



ESCUELA DE NEGOCIOS

MAESTRÍA EN INTELIGENCIA DE NEGOCIOS Y CIENCIA DE DATOS

**ANÁLISIS COMPARATIVO ENTRE LOS MODELOS SARIMAX Y STAR
PARA LA PREDICCIÓN DEL ÍNDICE DE PRECIOS AL CONSUMIDOR
DURANTE EL FENÓMENO DEL NIÑO EN EL PERIODO 2023-24**

**Profesor
Víctor Manuel González Holguín**

**Autor
Andrés Chasiliquin**

2024

RESUMEN

Los daños causados por eventos climatológicos extremos afectan al bienestar de la ciudadanía ecuatoriana. Uno de los más fuertes considerados en la historia del Ecuador es el fenómeno del Niño, el cual afectó negativamente a la economía ecuatoriana. El Niño tuvo un impacto importante en los precios nacionales de los alimentos durante el transcurso de los años 2015 y 2016. El presente proyecto se centra en generar una proyección sobre el índice de precios al consumidor durante la presencia del fenómeno del Niño en el periodo 2023-24.

Utilizando modelos predictivos de series de tiempo lineales y no lineales, se plantea realizar una comparación respecto al poder predictivo de estas dos técnicas. A través de medidas estadísticas como el coeficiente de determinación y el error cuadrático medio, se mide la precisión predictiva de los modelos y se define cuál tiene mejor rendimiento. Los resultados sugieren que un modelo no lineal tiene un mejor ajuste en la trayectoria del índice de precios, así como también un menor error en sus predicciones. Se concluye que la relación entre los precios de los alimentos y un evento meteorológico, asume un comportamiento no lineal. Esto causa que modelos que contemplan tanto la parte no lineal como lineal de estas relaciones sean los que tengan un mayor poder predictivo.

Palabras clave: Fenómeno del Niño, precios, SARIMAX, STAR

ABSTRACT

Damages caused by extreme weather events affect the well-being of Ecuadorian citizens. One of the strongest considered in Ecuador's history is the El Niño phenomenon, which negatively affected the Ecuadorian economy. El Niño had a significant impact on national food prices during the years 2015 and 2016. The present project focuses on generating a projection on the consumer price index during the presence of the El Niño phenomenon in the period 2023-24.

Using predictive models of linear and nonlinear time series, it proposes to make a comparison regarding the predictive power of these two techniques. Through statistical measures such as the coefficient of determination and the mean square error, the predictive accuracy of the models is measured and which one has the best performance is defined. The results suggest that a nonlinear model has a better fit in the trajectory of the price index, as well as a lower error in its predictions. It is concluded that the relationship between food prices and a meteorological event assumes a nonlinear behavior. This means that models that consider both the nonlinear and linear part of these relationship have the greatest predictive power.

Key words: El Niño phenomenon, price, SARIMAX, STAR

ÍNDICE DEL CONTENIDO

1.	RESUMEN.....	2
2.	ABSTRACT	3
3.	INTRODUCCIÓN.....	1
4.	REVISIÓN DE LITERATURA	2
4.1	Relación de ENOS con la economía alimentaria del ecuador.....	2
4.2	Modelos de series de tiempo lineales y no lineales	5
4.3	Casos de Uso, aplicación en la econometría	5
5.	IDENTIFICACIÓN DEL OBJETO DE ESTUDIO.....	9
6.	PLANTEAMIENTO DEL PROBLEMA.....	11
7.	OBJETIVO GENERAL.....	12
8.	OBJETIVOS ESPECÍFICOS	13
9.	JUSTIFICACIÓN Y APLICACIÓN DE LA METODOLOGÍA.....	14
9.1	Justificación de la metodología	14
9.2	Selección.....	14
9.2.1	Recolección de datos	14
9.3	Preprocesamiento	16
9.3.1	Limpieza y Preprocesamiento.....	16
9.4	Transformación	17
9.4.1	Transformación y carga de datos	17
9.4.2	Identificación y descripción de variables.....	18
9.4.3	Visualización de las variables	22
9.5	Minería de datos	25
9.5.1	Selección de modelo estadístico	25
10.	RESULTADOS	32
11.	DISCUSIÓN DE LOS RESULTADOS Y PROPUESTA DE SOLUCIÓN	39

11.1	Beneficios por la aplicación de la solución implementada	39
11.2	Diseño de estrategia organizacional	41
11.3	Estrategia organizacional, innovación y competitividad empresarial	42
12.	CONCLUSIONES Y RECOMENDACIONES.....	44
12.1	Conclusiones	44
12.2	Recomendaciones	45
13.	Referencias	46

ÍNDICE DE TABLAS

Tabla 1 Variación del precio de la división alimentos	4
Tabla 2 Diccionario de datos de las bases de datos IPC y ENOS	18
Tabla 3 Prueba de raíz unitaria	27
Tabla 4 Prueba LM No Linealidad para DIPC.....	28
Tabla 5 Resultados MSE y R2 para las variables IPC y ENOS.....	32
Tabla 6 RMSE del forecast para el IPC, modelo STAR	33
Tabla 7 Valor de los parámetros de suavización y ubicación del modelo ESTAR	34
Tabla 8 Estimaciones de parámetros y rendimiento en modelos SARIMAX	35
Tabla 9 Prueba de correlación serial	38

ÍNDICE DE FIGURAS

Figura 1 Matriz de correlación IPC vs ENOS	20
Figura 2 Anomalías de temperatura en la superficie del mar periodo 1996-22. 21	
Figura 3 Gráfico de líneas, comparaciones variables TSM de la región 1.2 y 3.4	22
Figura 4 IPC de legumbres, Alimentos y Alimentos y bebida no alcohólicas	23
Figura 5 IPC Alimentos y bebidas no alcohólicas y anomalías TSM 1.2 y 3.4..	24
Figura 6 Gráfico de dispersión entre el IPC Alimentos bebidas no alcohólicas y las anomalías TSM 1.2 y 3.4	25
Figura 7 Comparación entre las funciones ESTAR y LSTAR	29
Figura 8 Predicciones del IPC periodos 2015-16 y 2023-24	36
Figura 9 Forecast del índice de precio al consumidos de los alimentos	37

INTRODUCCIÓN

La creciente literatura entre las relaciones económicas y el clima, propone metodologías, técnicas e instrumentos de investigación que pretenden explicar la relación entre estos dos actores. Entender su comportamiento permite afrontar riesgos sociales, económicos o de seguridad que impactan a una nación. A pesar de los esfuerzos para analizar y predecir este sistema complejo, la utilización de modelos tradicionales tienen límites en cuanto estos pueden explicar. El manejo de modelos de previsión no tradicionales, examinará el vínculo entre los actores, permitiendo comprender su relación y a su vez generar predictores más precisos.

El presente proyecto plantea el desafío de crear una herramienta analítica, que permita predecir el comportamiento del índice de precios al consumidor nacional durante la presente temporada del fenómeno del Niño. Con este fin se pretende comparar dos modelos de previsión: autorregresivo de transición suave (STAR) y autorregresivo estacional integrando medias móviles y entradas exógenas (SARIMAX). Se espera que el modelo con el mejor rendimiento sea de utilidad para los sectores agropecuarios y de salud pública del gobierno ecuatoriano. Afectando a su variable riesgo, este brindará una mejora en la planificación y administración de los recursos del estado para el bienestar social.

El proyecto se encuentra estructurado en tres partes: revisión de la literatura, aplicación de metodología y la discusión de los resultados. En cada una de las secciones se revisará la relación entre los actores y se analizará el resultado final producto de la aplicación de los modelos predictivos.

REVISIÓN DE LITERATURA

4.1 Relación de ENOS con la economía alimentaria del Ecuador.

El Niño Oscilación del sur (ENOS) es un fenómeno meteorológico que trae consigo cambios en la temperatura del océano SST en la región del pacífico tropical (NOAA, 2024a). Las alteraciones de la temperatura en el mar varían entre cálidas y frías, denominado a las cálidas como fase del Niño y las frías como fase de la Niña. Estos eventos climatológicos se presentan en un transcurso de tiempo irregular, puede darse desde los 2 hasta 7 años después.

Los efectos que trae consigo ENOS varía según la época del año, intensidad, región afectada y otros factores ambientales como el calentamiento global. Los efectos de un prolongado avistamiento del Niño traen consigo aumentos de la temperatura y un bajo índice de precipitación. En comparación con la Niña, su efecto es el contrario del Niño, temperaturas más bajas y una mayor cantidad de precipitaciones. Bajo estas condiciones, sectores económicos de un país como el agropecuario, son sensibles a cambios climatológicos. Lluvias prolongadas o extensión en las sequías tienen consecuencias en la producción de las cosechas, que trae repercusiones en la asignación de un precio estándar a los alimentos en el tiempo.

En el Ecuador, la presencia de ENOS durante una época del año puede traer consigo efectos adversos a los que se espera por la temporada o incrementar los efectos de la temporada. La temporada invernal se caracteriza por las intensas lluvias en el transcurso del día, esto provoca desbordamiento de los ríos, inundaciones o movimientos en masa. Durante esta temporada, los alimentos se ven condicionados por el factor climatológico. Ahora con la presencia de ENOS, el factor climático puede agravarse o suavizarse. Como consecuencia los precios se pueden mantener o presentar una variación creciente, es decir una inflación respecto al índice del precio.

En los años 1997 a 1998 en el Ecuador, la presencia del fenómeno del Niño produjo pérdidas 2.882 millones de dólares, equivalentes al 14.58% del Producto Interno Bruto. En el sector productivo el monto de daños fue 1.516 millones de

dólares, el más afectado con un 53% del daño total. De este el 41% se asignó a daños de la producción agrícola (CAF, 2000).

En cuanto a los alimentos, los cultivos del litoral fueron los principales afectados por la presencia de inundaciones consecuencia del desbordamiento de ríos. No obstante, algunas de estas parcelas afectadas se verían con una mejora en la producción a mediano plazo por la cantidad de sedimentación trasladada por las aguas. Provocaría una tendencia decreciente en el precio de ciertos alimentos después del shock del Niño.

Si se remonta al fenómeno del niño más reciente, este se produjo en el periodo 2015-16. De igual manera que el año 1997, este afectó negativamente a la economía del país. Se estimaron 3.5 millones de dólares en pérdidas del sector agropecuario con más de 2 mil productores entre medianos y pequeños, afectados (FAO, 2023). Cosechas como la papa se vieron reducidas en este periodo por el incremento de la humedad y las altas temperaturas (Martínez et al., 2017), provocando el alza en los precios de la papa por una menor oferta en el mercado ecuatoriano. El fenómeno del Niño se avistaba principalmente en provincias de la costa como es el caso de Guayas, donde su incidencia se incrementa durante el primer trimestre del 2016 (SGR, 2015). No obstante, las predicciones sobre cómo afectaría esta anomalía climática a largo plazo aún eran desconocidas.

A medida que se desarrollaba el fenómeno del Niño, se presenciaron varios desastres naturales entre los más recurrentes se encontraban las inundaciones y los deslaves. Como resultado actividades agrícolas producidas en terrenos llanos de gran extensión, como el banano, afectaron al comercio exterior del país. En 2016, el Ecuador generaría por la exportación de banano un valor de \$2.742.005,20 millones (WITS, 2016), este decreció 2.76% respecto al año anterior por daños en los cultivos.

Estudios realizados por el Banco Interamericano de Desarrollo (BID), estimaron que los países más afectados por el Niño durante los periodos de noviembre 2015 a junio 2016, son Perú y Ecuador con un impacto en el PIB superior a 1.5%.

Por el lado de precios, se podría tener un aumento entre 1 y 5% del nivel de precio. El análisis es respaldado en base a los daños ocasionados por este fenómeno en los años 1982-83 y 1997-98 (Martin, 2016).

En la Tabla 1 se muestra el índice y la variación de los precios al consumidor para la división alimentos. Se tiene que el primer semestre del 2016 existe un incremento en la tasa de inflación mensual de los alimentos en comparación al segundo semestre 2015. Los mayores valores en la inflación mensual, se registran a partir de febrero del 2016, lo que respalda a lo mencionado por Martin y el BID. No obstante, el mayor porcentaje en la tasa de inflación mensual para el periodo 2015-16 se dio en abril 2015 con un valor igual a 1,63%. Cabe mencionar que la inflación anual tuvo un incremento constante durante este periodo, fechas donde se presenta el Niño. A pesar de que en 2016 la tasa mínima fue del 2,78%, menor que la tasa mínima del 2015, se puede constatar la creciente subida del precio de los alimentos.

Tabla 1 Variación del precio de la división alimentos.

Fuente: Índice de Precios al Consumidor (IPC), INEC

Mes	Índice	Inflación Mensual (%)	Inflación Anual (%)	Inflación Acumulada (%)
jun-15	105,08	0,42	6,17	3,65
jul-15	104,76	-0,30	5,13	3,34
ago-15	104,81	0,04	4,40	3,39
sep-15	104,96	0,15	4,54	3,54
oct-15	104,56	-0,39	3,87	3,14
nov-15	104,60	0,05	3,32	3,19
dic-15	105,06	0,44	3,64	3,64
ene-16	105,23	0,16	3,35	0,16
feb-16	106,02	0,75	3,69	0,91
mar-16	106,69	0,64	3,86	1,55
abr-16	107,36	0,62	2,83	2,19
may-16	107,54	0,17	2,78	2,36
jun-16	108,08	0,50	2,86	2,87
jul-16	107,78	-0,28	2,88	2,58

4.2 Modelos de series de tiempo lineales y no lineales

Los modelos de predicción autorregresivos y de medias móviles han evolucionado en el transcurso de los años, hasta convertirse en una de las herramientas más usadas en la predicción de series temporales. En un inicio, estos modelos analizan un evento sucesivo a lo largo del tiempo. Los modelos AR-MA combinaban dos partes, la parte autorregresiva generada de una combinación lineal de sus valores pasados y la parte media móvil que predice un valor a partir del ruido de pasadas predicciones (Micaela et al., 2023). Estos modelos llegaron a conseguir resultados satisfactorios en el campo empresarial al emplearse para la creación de forecast de ventas. No obstante, la necesidad de predecir variables más complejas con un mayor grado de asimetría y transitoriedad provocó que estos modelos AR-MA pierdan precisión. Como consecuencia se incorporó nuevas funcionalidades a este modelo como la adición de variables exógenas y nuevos órdenes para la no estacionalidad.

A pesar de todas las versiones extendidas de los modelos AR-MA, aun no se mejoraba la predicción de aquellas variables no lineales. De esta forma se crearon nuevos modelos que ajustaban su ecuación matemática para combinar la parte lineal con la no lineal de estas series temporales (Neog et al., 2022). Modelos de transición suave o de redes neuronales permitieron contemplar la linealidad y no linealidad de las variables para generar proyecciones más acertadas. A través del uso de métodos como el de suavización, estos modelos proporcionan un mejor ajuste en la trayectoria de la serie temporal.

4.3 Casos de Uso, aplicación en la econometría

La creciente literatura entre las relaciones económicas y el clima, propone metodologías que intentan explicar el vínculo entre estas variables. Por medio del uso de métodos matemáticos y estadísticos en el análisis de datos se pretende expandir el conocimiento de estas teorías económicas. A continuación se presentan investigaciones que aportan conocimiento sobre la relación entre el índice de precio al consumidor (IPC) y un evento meteorológico. También se detallan los métodos estadísticos y matemáticos recomendados para el tipo de relación descubierta en las variables.

Relación no lineal entre el fenómeno climático El niño y los precios de los alimentos colombianos (Abril-Salcedo et al., 2020)

El caso de estudio, plantea un modelo autorregresivo de transición suave (STAR) para analizar el impacto de ENOS en la variable crecimiento de la inflación en los alimentos (DINF). El DINF se interpreta como la inflación en los precios de los alimentos al consumidor, esta tiene una periodicidad mensual. Como representante de ENOS, se analiza la variable exógena anomalía de la temperatura en la superficie del mar (TSM) en la región 4.3. Utilizando un gráfico de dispersión se observa una tendencia polinómica en forma de U entre la inflación de los precios y ENOS. Usando pruebas LM de no linealidad, se corroboran los resultados gráficos y se concluye que un modelo no lineal (STAR) es el mejor candidato para evaluar el impacto de ENOS sobre DINF.

El conjunto de observaciones, comienzan en el mes de marzo de 1962 y termina en diciembre de 2018. El modelo se desarrolló con 682 observaciones, cada variable. El análisis matemático contempla tanto el componente lineal y no lineal entre ENOS y DINF. De esta manera se observa que ENOS tiene impactos asimétricos y transitorios sobre el crecimiento de la inflación. Es decir, a mayor impacto en el fenómeno del Niño, el DINF refleja cambios considerables en su tendencia. No obstante, el evento de la Niña manifiesta efectos ambiguos en la tendencia del precio (Abril-Salcedo et al., 2020).

Efecto de las precipitaciones en la predicción de los precios del tomate en la India: una aplicación del Modelo SARIMAX y NARX. (Kumar et al., 2022)

Se plantea una comparación entre dos modelos: Modelo autorregresivo estacional integrando medias móviles y entradas exógenas (SARIMAX) y el modelo exógeno autorregresivo no lineal (NARX). Durante su desarrollo se plantea estimar el modelo con el mejor pronóstico respecto al precio del tomate. A continuación se detallan las ventajas de cada modelo.

Por una parte, los modelos SARIMAX, versión extendida del tradicional ARIMA, incorporan nuevas funcionalidades en sus hiperparámetros para especificar la selección del orden estacional y no estacional del modelo. Útiles en series de

tiempo que presentan patrones estacionales. Además, este modelo puede adicionar variables exógenas que brindan un efecto en la serie al momento de realizar una predicción. A diferencia de su modelo tradicional, SARIMAX plantea una mejora al momento de brindar predicciones más precisas y completas considerando factores externos que afectan a la variable endógena.

Por otro lado, los modelos NARX, útiles para predecir series de tiempo no lineales, incluyen la relación con variables externas. Un ejemplo de esto último, son los fenómenos meteorológicos tales como precipitaciones, máximo y mínimo de temperatura, humedad, cantidad de viento y sequías. Al igual que SARIMAX, la introducción de variables externas o exógenas al modelo, tiene un efecto en el pronóstico de la variable endógena.

Para elegir el mejor modelo se especifica un criterio de evaluación. El criterio se basa en los resultados del pronóstico del precio del tomate. Medidas estadísticas como la raíz del error cuadrático medio RMSE y el error de porcentaje absoluto medio MAPE, son comparados entre los modelos para llegar a una conclusión. Los resultados revelaron que el modelo NARX tuvo menor raíz de error cuadrático medio respecto al modelo SARIMAX, de igual manera en el estadístico MAPE. A pesar de que el modelo SARIMAX no logró los efectos deseados, su desarrollo fue vital. A través de la prueba de Brock-Dechert-Scheinkman (BDS) para la no linealidad de los residuos, se demostró la no linealidad entre las variables precio y precipitación con el modelo SARIMAX.

El Niño y su pronóstico en la volatilidad observada en los precios del petróleo (Bouri et al., 2021)

Bouri, propone un modelo autorregresivo heterogéneo HAR-RV para la predicción del precio del petróleo bajo la influencia de un agente exógeno como lo es ENOS. Se trabajó con una serie de tiempo de periodicidad mensual en un rango de tiempo que comienza el 2 de enero de 1986, hasta el 31 de diciembre del 2020. Como variable exógena se utiliza el Índice de Oscilación del Sur Ecuatorial (EQSOI) para monitorizar el ciclo ENOS. Los resultados muestran que el modelo HAR-RV considerando la variable EQSOI, tiene un menor error

cuadrático medio de predicción (MSPE) en comparación al modelo HAR-RV sin esta variable exógena. Por lo tanto, se demuestra que el Índice de Oscilación del Sur Ecuatorial brinda predicciones más precisas en periodos de tiempo cortos y largos. Se concluye que ENOS tiene valor predictivo para datos dentro de la muestra. Sin embargo, no necesariamente estos resultados se reflejan en datos fuera de la muestra.

IDENTIFICACIÓN DEL OBJETO DE ESTUDIO

Los posibles impactos de ENOS respecto a la macroeconomía ecuatoriana en el periodo 2023-2024 se ven influenciados por la experiencia del pasado. Los hechos históricos resaltan que sectores como el agropecuario y la salud pública son los más afectados por el cambio climático.

En el sector agropecuario, alimentos agrícolas y ganaderos son sensibles a cambios en el precio al consumidor dependiendo del shock del fenómeno del Niño. Un shock fuerte puede traer repercusiones como el alza en los precios de los alimentos por una menor oferta de estos en el mercado, generando inflación en la variación del IPC. En consecuencia, pequeños y medianos productores agrícolas no dispondrán de los recursos necesarios para la inversión en la siguiente producción. Aumentando aún más la subida del precio por escasez de productos alimenticios.

En el sector salud, políticas como la protección social y alimentación serían afectadas por el cambio climático en el país. La seguridad alimentaria de los ciudadanos estaría comprometida, al existir más personas que no tienen los recursos económicos para obtener los alimentos suficientes.

Desarrollar predicciones más precisas y completas en el IPC de los alimentos permitiría descubrir los posibles escenarios que enfrentaría el país. De esta manera se pueden proponer alternativas políticas para mitigar y sobreponer los efectos negativos de ENOS en la industria alimentaria. No obstante, generar un pronóstico puede ser un desafío al momento de escoger el modelo indicado y las variables que tienen un efecto sobre la variación del precio.

El propósito del estudio se enfoca en comparar dos modelos econométricos con el fin de tener más alternativas al momento de predecir el precio de los alimentos. Con esto se pretende comparar el poder predictivo de un modelo lineal contra uno no lineal, bajo la influencia de una variable exógena, para obtener el resultado con mejor precisión. Respecto a la variable exógena se cuenta con varias alternativas. Cada una se comprobará con un análisis de correlación, error

cuadrático medio y ajuste de la serie para definir cuál es la que tiene un mayor impacto en la precisión del pronóstico.

Por último, si bien se plantea realizar una comparación utilizando como factor externo a ENOS, la metodología que se aplica, está abierta a ser reutilizada para otros eventos climatológicos de frecuencia diaria. De esta manera se puede evaluar el efecto de eventos meteorológicos diarios en la asimetría y transitoriedad del IPC de los alimentos.

PLANTEAMIENTO DEL PROBLEMA

El fenómeno del Niño, es una de las dos fases que forman parte del patrón climático definido como El Niño Oscilación del Sur, ENOS. Esto ha provocado millones de dólares en pérdida por daños en la historia del Ecuador. De acuerdo a la intensidad y prolongación de este suceso, sectores como el agropecuario y salud pública son los más afectados debido a su relación con el factor climático.

Inundaciones y deslaves por la intensidad de las precipitaciones ocasionan bajas en la producción, por consiguiente la oferta de productos alimenticios en el mercado disminuye. El resultado final de esta cadena de sucesos se ve reflejado en la variación del índice de precio al consumidor, donde los precios tienen una tendencia creciente. La inflación en el precio de los alimentos afecta principalmente a aquellas personas cuyos ingresos crecen menos que la inflación. Esto amenaza la seguridad alimentaria de la población nacional, ya que existirían más restricciones en la disponibilidad y acceso a los alimentos.

La capacidad de realizar predicciones eficaces sobre el IPC durante la temporada del fenómeno del Niño, permitirá identificar alertas en el crecimiento del IPC. De esta manera se puede crear políticas oportunas que mitiguen el proceso inflacionario de los alimentos. Lo cual conlleva a un bienestar para la población ecuatoriana.

OBJETIVO GENERAL

Desarrollar modelos de series de tiempo lineal y no lineal para la proyección del precio en los alimentos en Ecuador para el año 2023-24 bajo la influencia de El Niño Oscilación del Sur, con la intención de comparar el poder predictivo de estas dos técnicas respecto a la precisión en las proyecciones del índice del precio al consumidor, con el fin de generar alertas que ayuden en la toma de decisiones oportunas a nivel económico y político.

OBJETIVOS ESPECÍFICOS

1. Procesar y limpiar los datos históricos del índice de precios al consumidor en los alimentos nacionales, así como también de las variables relacionadas al Niño Oscilación del Sur.
2. Evaluar el rendimiento de los dos modelos a través del uso de estadísticos significativos como el error cuadrático medio MSE y el coeficiente de determinación.
3. Entrenar los dos modelos utilizando las mismas variables endógenas y exógenas con el propósito de comparar los mismos parámetros de inicio, de forma que los resultados de cada modelo se basen en una misma fuente de datos.
4. Documentar el ciclo de vida del desarrollo de los modelos analíticos con el fin que otras personas puedan replicarlos ya sea para validar los resultados o modificarlos de acuerdo a sus necesidades.

JUSTIFICACIÓN Y APLICACIÓN DE LA METODOLOGÍA

9.1 Justificación de la metodología

Se aplica Knowledge Discovery in Databases KDD, como metodología para el desarrollo del presente proyecto. KDD es un proceso iterativo que permite el descubrimiento de nuevo conocimiento a través de las bases de datos. Utilizando una serie de pasos ordenados, permite hallar patrones entre un conjunto de datos, con el fin de transformarlos en información útil para la organización. El proceso KDD consta de 5 pasos: Selección, Preprocesamiento, Transformación, Minería de datos e Interpretación y Evaluación (Nwagu et al., 2017). La metodología aplicada concuerda con el propósito del proyecto, ya que está enfocada en la analítica de datos aplicada en las organizaciones. Además, los pasos que utiliza KDD, se acoplan a la mayoría de las fases de un proyecto de analítica de datos y Big Data aplicado a una institución. No obstante, se resalta que para el paso Minería de datos del proceso KDD, se considerará a ésta como el diseño y selección del modelo estadístico.

9.2 Selección

9.2.1 Recolección de datos

Se emplean dos bases de datos para el desarrollo del proyecto, cada una pertenece a una distinta fuente de origen. El primer dataset corresponde al índice de precios nacional al consumidor (IPC), proporcionado por el Instituto Nacional de Estadística y Censos del Ecuador (INEC). El segundo conjunto de datos es suministrado por la Oficina Nacional de Administración Oceánica y Atmosférica de los Estados Unidos (NOAA), del cual se tiene los valores de las anomalías medias en la temperatura de la superficie del mar (TSM) y también el índice del niño oceánico (ONI). Este consolidado de indicadores compone el factor climático El Niño Oscilación del sur (ENOS). La juntura de las dos bases de datos genera el dataset final que servirá en el análisis cuantitativo.

Para el IPC se dispone de datos estadísticos mensuales de nueve ciudades del Ecuador y para las regiones Sierra y Costa. El mismo nivel de granularidad, se

tiene para los valores nacionales. La base de datos también cuenta con tres variables descriptivas asociadas a la Clasificación del Consumo Individual por Finalidades (CCIF) estas son: nivel, código y descripción de CCIF. Cada descriptivo del CCIF proporciona una clasificación y/o jerarquía de los bienes y servicios basados en los gastos realizados en el consumo diario de un hogar. Se dispone de 26.448 tomas de precio por artículo CCIF y área geográfica, entendiendo este último las ciudades, regiones y el total nacional. El histórico de los precios inicia en el año 2005 mes de enero y termina en diciembre 2023, vigente a la fecha actual. La actualización de la data se realiza cada mes, los precios de un mes atrás se registran hasta el término de la primera semana del mes actual. La base de datos del IPC, está disponible para todo público y puede ser descargada en la página oficial del INEC en formato Excel. Para el desarrollo del proyecto específicamente se utiliza el archivo Excel `ipc_ind_nac_reg_ciud_emp_clase_12_2023.xlsx` ubicado en la carpeta Series IPC Empalmadas (INEC, 2024).

Respecto al factor atmosférico ENOS, se tiene una base de datos con los valores TSM y sus anomalías. Cada valor TSM y anomalía representa a una de las regiones del Niño en el océano Pacífico tropical. Las regiones del Niño son cuatro: Niño 3.4, Niño 3, Niño 4 y Niño 1.2. Bajo estos mismos nombres se colocan los nombres de las variables TSM y sus anomalías en la base de datos. También se adjunta a este dataset información sobre el índice ONI con su respectiva anomalía. Todos los campos mencionados, tienen una periodicidad mensual y su unidad de medida está en grados centígrados. Para las variables TSM, los valores inician desde enero de 1982 y terminan en diciembre de 2023, vigente a la fecha actual (NOAA, 2024b). Por otro lado, la variable ONI tiene información desde enero 1950 hasta noviembre del 2023, vigente a la fecha actual (NOAA, 2024c). Por temas de empalme con el conjunto de datos de índice de precios, se delimita el año de inicio de la base de datos de ENOS a enero del 2005. La actualización de la data, tanto de las tomas TSM y el índice ONI, son las primeras semanas del mes actual. El dataset de ENOS, está disponible para

todo público y puede ser descargado en la página oficial del NOAA en un archivo de texto sin formato.

9.3 Preprocesamiento

9.3.1 Limpieza y Preprocesamiento

Debido a la utilización de fuentes distintas, se tiene dos bases de datos con diferente extensión, estructura y formato de codificación en los caracteres. Por lo tanto, se genera un proceso ETL que me permita unificar los dos conjuntos de datos en un solo dataset. El proceso ETL desarrollado contempla las tres funcionalidades que son extracción, transformación y carga de datos. Respecto a las reglas de limpieza y calidad en los datos, estas se aplican en la etapa de extracción. Para la creación del ETL y como base de datos destino se utilizará el programa Excel. La herramienta Power Query integrada en Excel servirá en el desarrollo del ETL. Mientras que las hojas de cálculo de Excel se usarán para el almacenamiento de los datos entregados por el ETL.

Extracción de datos:

Primero se define cuáles son las extensiones de las dos bases de datos a ser procesadas y la cantidad de observaciones de cada una. A base de esto se evalúa el motor de transformación y preparación de datos Power Query de Excel. Power Query abarca 4 fases: conectar, transformar, combinar y cargar, donde la fase conectar permite importar múltiples orígenes de datos a Excel. Entre los orígenes se encuentran archivos txt y xlsx y permite importar hasta 1,048,576 filas con un máximo de 16,384 columnas (Microsoft, 2024) por origen. Con estas cifras se decide que la herramienta cumple con nuestras necesidades y se procede con la extracción de las dos bases de datos.

Después se define las reglas de calidad y limpieza a ser ejecutadas para ambas bases de datos:

- Cantidad con valores nulos
- Cantidad con valores en blanco
- Cantidad con valores alfanuméricos, exclusivo de variables numéricas

- Cantidad con espacio en blanco al inicio o final de un valor.
- Igualdad en el número de dígitos de la parte decimal (Precisión)
- Validación del tipo de dato, variables numéricas
- Validación del formato de los campos tipo fecha, formato YYYY-MM-DD
- Validación del formato de los campos numéricos, separador decimal (,) y de miles (.)

De la lista anterior, una de las novedades es la validación del formato tipo fecha. Tanto para el conjunto de datos del IPC como de ENOS, la fecha está registrada en formato año y mes. Para el IPC se tiene la nomenclatura abreviatura del mes y año, mientras que en ENOS está en formato año y número del mes. Para no tener problemas con el tipo de nomenclatura de la fecha durante la lectura de las variables al modelo, se genera el formato YYYY-MM-DD. Como el día no se tiene, se toma la decisión de colocar el primer día del mes correspondiente. El cambio al nuevo formato fecha se aplica agregando una nueva columna en ambas bases de datos.

9.4 Transformación

9.4.1 Transformación y carga de datos

Aquí se realiza la combinación de los datos en una única estructura. La estructura final tiene un formato tabular donde cada columna representa a una variable IPC o ENOS y cada registro los valores observados.

Como primer paso en la tarea de combinación, se analiza la estructura inicial de los datos del IPC como ENOS para su posterior homologación a la estructura final. En el IPC, los datos se encuentran estructurados en forma de matriz donde las columnas representan las fechas y las filas los descriptivos de CCIF. Utilizando la función transponer de Excel, se cambia el sentido de las filas por columnas. Es decir, se colocan los descriptivos del CCIF en forma columnar de igual manera, se aplica una transpuesta a las columnas con los valores de fecha. Para el caso del conjunto de datos ENOS, la estructura está acorde a lo que se desea por ende no sufre cambio alguno.

Una vez terminada la transformación se procede con la carga de los datos a una base de datos unificada. Una vez homologados los datos y con sus respectivas validaciones, se realiza la unión de las dos estructuras IPC y ENOS. Se tiene como consideración que el rango de inicio y fin en las fechas debe ser al mismo corte. Como el dataset del índice de precios comienza en una fecha más temprana, se considera a enero 2005 como corte inicial en la juntura de los datos. La estructura final y sus datos son almacenados en una hoja de cálculo nueva en Excel.

9.4.2 Identificación y descripción de variables

A continuación, se muestra el diccionario de datos con la metadata de cada uno de los campos y la estructura final generada por el proceso ETL de la anterior sección. Respecto a las variables del IPC, se colocaron únicamente los descriptivos CCIF de la división alimentos.

Nombre del Archivo: NOAA_ENSO.xlsx

Fecha creación: 2024-01-31

Descripción: Base de datos que contendrá el histórico del IPC de los alimentos y los índices climatológicos de ENOS.

Tabla 2 Diccionario de datos de las bases de datos IPC y ENOS

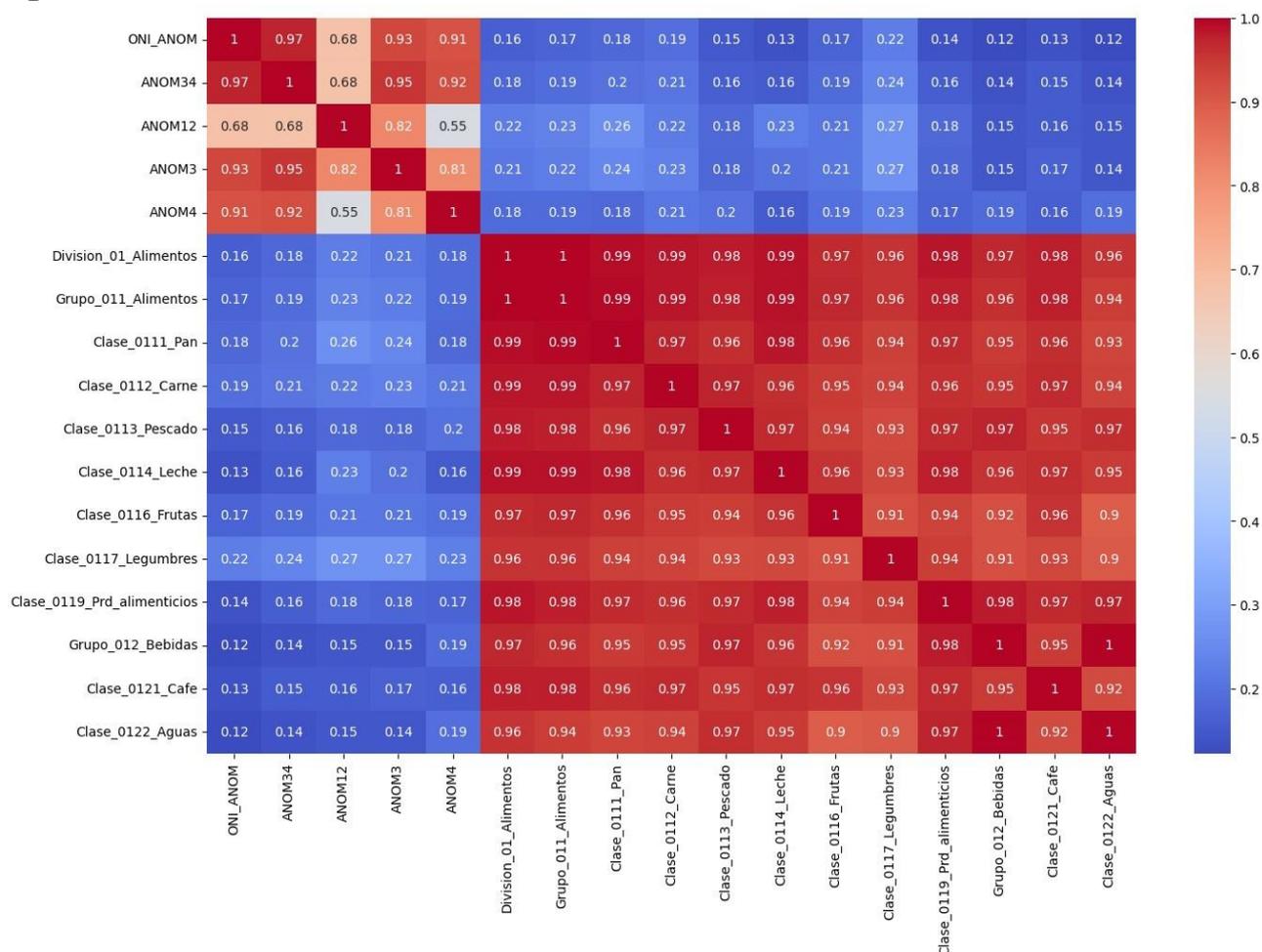
Campo	Tipo de datos	Tamaño	Descripción
Fecha_F1	Fecha	10	Fecha de la observación. Periodicidad mensual en formato dd/mm/yyyy. Se registra el primer día de cada mes.
ONI	Decimal	(10,2)	Índice del niño oceánico.
ONI_ANOM	Decimal	(10,2)	Anomalía del índice del niño oceánico.
NINO34	Decimal	(10,2)	TSM de la región 3.4 en grados centígrados.

ANOM34	Decimal	(10,2)	Anomalía TSM de la región 3.4 en grados centígrados.
NINO12	Decimal	(10,2)	TSM de la región 1.2 en grados centígrados.
ANOM12	Decimal	(10,2)	Anomalía TSM de la región 1.2 en grados centígrados.
NINO3	Decimal	(10,2)	TSM de la región 3 en grados centígrados.
ANOM3	Decimal	(10,2)	Anomalía TSM de la región 3 en grados centígrados.
NINO4	Decimal	(10,2)	TSM de la región 4 en grados centígrados.
ANOM4	Decimal	(10,2)	Anomalía TSM de la región 4 en grados centígrados.
Division_01_Alimentos	Decimal	(10,5)	IPC de alimentos y bebidas no alcohólicas
Grupo_011_Alimentos	Decimal	(10,5)	IPC de alimentos
Clase_0111_Pan	Decimal	(10,5)	IPC del pan y cereales
Clase_0112_Carne	Decimal	(10,5)	IPC de la carne
Clase_0113_Pescado	Decimal	(10,5)	IPC del pescado
Clase_0114_Leche	Decimal	(10,5)	IPC de la leche, queso y huevos
Clase_0115_Aceites	Decimal	(10,5)	IPC de los aceites y grasas
Clase_0116_Frutas	Decimal	(10,5)	IPC de las frutas
Clase_0117_Legumbres	Decimal	(10,5)	IPC de las legumbres y hortalizas
Clase_0118_Azucars	Decimal	(10,5)	IPC del azúcar, mermelada, miel, chocolate y dulces de azúcar
Clase_0119_Prd_alimenticios	Decimal	(10,5)	IPC de los productos alimenticios
Grupo_012_Bebidas	Decimal	(10,5)	IPC de las bebidas no alcohólicas
Clase_0121_Cafe	Decimal	(10,5)	IPC del café, té y cacao
Clase_0122_Aguas	Decimal	(10,5)	IPC de las aguas minerales, refrescos,

			jugos de frutas y de legumbres
--	--	--	--------------------------------

A continuación, se presenta la matriz de correlación con las variables de interés de ENOS y el IPC de la división alimentos y bebidas no alcohólicas:

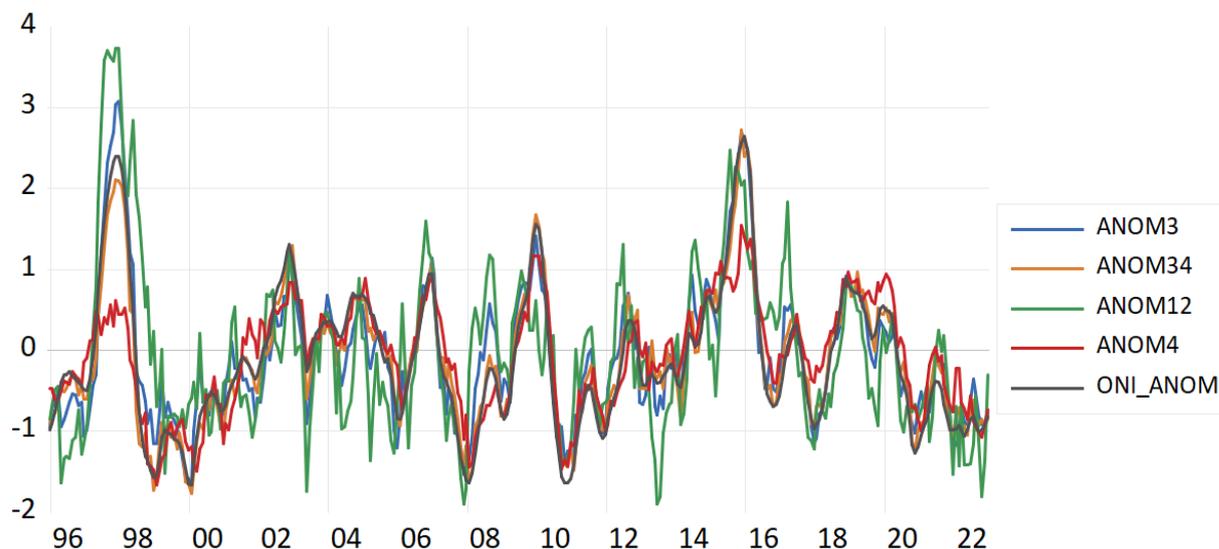
Figura 1 Matriz de correlación IPC vs ENOS



De acuerdo a la matriz, se observa que las variables climatológicas no sobrepasan de una correlación de 0.27 respecto a las variables del IPC. La razón de este valor bajo es porque ENOS tiene una ocurrencia transitoria y asimétrica en el tiempo que puede ocasionar no linealidad en la variable de interés (Abril-Salcedo et al., 2020). El fenómeno del Niño es un evento que puede ocurrir entre 2 a 7 años respecto a su última presencia. Sin embargo, cuando este se presenta, genera variaciones en el precio tal como sugiere la historia del Ecuador durante el fenómeno del Niño. Por otro lado, cuando el evento termina se muestra un efecto ambiguo sobre los precios.

Otro factor a considerar de los bajos coeficientes de ENOS sobre el IPC en la matriz de correlación es su definición. Esta mide la asociación lineal entre dos variables de escala (IBM, 2023), para asociaciones no lineales el coeficiente de Pearson baja.

Figura 2 Anomalías de temperatura en la superficie del mar periodo 1996-22



En la Figura 2 se muestra que los picos más altos en las anomalías TSM relacionadas a ENOS son en los años 1997 y 2015-16. Periodos de tiempo donde la presencia del fenómeno del Niño produjo cambios en los precios nacionales del Ecuador. El presente trabajo, se plantea realizar una predicción del IPC de los alimentos durante la presencia del fenómeno del Niño en el Ecuador en el periodo 2023-24. Por lo tanto, se consideran variables significantes a las anomalías TSM y ONI a pesar de tener un coeficiente bajo de correlación.

Para la elección de la variable endógena y exógena de nuestro modelo de predicción se toma como criterio los valores de la matriz de correlación y el nivel de impacto de ENOS. En la Figura 2 se puede ver que en el año 1997, la anomalía 1.2 fue la que tuvo mayor impacto, sin embargo, para el periodo 2015-2016 la anomalía 3.4 presentó el mayor pico. Respecto a estas variables, anomalía 3.4 y 1.2, se revisa en la matriz de correlación las que tienen mayor fuerza por división, grupo y clase IPC. Como variables endógenas se tiene: en la

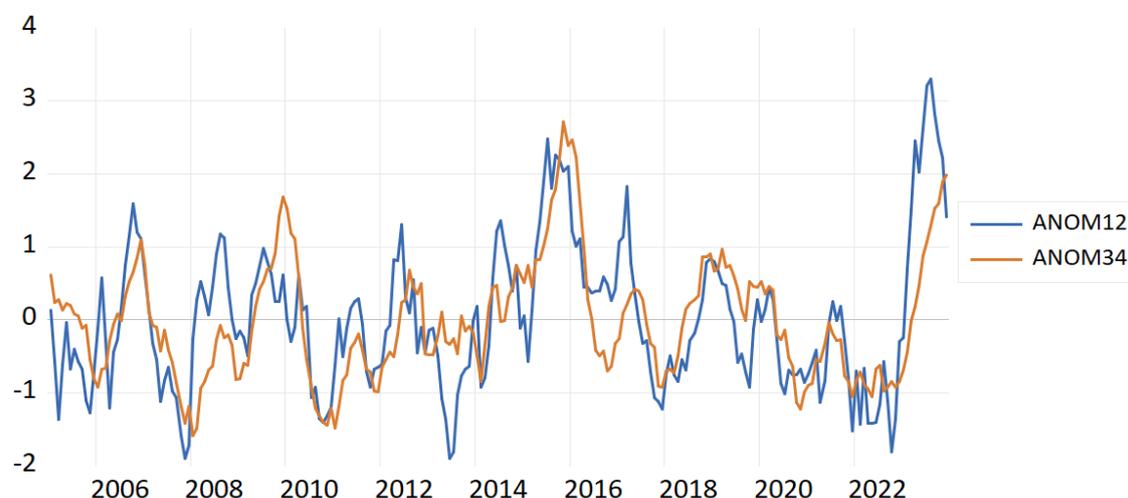
sección división a Division_01_Alimentos, sección grupo a Grupo_011_Alimentos y en la sección clase a Clase_0117_Legumbres.

Se tomarán como referencia inicial en las siguientes etapas a estas cinco variables. Posteriormente se someterán a otros tipos de análisis y pruebas hasta dejar una única variable endógena y exógena.

9.4.3 Visualización de las variables

En la Figura 3 se puede observar las anomalías de la temperatura en la superficie del mar de las regiones 1.2 y 3.4 desde el año 2005 hasta el 2023. Se aprecia un mayor cambio de tendencia en la anomalía 1.2, también se sugiere que esta posee los picos más altos y bajos llegando a ser de 3.3°C y -1.9°C respectivamente. En comparación de la región 3.4 donde su pico máximo es de 2.72°C y el más bajo es de -1.59°C . También se denota que los periodos de tiempo donde se tiene una mayor temperatura son en 2015-16 y a mediados del 2023. Esto tiene concordancia con las fechas del último fenómeno del niño en el Ecuador, y el avistamiento del Niño en el presente periodo 2023-24.

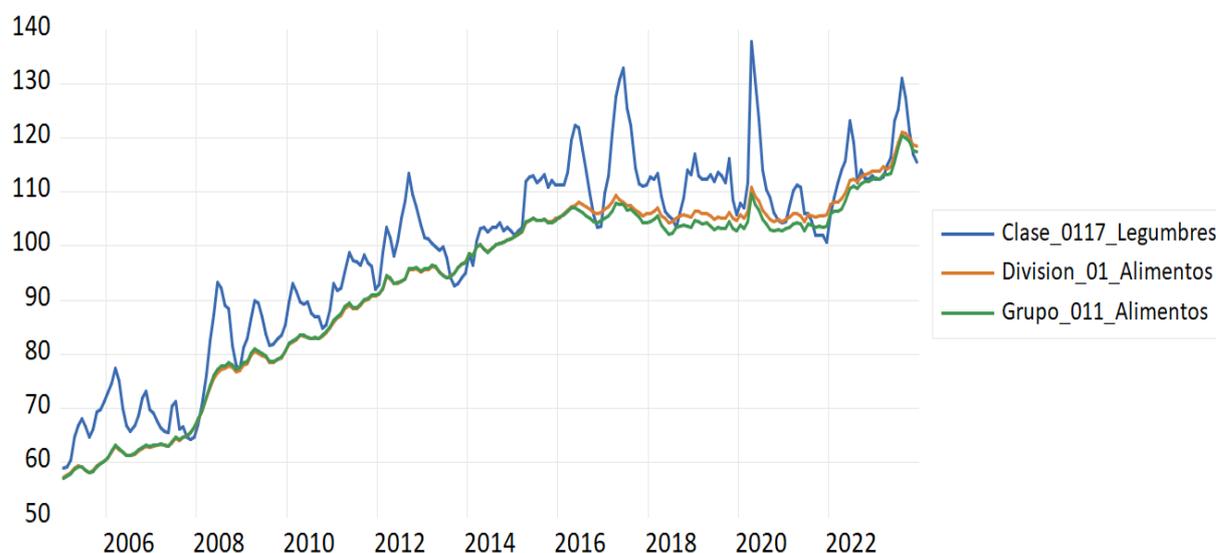
Figura 3 Gráfico de líneas, comparaciones variables TSM de la región 1.2 y 3.4



En la Figura 4 se muestra la evolución del IPC de cada representante por nivel CCIF desde el año 2005 hasta 2023. Se puede apreciar que las legumbres tienen una mayor variabilidad en su índice. También se observa que todas las variables tienen una tendencia creciente durante shocks moderados del Niño Figura 3.

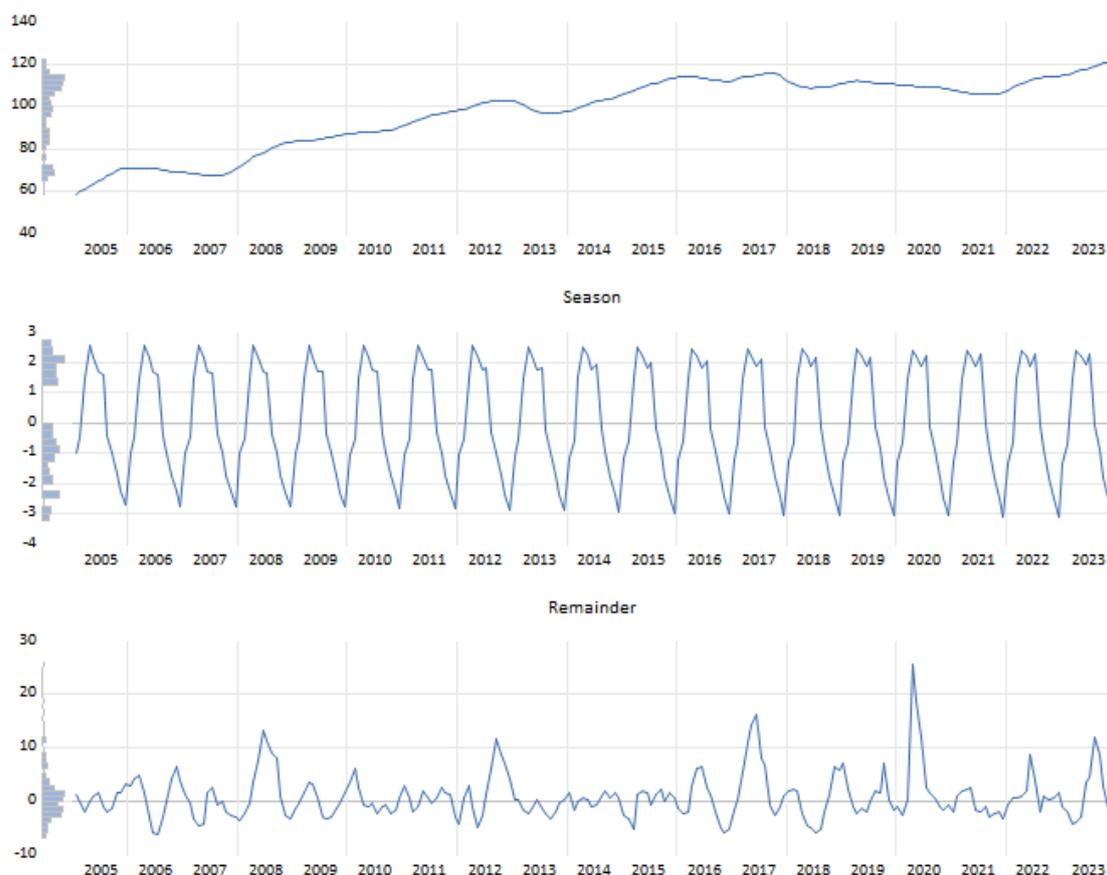
Entendiéndose moderado a aquellas temperaturas superiores o iguales a 1.5°C . Otra observación resaltante es la semejanza en las tres series de tiempo, esto se debe porque la clase Legumbres es un subconjunto del grupo Alimentos. De igual forma el grupo Alimentos es un subconjunto de la división Alimentos y bebidas no alcohólicas. A pesar que existen más grupos y clases que constituyen la división de Alimentos, se puede concluir que la clase y grupo seleccionado es una muestra representativa de la división Alimentos.

Figura 4 IPC de legumbres, Alimentos y Alimentos y bebida no alcohólicas



Otro aspecto en común en las tres variables es su tendencia creciente. Si bien existen rangos cortos de tiempo donde el IPC disminuye, de forma general se observa inflación en el precio de los alimentos y de las legumbres. Como se puede observar en la Figura 5, se tiene una serie de tiempo con tendencia creciente. También se puede apreciar la estacionalidad, esta sigue el mismo patrón cada año. Exactamente pasa lo mismo con las series temporales grupo alimentos y división alimentos y bebidas alcohólicas. Estas tienen una tendencia creciente y su estacionalidad tiene la misma forma en un periodo de doce meses. Se concluye que las tres variables IPC son series no estacionarias porque tienen una tendencia creciente y gráficamente no tienen una media constante.

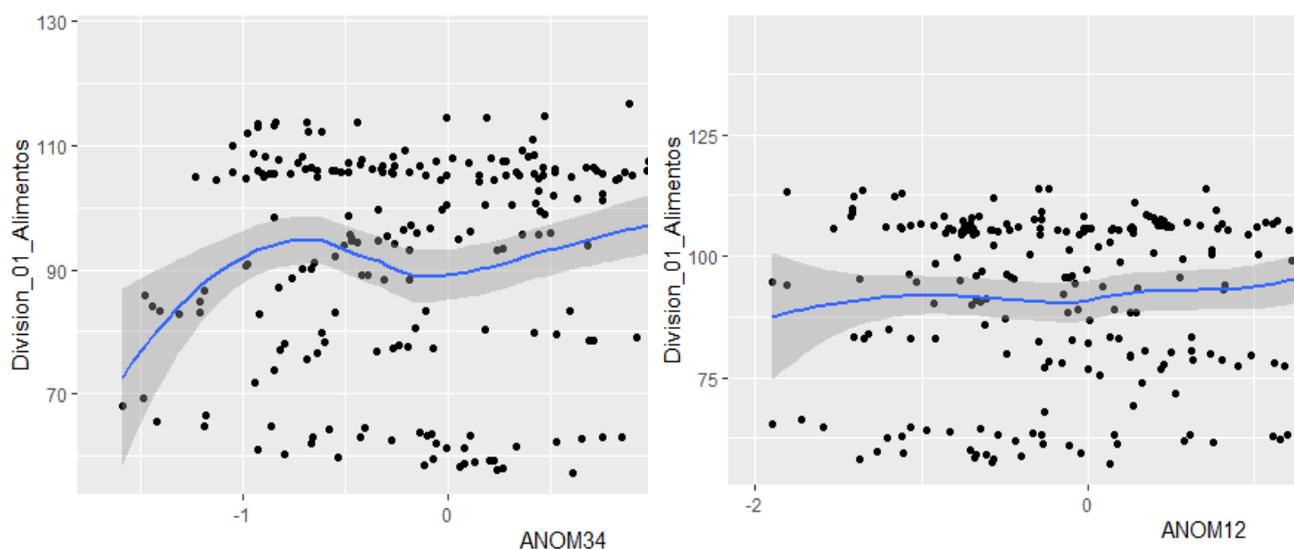
Figura 5 IPC Alimentos y bebidas no alcohólicas y anomalías TSM 1.2 y 3.4



Otros análisis como el de no linealidad entre las variables IPC y ENOS proporciona información clave sobre la asimetría y el grado de correlación. En una evaluación inicial se utilizó el gráfico de dispersión para observar el comportamiento de la variable división de Alimentos y ENOS como se muestra en la Figura 6. Se traza una línea de tendencia polinómica para tener una mejor visualización en la determinación de las fluctuaciones en la curva. En la parte izquierda de la Figura 6 se tiene que la anomalía 3.4 forma una función polinómica ya que se aprecia elevaciones curvas entre los valores -1°C y 0°C . A comparación de la anomalía 1.2, su curva presenta una menor pendiente. No obstante, se pueden apreciar elevaciones no tan significativas en forma de U alrededor de los 0°C . De forma homóloga esta situación sucede con las variables clase y grupo. Se puede concluir la presencia de no linealidad en algunas de las variables ENOS, como es el caso de la anomalía 3.4. En consecuencia, valores

como el coeficiente de pearson no se ajustan a este tipo de asociaciones entre los elementos.

Figura 6 Gráfico de dispersión entre el IPC Alimentos bebidas no alcohólicas y las anomalías TSM 1.2 y 3.4



9.5 Minería de datos

9.5.1 Selección de modelo estadístico

Desde un enfoque econométrico, los modelos matemáticos utilizados en la economía permiten entender, evaluar y predecir el comportamiento de sistemas complejos. Utilizado como insumo datos transversales o series de tiempo, los modelos permiten analizar la naturaleza de las variables y su interacción entre ellas. Un mejor entendimiento sobre estas, permiten formular posibles escenarios donde se evalúe su impacto en un ambiente controlado. No obstante, con el aumento exponencial del volumen de los datos, Big Data, se generan nuevas relaciones entre las variables. Dónde modelos tradicionales pueden distorsionar la correlación entre las variables generando degradaciones en su resultado final. Por esta razón se plantea evaluar dos tipos de modelos, el modelo autorregresivo estacional integrando medias móviles y entradas exógenas (SARIMAX), extensión del modelo tradicional ARIMA, y el modelo autorregresivo de transición suave (STAR). Considerando como factor de relación, el comportamiento lineal y no lineal entre las variables de interés.

Modelo autorregresivo de transición suave (STAR)

Es un modelo que permite combinar la parte lineal y no lineal en la relación de dos variables a través de una función matemática. Para el proyecto, se analiza cómo afecta las anomalías TSM en la variación del IPC. Como el modelo explica el comportamiento no lineal de las variables se plantea la utilización de la anomalía 3.4 como variable exógena y el IPC de alimentos y bebidas no alcohólicas como variable endógena.

La ecuación matemática para un modelo STAR es:

$$y_t = (\phi_{10} + \sum_{i=1}^r \phi_{1i} y_{t-i}) + (\phi_{20} + \sum_{i=1}^r \phi_{2i} y_{t-i}) G(s_t; \gamma, c) + \varepsilon_t \quad (1)$$

Donde y_t es la variable dependiente o endógena y y_{t-i} es un vector de variables explicativas compuesto por los rezagos de la variable endógena y_t y rezagos de la variable exógena. Los valores ϕ_1 y ϕ_2 son los vectores de coeficientes a estimar, en cuanto ϕ_{10} y ϕ_{20} con constantes. La función de transición está dada por $G(s_t; \gamma, c)$ donde s_t es la variable de transición. El valor γ es el parámetro de suavidad y c el parámetro de ubicación. Por último ε_t es el término del error, en este caso se considera como ruido blanco.

Existen variaciones respecto a la función de transición, las más comunes son la logarítmica y exponencial. Estas se expresan de la siguiente forma:

$$G(s_t; \gamma, c) = \{1 + \exp[-\gamma(s_t - c)]\}^{-1} \quad (2)$$

$$G(s_t; \gamma, c) = 1 - \exp[-\gamma(s_t - c)^2] \quad (3)$$

Combinando la ecuación (1) con (2) o (3) se tienen los modelos Logistic Smooth Transition Autoregressive (LSTAR) y Exponential Smooth Transition Autoregressive (ESTAR).

Utilizando las variables de interés IPC y ENOS se reemplazan en ecuación (1) de la siguiente expresión:

$$DIPC_t = \phi_{10} + \sum_{i=1}^r \phi_{1i} DIPC_{t-i} + \sum_{i=1}^p \psi_{1i} ENOS_{t-i} + (\phi_{20} + \sum_{i=1}^r \phi_{2i} DIPC_{t-i} + \sum_{i=1}^p \psi_{2i} ENOS_{t-i}) G(ENOS_{t-d}; \gamma, c) + \varepsilon_t \quad (4)$$

En la ecuación (4), DIPC es la variable alimentos y bebidas no alcohólicas (IPC) diferenciada una vez. El cálculo se realizó debido a la presencia de raíces unitarias durante las pruebas de estacionariedad.

Se ejecutaron tres pruebas de raíz unitaria: Dickey-Fuller (ADF), Phillips-Perron (PP) y Kwiatkowski-Phillips-Schmidt-Shin (KPSS). Respecto a las pruebas ADF y PP se considera como hipótesis nula: Tiene una raíz unitaria. Mientras que en la prueba KPSS se considera como hipótesis nulas: Es estacionaria.

Los resultados de las tres pruebas se muestran en la Tabla 3. Se puede observar que la variable IPC no es estacionaria en las tres pruebas con un p-value de 0.268 en ADF, 0.43 en PP y 0.01 en KPSS. Esto comprueba lo revisado en la sección de visualización de las variables, donde se observó que el IPC tenía una tendencia creciente y una media irregular. Con este primer resultado se realizó una primera diferenciación en la variable IPC y se la denominó DIPC. Esta variable se interpreta como la inflación o deflación del precio de los alimentos. De igual forma se comprueba la presencia de raíces unitarias en la nueva variable. Se concluye que ENOS y DIPC son series estacionarias mientras que el IPC es una variable no estacionaria.

Tabla 3 Prueba de raíz unitaria

	Augmented Dickey-Fuller (ADF)	Phillips-Perron (PP)	KPSS
IPC (t-statistic)	-2.042	-1.699	2.072
IPC Prob(t-statistic)	0.268	0.43	0.01
ENOS (t-statistic)	-4.477	-3.324	0.109
ENOS Prob(t-statistic)	0.000	0.014	0.1
DIPC (t-statistic)	-6.067	-10.686	0.28
DIPC Prob(t-statistic)	0.000	0.000	0.1

Para el desarrollo del modelo se selecciona las variables estacionarias obtenidas del paso anterior, es decir se utiliza DIPC y ENOS como variable endógena y exógena respectivamente. Como siguiente paso se plantea el número de rezagos de la variable DIPC y ENOS. Se utiliza el criterio bayesiano de Schwarz para la selección del número de rezagos r y p . Se estable un $r = 26$ y $p = 11$ como el máximo número de rezagos para la variable endógena y exógena de la

ecuación (4). Nótese que la asignación de r y p corresponde tanto para los coeficientes lineales como no lineales en la ecuación del modelo STAR.

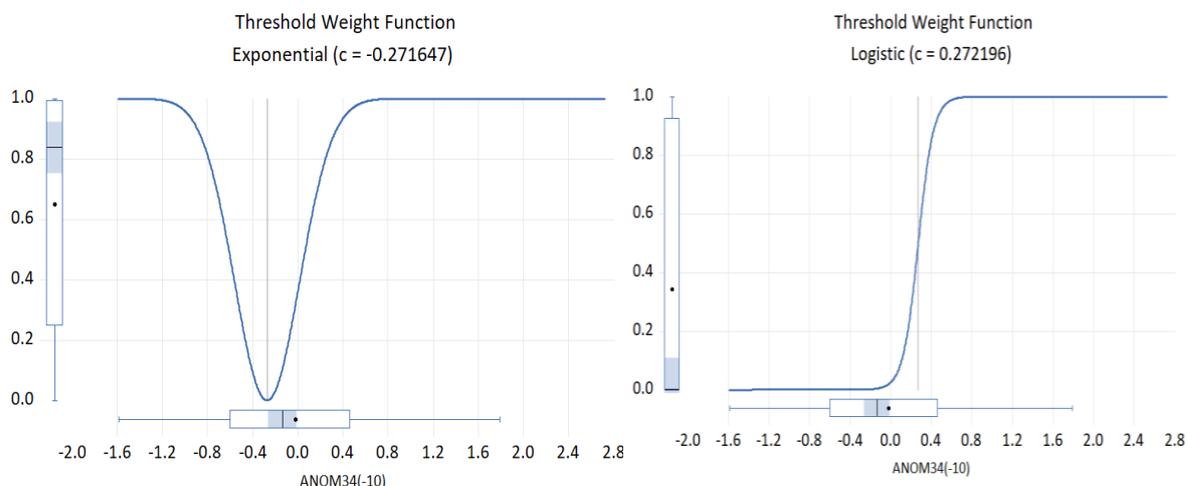
Como siguiente paso se analiza el rezago $ENOS_{t-d}$ que se utilizara en la función de transición. Para su elección, se realiza una prueba LM no linealidad (Terasvirta Sequential Test) en los p rezagos de ENOS. La prueba de Terasvirta da como respuesta el tipo de modelo a utilizar en el rezago evaluado. Bajo la hipótesis nula, sigue un modelo lineal y la alterna sigue un modelo LSTAR o ESTAR se resuelve ejecutar para los once lags de ENOS como se muestra en la Tabla 4. Los resultados revelan que el rezago 10 y 11 se ajustan a un modelo ESTAR, los demás rezagos son irrelevantes para el modelo porque se alinean a una función lineal. La idea de la función de transición es suavizar las irregularidades dadas en los picos altos o bajos de la serie temporal. Para el modelo se escoge el menor p-value de la hipótesis H2, debido a esta evalúa si el modelo se ajusta a una función exponencial.

Tabla 4 Prueba LM No Linealidad para DIPC

Variable Transición	H1 (p-value)	H2 (p-value)	H3 (p-value)	Modelo
$ENSO_t$	0.276	0.431	0.842	Lineal
$ENSO_{t-1}$	0.07	0.889	0.192	Lineal
$ENSO_{t-2}$	0.22	0.931	0.277	Lineal
$ENSO_{t-3}$	0.366	0.844	0.159	Lineal
$ENSO_{t-4}$	0.32	0.739	0.108	Lineal
$ENSO_{t-5}$	0.431	0.677	0.449	Lineal
$ENSO_{t-6}$	0.862	0.591	0.025	Lineal
$ENSO_{t-7}$	0.947	0.339	0.707	Lineal
$ENSO_{t-8}$	0.943	0.769	0.411	Lineal
$ENSO_{t-9}$	0.737	0.779	0.839	Lineal
$ENSO_{t-10}$	0.357	0.027	0.095	Exponencial
$ENSO_{t-11}$	0.254	0.041	0.052	Exponencial

A pesar de escoger una función exponencial, como prueba adicional del modelo STAR se graficó las función logarítmica y exponencial para ver su influencia en el IPC. En la Figura 7, se muestra el grafico de la función logarítmica y exponencial en la transición del estado 0 a 1.

Figura 7 Comparación entre las funciones ESTAR y LSTAR



Limitaciones:

El modelo STAR explica comportamientos no lineales entre dos variables. Su robustez está sujeto al cumplimiento de los supuestos de no linealidad de los rezagos asociados a las variables de interés. No obtener un valor de probabilidad significativo en las pruebas de no linealidad, impide la utilización de dicho rezago y se tiene que buscar o agregar otro. El replanteamiento del número de rezagos especificados en un inicio por el criterio bayesiano de Schwarz, afecta el rendimiento del modelo. La agregación de una mayor cantidad de rezagos genera un menor rendimiento en el modelo. Se debe tener un equilibrio entre rendimiento y el número de rezagos necesarios para cumplir con los supuestos del modelo. En el caso de que ningún rezago supere las pruebas de no linealidad, el modelo STAR no se ajusta con la trayectoria de las variables analizadas.

Modelo autorregresivo estacional integrando medias móviles y entradas exógenas (SARIMAX)

El modelo SARIMAX es una versión extendida de los modelos ARIMA. En esta última versión se conserva la base matemática de los modelos ARIMA y se adicionan dos nuevas funcionalidades. Con el fin de optimizar la precisión en los pronósticos y disminuir el intervalo de confianza de la variable de interés.

Respecto a la base de los modelos ARIMA, esta se resume como la combinación de varios modelos estadísticos como son: el modelo Autorregresivo (AR) y el modelo de medias móviles (MA). Donde la parte autorregresiva se refiere a la combinación lineal de sus valores pasados para establecer el valor de un periodo posterior. La parte media móvil que predice un valor a partir del ruido de pasadas predicciones (Micaela et al., 2023). Adicional a estos modelos, ARIMA incorpora el componente integrado (I) que indica el número de diferenciaciones para que la serie temporal se convierta en estacionaria. El conjunto de estos tres componentes se definen en los parámetros (p, d, q) donde p indica el número de rezagos de la parte autorregresiva, d el número de diferenciaciones y q el tamaño de la media móvil

En cuanto a las nuevas funcionalidades de SARIMAX estas son: la adición del orden estacional (P, D, Q)_m y la inclusión de variables exógenas. Donde P indica el número de rezagos de la parte estacional, D el número de diferenciaciones de la parte estacional, Q el tamaño de la media móvil de la parte estacional y m la periodicidad de la parte estacional.

De igual forma que el modelo STAR, se realizó una diferenciación sobre la variable alimentos y bebidas no alcohólicas, para cumplir con el supuesto de estacionariedad. Es decir, se definió el parámetro d con un valor igual a 1. Los demás parámetros se seleccionaron en base a la magnitud del error cuadrático medio y el coeficiente de determinación del modelo.

Limitaciones:

El modelo SARIMAX necesita de las proyecciones de la variable exógena ENOS para obtener las predicciones del IPC. Si se quiere evaluar el estado de los precios en el segundo y tercer trimestre del 2024, necesariamente se deben conocer los valores de ENOS de ese periodo. SARIMAX no puede extrapolar los valores del IPC sin conocer primero el valor de ENOS para cada periodo de tiempo. Esto genera una dependencia en la variable exógena, es decir, se necesita de un modelo predictivo o fuente externa donde se puedan obtener las proyecciones de la variable ENOS. Estos datos a su vez servirán de insumo en

el modelo SARIMAX para que este pueda generar las predicciones correspondientes al periodo de tiempo evaluado.

RESULTADOS

Análisis de modelos estadísticos e Interpretación de Resultados

Se detallan los resultados obtenidos de los modelos SARIMAX y STAR respecto al error medio cuadrático (ECM) y el coeficiente de determinación (R^2).

Tabla 5 Resultados MSE y R^2 para las variables IPC y ENOS

Modelo	MSE	R^2
STAR	1.04	0.85
SARIMAX	3.51	0.5

El MSE, nos detalla el promedio de los errores al cuadrado para los valores pronosticados y los valores reales. Lo relevante de esta fórmula es la penalización de los errores al elevarlos al cuadrado. Diferencias grandes de una predicción y su correspondiente valor real produce un mayor error promedio. Por lo tanto, se infiere que magnitudes cercanas a cero generan predicciones más precisas, por consiguientes el modelo posee un mejor rendimiento.

Por otro lado, el coeficiente de determinación indica que tanto se ajusta el modelo con las observaciones de la serie temporal. La magnitud R^2 tiene una variación de 0 a 1, donde un valor cercano a cero nos indica que el modelo no se ajusta con las observaciones analizadas. Mientras que un valor próximo a uno, indica un ajuste perfecto. No obstante, esta medida estadística, debe ser debidamente interpretada según el modelo a utilizar.

El periodo de tiempo aplicado para las predicciones fuera de la muestra del IPC en ambos modelos, es enero 2023 hasta enero 2024. La razón, el avistamiento del fenómeno de niño y el levantamiento de alerta naranja en el Ecuador para diciembre 2023 (ERFEN, 2023). En la Tabla 5 se detallan los resultados para los estadísticos R^2 y MSE de los dos modelos. Como nota, se aclara que los valores mostrados para el modelo STAR se obtuvieron a partir del promedio. La razón, la variación del MSE al momento de ejecutar varias veces la misma predicción bajo condiciones iguales en el modelo STAR. Para generar el resultado final se realizó la toma de 20 pronósticos del IPC y se registró el estadístico Raíz del

Error Medio Cuadrático, Tabla 6. Después se obtuvo el promedio de este último y se elevó al cuadrado para obtener el estadístico MSE. Para el R^2 se revisó el historial de predicciones tomadas en el paso anterior, el error cuadrático medio con mayor proximidad al MSE promedio. A partir de este pronóstico se calculó el coeficiente de determinación.

Tabla 6 RMSE del forecast para el IPC, modelo STAR

N°	RMSE	N°	RMSE	N°	RMSE	N°	RMSE
1	1,012	6	1,019	11	1,008	16	1,048
2	1,01	7	1,023	12	1,026	17	1,029
3	1,019	8	1,027	13	1,014	18	1,017
4	1,005	9	1,02	14	1,004	19	1,015
5	1,005	10	1,013	15	1,014	20	1,04

Se concluye que el modelo con mejor precisión en el pronóstico del IPC división alimentos, es el modelo STAR con un MSE igual 1,04 dólares por unidad. Esto quiere decir que en promedio la diferencia entre los valores pronosticados y los reales es de 0,08 dólares por unidad. De igual manera respecto al R^2 , el modelo STAR posee un mayor ajuste en los datos que el modelo SARIMAX, con un valor igual al 0,85. El modelo STAR explicó un 71,13% más de variación en los datos del IPC que el modelo SARIMAX.

Con base en los resultados estadísticos, un modelo autorregresivo de transición suave es el más indicado para explicar el IPC de los alimentos durante la presencia del fenómeno del Niño. Variables que exhiben un comportamiento no lineal entre ellas, como los precios y factores climatológicos (Figura 6), poseen predicciones más precisas cuando se utiliza un modelo STAR. No obstante, cabe resaltar que la utilización de una función exponencial ESTAR, para la suavización de la serie, contribuyó en la reducción del MSE y aumento del R^2 . Por otro lado, la utilización de una función logarítmica LSTAR, genera degradaciones en los valores de los estadísticos. Una de las razones es el tipo de relación entre la variable IPC y la anomalía TSM en la región 3.4. Gráficamente, se observó una curvatura cóncava y convexa similar a una ecuación de segundo grado. Sin embargo, se justificó cuantitativamente la utilización de un modelo ESTAR con

la prueba LM No linealidad de Terasvirta (Tabla 4) y el valor p-value de los parámetros de suavidad y ubicación. En la Tabla 7 se muestra el p-value para los parámetros de la función de transición de los modelos LSTAR y ESTAR. Si bien no existe una diferencia con el parámetro de ubicación, el p-value del parámetro γ del modelo ESTAR es más significativo que el modelo LSTAR. Lo que sugiere la utilización de una función de transición de tipo exponencial.

Tabla 7 Valor de los parámetros de suavización y ubicación del modelo ESTAR

Parámetro	Modelo	Coefficiente	t-Statistic	P-value
γ	ESTAR	6.018	3.343	0.00
c	ESTAR	-0.271	-10.233	0.00
γ	LSTAR	13.434	1.461	0.146
c	LSTAR	0.272	3.465	0.00

También se resalta la magnitud del coeficiente en el parámetro de suavidad, el cual indica el grado de no linealidad durante el cambio de dos regímenes. Para el modelo ESTAR se muestra un grado de suavización más alto que LSTAR, debido a que posee coeficientes más bajos que derivan en una función más curva.

Respecto al modelo SARIMAX, se realizó varias pruebas respecto a sus parámetros Tabla 8, nótese que el componente de integración tiene un valor constante igual a uno. Su valor se relaciona con la presencia de raíces unitarias en la serie temporal del IPC. La elección del orden estacional y sus componentes autorregresivo y medias móviles fue respecto al menor MSE y mejor ajuste R^2 obtenidos de las predicciones fuera de la muestra. Cabe resaltar que se fueron descartando modelos que no cumplieron con el supuesto autocorrelación en los residuos o ruido blanco. De esta forma se logró el orden (2,1,2)(2,0,2,12) (Chasiliquin, 2024). A pesar de la utilización de una variable exógena, anomalía TSM de la región 3.4, esta no tuvo el impacto esperado en el entrenamiento del modelo. Con un p-value igual a 0.09 se interpreta como un coeficiente no significativo, lo cual baja la precisión de la predicción del índice de precios al consumidor.

Tabla 8 Estimaciones de parámetros y rendimiento en modelos SARIMAX

Orden (p,d,q)(P,D,Q,M)	MAE	MSE	R ²
(1, 1, 1) (1, 0, 1, 12)	2	5,6	0,2
(1, 1, 1) (2, 0, 2, 12)	1,75	4,28	0,39
(2, 1, 1) (1, 0, 1, 12)	1,83	4,83	0,31
(2, 1, 2) (1, 0, 1, 12)	1,76	4,55	0,35
(2, 1, 2) (1, 0, 1, 12)	1,76	4,55	0,35
(2, 1, 2) (2, 0, 2, 12)	1,51	3,51	0,5
(3, 1, 2) (2, 0, 2, 12)	1,98	5,64	0,19
(3, 1, 0) (2, 0, 2, 12)	1,71	4,11	0,41
(5, 1, 2) (2, 0, 2, 12)	1,79	4,69	0,33

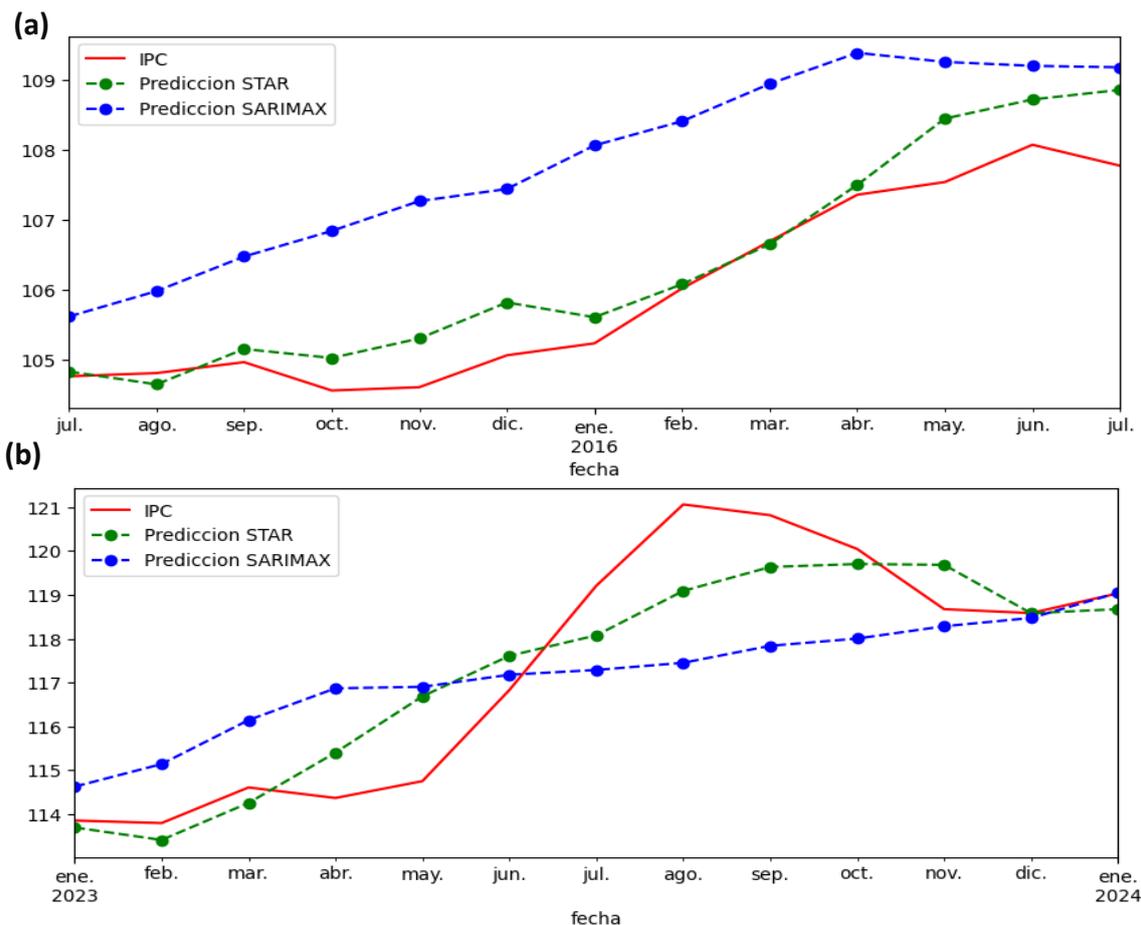
Predicciones modelo STAR y SARIMAX

Según la Organización de las Naciones Unidas ONU, indica que el avistamiento del Niño en el Ecuador ya está presente y se espera que el shock más fuerte sea diciembre 2023 y el primer trimestre del 2024 (ONU, 2023). No obstante, la Secretaría de Gestión de Riesgos (SGR) expresó un fuerte shock del fenómeno del Niño a partir de la próxima época lluviosa, finales de noviembre 2023 (OIM, 2023).

En la Figura 8 (b) se detalla el comportamiento del IPC en los modelos STAR y SARIMAX en predicciones fuera de la muestra. Se realiza una segunda valoración de los modelos SARIMAX y STAR considerando shocks fuertes del Niño, es decir a partir de noviembre 2023. Durante los tres últimos meses, se puede observar en el gráfico de líneas que el IPC de los alimentos tiene variaciones en su tendencia. Comienza con una tendencia decreciente en noviembre y cambia de dirección hasta enero 2024. Este tipo de tendencia también se refleja en las predicciones del modelo STAR, no obstante, con diferente magnitud en la pendiente. Para el modelo SARIMAX se nota una tendencia creciente durante los últimos tres meses, propio de un modelo lineal. Una observación resaltable en el modelo STAR es que a pesar del impacto fuerte de ENOS, se produce una baja en el precio de los alimentos en la transición del mes de noviembre a diciembre. La razón, la relación de no linealidad entre las

variables endógena y exógena. En otras palabras, cuando se duplica el impacto del ENOS, la proyección del IPC no necesariamente tiene que duplicarse (Abril-Salcedo et al., 2020). Esta característica hace que la predicción en este punto de tiempo tenga un menor margen de error y a su vez proporciona una mejor colocación en el cálculo de la siguiente predicción.

Figura 8 Predicciones del IPC periodos 2015-16 y 2023-24

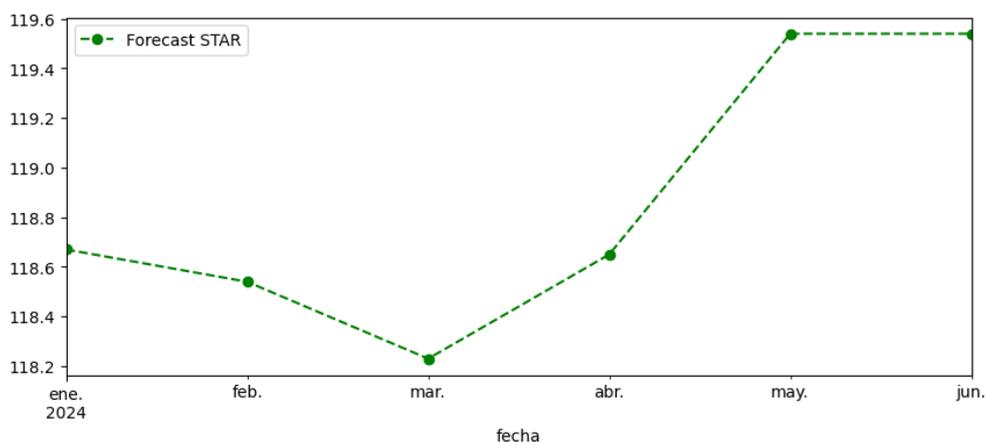


Utilizando los datos de la serie IPC ajustada que entrega los modelos, se puede analizar el grado de ajuste de estos respecto al periodo de tiempo donde el shock ENOS fue moderado o fuerte. Si analizamos el último avistamiento del Niño de intensidad fuerte, este se produjo a mediados de noviembre-diciembre de 2015 y enero de 2016 (Martínez et al., 2017). En la Figura 8 (a) se muestra una tendencia alcista en el IPC de los alimentos. Este mismo patrón se puede observar en los valores del modelo SARIMAX. En cuanto a STAR se muestra con una variación en el mes de enero de 2024. No obstante, sus valores son menos distantes al IPC real que SARIMAX.

Si comparamos la evolución del IPC tanto en el periodo 2025-16 y 2023-24, se puede ver una asimetría en su valor. Mientras que en el último trimestre del 2015 se tiene una tendencia alcista en el precio, el último trimestre del 2023 es lo contrario. Esto a pesar de que en ambos trimestres se considera un mayor impacto del Niño. Otra observación resaltable, es la continuidad de una tendencia creciente en el IPC después del impacto de ENOS. En el modelo 2015-16 se muestra inflación en el precio durante los meses de febrero, marzo y abril después del impacto del fenómeno del Niño.

Como siguiente paso, se realiza la proyección del IPC para meses posteriores a enero del 2024, utilizando el modelo con mejor precisión en la predicción del IPC. En la Figura 9, se muestra la extrapolación de los valores de la variable IPC del modelo STAR en el primer semestre del 2024. Este resultado se puede dividir en dos secciones tomando como criterio la tendencia de la variable. La primera sección se produce a partir de febrero y marzo, donde se espera que el índice de precio disminuya. La segunda sección sería en los meses posteriores a marzo, donde se supondría un aumento en la tasa de inflación.

Figura 9 Forecast del índice de precio al consumidos de los alimentos



Por último, en la Tabla 9 se visualiza el valor p-value del test de autocorrelación en los residuos de acuerdo al orden y número de rezagos seleccionados en cada modelo. Para el modelo STAR se utilizó la prueba de Breusch-Godfrey, mientras que en SARIMAX se utilizó la prueba Ljung-Box. En ambas pruebas se tiene

como hipótesis nula: No existe autocorrelación en los residuos, por lo tanto, se concluye que no hay especificaciones equívocas en los resultados detallados.

Tabla 9 Prueba de correlación serial

N° lags	SARIMAX (p-value)	STAR (p-value)
1	0.901	0.112
2	0.989	0.181
6	0.999	0.117
9	1.0	0.081
40	1.0	0.615
60	1.0	0.807
78	1.0	0.179

DISCUSIÓN DE LOS RESULTADOS Y PROPUESTA DE SOLUCIÓN

La utilización de un modelo STAR, para la predicción del precio de los alimentos bajo la influencia de un evento climatológico transitorio, genera proyecciones más precisas. De esta forma se brinda una solución al problema central, descubrir los posibles escenarios que enfrentaría el país en razón a la variación de los precios en los alimentos durante la temporada del fenómeno del Niño.

En la economía nacional, sectores como el agropecuario y de salud pública se verían principalmente beneficiados del uso de esta herramienta analítica. Al adicionar un nuevo elemento que afecte a su variable riesgo, permitirá mejorar la planificación y administración de sus recursos. Disminuyendo el impacto del fenómeno del Niño en el bienestar social y calidad de vida de la población ecuatoriana.

11.1 Beneficios por la aplicación de la solución implementada

Optimización en los bonos de Seguridad Alimentaria

Predecir la tasa de inflación de los precios de los alimentos durante eventos climáticos de impacto asimétrico, como lo es ENOS, mejorará las políticas: Protección social y alimentación.

Las políticas de protección social y alimentación detallan las acciones dirigidas sobre el acceso a los alimentos para la población, una acción es la transferencia de recursos monetarios. Estimar el precio de los alimentos para los siguientes meses dará una idea sobre la revalorización del monto de dinero a entregar de parte de las entidades gubernamentales. Programas estatales como el Bono de Desarrollo Humano, Aliméntate Ecuador, Plan Intersectorial de Alimentación-Nutrición Ecuador y otros bonos de contingencia tendrán una mejor percepción sobre las compensaciones monetarias que se debe entregar a los hogares vulnerables. De forma que se garantice la seguridad alimentaria a pesar de la variación en precio de los comestibles de la canasta básica.

El gobierno nacional proyectó un presupuesto de 1.056.000 dólares para la entrega de bonos de cobertura, parte del plan de contingencia respecto al

fenómeno del niño 2023-24. Estos bonos de contingencia tendrán un valor de 257 dólares y serán entregados a personas afectas por este suceso climático (El Comercio, 2023). Si bien este bono cubre varios gastos como vivienda, salud y alimentación, las sugerencias y comentarios que se detallan están orientados al servicio de alimentación. De acuerdo con las proyecciones detalladas en la sección de resultados, el bono de contingencia disminuiría en los meses de febrero y marzo 2024 respecto a su valor inicial, lo cual generaría un ahorro durante este periodo. Para meses posteriores a marzo se debería generar una mayor inversión o utilizar el ahorro de anteriores meses para solventar la demanda de este bono. Se prevé que durante los meses de marzo, abril y mayo del 2024, el número de personas que recurrían a este bono aumentaría. Debido a la dificultad de las personas para pagar los nuevos precios de los productos alimenticios.

Ajuste de costos del seguro agrícola y precios de sustentación

Los resultados del modelo influyen en la medición de los porcentajes de subsidio de los alimentos a causa de cultivos afectados por fenómenos climáticos. Cultivos como el trigo, café y cacao son los analizados en el modelo. A medida que su precio suba durante el fenómeno del Niño, se interpretará como una menor producción de alimentos por daños en las cosechas. Ajustar la tasa de costo en los seguros agrícolas teniendo en consideración el factor inflación en el IPC, ofrecerá facilidades en la inversión para la siguiente producción. Incentivar la producción agrícola después de eventos climáticos fuertes como ENOS, ayudará a que la estabilización del precio sea más rápida.

De igual forma, el modelo predictivo también ayudará al replanteamiento del precio de sustentación en los alimentos. Este último más orientado a medidas que buscan mejorar las políticas de pequeña agricultura y desarrollo rural. Estas políticas enfatizan el apoyo directo a los productores agropecuarios, esto se realiza a través de compras desde el estado a pequeños agricultores. Dependiendo del incremento de la tasa de inflación entregado por el modelo analítico, precios muy elevados en alimentos derivaran que el estado necesite más dinero. Dejando a los pequeños agricultores en la decisión de bajar sus

precios para cumplir con las demandas. Un precio de sustentación más ajustado a la realidad durante eventos como El Niño, beneficiaran a huertos familiares y pequeños productores. Obtener ingresos sin reducir el precio de sus bienes por debajo del precio de coste permite garantizar la seguridad alimentaria en sus hogares.

Adquisición de créditos agropecuarios más fáciles

Es un beneficio otorgado por el gobierno a aquellos productores agrarios que no cuentan con un seguro agrícola. De igual forma que el seguro agrícola, se trabaja con el mismo supuesto, una menor producción de un producto genera un aumento en el precio ofertado. En este caso las predicciones sobre IPC de los alimentos se transmitirán a un aumento en la cantidad de créditos que una institución financiera puede conceder. Así como también en un ajuste en las tasas de interés por la concesión de este tipo de crédito. Promover la producción de productos agrícolas que han sido afectados por el Niño ayudará a un restablecimiento del índice de precio.

11.2 Diseño de estrategia organizacional

Se elabora una estrategia inclinada a las organizaciones gubernamentales del sector agropecuario y de salud. La razón es porque el tema central del proyecto afecta directamente a estas instituciones como se detalla en el apartado anterior. Crear una estrategia enfocada al logro de objetivos sociales permite justificar al estado su implementación y la recolección de recursos para su ejecución.

El diseño de nuevas estrategias desde una perspectiva social puede abordarse dentro un cuadro de mando integral (Cabrera Arcos et al., 2011). Se detalla las estrategias desde el punto de vista de los indicadores Impacto al ciudadano, Recurso Financieros y Aprendizaje y Crecimiento.

Indicador Impacto al ciudadano: Mejorar la cobertura de los servicios del sector agropecuario y de salud tales como bonos de apoyo, seguros contra incidentes en los cultivos, créditos agrícolas, distribución de alimentos y

protección de precios. Con el fin de garantizar el bienestar social y seguridad alimentaria de los ciudadanos.

Indicador Recurso Financieros: Presupuestos para programas estatales más ajustados a la situación actual del país. Un uso eficiente del presupuesto nacional permitirá un ahorro o un dispendio de este para combatir problemas agrícolas y de salud que acontece en la población ecuatoriana.

Indicador Aprendizaje y Crecimiento: Fomentación de programas o capacitaciones de calidad. Adherir y entender nuevos modelos analíticos para la predicción de precios aparte de las herramientas tradicionales utilizadas, mejorará la base de conocimientos de la organización. Funcionarios públicos con mayor grado de preparación en sus conocimientos producirán resultados más eficientes.

11.3 Estrategia organizacional, innovación y competitividad empresarial

Alineación de la estrategia:

La presente estrategia trata de abrir paso a la utilización de modelos no tradicionales para el consumo y análisis de datos. La inexactitud de predecir los escenarios que se enfrenta el país respecto a un evento externo genera una mala administración de los recursos estatales. Como consecuencia afecta al buen vivir de la ciudadanía. La presente propuesta pretende reducir este impacto hacia la ciudadanía realizando predicciones más precisas para la toma de decisiones oportuna. Refiriéndose esta última al correcto uso de los recursos del gobierno para salvaguardar a la ciudadanía desde un eje económico, social y de seguridad.

Innovación:

La aplicación de modelos no tradicionales, como el autorregresivo de transición suave, para la predicción del IPC brinda una solución nueva a problemáticas agrícolas y de salud. Estos desafíos pueden surgir por un acontecimiento externo como desastres naturales o cambio climático y se producen en distintos periodos de tiempo. La solución propuesta pretende anticipar y abordar estos escenarios

cambiantes para evaluar sus resultados. Lo que se intenta hacer es que los actores involucrados tengan un nuevo instrumento que les ayude en la toma de decisiones con el fin de mitigar un riesgo eventual. Otra política que se plantea es promover la utilización de estos modelos analíticos en otras áreas del gobierno. En especial aquellas instituciones que se especializan en proyectar el crecimiento de la inflación en los alimentos o el cambio climático.

Desafíos en la implementación:

Una dificultad que se puede producir en la implementación del modelo STAR, está en el cambio de la variable exógena relacionada con el factor climatológico ENOS. Si bien se utilizó un estándar internacional que es la temperatura en la superficie del mar de la región 3.4, la cual los científicos la han utilizado como un indicador de intensidad del Niño. Esta variable puede ser cambiada por estándares nacionales dados por entidades gubernamentales nacionales como el INAMHI o el INOCAR. El cambio de la variable TSM en la región 3.4 por otra tiene un impacto alto con lo publicado en este proyecto. No obstante, se mantiene la conclusión principal: las relaciones entre variables económicas y el clima pueden ser analizadas, proyectadas y comprendidas de mejor manera a través de un modelo no lineal, como el autorregresivo de transición suave. Partiendo de este enunciado, lo que se procedería con el cambio de la variable exógena, es realizar un nuevo análisis bajo los mismos procedimientos y pruebas descritas en este trabajo.

CONCLUSIONES Y RECOMENDACIONES

12.1 Conclusiones

En resumen, basándose en los estadísticos, error cuadrático medio y coeficiente de determinación se evidencia que el modelo STAR brinda predicciones más precisas que el modelo SARIMAX. Con MSE y R^2 igual a 1.04 y 0.85, se tiene que el modelo STAR explicó un 71,13% más de variación en los datos del IPC de los alimentos que el modelo SARIMAX. En cuanto al error, se tiene que el modelo STAR dista de 2.47 dólares por unidad que el modelo SARIMAX. Además, por cada valor del IPC pronosticado en el modelo STAR, se tiene un margen de error de ∓ 1.04 dólares por unidad.

Se demuestra que un modelo autorregresivo de transición suave explica de mejor manera el comportamiento entre las variables índice de precio nacional al consumidor de los alimentos y anomalía de la temperatura en la superficie del mar de la región 4.3. Parte de la contribución del modelo con la mejora de las predicciones, es la integración del componente no lineal dentro de su ecuación matemática. Dado que las pruebas realizadas entre las variables endógenas y exógenas demostraron una tendencia polinómica, el modelo demostró ajustarse de mejor forma a la serie IPC. Utilizando una función de transición de tipo exponencial se evidenció cómo la suavización permitió obtener pronósticos con menor error.

Con respecto al modelo autorregresivo estacional integrando medias móviles y entradas exógenas, se observó ambigüedad en predecir ciertos periodos de tiempo. En especial aquellos donde el impacto del Niño era moderado o fuerte. Impactos fuertes de la variable climatológica, las predicciones del modelo mostraron una correlación positiva con el IPC. A medida que la variable TSM aumentaba, el valor del IPC de los alimentos también aumentaba, lo esperado por un modelo lineal. Si bien este comportamiento se reflejó en el periodo 2015-16, no obtuvo la misma tendencia para el periodo 2023-24. Aquí se observó una caída del IPC en la transición de noviembre a diciembre del 2023, lo que generó un aumento del error en este punto.

12.2 Recomendaciones

Utilizando el modelo STAR, analizar el IPC de todas las clases de la división alimentos según la Clasificación del Consumo Individual por Finalidades, CCIF. Si bien en este proyecto se evidenció la utilidad del modelo STAR para predecir el precio de los alimentos, no se puede dar una respuesta clara a qué alimentos son los más afectados por ENOS. Dentro del CCIF, se tiene una jerarquía de los productos de la canasta básica entre otros de consumo masivo. Priorizar cual producto debe considerarse primero para la toma de acciones, genera valor agregado.

El análisis de los modelos se realizó con base en la variable TSM de la región 4.3. Sin embargo, existen otros factores climatológicos propios de ENOS que podrían valorarse con el fin de mejorar la proyección del IPC de los alimentos. También se puede considerar el efecto de la Niña dentro de los siguientes análisis. Con el propósito de generar estudios más completos sobre el comportamiento global de ENOS sobre los precios nacionales.

Utilizar otros estadísticos significativos y métodos de suavización para mejorar y medir el rendimiento de los modelos analíticos. Entre los métodos se puede mencionar a Holt Winters o el suavizado exponencial simple. Comparar el resultado de estos métodos contra la función de transición del modelo STAR, brindará un mayor conocimiento sobre cual tiene mejores resultados. Para los estadísticos se destacan el error absoluto medio y el coeficiente de desigualdad Theil. Este último está orientado a medir la precisión del conjunto de predicciones entregadas por los modelos.

Para el modelo STAR, se recomienda realizar la función impulso respuesta para observar el comportamiento entre las variables. Examinar este tipo de relación en función del tiempo, permite entender los vínculos entre el fenómeno del Niño y los precios de los alimentos. Dado que el fenómeno del Niño puede tener diferentes niveles de impacto durante su presencia, la función impulso respuesta dará una visión más clara de cómo estos influyen en los precios.

REFERENCIAS

- Abril-Salcedo, D. S., Melo-Velandia, L. F., & Parra-Amado, D. (2020). Nonlinear relationship between the weather phenomenon El niño and Colombian food prices. *Australian Journal of Agricultural and Resource Economics*, 64(4), 1059–1086. <https://doi.org/10.1111/1467-8489.12394>
- Bouri, E., Gupta, R., Pierdzioch, C., & Salisu, A. A. (2021). El Niño and forecastability of oil-price realized volatility. *Theoretical and Applied Climatology*, 144(3–4), 1173–1180. <https://doi.org/10.1007/s00704-021-03569-1>
- Cabrera Arcos, C., Alvarado Espinel, V., Villavicencio Vivar, A., Parra, F. L., Romero, C. M., & Toscano, G. S. (2011). *La Gestión Estratégica en Organizaciones del Sector Público*.
- CAF. (2000). EL FENOMENO EL NIÑO MEMORIA, RETOS Y SOLUCIONES. *Corporación Andina de Fomento - CAF, IV*, 14–311.
- Chasiliquin, A. (2024, February 21). *PROYECTO MBD, IMPLEMENTACIÓN DE MODELOS*. <https://github.com/AndresChsl/PROYECTO-MBD>
- El Comercio. (2023, June 15). *Los bonos que se plantean por el fenómeno de El Niño*. <https://www.elcomercio.com/actualidad/ecuador/bonos-plantean-fenomeno-nino.html>
- ERFEN. (2023). *COMITÉ NACIONAL PARA EL ESTUDIO REGIONAL DEL FENÓMENO EL NIÑO*.
- FAO. (2023). *El llamado a la acción frente al pronóstico de El Niño*. FAO En Ecuador. <https://www.fao.org/ecuador/noticias/detail-events/es/c/1643348/>
- IBM. (2023, March 3). *Matriz de correlaciones - Documentación de IBM*. <https://www.ibm.com/docs/es/spss-statistics/29.0.0?topic=efficiency-correlation-matrix>

- INEC. (2024, February 20). *Índice de Precios al Consumidor*.
https://www.ecuadorencifras.gob.ec/documentos/web-inec/Inflacion/2024/Enero/Tabulados_y_series_historicas_Excel.zip
- Kumar, S., Gupta, A. K., & Vatta, K. (2022). Effect of Rainfall in Predicting Tomato Prices in India: An Application of SARIMAX and NARX Model. In *Journal of Agricultural Development and Policy* (Vol. 32, Issue 2).
<https://www.researchgate.net/publication/366985505>
- Martin, L. (2016). ¡Es-Niño!-Impacto-económico-en-la-Región-Andina. *Banco Interamericano de Desarrollo, Nota técnica del BID; 951, 2–23*.
- Martínez, R., Zambrano, E., Nieto López, J. J., Hernández, J., & Costa, F. (2017). Evolución, vulnerabilidad e impactos económicos y sociales de El Niño 2015-2016 en América Latina. *Investigaciones Geográficas, 68, 65*.
<https://doi.org/10.14198/ingeo2017.68.04>
- Micaela, A., Rubio, M., & Cañas, J. C. (2023). *ESTIMACIÓN DE MODELOS ARMA USANDO EL MODELO ESPACIO DE LOS ESTADOS*.
- Microsoft. (2024). *Power Query specifications and limits in Excel*.
<https://support.microsoft.com/en-gb/office/power-query-specifications-and-limits-in-excel-5fb2807c-1b16-4257-aa5b-6793f051a9f4>
- Neog, B., Gogoi, B., & Patowary, A. N. (2022). DEVELOPMENT OF HYBRID TIME SERIES MODELS FOR FORECASTING AUTUMN RICE USING ARIMAX-ANN AND ARIMAX-SVM. In *Ann. For. Res* (Vol. 65, Issue 1). <https://www.e-afr.org/>
- NOAA. (2024a). *ENSO: Recent Evolution, Current Status and Predictions*.
- NOAA. (2024b, February 20). *Niño Regions Sea Surface Temperatures (SST)*.
<https://www.cpc.ncep.noaa.gov/data/indices/sstoi.indices>
- NOAA. (2024c, February 20). *Oceanic Niño Index (ONI)*.
<https://www.cpc.ncep.noaa.gov/data/indices/oni.ascii.txt>

- Nwagu, C., Omarkwu, O., & Inyama, H. (2017). *Knowledge Discovery in Databases (KDD): An Overview*. <https://sites.google.com/site/ijcsis/>
- OIM. (2023). *FENÓMENO EL NIÑO (ENOS)*. <https://www.inamhi.gob.ec/>
- ONU. (2023). *Plan de Contingencia Fenómeno de El Niño 2023*. 1–27.
- SGR. (2015). *SGR confirma la presencia del Fenómeno del Niño en Ecuador*. Secretaría de Gestión de Riesgos. <https://www.gestionderiesgos.gob.ec/sgr-confirma-la-presencia-del-fenomeno-del-nino-en-ecuador/>
- WITS. (2016). *Ecuador | Resumen del comercio*. World Integrated Trade Solution. <https://wits.worldbank.org/CountryProfile/es/Country/ECU/Year/2016/Summarytext>