Finance: An empirical*

- analysis of the time-varying relationship between macroeconomics, fundamentals and Pan-European industry portfolios*. Universitätsverlag Göttingen.
- [21] Nason, G. P. (2003). *Adaptive methods for modelling, estimating and forecasting locally stationary processes*. PhD thesis, Université catholique de Louvain.
- [22] Padilla, L. J. (2018). Implementación de un método de estimación y predicción espacio-tiempo con datos faltantes basado en filtro de kalman.
- [23] Palma, W. and Olea, R. (2010). An efficient estimator for locally stationary gaussian long-memory processes. *The Annals of Statistics*, 38(5):2958–2997.
- [24] Palma, W., Olea, R., and Ferreira, G. (2013). Estimation and forecasting of locally stationary processes. *Journal of Forecasting*, 32(1):86–96.
- [25] Pascual, L., Romo, J., and Ruiz, E. (2004). Bootstrap predictive inference for arima processes. *Journal of Time Series Analysis*, 25(4):449–465.
- [26] Pfeffermann, D. and Tiller, R. (2005). Bootstrap approximation to prediction mse for state–space models with estimated parameters. *Journal of Time Series Analysis*, 26(6):893–916.
- [27] Priestley, M. and Tong, H. (1973). On the analysis of bivariate non-stationary processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 35(2):153–166.
- [28] Priestley, M. B. (1965). Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 27(2):204–229.
- [29] Quenneville, B. and Singh, A. C. (2000). Bayesian prediction mean squared error for state space models with estimated parameters. *Journal of Time Series Analysis*, 21(2):219–236.
- [30] Rodriguez, A. and Ruiz, E. (2009). Bootstrap prediction intervals in state–space models. *Journal of time series analysis*, 30(2):167–178.
- [31] Rodriguez, A. and Ruiz, E. (2012). Bootstrap prediction mean squared errors of

- unobserved states based on the kalman filter with estimated parameters. *Computational Statistics & Data Analysis*, 56(1):62–74.
- [32] Solanas, A. and Olivera, V. (1992). Bootstrap: fundamentos e introducción a sus aplicaciones. *Anuario de psicología/The UB Journal of psychology*, pages 143–154.
- [33] Stoffer, D. S. and Wall, K. D. (1991). Bootstrapping state-space models: Gaussian maximum likelihood estimation and the kalman filter. *Journal of the american statistical association*, 86(416):1024–1033.
- [34] Tong, H. (1973). Some comments on spectral representations of non-stationary stochastic processes. *Journal of Applied Probability*, 10(4):881–885.
- [35] Wall, K. D. and Stoffer, D. S. (2002). A state space approach to bootstrapping conditional forecasts in arma models. *Journal of Time Series Analysis*, 23(6):733–751.
- [36] Wi, S., Valdés, J. B., Steinschneider, S., and Kim, T.-W. (2016). Non-stationary frequency analysis of extreme precipitation in south korea using peaks-over-threshold and annual maxima. *Stochastic environmental research and risk assessment*, 30(2):583–606.