CESAR, NECESITO QUE LA PRÓXIMA VEZ, POR FAVOR, ME ENTREGUE UN .DOCX CON BUENOS TÍTULOS Y CON UN FORMATO MÁS ADECUADO. UD LO PODRÍA HACER EN 10 MIN DE FORMA MANUAL UNA VEZ LO PROCESA EN MARKDOWN. ESTOY CONSCIENTE QUE UD QUIERE VER SOBRE TODO EL FONDO, PERO YO NECESITO IR VIENDO FONDO Y COMO ESTRUCTURA EL TRABAJO. ADEMÁS ME ES INCOMODO VER DE ESTA FORMA LOS PÁRRAFOS. ME COSTO BASTANTE CORREGIRLO ASÍ, ADEMAS DE QUE NO ES APRIADO COMO ENTREGA DE SEGUIMIENTO.

\subsection{Antecedentes}

Conocer el estado pasado, actual y futuro tanto de un mundo como de una economía a su vez tan volátil, ha producido un aumento en el análisis de los datos referidos en el tiempo . Este tipo de datos se encuentra en diferentes áreas, tanto en investigación académica como en el análisis de datos para la toma de decisiones. En el campo financiero es común hablar de la devaluación del colón con respecto al dólar, cantidad de exportaciones mensuales de un determinado producto o las ventas de este [@oscarh-1]. El estudio de las series cronológicas posee una particular importancia en el análisis de datos en la investigación de mercados bancarios y proyecciones demográficas que de manera conjunta apoyan la toma de decisiones para la aprobación presupuestaria en distintas áreas.

La información temporales tiene un peso muy relevante en la actual sociedad: el [Banco Mundial] (https://databank.worldbank.org/home.aspx)[^1] cuenta en su sitio web con datos para el análisis de series cronológicas de indicadores de desarrollo, capacidad estadística, indicadores educativos, estadísticas de género, nutrición y población. De manera similar al Banco Mundial, uno de los sitios más populares relacionados con el análisis de información es [Kaggle](https://www.kaggle.com/)[^2]. En su sitio web, Kaggle ofrece una gran cantidad de conjuntos de datos para poner a prueba distintas formas de análisis o bien, para participar en competencias. Actualmente, para el análisis de series cronológicas, Kaggle cuenta con más de 30 competiciones[^3]. Estos son unos de muchos casos donde se aplica de forma exhaustiva los análisis de series de tiempo y que evidencia su relevancia en el entorno actual.

Asimismo, los pronósticos (estimación futura en una particula serie temporal) son también utilizados en instituciones públicas, gobiernos municipales, instituciones del sector privado, centros académicos, población civil, centros nacionales o regionales de investigación y organizaciones no gubernamentales dedicadas al desarrollo social. Si las entidades previamente mencionadas cuentan con proyecciones de calidad, la puesta en marcha de sus respectivos planes tendrá un impacto mayor y más efectivo.

Los métodos existentes para llevar a cabo un análisis de series cronológicas son diversos, y responden al propio contexto y tipo de datos: en la actualidad obtener buenos pronósticos o bien explicar el comportamiento de un fenómeno en el tiempo sigue siendo un tema recurrente de investigación. Generar una adecuada estimación es fundamental para obtener un pronóstico de confianza, y así asegurar la confianza en el proceso o investigación que se esté utilizando el método para dicho fin. Resulta importante mencionar una diferencia clave entre los dos modelos clásicos más utilizados: los modelos de suavizamiento exponencial y los modelos ARIMA. Ambos representan enfoques complementarios a un problema, pues los modelos de suavizamiento exponencial se fundamentan en un enfoque más descriptivo de los componentes de la serie cronológica en estudio, mientras que los modelos ARIMA tienen como objetivo explicar las relaciones pasadas de ésta [@hyndman2018forecasting].

El método ARIMA se fundamente en las autocorrelaciones pasadas, y contempla un proceso iterativo para identificar un posible proceso óptimo a partir de una clase general de modelos… El teorema de Wold sugiere que todo proceso estacionario puede ser determinado de una forma específica y cuya ecuación posee, en realidad, infinitos coeficientes, pero que debe ser reducido a una cantidad finita para luego evaluar su ajuste sometiéndolo a diferentes pruebas y medidas de rendimiento.

Al trabajar con la metodología de Box-Jenkins, uno de los pasos a concretar es identificar los parámetros autorregresivos y de medias móviles que gobiernan la serie. Para indagar los términos en el proceso de investigación se ha utilizado la identificación de parámetros mediante autocorrelogramas parciales y totales, sin embargo, los autocorrelogramas formados no analizan de forma expedita, exhaustiva y óptima los posibles coeficientes que podrían contemplarse en la la mencionada ecuación de Wold, pues como se mencionó, esta posee infinitos coeficientes, razón por la cual el método propuesto busca aproximar de una mejor manera la identificación del modelo cubriendo un mayor número de posibilidades de dichos parámetros mediante lo que suele considerarse un problema: la sobreparametrización.

[^1]: [https://databank.worldbank.org/home.aspx](https://databank.worldbank.org/home.aspx)

[^2]: Se trata de una subsidiaria de la compañía Google que sirve de centro de reunión para todos aquellos interesados en la ciencia de datos.

[^3]: Muchas de ellas incluyen recompensas económicas que van desde los \$500 hasta los \$100,000 para aquellos que logren obtener los mejor pronósticos.

\subsection{El problema}

Al identificar un proceso ARIMA, el criterio de los autocorrelogramas aporta una aproximación al proceso que gobierna la serie, representando en muchos casos una pobre alternativa dado que subestiman o sobre estiman el número óptimo de parámetros que se debería de contemplar en el proceso de estimación. Aunado a esto, existirá el problema de la subjetividad, pues a pesar de que alguien proponga un patrón que gobierne la serie, otra persona podría tener otra interpretación del proceso, proponiendo así una especificación diferente. Esto sucede incluso con cantidades moderadas de series cronológicas a analizar, por lo que la generación de algoritmos que ayuden a esta identificación se vuelve cada vez más necesaria [@auto.arima], siendo estos métodos una valiosa alternativa ante los muchas veces pobres resultados en la identificación correcta del modelo.

Han sido varias las aproximaciones a un método que genere de manera automática un modelo ARIMA, siendo uno de los métodos automatizados de estimación más populares es el que ofrece el paquete `forecast` [@auto.arima] del lenguaje de programación R[^4] y que permite hacer uso de la función `auto.arima()` para estimar un modelo ARIMA basado en pruebas de raíz unitaria y minimización del AICc [@burnham2007model]. Así se obtiene un modelo temporal definiendo las diferenciaciones requeridas en la parte estacional $d$ mediante las pruebas KPSS [@doi:10.1111/1467-9892.00213] o ADF [@fuller1995introduction], y la no estacional $D$ utilizando las pruebas OCSB [@Osborn2009SEASONALITYAT] o la Canova-Hansen [@10.2307/1392184], seleccionado el orden óptimo para los términos $ARIMA(p, d, q)(P, D, Q)\_s$ para una serie cronológica determinada. Sin embargo, alternativas como la función `auto.arima()` no someten a prueba las posibles especificaciones de un modelo en un rango determinado, concepto conocido como sobreparametrización, dejando así un vacío en el cual se corre el riesgo de no seleccionar un modelo que ofrezca mejores pronósticos. Acá debe poner algo como: “Por lo tanto, un método fundamentado en la sobreparamestrización… ” y bla bla bla y las flores del caso…

[^4]: Descarga gratuita en [https://cran.r-project.org/](https://cran.r-project.org/)

\subsection{Objetivos del estudio}

El objetivo del presente trabajo es diseñar un algoritmo para la selección de modelos ARIMA según la temporalidad de la serie vía sobreparametrización.

Para lograr esto, se prente: :

Acá no falta uno que sea “Crear algo de permutación que evalúe las posibles cominaciones de términos AR y MA en la identificación de un proceso”.

\*\*1.\*\* Aplicar validación cruzada en distintos horizontes de pronóstico para identificar la mejor especificación de un modelo ARIMA.

\*\*2.\*\* Integrar el desarrollo de la metodología de análisis de series temporales en una librería del lenguaje estadístico R.

3Contrastartantode la estimación así como la generación de otros,

OJO: “como el propuesto por Rob Hyndman, de la Oficina de Censos de los Estados Unidos, entre otros”, eso no hace falta decirlo acá, luego en el capítulo de metodología se abordará.

\subsection{Metodología en la aplicación de series cronológicas}

La aplicación de las series cronológicas persigue tres objetivos: 1) el análisis exploratorio de la serie en cuestión, 2) estimar modelos de proyección, y, 3) generar pronósticos para los posibles valores futuros que tomará el problema en cuestión. Los tres objetivos anteriores se pueden trabajan de manera conjunta y secuencia, pues es necesario realizar primero el análisis exploratorio de los datos para tener una noción global del panorama y así conocer la serie cronológica con la que se está trabajando. Asimismo, existen múltiples formas de proceder mediante la etapa de estimación, como lo son los métodos de suavizamiento exponencial [@brown], modelos de regresión para series temporales [@kedem], redes neuronales sequenciales aplicadas a datos longitudinales [@redes], estimaciones bayesianas [@bayes], y finalmente, los procesos autorregresivos integrados de medias móviles o ARIMA por sus siglas en inglés [@box-jenkins], siendo estos últimos el foco de interés en este estudio, pues se trata de modelos que se basan en las relaciones pasadas de la propia serie cronológica; es decir, toman como referencia las correlaciones entre los valores actuales y pasados de la serie para entender el comportamiento de la misma en el futuro.

Como menciona Rob. Hyndman [@hyndman\_box-jenkins], la metodología de Box-Jenkins difiere a los demás métodos porque no supone un determinado patrón en la serie cronológica, sino que parte de un proceso iterativo para identificar el modelo de un gran grupo de estos para luego ponerlo a prueba según varias medidas de rendimiento. Un proceso ARIMA es caracterizado por dos funciones: la autocorrelación y la autocorrelación parcial; es mediante la comparación de dichas funciones que la metodología Box-Jenkins busca la identificación el proceso que describa de manera adecuada el comportamiento de una serie cronológica.

Con tal de mejorar la precisión y calidad del modelo estimado es que el presente trabajo propone una metodología para la estimación un modelo ARIMA de una serie cronológica determinada abarcando más posibilidades que los enfoques tradicionales. Se consideran temporalidades mensuales, bimensuales, trimestrales o cuatrimestrales, mediante un proceso de selección fundamentada en las permutaciones de todos los parámetros de un modelo ARIMA hasta en un rango determinado, considerando la inclusión semiautomática de intervenciones en periodos específicos y la validación cruzada para evaluar la calidad de las particiones de la base de datos en conjuntos para entrenar y probar el rendimiento del modelo. Dichas pruebas involucran criterios de información como el AIC, el AICc y el BIC, además de medidas de rendimiento como el MAE, RMSE, MAPE y MASE, las cuales sirven de insumo para utilizar un método de consenso entre ellas y seleccionar el modelo más adecuado mediante la sobreparametrización: se comparan todos los posibles en in intervalo específico de términos definiendo una diferenciación adecuada para la serie y permutando hasta un máximo definido para los términos autorregresivos y de medias móviles especificados para así seleccionar la especificación que ofrezca mejores resultados al momento de pronosticar valores futuros de la serie cronológica.

CÉSAR, CREO QUE LA SUB SECCIÓN “Metodología en la aplicación de series cronológicas”, NO ESTÁ FORMULADA DE FORMA EXPEDITA. ACÁ NO DEBEMOS HACER MUCHO ÉNFASIS DE ESTADÍSTICOS Y OTROS. SERÍA PRESENTAR DE FORMA GENERAL LA FORMA DE PROCEDER, Y LUEGO ATACAR QUE EN UN DETERMINADO “PUNTO” DE LA METODOLOGÍA DE UNA SERIE, ES QUE SE QUIERE INVESTIGAR LA PROPUESTA EXPUESTA EN LA PROBLEMÁTICA.

TAMBIÉN CUIDADO CON LOS GRANDES PÁRRAFOS CON TANTAS IDEAS, Y ADEMÁS DE LA FORMA QUE SE EXPRESAN. APRECIAR EL FONDO CON FALENCIAS DE EXPRESIÓN SE HACE COMPLICADO.

\subsection{Justificación del estudio}

El accionar de políticas gubernamentales, así como de otro tipo de sectores, se apoya cada vez más en un acertado análisis de la información temporal de los hechos observados hasta una fecha determinada, y por ende sus posibles evoluciones en el futuro. En el campo la demografía, uno de los principales temas de investigación son las proyecciones de población; durante una emergencia, conocer la posible cantidad de población que habita una zona es clave para la rápida reacción de las autoridades en el envío de ayuda o en la ejecución de planes de evacuación. Este tipo de situaciones ocurren en Costa Rica y en muchos lugares del mundo ante condiciones climáticas adversas como huracanes, inundaciones, o más recientemente ante la pandemia del COVID-19. También, el campo actuarial se ve beneficiado al mejorar sus métodos de pronóstico, pues una de sus principales áreas de estudio es la mortalidad, ya que representa un insumo de vital importancia para la planificación y sostenibilidad de los sistemas de pensiones, servicios de salud tanto pública como privada, seguros de vida y asuntos hipotecarios [@supenprodc].

El principal aporte de este estudio es, por medio de un proceso de simulación, brindar evidencia sobre cómo la sobreparametrización puede contribuir a definir la especificación de un modelo ARIMA que genere pronósticos de mayor calidad, contrastando la calidad de estos con respecto a otros métodos similares, como lo son las funciones `auto.arima()` o `seas()`.

Tras desarrollar y probar el método mediante datos simulados, la aplicación real del algoritmo mostrará el potencial de la sobreparametrización en la estimación de modelos ARIMA. En el campo demográfico, por ejemplo, las estadísticas vitales son sistematizadas y divulgadas año tras año, por tanto, revelan los cambios acontecidos durante este periodo. Esta información junto con la proveniente de los censos de población constituye la base para construir los diferentes índices, tasas y otros indicadores que revelan la situación demográfica del país, información de gran relevancia para la planificación nacional, regional y local en diversos campos. Uno de estos principales campos de acción es la salud pública, para la cual la tasa de mortalidad infantil se considera uno de los indicadores prioritarios dado que refleja no solo las condiciones de salud de la población infante, sino también los niveles de desarrollo del país, pues depende de la calidad de la atención de la salud, principalmente de la prenatal y perinatal, así como de las condiciones de saneamiento. Por tanto, su continuo monitoreo es fundamental para diseñar, implementar y evaluar políticas de salud pública orientadas a disminuir y erradicar aquellas que son prevenibles [@calidad\_vitales].

CREO QUE LA JUSTIFICACIÓN ES BIEN DECIR QUE EN MUCHOS CAMPOS SE LLEVAN A CABO ESTIMACIÓN DE SERIES DE TIEMPO PARA PODER PRONOSTICAR LO QUE SUCEDERÁ EN LOS PRÓXIMOS “H” PERIODOS. A PARTIR DE AHÍ DEBE DECIR CORRECTAMENTE QUE EN LA ACTUALIDAD EXISTEN MÉTODOS AUTOMÁTICOS COMO EL AUTO.ARIMA() Y OTROS QUE LO HACEN, PERO QUE POR SU MODALIDAD O ALGORITMO DE IMPLEMENTACIÓN, NO TOMAN EL O LOS MÁS ADECUADOS PROCESOS, Y QUE ENTONCES SE PODRÍA SUB O SOBRE ESTIMAR EL VALOR DEL PRONOSTICO COMO TAL. A PARTIR DE ACÁ MENCIONAR QUE EL MÉTODO PROPUESTO DE ESTIMACIÓN POR PERMUTACIÓN ES UNA FORMA COMPLEMENTARIA Y MÁS EXHAUSTIVA DE ESTIMACIÓN, QUE BUSCA MITIDAR LOS PROBLEMAS DE LOS ACTUALES MÉTODOS DE ESTIMACIÓN. FINALMENTE, UNA MEJOR ESTIMACIÓN SE TRADUCE POR MEJORES PRONÓSTICOS, Y POR ENDE POR POLITICAS PÚBLICAS, PROYECCIONES DEMOGRÁFICAS Y OTROS, MÁS ACERTATOS. EL FOCO DE LA INVESTIGACIÓN ESTÁ EN DECIR QUE ESTE NUEVA FORMA DE ESTIMAR POR SOBRE PARAMETRIZACIÓN MEJORA EL PROCESO DE PRONÓSTICO.

LA ESENCIA ESTÁ, COMO SE EXPRESA NO LO HACE VER CLARAMENTE. TIENE QUE SER MÁS EXPEDITO A LA HORA DE PLASMAR LAS IDEAS, ADEMÁS DE EVITAR PONAR BLA BLA INNECESARIO.

\subsection{Organización del estudio}

EL presenta trabajo de investigación consta de cinco capítulos. El presente capítulo ofrece una contextualización del uso de las series de tiempo, así como la importancia de poder contar con pronósticos de calidad. Se presentó el objetivo del estudio, así como una breve descripción de la metodología empleada en la aplicación de series temporales, y cómo se planea modificar el método de estimación en los modelos ARIMA. Se concluye esta sección con hechos que justifican la importancia de esta investigación.

El siguiente capítulo consiste en el marco teórico, abarcando aspectos fundamentales de la ecuación de Wold, la metodología Box-Jenkins, la selección de los procesos que gobiernan la serie, la descripción del proceso iterativo, entre otros.

En el tercer capítulo se describe en detalle toda la metodología relacionada al estudio, iniciando con una descripción global de los conceptos más fundamentales del análisis de series cronológicas, pasando por los componentes fundamentales de las mismas: tendencia, estacionalidad, ciclos e irregularidades. En este capítulo se discuten también los supuestos clásicos del análisis de series cronológicas, los distintos tipos de modelos, el análisis de intervención, la validación cruzada y las medidas de rendimiento; aspectos cruciales para obtener un modelo ARIMA vía sobreparametrización. La sección metodológica culmina con la descripción del proceso de simulación que se utilizará, así como la discusión del método propuesto.

El capítulo cuatro consiste trata la presentación de los resultados, tanto en los datos simulados como en la aplicación a datos costarricenses y se contrastarán contra los obtenidos por otros métodos como el de la función `auto.arima()`, entre otros.

El último capítulo busca discutir los principales resultados, así como señalar las conclusiones más importantes y ofrecer algunas recomendaciones que orienten futuros estudios relacionados.

Finalmente se ofrece en la sección de anexos los códigos en lenguaje R más importantes utilizados en el estudio, así como una sección dedicada a las referencias bibliográficas obtenidas para esta investigación.