

DISEÑO Y APLICACIÓN DE UN MODELO PREDICTIVO DE PRECIOS DE FRUTAS Y HORTALIZAS FRESCAS PARA EL COMERCIO MAYORISTA DE URUGUAY



ANII

AGENCIA NACIONAL
DE INVESTIGACIÓN
E INNOVACIÓN



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY



Ministerio
de Ganadería,
Agricultura y Pesca

Diciembre 2022

ÍNDICE

Equipo	5
1. Resumen ejecutivo	6
2. Objetivos.....	7
3. Introducción general	7
4. Objetivo i: Modelo de predicción	12
4.1 Antecedentes.....	12
4.2 Metodología	15
4.2.1 Rubros elegidos	15
4.2.2 Bases de datos	15
a) Precios de frutas y hortalizas.....	15
b) Series de variables auxiliares.....	18
4.2.3 Caracterización de rubros.....	18
a) Rubro Papa	18
b) Rubro Manzana	21
c) Rubro Tomate.....	24
d) Rubro Cebolla	25
e) Rubro Naranja.....	27
4.3 Resultados	29
4.3.1 Métodos estadísticos.....	30
a) Modelos de espacio-estado.....	30
a) Evaluación de predicciones	31
i. Muestra de entrenamiento y de prueba	31
i. Métricas de evaluación.....	31
4.4 Resultados para el rubro papa.....	33

a)	Descomposición de la serie de precios.....	33
b)	Ajuste de modelos, selección y diagnóstico	33
c)	Modelo de mejor desempeño predictivo.....	36
4.5	Resultados para el rubro manzana	36
a)	Descomposición de la serie de precios.....	36
b)	Ajuste de modelos, selección y diagnóstico	38
c)	Modelo de mejor desempeño predictivo.....	39
4.6	Resultados para el rubro tomate.....	40
a)	Descomposición de la serie de precios.....	40
b)	Ajuste de modelos, selección y diagnóstico	41
c)	Modelo de mejor desempeño predictivo.....	43
4.7	Resultados para el rubro cebolla	44
a)	Descomposición de la serie de precios.....	44
a)	Ajuste de modelos, selección y diagnóstico	44
b)	Modelo de mejor desempeño predictivo.....	46
4.8	Resultados para el rubro naranja	47
a)	Descomposición de la serie de precios.....	47
b)	Ajuste de modelos, selección y diagnóstico	48
c)	Modelo de mejor desempeño predictivo.....	50
4.9	Conclusiones y pasos futuros	50
5.	Objetivo ii: Desarrollo de plataforma unificada de información de mercado de Frutas y hortalizas.....	52
5.1	Descripción general de la plataforma e ingresos a la UAM y precios mayoristas	52
5.2	Portal de visualización de modelo de predicción.....	71
a)	Estructura de la solución y sistemas.....	71

b) Especificaciones técnicas de los modelos de predicción.....	73
5.3 Comentarios finales.....	74
6. Objetivo iii: Indicador de comportamiento anómalo de precios en frutas y hortalizas (ICAP) 74	
6.1 Antecedentes.....	74
6.2 Metodo	75
6.3 Resultados	77
6.4 Automatización del calculo	80
6.5 Comentarios finales:.....	80
7. Comentarios finales del proyecto	80
8. Bibliografía	81

Equipo

Responsable técnico-científico: *Juliana Bruzzone UAM*

Apoyo a la Gestión del proyecto, coordinación de actividades: *Pablo Pacheco UAM*

Desarrollo del modelo predictivo y redacción de informes: *Maximiliano Saldaña*

Desarrollo del modelo predictivo y tutoría: *Ignacio Álvarez-Castro IESTA*

Desarrollo del modelo predictivo y redacción de informes: *Juan Millán*

Desarrollo del modelo predictivo, ICAP y redacción de informes: *Andrés Díaz, DIGEGRA*

Desarrollo del modelo predictivo, técnico de apoyo: *Diego Romero, UAM*

Desarrollo de ICAP y redacción de informes: *Maria Noel Ackerman, y Angela Cortelezzi OPYPA*

Técnicos de apoyo UAM, DIGEGRA: *Alejandra Pallante, Luigi Marveglio, Ana Clara Malaquina, Pilar Acosta, Federico Acosta*

IT: *Javier Ferrer y Nicolas Isnardi UAM*

Consultor: *Matías Carambula, UDELAR y Joaquín Arias*

IICA: *Alexandra Chávez y Oscar Hermida*

1. Resumen ejecutivo

Las frutas y hortalizas para consumo en fresco son de los rubros más complejos en lo que a la formación de precios de venta concierne, de hecho, son uno de los grupos de productos transables que presentan mayor dificultad de predicción de precios en el corto y mediano plazo. Los dos principales componentes en la formación de precios que caracterizan a este sector y que lo diferencian de otros, son la marcada estacionalidad y la alta volatilidad de los precios. Determinadas por las características fisiológicas de las diferentes especies y variedades que componen la oferta, así como las diferentes tecnologías productivas y de conservación disponibles y la gran dependencia de las cada vez más variables e impredecibles condiciones meteorológicas. Todo esto en un contexto de alimentos que representan una pieza fundamental para la nutrición adecuada según el acuerdo unánime de las instituciones de referencia internacional en alimentación y salud. En situaciones de crecimiento sostenido, y en particular en el caso de rubros alimentarios, los precios suelen ser objeto de análisis debido a su incidencia sobre los registros inflacionarios y la seguridad alimentaria de la población. Por otro lado, bajas sustanciales y persistentes de los precios, compromete la rentabilidad de quienes producen alimentos. El diseño y desarrollo de un modelo de predicción de precios mayoristas de frutas y hortalizas frescas y la consolidación del ICAP, aporta una herramienta de análisis económico, que permitirá a los actores vinculados a la cadena comercial frutihortícola (productores, operadores mayoristas, minoristas, etc.) y a los actores vinculados a la institucionalidad oficial y agropecuaria, hacer ajustes en la toma de decisiones, emprender acciones más consistentes, y reducir costos. Empleando los datos relevados por el Observatorio Granjero en la Unidad Agroalimentaria Metropolitana, en este trabajo se utilizan modelos de espacio estado para la predicción de los precios de un conjunto acotado de rubros, llegándose a la conclusión de que los diversos rubros tienen distintos grados de predictibilidad y no hay un modelo en particular que tenga aplicabilidad general, sino que para cada caso particular se debe considerar una especificación. El impacto de la pandemia de COVID-19 y la inclusión de variables auxiliares fueron las cuestiones de mayor dificultad, esto último remarcando la necesidad de una mayor sistematización de los datos en el área Agropecuaria.

Este proyecto contribuyó en el desarrollo de una plataforma diseñada específicamente para el registro de volúmenes de ingreso de mercadería a la UAM y de precios mayoristas de frutas y hortalizas, cálculos de indicadores y de precio predicho en base a los modelos desarrollados. Además, se avanzó en la incursión de una declaración digital del ingreso de mercadería por parte de las empresas operadoras.

En esta publicación se presentan los resultados obtenidos en el marco del proyecto ANII Diseño y aplicación de un modelo predictivo de precios de frutas y hortalizas frescas para el comercio mayorista de Uruguay. Se presentan las herramientas desarrolladas para el almacenamiento y consulta de datos, además del desarrollo del modelo de predicción.

Se emplean modelos de espacio estado para la predicción de los precios de un conjunto acotado de rubros, llegándose a la conclusión de que no hay un modelo en particular que tenga aplicabilidad general, sino que para cada caso particular se debe considerar una especificación. La inclusión de variables auxiliares fue una de las cuestiones de mayor dificultad y que remarca la necesidad de una mayor sistematización de los datos.

2. Objetivos

- i. Desarrollar modelos de predicción de precios para productos frutihortícolas, que permita su utilización por parte de actores claves en la cadena comercial, así como por parte de decisores para la definición de políticas comerciales y de abastecimiento.
- ii. Desarrollar una plataforma unificada de información de mercado de frutas y hortalizas que permita optimizar el uso de los datos de precios mayoristas y volúmenes de ingreso al Mercado Modelo/UAM e incorporar información relevante para el desarrollo de los modelos de predicción.
- iii. Ajustar el indicador de comportamiento anómalo de precios en frutas y hortalizas (ICAP), ampliar su alcance a más productos de la canasta y a otras variables de manera de consolidar su uso como herramienta de decisión.

3. Introducción general

Las frutas y hortalizas constituyen productos de especial importancia en la canasta familiar, por lo que resulta de interés generar herramientas para la predicción de sus precios, dado que se espera que sus variaciones tengan incidencia sobre la seguridad alimentaria de los hogares, la rentabilidad para los productores y los registros inflacionarios. En el caso de Uruguay, este rubro representa un sector de la economía que depende fuertemente de la producción nacional, donde una proporción considerable de la misma (60% de la producción nacional aproximadamente¹) era volcada al Mercado Modelo y en la actualidad es canalizada a través de la Unidad Agroalimentaria Metropolitana (UAM). A su vez, un alto porcentaje de las importaciones de estos productos son comercializadas a nivel mayorista en esta plaza comercial. Tomando en cuenta esto, se considera que la información de precios y volúmenes a nivel mayorista recabada en el

¹ Estimación de los técnicos del Observatorio Granjero

marco del Observatorio Granjero resulta de utilidad para contar con datos que representen el sector de frutas y hortalizas, a partir de los cuales se estudien alternativas para la predicción de los precios.

Las bases de datos pilares para este proyecto son: los datos de ingresos de mercadería al Mercado Modelo/UAM y precios mayoristas de frutas y hortalizas relevados en el mismo. La información de ingresos al Mercado Modelo es obtenida mediante una guía de ingreso en formato papel que es entregada por cada vehículo, la cual luego es digitalizada. La información es almacenada en una base de datos relacional. La información que es registrada consta de: datos del vehículo, propietario de la mercadería, zona de producción y volumen por producto. Un punto crítico de este proceso es la digitalización de la información, por la informalidad de la declaración y dificultad de interpretación de los datos, que pueden llevar a errores tanto de digitación como de interpretación de la información. Además, el tiempo dedicado a esta tarea es importante, ya que requiere varias horas de trabajo al día. Además del valor del volumen por producto frutihortícola ingresado al Mercado Modelo/UAM, puede ser un insumo importante para el modelo de predicción el número de vehículos ingresado y la zona de producción ya que esto brinda información complementaria que hace al proceso de formación de precios, ya que por ejemplo se puede determinar el grado de atomización de un producto al establecer las cantidades que comercializa cada operador; y mediante el origen se puede determinar en parte la calidad ya que en ciertas épocas del año las diferentes zonas de producción pueden relacionarse con diferentes calidades comerciales. La mejora de este sistema de declaración de la información, por ejemplo, con una guía digital permitiría obtener un dato inmediato del ingreso en el momento de operativa plausible de ser utilizado por los comerciantes en la definición del precio en ese mismo día, ya que hasta ahora esa información es utilizada por operadores comerciales del Mercado, pero de forma más empírica u observacional, al determinar el nivel de oferta de los rubros en forma más “artesanal”, esto puede significar una mejora sustancial en la calidad y precisión de esta información que aporte transparencia al negocio y mejore las decisiones de todas las partes involucradas. La información de precios mayoristas de referencia de frutas y hortalizas es relevada dos veces por semanas utilizando un sistema ajustado para garantizar la confiabilidad del dato. El dato se recaba los lunes y jueves, por ser los días de mayor nivel de operativa comercial y por ser los dos días en que los precios varían generalmente. El sistema consiste en el relevamiento por un equipo de técnicos con el suficiente *expertise* para determinar los precios de venta de las diferentes categorías comerciales ofertadas en la plaza. Estos relevadores recorren los distintos puestos comerciales entrevistando a los llamados “informantes calificados” que son operadores que otorgan información confiable sobre el precio y situación de mercado. El técnico consulta el precio de venta, e inspecciona los atributos comerciales del producto ofrecido, registrándolo en una planilla en formato papel. En un mercado donde la calidad de los productos tiene una variabilidad muy grande, es necesario la observación en detalle de la calidad y calibre del producto ofrecido. Para hacer esta observación el técnico se basa en el "Manual de especificaciones técnicas de calidad" donde se especifican tres categorías de calidad Extra, categoría I y categoría II, y cuáles son las características para estos productos en cuanto a daños críticos y no críticos. Luego de hacer una recorrida por no

menos de 10 informantes calificados por día, el equipo de relevadores se reúne para intercambiar los valores relevados. Para la determinación del precio de referencia de cada una de las categorías relevadas se toman en cuenta diversos factores por ejemplo como estuvo la operativa, el tipo de informante, el día de la semana, todo esto influye en la determinación del rango de precios. Este precio refleja el rango donde se realizan la mayoría de las transacciones para cada uno de los calibres y calidades de cada uno de los rubros ofertados más significativos (que llegan a ser mas de 150 renglones), este precio es el precio sobre el final de la venta, evitando así las fluctuaciones que pueden presentar los valores a lo largo de la jornada comercial. En base a la información obtenida se elabora el boletín de precios mayoristas donde se indica un valor máximo y mínimo del precio de comercialización para cada producto de determinado calibre y categoría de calidad. Esta información es almacenada en una base de datos y publicada en la página web de forma de que sea accesible a cualquiera que así lo disponga. Además del boletín de referencia de precios mayoristas se elaboran una serie de informes complementarios con miras de enriquecer la información generada en los diferentes días. Al cierre de cada semana se elabora un informe que establece las variaciones de ingresos de mercadería y las variaciones de precios de los principales productos ofertados. Se utilizan los valores de los jueves de manera de observar la variación de precios al cierre de la semana. Además de la información cuantitativa, el reporte se enriquece con una breve reseña de lo acontecido en la semana para los rubros más significativos. Esa información es en gran parte explicaciones del comportamiento comercial de esa semana, o posibles repercusiones en el corto y mediano plazo. Cada mes se realiza un informe mensual, donde para una canasta fija de los principales 30 rubros ofertados se analiza la información del total de ingresos para cada uno de ellos en ese mes y para el precio promedio de cada especie, los datos son comparados tanto con los datos del mes anterior como con los datos del mismo mes del año anterior y del promedio de los últimos años para ese mes. A su vez se realizan los cálculos del llamado Valor Kilo Canasta (VKC), el cual intenta mostrar una variación general de los precios mayoristas al ponderar los precios de cada uno de los 30 rubros por sus respectivos volúmenes ingresados al Mercado para ese mes, en el entendido que la gran mayoría de lo que es comercializado en el allí es consumido y que al concentrar el Mercado el 60% del volumen nacional de frutas y hortalizas, puede tomarse lo que sucede aquí como un reflejo del consumo a nivel nacional. Se elaboran también con los registros de ingresos y precios un anuario, donde se condensa los datos del año y se compara con los anteriores. Se realizan índices económicos, de uso particular para un publico mas especifico donde se calculan los índices de precios y de valor. Si bien el alcance de estos se circunscribe a la actividad comercial desarrollada dentro del recinto del Mercado Modelo/UAM de Montevideo, como este es la principal central de distribución mayorista en todo el país, se puede extrapolar al comercio nacional de frutas y hortalizas. Es oportuno ejemplificar la complejidad de factores que involucran a la formación de precios, donde en manzana y otros frutos de hoja caduca, parte del precio de una temporada es determinado en el invierno anterior, donde por la naturaleza de estas plantas comienza a definirse la cosecha futura, esto puede medirse a través de la cantidad de frío que acumulan estas plantas, es decir la cantidad de horas de frío de un invierno puede determinar el volumen a cosechar la temporada siguiente y por ende ya anticipar un escenario de precios altos o bajos en

función de dicha acumulación de frío. Además, en algunos años se observa un fenómeno que a primera instancia puede parecer contradictorio y puede determinar cambios en el escenario de precios. En año de alta productividad se han constatado volúmenes muy altos de ingresos de mercadería en diciembre, donde ya han pasado alrededor de 10 meses de conservación desde la cosecha, llegando incluso a superar los ingresos en meses de plena cosecha, así que determinar los volúmenes conservados y su estado de calidad en este tipo de rubros permite mejorar la precisión con que se estiman los precios, ya que luego de la cosecha esta fruta es almacenada y paulatinamente volcada a la oferta hasta la nueva cosecha. Para obtener la información meteorológica se utilizó la base de datos climáticos de estaciones meteorológicas de INIA que cuentan con datos de varias zonas del país. En el transcurso de este proyecto, se estudiarán las variables más adecuadas a ser consideradas. Es de esperar que, dependiendo del producto en estudio, determinadas variables meteorológicas tengan una mayor incidencia en momentos particulares del año, y por esto es necesario un análisis especial para la definición de las variables a utilizar. Por ejemplo, en el caso de cultivos a campo de cebolla y ajo si se registra lluvia excesiva en el momento en que se determina la calidad y el potencial de conservación de estos, generara un impacto diferente que si se está previo a cosecha; donde también la ocurrencia de lluvias en este momento puede indicar un menor potencial de conservación de estos y por ende un escenario de precios anticipadamente altos. A las bases de datos mencionadas, se agregan otras que podrían resultar complementarias y eventualmente podrían utilizarse en la contextualización y caracterización del sector. Dentro de estas se encuentran los Censos Agropecuarios y las encuestas agroindustriales y hortofrutícolas realizadas por DIEA-MGAP. Estas últimas para el sector cítrico y de la papa, se cuenta con continuidad de la información hasta el presente año, en el caso del sector frutícola y hortícola se discontinuó en 2017 por la incorporación del Registro Nacional Frutihortícola. Por ello, en el proyecto también se utilizarán los datos del Registro Nacional Frutihortícola (RNFH). Este es un registro que fue implementado a partir del año 2015 y la inscripción es de carácter obligatorio según dispone la Ley N° 19.355 del mismo año. En éste corresponde que se inscriban aquellos productores o empresas con actividad comercial dedicada a la horticultura, fruticultura o producción de flores de corte. Del mismo se puede extraer información sobre número de productores, superficie cultivada por especie y variedad, tipo de protección, rendimientos y zona de producción para los rubros analizados en este proyecto. Asimismo, para una contextualización más general, se utilizará información macroeconómica del Sistema de Cuentas Nacionales que publica el Banco Central de Uruguay, de precios minoristas que releva y publica el Instituto Nacional de Estadísticas (en particular en lo referente a precios de comercialización minorista de frutas y hortalizas) y de evolución de ingresos de los hogares a partir de las Encuestas Continuas de Hogares.

La elaboración de herramientas de análisis y prospección permitirá a los actores claves de la institucionalidad pública poder interpretar de mejor forma, cuáles son los motivos que generan fluctuaciones de precios en el mediano y largo plazo, y si ese efecto es perdurable en el tiempo. Podrían representar una herramienta para el análisis económico que puede aportar a la toma de decisiones de política en base a la perspectiva brindada sobre futuros escenarios. La información generada por un análisis predictivo puede contribuir a la creación de nuevas oportunidades de

negocio para la agroindustria, por ejemplo, optimizando recursos y orientando sus acciones hacia procesos identificados como prometedores.

4. Objetivo i: Modelo de predicción

El objetivo ante todo es la predicción con el menor error, siendo definido este último en base a un conjunto de métricas delineadas en Hyndman (2006). El tipo de modelos que se aplican están enmarcados en el contexto de los modelos de espacio estado, tomándose como referencia los desarrollos de Durbin & Koopman (2012) y Petris et al. (2009). Toma particular importancia enfrentar las complicaciones que genera la frecuencia semanal de las series trabajadas, para lo cual se recurre a las especificaciones presentadas en De Livera et al. (2011) y Hyndman & Athanasopoulos (2021), que hacen uso de modelos de espacio estado donde la estacionalidad se modela con términos de fourier. La inclusión de variables auxiliares también es evaluada, pero muchas de las series propuestas no contaban con registros históricos o su efecto no fue claro sobre la predicción.

Como se indica en los objetivos originales del proyecto, la información que darían las predicciones sería útil para la toma de decisiones en los distintos niveles de la cadena productiva. Sin embargo, se deberá tener cuidado con el alcance de las predicciones y su publicación. Puede darse que una vez se comiencen a publicar las predicciones, los operadores que se desempeñan en la UAM incorporen dicha información en la fijación de precios y sirva como incentivo para la especulación. Esta influencia invalidaría cualquier tipo de modelo predictivo, al estar este interfiriendo con su objeto de predicción (Hyndman & Athanasopoulos, 2021, sec. 1.1). No obstante, debe ser tenido en cuenta que aun existiendo y siendo publicadas estas predicciones hay una plétora de factores que inciden en los precios hortifrutícolas, muchos de ellos difícilmente controlables, como son los climáticos. A su vez, hay un fuerte elemento de negociación en las transacciones del mercado, el cual depende de toda la información que puedan contar las partes, no solo pronósticos de precios, sino también la que circula durante el funcionamiento de la UAM (en general más rápido que el tiempo en el que se relevaría la información y se integraría a un modelo). A pesar de estas últimas consideraciones, se recomienda que las predicciones sean de uso interno de las instituciones asociadas al proyecto.

4.1 Antecedentes

La predicción de precios de frutas y hortalizas ha sido afrontada de varias formas, en particular, los enfoques paramétricos y los no paramétricos han resultado de interés. Los primeros constituyen un acercamiento al problema de predicción ya clásico, donde se plantea una estructura del modelo en la cual se asume una distribución para los datos y hay parámetros a estimar a partir de los mismos. Los modelos de espacio-estado sistematizados en el marco de la estadística y la econometría en la obra *Forecasting, Structural Time Series Models and the Kalman Filter* (Harvey (1990)) y el proverbial caso particular de ellos, los modelos Autorregresivos integrados de medias móviles (ARIMA), sistematizados por George Box y Gwylim Jenkins en los

años setenta, son ejemplos claros del enfoque paramétrico. Por otro lado, las técnicas no paramétricas son relativamente más recientes y no asumen una distribución para los datos, generalmente optando por resoluciones donde el componente computacional juega un papel de suma importancia. Dentro de estas últimas metodologías está enmarcada la de *Artificial Neural Networks* (ANN)², que consiste en un sistema de nodos interconectados, los cuales intercambian información entre sí (Peng et al., 2015). Aunque las metodologías no paramétricas cuentan con el atractivo de potencialmente predecir mejor las series de comportamiento volátil, los modelos de tipo paramétrico presentan la ventaja de que, si los datos se ajustan a los supuestos que el modelo requiere, será más sencillo desarrollarlo y emplearlo.

Yendo a trabajos previos particulares, en Dieng (2008) se predicen precios minoristas de vegetales en Senegal, haciendo uso de las metodologías ARIMA, Alisado exponencial (una técnica en la cuál subyace un modelo de espacio-estado), y del análisis espectral. Analizando cuál método da mejores resultados el autor concluye que el modelo ARIMA es el que resulta mejor para el problema planteado.

Por otro lado, en Li et al. (2010) las metodologías empleadas para la predicción son ANN y ARIMA y se hace énfasis en la predicción a corto plazo de precios mayoristas del tomate en China. Los resultados que obtuvieron los autores los llevaron a concluir que el modelo no paramétrico resultó superior. También en Luo et al. (2011) se hace uso de ANN para predecir precios mayoristas de China, en este caso de hongos comestibles. El uso de esta metodología es justificado por los autores en que es particularmente apropiada cuando la información de factores del mercado y el cumplimiento de los supuestos de los modelos no resulta clara.

Otro caso de empleo de metodologías no paramétricas se expone en Peng et al. (2015). Se busca predecir precios diarios de un conjunto de hortalizas y frutas, empleando las metodologías ARIMA, mínimos cuadrados parciales (PLS, por su sigla en inglés) y ANN, siendo las dos últimas las que cuentan con el mejor desempeño. Los autores hacen uso del concepto de *online learning* en sus algoritmos predictivos, actualizan el modelo predictivo con la información nueva que se va recabando.

En Mitra & Paul (2017) también se hace uso de metodologías mixtas, al considerarse que para capturar la complejidad de los fenómenos reales es necesario considerar un elemento no lineal además del lineal, dado que la complejidad hace que estos últimos pierdan capacidad explicativa. Se predicen precios del mercado de Angra en India mediante un modelo aditivo, donde se combina el uso de un modelo ARIMA y ANN.

A nivel nacional, a pesar de que ha resultado de interés la predicción del Índice de Precios al Consumo y de sus componentes, el sector frutas y hortalizas no ha sido un punto de foco en la literatura. Entre los trabajos previos que se enfrentan a la predicción de precios se encuentran

² Redes Neuronales Artificiales

Güenaga (2014), donde la autora emplea la metodología ARIMA para predecir precios de un conjunto de rubros, tanto por separado como en grupos de series agregados, ya que esto puede resultar en una mejor predicción si las series presentan un comportamiento similar. Por otro lado, en Millán & Romero (2019) se sistematizan los factores que inciden en la formación de los precios de un conjunto de frutas y hortalizas, estudiándose en particular aquello que incide en la tendencia, la estacionalidad y la variabilidad de las series. En Díaz (2018) se realiza una caracterización del mercado de frutas del Uruguay, en particular de los rubros manzana y pera, aunque con la finalidad de estudiar los determinantes del consumo de frutas y hortalizas.

En lo que respecta al modelado de series semanales en general, múltiples propuestas han sido presentadas.

En Harvey et al. (1997) se considera el enfoque de los modelos estructurales (modelos de espacio-estado). Se formula el modelo con una frecuencia diaria, donde la estacionalidad es modelada empleando dos componentes; una función de la fecha del año y una colección de efectos asociados a feriados móviles. El patrón de estacionalidad anual se expresa de forma trigonométrica. Luego, es posible convertir el modelo a una forma semanal, modificándose las ecuaciones apropiadamente.

Por otro lado, en Hyndman & Athanasopoulos (2021), se presentan distintas alternativas que se han empleado para enfrentar el problema. En particular, dentro del marco de los modelos de espacio-estado se propone la metodología de regresión dinámica armónica, propuesta en Young et al. (1999), la cual consiste en el modelado de la estacionalidad de manera determinística mediante términos de Fourier y de la dinámica a corto plazo mediante un ARMA. En Cleveland et al. (2014) se busca flexibilizar la representación de la estacionalidad empleando una regresión con pesos locales, lo cuál permite que dicho componente varíe en el tiempo. Otra opción presentada es la de los modelos TBATS³, la cual se introduce en De Livera et al. (2011) y que comprende también el uso de términos de fourier, pero de manera estocástica, entre otras características particulares de la especificación. Esta característica hace que este modelo pueda también modelar una estacionalidad que evoluciona en el tiempo.

Las metodologías anteriores se aplican en conjunto en el trabajo de Godahewa et al. (2020), en conjunto con Redes Neuronales y el método Theta (una extensión del alisado exponencial), con el propósito de contar con una metodología exhaustiva de modelado de series semanales.

³ Sigla en inglés de Estacionalidad trigonométrica con transformación Box-Cox, errores ARMA, tendencia y componentes estacionales

4.2 Metodología

4.2.1 Rubros elegidos

Los rubros tomados en cuenta para el análisis son: manzana, papa rosada, tomate (redondo), cebolla blanca (incluyendo la cebolla temprana) y naranja. Para su selección se tomaron en cuenta varios criterios.

En primera instancia se consideró que debían ser productos dentro de dos principales grupos comerciales; las frutas y las hortalizas, y que tuvieran peso en la canasta alimenticia.

En segundo lugar, se buscó un conjunto de rubros que entre sí sean claramente contrastantes. Esto refiere a que tengan diferencias sustanciales en las siguientes áreas: número de productores, tamaño medio de la explotación, tecnología utilizada, dinámica comercial, estacionalidad, perecebilidad y características de comercialización. La finalidad de esto es captar la diversidad de situaciones que pueden plantear distintos mercados.

Como tercer criterio de selección se consideró que la evolución de los precios pueda tener cierta caracterización, teniéndose conocimiento sobre los factores que afectan la producción y consecuentemente los precios, como los ambientales y el potencial de conservación.

Por último, se tomó en cuenta la volatilidad a corto plazo de la evolución de precios en el mercado, dado que esta última puede comprometer la predicción si es excesiva. En cuanto a esto último es excepción el tomate, que cuenta con una mayor variación en sus precios en marcos temporales reducidos, al ser un rubro de cosecha continua y particularmente susceptible al clima. A pesar de esto se incluye este producto al ser la segunda hortaliza más consumida en el país (Observatorio Granjero, 2019) y debido a que sus cambios en precios pueden ser caracterizados en el sentido del criterio previamente mencionado.

4.2.2 Bases de datos

a) Precios de frutas y hortalizas

Se cuenta con series digitalizadas de precios por kilo e ingresos en kilos para los rubros seleccionados, que van desde enero de 2013 hasta junio de 2022 y serán las empleadas para los análisis. A futuro es posible actualizar las bases con la información que se releve en adelante e incorporarla dentro de los modelos. Las series de precios originales tienen una frecuencia bisemanal, dándose los relevamientos en el marco del Observatorio Granjero los lunes y los jueves, días de mayor actividad en el Mercado Modelo/UAM (aproximadamente el 40% del volumen semanal comercializado se acumula en esos días). Para dicho relevamiento de datos se considera la metodología presentada en Comisión Administrativa del Mercado Modelo (2021), donde se especifica una tipificación de calidad. Si bien es posible obtener datos previos a 2013, no son considerados del todo fiables, por lo que no se tomarán en cuenta. Además, se acota el

problema a los rubros de origen nacional, dado que para los productos de origen extranjero no se cuenta con los volúmenes semanales de mercadería, sino con un agregado mensual.

El traslado de la operativa del Mercado Modelo a la UAM en febrero de 2021, el cual a su vez fue coetáneo con la pandemia de COVID-19, puede representar un cambio estructural en los registros, si bien la metodología de relevamiento no cambió fundamentalmente. Ante este hecho deberá posteriormente evaluarse si la situación amerita un cambio en los modelos.

Se cuentan con series de precios para los distintos productos y sus variedades, distinguiéndose en cada caso tres calidades (en orden descendiente): Extra, I y II. Para contar con una noción general del comportamiento de los precios, los técnicos del Observatorio Granjero recurren a cuestionar un conjunto selecto de informantes, que consiste en vendedores mayoristas (operadores) del mercado cuya información se considera fiable y “representativa”. Esta información se releva cuando ya las transacciones han comenzado a disminuir y consecuentemente el clima de precios ya se definió. Posteriormente, se discute la información obtenida y se llega a un consenso de cuáles son los precios puntuales máximos y mínimos observados para cada rubro de mayor veracidad. La representatividad no es en términos de muestreo probabilístico, debido a que no hay un diseño muestral subyacente en el relevamiento de la información, sino que se considera que el proceso permite captar con un buen grado de aproximación las señales de precios. Esto es justificado en la dificultad que representaría obtener información fidedigna de una muestra probabilística de todos los operadores; la selección de estos es deliberada y se toma en cuenta a aquellos considerados informantes calificados.

Esta metodología de muestreo no probabilístico se enmarca en el llamado *muestreo de juicio o de propósito*, en el cual la selección sigue un juicio o ideas arbitrarias de los técnicos en su búsqueda por una muestra representativa (Wolf et al., 2016, p. 328). Es susceptible a sesgos, debido al grado de subjetividad que implica, e ir realizando ajustes con la experiencia resulta vital para obtener buenos resultados. A modo de ejemplo, en el Observatorio Granjero es conocido que hay diferencias sistemáticas entre los relevamientos de algunos técnicos con respecto a otros, lo cual es tenido en cuenta a la hora de conciliar los distintos precios obtenidos. Esta característica del problema puede ser enmarcada dentro del contexto del modelado de espacio-estado, y en particular de sus casos lineales, donde se considera que se trabaja con observaciones ruidosas de una señal subyacente, siendo tal señal en este caso el precio verdadero. La magnitud inobservable del sesgo será en última instancia otro factor que distancie el modelo del proceso real de precios.

Se consideró deflactar las series de precios, con la finalidad de poder observar con mayor claridad tendencias sin el efecto de la inflación, pero esto resulta en una mayor complejización ya que al realizar predicciones con el modelo dicho efecto deberá ser reincorporado (con una variable que refiera a las expectativas de inflación, por ejemplo). De esta manera se estaría incorporando mayor incertidumbre en la modelización. Se considera que, al tomar los precios a valores corrientes, además de representar las verdaderas variables de interés, permiten incorporar

naturalmente el fenómeno inflacionario en el modelo. En caso de que haya una tendencia, esta puede ser incorporada dentro de las especificaciones del modelo.

Según los técnicos del Observatorio Granjero, la calidad Extra en general representa menos del 5% del total en los distintos rubros. Por este motivo se la excluye del análisis, considerándose de mayor importancia capturar el comportamiento de las otras dos calidades que además se toman como más representativas de los productos transados en el mercado y sus variaciones de precios. Las series de precios de los rubros separadas por calidad comparten un comportamiento similar, por lo que se considera que promediar en cada rubro las categorías I y II resulta un buen resumen de su evolución en cuanto precios. Dichas series de precios promedios serán las empleadas en los análisis.

Las series de precios son empleadas con frecuencia semanal, considerándose para cada semana el precio promedio de los dos precios relevados. A pesar de la simplificación que la serie semanal supone, permite apreciar la evolución de precios e ingresos y se gana en una mayor facilidad de manejo de las series respecto a los datos originales. El promedio se hace dentro de las semanas correspondientes al sistema de fechas ISO 8601, un sistema estándar definido sobre el calendario gregoriano de uso cotidiano. Un año tiene 52,1775 semanas, el sistema ISO opta por definir años con un número entero de semanas completas que comienzan el lunes y terminan el domingo, pudiendo ser este 52 o 53, para un total de 364 o 371 días por año. Bajo este esquema, en general los años cuentan con 52 semanas, a excepción del caso de los años donde el 1° de enero es jueves y los años bisiestos donde el 1° de enero es miércoles.

Como resultado de realizar los promedios para las series de cada rubro, se termina con 493 observaciones, con 52 observaciones en cada año, con excepción de los años 2015 y 2020, que cuentan con 53 semanas. Este número de semanas que cambia imposibilita el uso de metodologías clásicas de representar la estacionalidad de la serie (como las diferencias estacionales), que dependen de que el periodo estacional sea constante.

La muestra de entrenamiento para los rubros cuenta con 441 observaciones y va de la semana que comenzó el 3 de enero de 2013 a la que comenzó el 7 de julio de 2021. La de prueba tiene 52 observaciones y va de la semana que comenzó el 14 de junio de 2021 hasta la que comenzó el 6 de junio de 2022. La excepción de esto es la papa, donde el comienzo de la muestra se recortó debido a tener un comportamiento notoriamente distinto al resto de los datos. Debido a la cantidad de datos con la que se cuenta y que se están removiendo del comienzo de la serie, esto no afecta negativamente el análisis. Para este rubro la muestra de entrenamiento va de la semana que comenzó el 30 de diciembre de 2013 hasta la que comenzó el 7 de junio de 2021 (389 observaciones). Por otro lado, la muestra de prueba va de la semana que comienza el 14 de junio de 2021 y termina en la semana que va de 6 de junio de 2022 (52 observaciones).

b) Series de variables auxiliares

Otras series disponibles para el periodo considerado son las de ingresos de los distintos rubros al Mercado Modelo y la UAM, es decir, el volúmen de mercadería que ingresa a plaza. Se cuenta con estimaciones semanales, dado que no todos los transportistas declaran el volumen de productos con el que cuentan a su ingreso. Esta información no se encuentra desglosada por calidades y también es susceptible a errores de estimación, aunque con el traslado de las actividades a la UAM los registros presentan una mayor fiabilidad al haber una única entrada para los transportes (en el antiguo Mercado Modelo había 13 entradas, dificultando el registro).

También se seleccionó un conjunto de variables auxiliares consideradas de importancia para la explicación de fenómenos que afectan los precios de los rubros, las cuales se esperaba que mejoren la capacidad de predicción al incluirse en una lógica de regresión. Tal conjunto fue sopesado en varias reuniones con los técnicos del proyecto, considerándose inicialmente un grupo deseable amplio, más allá de consideraciones de disponibilidad de los datos.

Posteriormente, se procedió a la búsqueda de las distintas series especificadas. En algunos casos fue posible encontrar series completas para el periodo considerado, recurriendo a informes de organismos nacionales como la Dirección de Estadísticas Agropecuarias (DIEA) del Ministerio de Ganadería, Agricultura y Pesca (MGAP), las mediciones meteorológicas del Instituto Nacional de Investigación Agropecuaria (INIA), el Banco Central del Uruguay (BCU), e internacionales como la *Centrais de Abastecimento do Rio Grande do Sul*, entre otros. No obstante, no fue posible construir muchas de las variables que en un principio se habían planteado debido a la ausencia de datos, o si los había, eran pocas las observaciones en el tiempo lo que imposibilitaba la construcción de una serie.

4.2.3 Caracterización de rubros

a) Rubro Papa

La papa es el producto hortifrutícola de mayor consumo en fresco en nuestro país, produciéndose 81.700 toneladas anuales (promedio últimas tres zafras a 2019/2020), que se destinan en su mayoría para consumo en fresco, un 85% promedio entre los años 2014-2016. Se realizan dos ciclos de producción al año: en primavera, que va de agosto a diciembre, y en verano-otoño, que va de mediados de enero a mayo (aunque según ocurrencia de heladas puede terminar más tarde). La papa verano-otoño es la de mayor peso en la oferta nacional, con una participación de 69% en el volumen total, en el promedio de las últimas tres zafras cerradas a 2019/2020 (Freiría, 2020) y (Casares, 2020). Es un rubro de alta concentración donde 21 productores de más de 50 hectáreas ocupan el 67% del área total del cultivo asociado a dicha zafra. La producción se centra en el Sur del país (San José, Canelones, Montevideo, Florida, Colonia y Soriano) con el 92% de la superficie en la zafra de otoño 2020 (Casares, 2020).

Las cosechas de las dos grandes zafras, otoño y primavera, se dan de febrero a abril y de octubre a diciembre respectivamente. La principal diferencia entre estos ciclos es que el área destinada

al cultivo de primavera es significativamente menor que el área destinada en otoño, debido a que existen opciones de otros cultivos de verano (soja, maíz, etc), que en función de los precios internacionales de los *commodities* y la cotización del dólar pueden ser opciones atractivas para sustituir área de papa.

A su vez, al ser un rubro donde existe una gran concentración, el conjunto de expectativas sobre los resultados productivos es el primer factor que influye en la formación del precio. Todos los aspectos que modifiquen los resultados productivos (Área sembrada, rendimiento por hectárea), los que afecten la calidad y su capacidad de conservación (Calidad en momento de cosecha, volumen guardado en cámaras, calidad en conservación) son determinantes en la cotización del producto. Otros factores que determinan la cotización son los que provienen del lado de la demanda mayorista, que van desde la percepción en la plaza, condiciones meteorológicas para el fin de semana, momento del año, etc. Es complejo generar indicadores que permitan la inclusión de estas variables en las especificaciones de los modelos ya que no existen bases de datos que releven esta información de forma periódica y consistente, salvo la relevada en la *Encuesta de Papa de DIEA*, a partir de la cual se obtiene información de la producción, el área sembrada y el rendimiento. Adicionalmente, se dispone de las series de precipitaciones y de cotización del dólar, obtenidas del GRAS⁴ del INIA y el INE⁵, respectivamente.

⁴ Unidad de Agro-clima y Sistemas de información

⁵ Instituto Nacional de Estadística



Figura 4.1: Series de precios semanales promedio por kg. Rubro: papa.

Como se puede apreciar en la [Figura 4.1](#), la oferta entre los años 2014 a 2019 se mantuvo en una banda de precios, donde han ocurrido muy pocos eventos de fuertes alteraciones de precios. Es probable que esta tendencia se deba a que es un sector productivo con pocos productores que poseen capacidad logística y productiva para regular la oferta, y a la integración vertical muy fuerte por parte de los productores de mayor porte, lo cual hace que tengan la información y la capacidad de articular la fijación de precios. Además, debe tenerse en cuenta que el elevar los precios por encima de un llamado “precio de referencia” implica que se importe el producto lo cual también puede estar teniendo un factor disciplinante sobre el precio. Este “precio de referencia” es una cotización construida con el fin de que el productor tenga un margen de ganancia razonable y el consumidor enfrente un precio que no es excesivo; mantenerse debajo de ese precio implica que no se de importación. Adicionalmente la papa industrializada, que en su mayoría se importa, es un sustituto fuerte de la papa fresca, lo cual también puede estar contribuyendo en el efecto disciplinante sobre el precio. Esta condición de concentración de la producción dificulta la predicción de precios, ya que estos terminan dependiendo principalmente de pocos actores, por encima de otros factores o una lógica de mercado.

A mediados de 2013 hubo una caída fuerte de precios y desde 2020 hay un aumento persistente hasta 2021, el cual se puede deber a las dificultades para la importación en el contexto de la pandemia de COVID-19. La atipicidad del año 2020 deberá ser evaluada a la hora de modelar ya

que podría llevar a malas predicciones, debido a que puede considerarse un quiebre estructural transitorio.

El volumen de ingresos puede ser una variable *proxy* de la oferta de los rubros y a priori se esperaría que pudiera tener un efecto considerable sobre la fijación de precios, por lo que es de las primeras variables auxiliares a considerar. Los precios promedios semanales y los ingresos presentan una correlación lineal simple de -0,23, que si bien tiene signo negativo como es de esperarse, no es de gran magnitud.

b) Rubro Manzana

El volumen de producción ronda las 53.333 toneladas anuales promedio para las zafas comprendidas entre 2017/18-2019/20. Se estima que, entre 2013 y 2015, de la oferta local de manzana (producción nacional más importaciones) se destinó a consumo en fresco un promedio del 78%, para industrialización el 17% y para exportación el 5%. La manzana es una de las pocas frutas y hortalizas producidas localmente que destina un volumen relevante para la exportación. El proceso de producción de la manzana involucró unos 472 productores en la zafra 2015/16⁶. La superficie destinada a este cultivo se estima que alcanzó las 3.059 hectáreas promedio en las tres zafas culminadas en 2019/20. Dicho cultivo representa aproximadamente el 60% de la producción total de Frutales de Hoja Caduca destinada a cultivos frutícolas en Uruguay, según Bregante (2016).

La producción está concentrada en los departamentos del Sur del país: Montevideo (fundamentalmente en la zona de Melilla), Canelones, San José y Colonia. Al ser un cultivo plurianual los cambios en el volumen de la producción están más asociados a factores que afectan la productividad de una zafra que a cambios en su superficie, como pasa en la mayoría de las hortalizas de ciclos productivos cortos. De todas formas, analizando el periodo 2001/02 al 2015/16, se constata que la superficie total pasó de 3.716 a 2.677 hectáreas (Bregante, 2002, 2016), por lo cual no se descarta que a largo plazo tal reducción haya tenido un impacto en la producción.

Regularmente, el periodo de cosecha se concentra de enero a abril, meses en que predomina en el mercado el producto recién cosechado. Hacia los meses de invierno, se conserva en cámaras comunes y entrada la primavera el producto proviene de cámaras de atmósfera común. Hacia fin de año se incorpora (generalmente sobre noviembre y diciembre) producto importado, que complementa la oferta en casos de bajo stock o baja proporción de frutas de calidad comercial.

⁶ En 2016 se publicó la última encuesta Frutícola de Hoja Caduca elaborada por DIEA-MGAP, en la cual calculaba la superficie, producción, número de productores. Luego de esos años, los datos mencionados refieren a estimaciones realizadas por técnicos del MGAP con base a datos del Registro Nacional Frutihortícola, registros de ingresos de la Unidad Agroalimentaria de Montevideo y volúmenes de exportación de *Penta Transaction*.

La distribución comercial a lo largo del año según las variedades más importantes localmente es la siguiente:

- Enero a abril: grupo Gala
- Mediados de febrero a diciembre: Red Delicious y mutaciones
- Marzo a diciembre: Granny Smith
- Abril a diciembre: Fuji y Cripps Pink

La lógica de la conformación de precios, en la que operadores comerciales y productores establecen el envío de producción al mercado, opera en función de las expectativas ya que son rubros de cosecha anual, donde operan pocos actores y existe posibilidad de regular *stocks*. Los resultados en cuanto a cantidades producidas, y la consecuente incidencia en los precios, dependen de un gran cúmulo de factores que van desde las horas de frío acumuladas durante el receso invernal, nivel de precipitaciones, ocurrencia de fenómenos climáticos extremos, entre otros. La manzana es un cultivo que depende particularmente de los factores climáticos anteriormente mencionados, cuya inclusión deberá ser evaluada a la hora de especificar los modelos. A su vez, se debe valorar los correspondientes rezagos temporales necesarios, ya que si bien se espera que los fenómenos citados afecten los precios no lo harán en el periodo que ocurren, sino cuando la zafra afectada tenga que entrar en el mercado.

Los factores mencionados son los que definen el volumen existente y su calidad durante la zafra, determinando en la plaza comercial la cotización del producto. También influye al momento de la determinación de los precios los costos de producción, que si bien son variables en función de la escala de producción y la tecnología aplicada (por ejemplo, el costo del combustible puede afectar más o menos), tienen un fuerte componente de mano de obra (condicionado su costo por el mercado laboral) y de los insumos (condicionados por la cotización del dólar). Los costos de conservación asimismo influyen al momento de fijar precios del producto, en la medida que la zafra avanza alejándose del momento de cosecha y resulta necesario mantener el producto en cámaras de frío. En momentos en que se constatan problemas de calidad organolépticos y los *stocks* remanentes son bajos, se recurre a la importación de manzana para complementar la oferta nacional, por lo que en esta coyuntura un factor que condiciona el precio es la cotización de manzana a nivel regional. Una problemática de incluir los precios en la región como variable auxiliar es la identificación del rezago con que se debería incluir esta variable, debido a la complicación de identificar el momento en que se estudian para evaluar la importación. En este rubro fue posible encontrar bases de variables auxiliares para la producción, las horas de frío acumuladas y las precipitaciones, además del ingreso de mercadería.



Figura 4.2: Series de precios semanales promedio por kg. Rubro: manzana.

Se puede apreciar un comportamiento fuertemente estacional en el precio de la manzana, en general comenzando bajo a principio del año (siendo marzo el mes que en general registra las cotizaciones más bajas, luego del inicio de cosecha del grupo Gala de manzanas a mediados de enero) y aumentando a medida que avanza el año. La caída en los precios que se da luego del aumento a lo largo del año suele ser bastante abrupta. De marzo a julio, generalmente las manzanas almacenadas en “cámara común” mantienen buenas características de calidad (se conocen como atmósfera “común” o convencional en la jerga de los operadores a aquellas cámaras frigoríficas con menor nivel de control de las condiciones de su atmósfera). Posteriormente, hacia finales del invierno e inicios de primavera comienzan a intensificarse los problemas de conservación en parte de la oferta, lo que genera que se incrementen los valores mayoristas. También es generalmente el inicio de las aperturas de cámaras de “atmósfera controlada”, como se denominan a aquellas cámaras de conservación que presentan mayor control sobre las condiciones de sus gases, permitiendo generalmente, mejor conservación de la calidad. El producto de mayor calidad es el que tiene mayor potencial de conservación y el que se pone a disposición del público hacia el final del año, siendo su precio mayor.

Hay un aumento del rango de la serie del precio promedio hacia los últimos años y una tendencia a que el pico de precios se vuelva mayor, siendo el de fines de 2020 y principios de 2021 el más

elevado, cuya atipicidad a su vez puede que tenga que ser evaluada (ver [Figura 4.2](#)). Esto puede indicar la necesidad de considerar que hay un elemento de tendencia.

c) Rubro Tomate

El tomate es la segunda hortaliza más comercializada en Uruguay (Observatorio Granjero, 2022). A nivel nacional se produjeron unas 36.000 toneladas aproximadamente en 2021 según estimaciones del observatorio granjero. En los últimos años se ha dado una tendencia a la disminución de los productores, llegando a unos 855 según los datos de DIEA, pero con un importante aumento de la productividad por superficie (Millán & Romero, 2019). Hay dos zonas de producción en el país, siendo la primera el litoral norte (Salto, Constitución y Bella Unión); que acumula aproximadamente 60% de la producción nacional y donde casi la totalidad se produce bajo invernáculo. La segunda es la zona sur (Canelones, Montevideo, San José), donde se produce el 40% restante bajo la modalidad de invernáculo, que año a año gana participación, y “a campo”, donde se producen variedades donde una porción de lo cosechado tiene como destino la industria, como el tomate perita. El avance tecnológico y la concentración del rubro determinan que la modalidad “a campo” cada vez sea menos relevante en la composición de la oferta.

La producción de tomate es esencialmente no transable con el exterior durante todo el año, está destinada al abastecimiento del mercado doméstico y sólo se realizan importaciones en momentos puntuales de baja producción. La perecibilidad y escasa vida poscosecha que posee este fruto, hacen que sea poco atractivo para los importadores por el alto riesgo que tiene su importación. Además, el hecho de no tener un flujo comercial establecido hace que normalmente demore varios días hasta que finalmente llega la partida a la venta.

El tomate es un artículo que presenta una particular sensibilidad a los cambios de clima (temperatura, precipitaciones e intensidad de luz) tanto a nivel de volumen ofertado en la plaza comercial como en la demanda por parte de los consumidores (que es mayor en verano). Se trata de un cultivo muy susceptible y cuando se conjugan algunas condiciones ambientales como son la alta humedad relativa, temperaturas altas y el exceso de precipitaciones suelen presentarse problemas sanitarios severos, lo que redundaría en una baja calidad de producto en general, dado que no hay buenas condiciones de cuajado. Esto hace que aquellas partidas que presentan buenas condiciones para la comercialización se coticen a mayor precio. Otro factor que es determinante en el precio es la superficie sembrada; cuando hay años de precios altos y menor superficie implantada tienden a seguir años en que se aumenta la intención de siembra lo que generalmente repercute a la baja en los precios.

Las variables auxiliares con las que se cuentan para este rubro son una estimación del nivel de producción y la estacionalidad de la demanda por el verano. Si bien es posible obtener los valores para las variables ambientales, resulta difícil determinar su conjugación y valores exactos para definir la incidencia en las condiciones de cuajado, por lo que no son incluidas. También fue discutida la inclusión de una variable que refiera a la concentración de la producción, pero no hay datos disponibles sobre este factor.

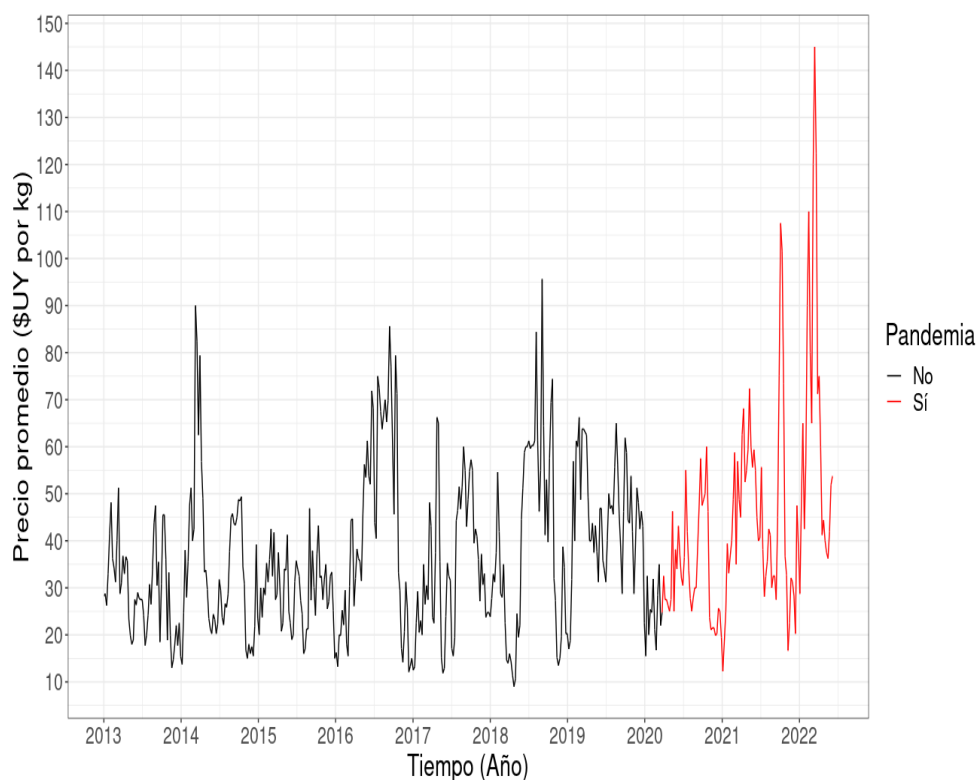


Figura 4.3: Series de precios semanales promedio por kg. Rubro: tomate.

Se presenta el gráfico de la evolución de los precios promedios en la [Figura 4.3](#). Los precios generalmente están sujetos a comportamientos irregulares, debido a que la oferta es muy sensible a las condiciones ambientales y cualquier eventualidad climática puede ocasionar escasez o sobreoferta. Lo mismo sucede con la demanda que es altamente variable y dependiente en gran medida de la temperatura y condiciones atmosféricas, lo cual también puede estar incidiendo en el precio. El comportamiento que ocurre en forma sistemática y la estacionalidad no son evidentes a primera vista y deberán ser estudiados con el conjunto de metodologías que se trabajarán más adelante.

En marzo de 2022 se registraron precios atípicamente altos, incluyendo el máximo histórico de la serie en la semana 11 (\$145 por kilogramo de tomate). Dado que este valor se encuentra al final de la serie habrá que evaluar su intervención, dado que se espera que tenga un efecto sobre las predicciones, en comparación si este atípico se diera al comienzo de la serie, por ejemplo.

d) Rubro Cebolla

La producción cosechada de cebolla se estima en unas 32.650 toneladas de acuerdo con lo relevado por DIEA en promedio en las últimas tres safras y es realizada por entre 800 y 900 productores aproximadamente cada año (Millán & Romero, 2019). Se distinguen dos zonas de producción: la zona sur, desde donde proviene el 77% del volumen total y la zona norte, que

provee el 23% restante. Las tareas culturales que generalmente implica este cultivo, desde la etapa de siembra hasta la comercialización mayorista, son: elaboración del almácigo solarizado, trasplante, cosecha y engavillado, descolado y almacenamiento en *bins*, y *packing*.

Es un cultivo que tiene una gran demanda de mano de obra, sobre todo zafral en momentos claves del cultivo como el trasplante y la cosecha. Esto es evidenciado cuando se analiza la estructura de costos de producción donde se estima que la mano de obra tiene un peso aproximado de 45% del total. Dentro de esta, un 36% corresponde al costo de la mano de obra zafral por la cosecha y *packing* (e incluye servicios contratados), en tanto que el restante 9% es por la mano de obra permanente. La cebolla es un producto de “guarda”, en la cual el momento de cosecha y el potencial de conservación está determinado por la variedad que se cultiva y la zona de producción.

Existen básicamente tres tipos de cebolla; las tempranas, las intermedias y las de ciclo largo. Las tempranas son las que se cultivan principalmente en el litoral norte, que tiene un ciclo de producción corto y un potencial de conservación escaso (menor a 2 meses). Por otro lado, las intermedias son caracterizadas muchas veces como cebollas tipo “pantanosas”, cosechadas en diciembre y con un potencial de conservación de hasta 8 meses, siendo la de mayor volumen producido. Finalmente, las de ciclo largo tienen mayor grosor de cáscara y son las de mayor potencial de conservación, cuyo momento de cosecha es entre enero y febrero.

Estas particularidades son las que determinan que la zafra inicie en setiembre con las primeras partidas de cebollas provenientes del litoral norte. Hacia febrero cobra relevancia la oferta de cebollas intermedias y hacia agosto ingresan las últimas partidas de cebolla tardía. Por lo tanto, generalmente queda un periodo ventana donde la oferta nacional debe complementarse con importaciones de la región para satisfacer la demanda. Estas adquieren mayor relevancia en años en que se verifican problemas de conservación a raíz de los cuales el periodo ventana puede extenderse, debido a que los menores volúmenes de oferta nacional no satisfacen la demanda doméstica.

Entre las variables que inciden en el precio se encuentran el área sembrada y el rendimiento global del cultivo. La intención de siembra de un año depende muchas veces de los resultados del ejercicio o la zafra anterior. A su vez, la productividad depende del régimen hídrico, esto es, el exceso o falta de precipitaciones al momento de la cosecha, años de mayor o menor incidencia de las enfermedades que mayor afectan el cultivo y la capacidad de conservación. La cantidad de semillas vendidas tiene también efecto sobre el precio al tener efecto sobre la producción, y finalmente, el precio en la región, dado que cuando hay escasez se importa, también tienen efecto sobre los precios. Del conjunto de variables fue posible contar con la producción por zafra y las precipitaciones.

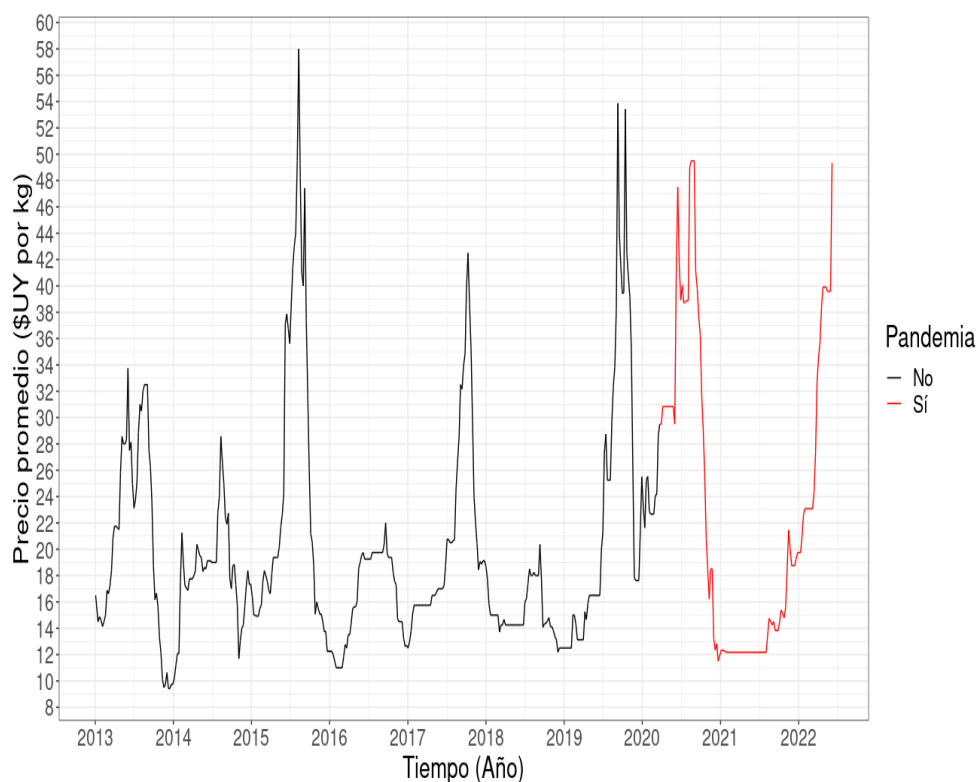


Figura 4.4: Series de precios semanales promedio por kg. Rubro: cebolla.

En la Figura 4.4 se puede observar la evolución del precio de la cebolla, resaltando en primera instancia los picos que se dan en algunos de los años como 2015, 2017, 2019, 2020 y 2022. De los años pasados, 2020 es el que cuenta con el mayor precio promedio anual, mientras que 2021 cuenta con el menor precio promedio anual, lo cual puede interpretarse como una manifestación del shock causado por la pandemia. A primera vista se puede apreciar cierto comportamiento estacional en los precios, donde hacia mitad del año hay un aumento sostenido que para mitad de año ya se ha revertido.

e) Rubro Naranja

La naranja es la principal especie de cítricos cultivada en el país. La producción anual con calidad comercial es de 122.000 toneladas en promedio en los últimos años. La oferta casi en su totalidad se conforma de producto nacional, registrándose escasas importaciones que generalmente no superan el 1% y ocurren en años que se detectan severos problemas de calidad en la fruta. El principal destino de todos los cítricos es la exportación y en particular para la naranja este destino capta aproximadamente 54.700 toneladas por año (45% de la oferta). La industria absorbe 34.600 toneladas (28%) y las restantes 32.700 toneladas (27%) quedan disponibles para el consumo en fresco en el mercado interno.

Según Casares (2017) la producción citrícola del Uruguay ocupa una superficie total de 14.848 hectáreas e intervienen 423 productores. Actualmente cuatro empresas con más de 200.000 plantas producen el 61% del volumen total; por otra parte, los citricultores de menos de 5.000 plantas son el 65% del total, pero explican solamente el 4% de lo producido. Las empresas de mayor tamaño son las que se han establecido en el mercado como los referentes de comercialización mayorista a nivel de mercado interno y a su vez son las empresas que operan en la exportación de la fruta.

En lo que refiere a la distribución del sector a nivel nacional, aproximadamente el 92% de los productores se encuentran ubicados en la zona norte del país (Salto, Paysandú y Rio Negro) y el resto en el sur (San José, Colonia y Montevideo).

La naranja es un producto que se oferta durante todo el año, aunque la disponibilidad disminuye en los últimos meses de verano y otoño. Básicamente existen dos tipos de naranjas, el grupo Valencia, cuya oferta va de noviembre a abril, y el grupo Navel (ombligo), de mayo a octubre.

Dentro de los factores que inciden en la cotización del producto destinado para consumo en fresco, al igual que la mayoría de los productos de la canasta frutihortícola, se encuentra el volumen de producción. La calidad del producto con categoría comercial está determinada por la estacionalidad del producto y las condiciones climatológicas y fitosanitarias del cultivo. Adicionalmente, incide también en el precio la estacionalidad de la demanda. Este rubro presenta la particularidad de que el principal destino comercial es la exportación; los mercados externos absorben la fruta con mejor calidad comercial, por lo que el mercado doméstico ocupa un lugar secundario en el caso de grandes empresas exportadoras. Es así como los mercados internacionales y los precios de venta internacional pueden influenciar en el nivel de oferta de fruta en el mercado interno. De estas variables se cuenta con la producción, proveniente de la encuesta citrícola y la estacionalidad de la demanda. Si bien se pudieron obtener precios de exportación, surgió la problemática de definir cuándo estos precios afectarían la formación de precios nacional, situación análoga al caso de la manzana.

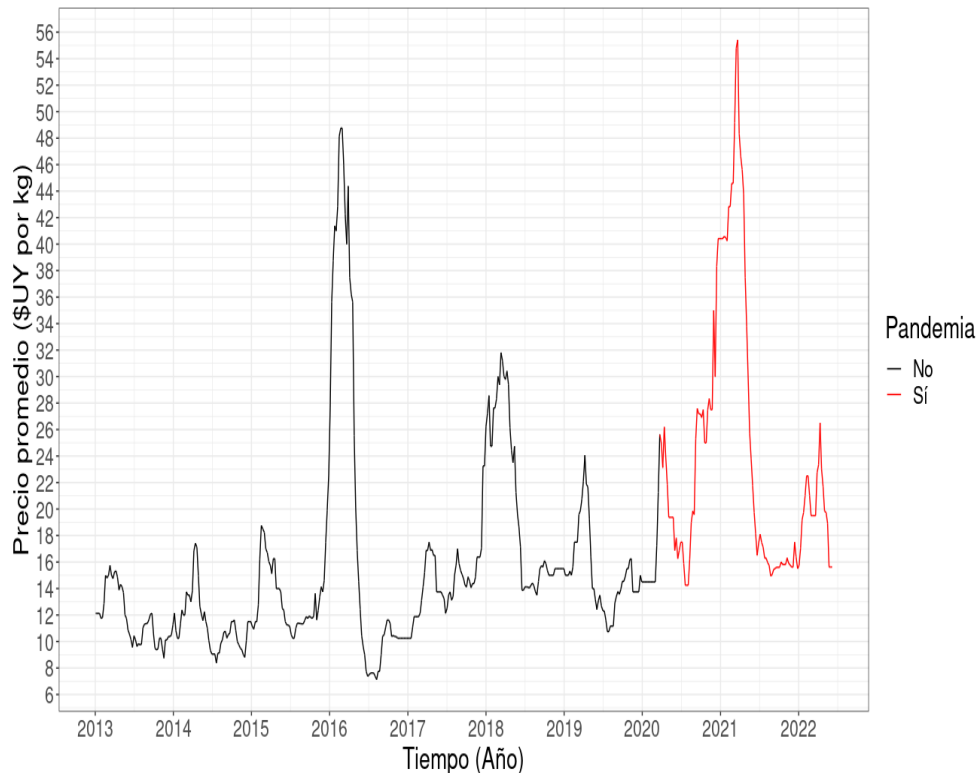


Figura 4.5: Series de precios semanales promedio por kg. Rubro: naranja.

En términos generales, el precio de este rubro sigue un patrón donde los picos se dan en torno al comienzo del año (Figura 4.5). En los años 2016 y 2021 estos picos tomaron valores mucho mayores a los de los otros, lo cual puede dificultar el modelado, en particular el último debido a que se encuentra a finales de la muestra y esto tendrá consecuencias sobre las predicciones.

4.3 Resultados

El acercamiento general al problema para los distintos rubros es el mismo. Se estiman un conjunto de modelos de los que son desarrollados en el Sección 5.1, se evalúa su desempeño predictivo dentro de la muestra de entrenamiento y de la de prueba, considerándose en esta última un horizonte temporal de corto plazo (12 semanas) y otro de largo plazo (52 semanas). Posteriormente, se evalúa la predicción puntual mediante las múltiples medidas de error presentadas en la siguiente subsección. Si bien las metodologías son las mismas para cada cultivo, se estudian las especificaciones de los modelos y obviamente la estimación de los parámetros para cada rubro particularmente.

4.3.1 Métodos estadísticos

a) Modelos de espacio-estado

Los modelos que son el foco de atención en este trabajo son los modelos de espacio-estado, una metodología flexible y general que permite tratar una variedad de problemas de análisis de series de tiempo (Durbin & Koopman, 2012). Este enfoque tuvo sus orígenes en el campo de la ingeniería, para luego ser aplicado en la estadística y la econometría, siendo abordado en libros como Harvey (1990) y West & Harrison (1997).

Bajo este marco, se considera que el sistema o proceso que se estudia es determinado por una serie de vectores no observables $\theta_1, \dots, \theta_n$ (proceso de estado), que se asocian a una serie de vectores que sí es observada y_1, \dots, y_n (proceso observado). La estructura del modelo es definida en base a estos conjuntos de vectores, matrices que definen la especificación y vectores de términos de error; todos ellos pudiendo depender de parámetros desconocidos a ser estimados.

Uno de los atractivos de esta clase de modelos es su flexibilidad, que permite trabajar directamente con series de media y varianza inestable (no estacionarias) y cambios de nivel, sin la necesidad de realizar transformaciones para lograr la estacionariedad como las que son necesarias para trabajar con los modelos del tipo *ARIMA* tradicionales (los cuales de hecho son un caso particular de los modelos de espacio-estado lineales). Asimismo, permite la inclusión de regresores, lo cual puede ser de utilidad para lograr una mejor predicción o dar cierta explicación al fenómeno estudiado.

La estructura de un modelo de espacio-estado lineal se puede describir de la siguiente manera:

$$Y_t = F_t \theta_t + v_t, v_t \sim N(0, V_t)$$
$$\theta_t = G_t \theta_{t-1} + w_t, w_t \sim N(0, W_t)$$

Donde:

- Y_t es el vector m -dimensional de los valores de las m series de las variables consideradas en el momento t .
- θ_t es un vector p -dimensional de procesos no observables.
- F_t y G_t son matrices de dimensiones $m \times p$ y $p \times p$ respectivamente que permiten caracterizar la especificación del modelo.
- v_t y w_t son perturbaciones de los procesos observados y de estado respectivamente.

En particular, en este trabajo se emplean las especificaciones descritas en Petris et al. (2009) y De Livera et al. (2011).

a) Evaluación de predicciones

Dado que el objetivo de este trabajo es la predicción, resulta necesario definir una metodología para evaluar el desempeño predictivo de las especificaciones de los modelos que se planteen.

A partir de esto surgen dos cuestiones a tratar: en qué datos se evalúan las predicciones y la o las métricas que se emplean para hacerlo.

i. Muestra de entrenamiento y de prueba

Es de interés evaluar predicciones de datos verdaderamente futuros, es decir, si tenemos una serie de tiempo con T datos, queremos evaluar las predicciones para los datos en un momento $T + h$, siendo h el horizonte predictivo. Esto se debe a que en general si se realizan predicciones con el modelo dentro de la muestra con la cual se estimó, su desempeño será en general mejor que si se realizan predicciones verdaderas.

Obviamente, no es posible realizar tal evaluación, dado que los verdaderos datos futuros no pueden ser conocidos. Sin embargo, podemos buscar simular esta evaluación con los datos con los que sí contamos, separando la muestra en una sección de entrenamiento y otra de prueba. En la primera se ajustará el modelo a partir del cual se realizan predicciones a un determinado horizonte predictivo. Luego, comparamos nuestras predicciones con los respectivos valores que están en la muestra de prueba, que hacen las veces de valores “futuros”.

Como se indica en Hyndman & Athanasopoulos (2021), cuando la cantidad de datos lo permite el tamaño de muestra de prueba se fija en un 20% del total de datos e idealmente tantos datos como lo requiera el horizonte predictivo deseado.

Una vez que fueron separados los datos en estas muestras, se tiene que decidir qué métrica se emplea para comparar predicciones con valores reales.

i. Métricas de evaluación

Las métricas de evaluación del error de predicción son medidas de resumen de cuánto se equivoca un modelo en predecir determinada variable en un conjunto de datos dado. El error se define como:

$$e_{T+h} = y_{T+h} - \hat{y}_{T+h|T}$$

donde y_{T+h} es la observación de la variable en el momento $T + h$ y $\hat{y}_{T+h|T}$ la predicción de dicha variable en ese momento.

Sea $\{y_1, y_2, \dots, y_T\}$ la muestra de entrenamiento y $\{y_{T+1}, y_{T+2}, \dots, y_{T+m}\}$ el conjunto de prueba de un largo m . En general las medidas son un promedio de algún tipo de los errores en la muestra de prueba, las empleadas en este trabajo son:

- Raíz del error cuadrático medio: $RMSE = \sqrt{\sum_{i=T+1}^m e_i^2}$
- Error absoluto medio: $MAE = \sum_{i=T+1}^m |e_i|$

Estas medidas son dependientes de la escala y unidad de medida de la variable en cuestión, por lo que sirven para comparar predicciones que sean de una misma variable.

Ahora, sea $p_t = 100 e_t/y_t$. Una medida de error porcentual es:

- Error porcentual absoluto medio: $MAPE = \sum_{i=T+1}^m |p_i|$

Tiene la ventaja de ser independiente de la escala y tener la interpretación simple de ser el porcentaje promedio de los errores. Por otro lado, tiene desventajas, como el hecho de que penaliza más severamente los errores negativos que los positivos, no está definido cuando la variable evaluada toma el valor 0 y toma valores extremos cuando dicha variable toma valores cercanos a 0.

En Hyndman (2006) se propone el error absoluto escalado medio (MASE). Un error escalado se puede definir como:

$$q_j = \frac{e_j}{\frac{1}{T-1} \sum_{t=2}^T |y_t - y_{t-1}|}$$

Como se puede apreciar, como la escala de los datos está tanto en el numerador como el denominador, q_j es independiente de la escala de los datos. Este tipo de error tiene la interpretación de que será menor a uno si la predicción es mejor que la predicción *naive*⁷ promedio en los datos de entrenamiento. Luego, la medida de error para el conjunto de las predicciones se define como:

$$MASE = \sum_{i=T+1}^m |q_j|$$

Alternativamente, el denominador de los errores escalados puede ser remplazado por el error de la predicción *naive* estacional o por el error de la predicción *naive* en el conjunto donde se está evaluando la predicción, caso en el que se contaría con una medida de error relativa.

⁷ Es la que surge de predecir el próximo valor de la variable con el valor actual

4.4 Resultados para el rubro papa

a) Descomposición de la serie de precios

Como se mencionó anteriormente, este rubro presenta características que hacen difícil su predicción mediante el herramental de modelado de espacio-estado, dado que el elevado grado de concentración del rubro hace que haya una definición de los precios desde el lado de la oferta.

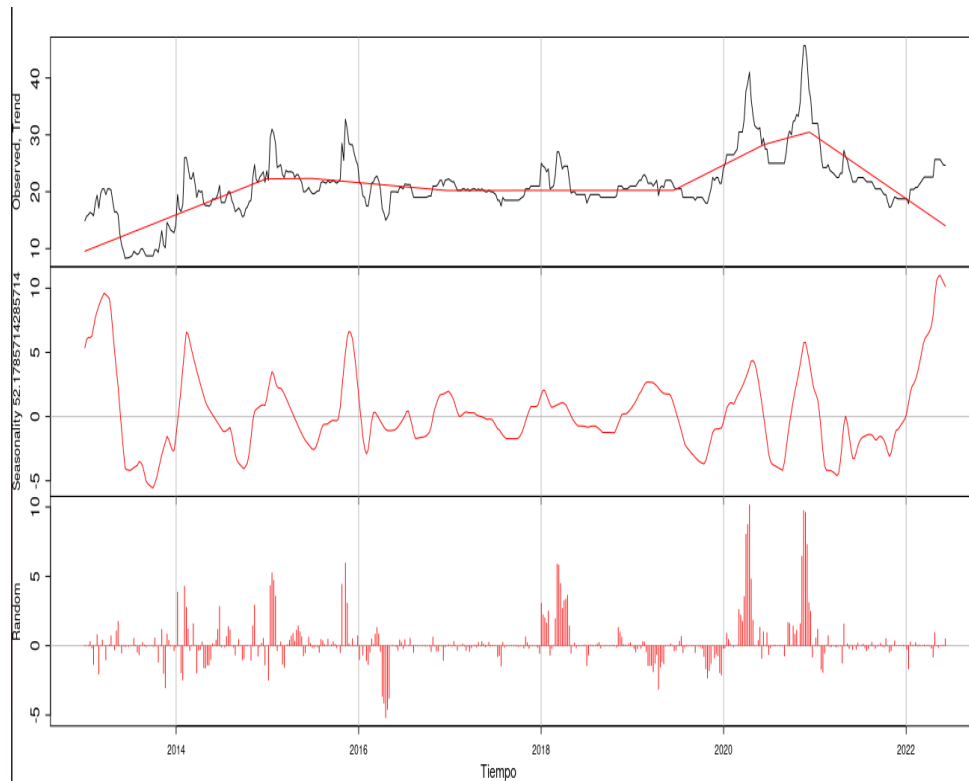


Figura 5.1: Descomposición STR robusta de la serie de precios de papa.

En la Figura 5.1 se muestra el gráfico de los componentes de la serie de precios promedio de papa resultado de la descomposición STR. En primera instancia se observa como el elemento de tendencia permanece relativamente constante de 2014 a 2020, y cuando se ve alterado es por efecto de perturbaciones considerables más que por una tendencia en sí misma. Luego, el elemento estacional no parece presentar un patrón claro año a año o dentro de los mismos y en el elemento aleatorio o irregular es donde caen las anomalías correspondientes a los años 2020 y 2021.

b) Ajuste de modelos, selección y diagnóstico

Se ajustan los siguientes métodos y modelos presentados en el Sección 5.1: promedio simple, caminatas al azar simples (*Random walks*) y con regresores, modelo SUTSE, alisado exponencial

estacional, ARIMA simple, ARIMA con estacionalidad representada con términos de Fourier y el modelo TBATS. Los modelos de corte estocástico están enmarcados en los modelos de espacio estado y desde esta perspectiva es que son estimados sus parámetros.

Debe evaluarse el cumplimiento de los supuestos realizados a la hora del planteo de los modelos para constatar su validez. Al realizarse las pruebas de diagnóstico de autocorrelación nula y normalidad de los residuos para los modelos de corte probabilístico, para todos ellos se rechaza la hipótesis de normalidad; lo cual puede deberse a los posibles quiebres estructurales que se pudieron observar en los gráficos de la serie y/o que el proceso generador de datos no siga este supuesto distribucional. Observando los residuos, en general se puede apreciar heterocedasticidad y varios residuos atípicos, factores que en general comprometen el cumplimiento del supuesto de normalidad. Debido a que varios componentes de la inferencia se fundamentan en el cumplimiento de dicho supuesto, como por ejemplo los intervalos de confianza asociados a la predicción, resulta necesario encontrar una manera de subsanar este problema. Una manera de hacerlo es emplear técnicas de bootstrap a partir de la cual se estiman intervalos de manera no paramétrica. Adicionalmente, se debe tener en cuenta que esta falla de la hipótesis distribucional afecta incluso la estimación de los parámetros involucrados en el modelo, que también se fundamenta en ella. Empero, como se expresa en Hamilton (1994, p. 117), las estimaciones realizadas bajo este supuesto serán razonables, aunque no se cumpla.

El cumplimiento del supuesto de autocorrelación conjunta nula de los errores resulta clave, ya que en caso de que no se dé esto implicará que la dinámica de dependencia temporal y su estructura no está siendo captada correctamente, lo cual se ve reflejado en los residuos. Para el caso de los modelos planteados y considerándose el test conjunto para los primeros 99 rezagos (cantidad elegida con el criterio de Hyndman & Athanasopoulos (2021)), para todos los modelos no se rechaza la hipótesis de autocorrelación conjunta nula.

Tabla 5.1: Tabla de errores de predicción de precios a 12 pasos. Papa.

Método	RMSE	MAE	MAPE
Random Walk	0.6016	0.4375	2.007
Random Walk + Reg.	0.6016	0.4375	2.007
SUTSE, Random walk	0.6023	0.4396	2.0166
Promedio simple	0.7983	0.5965	2.7955
SES	3.8182	3.1781	17.0481
ARIMA(2,1,2)	0.6303	0.5128	2.3449
ARIMA + Fourier	1.1107	1.027	4.653
TBATS	0.6016	0.4375	2.0071

Dado que el objetivo es la predicción de valores futuros, este será el criterio principal de selección de modelos. Se opta por acotar el horizonte predictivo a 12 semanas para este rubro, debido a la falta de comportamiento estacional en los precios que permita tal horizonte. Las medidas del error de la predicción a 12 semanas correspondientes a cada uno de los modelos aplicados se presentan en la [Tabla 5.1](#). Estas predicciones son realizadas empleando la esperanza condicionada a los datos de la muestra de entrenamiento, con los que son estimados los parámetros de los distintos modelos. Los errores de predicción se calculan empleando estas predicciones y comparándolas con los verdaderos valores, correspondientes a la muestra de prueba. Se puede apreciar que el modelo de caminata al azar con regresores y el TBATS son los que tienen el menor error de predicción en todas las métricas, habiendo diferencias negligibles entre ellos. La predicción puntual de las caminatas al azar es la predicción *naive*, donde se predice con el último valor de la muestra de entrenamiento para todo el horizonte predictivo.

Tabla 5.2: Tabla de errores de predicción a 12 pasos, validación cruzada.

Método	RMSE	MAE	MAPE
Promedio simple	0.8247	0.6076	2.8498
Random Walk	0.368	0.2292	1.0633
Random Walk + Reg.	0.3681	0.2293	1.0637
SUTSE, Random walk	0.3688	0.2759	1.2726
SES	3.9658	3.2614	17.7376
ARIMA(2,1,2)	0.3521	0.3117	1.4315
ARIMA + Fourier	0.404	0.3573	1.6415
TBATS	0.3689	0.2382	1.1033

Para simular un proceso de predicción iterativa, en los primeros 12 valores de la muestra de prueba se realiza validación cruzada con las predicciones a un paso, agregándose a cada paso el dato observado que antes había sido predicho y calculándose las métricas de error en el conjunto de iteraciones. En la [Tabla 5.2](#) se presentan las métricas de error para los modelos anteriormente planteados, pudiéndose observar que, si bien el modelo ARIMA estimado cuenta con un menor RMSE, para las otras métricas planteadas la predicción de la caminata al azar es la de menor error. Tanto el modelo TBATS como los modelos que incluyen variables auxiliares tienen un desempeño similar a esta última metodología.

La conclusión principal de estos resultados es que los intentos de modelización estocástica más elaborados que se aplicaron no dieron resultados substancialmente mejores que simplemente emplear el último precio con el que se cuenta como predicción, es decir, la predicción *naive* que surge de una caminata al azar. Como se adelantó anteriormente, el rubro cuenta con características que pueden estar dificultando el modelado mediante las metodologías

empleadas, como por ejemplo el grado elevado de concentración de la producción del rubro. Esto hace que no se esté ganando en desempeño predictivo respecto a interpretar el proceso de formación de precios de la papa como una caminata al azar. Además de esto, la formación de precios fue afectada por la pandemia. Teniendo en cuenta esto, se presentan las predicciones de la caminata al azar.

c) Modelo de mejor desempeño predictivo

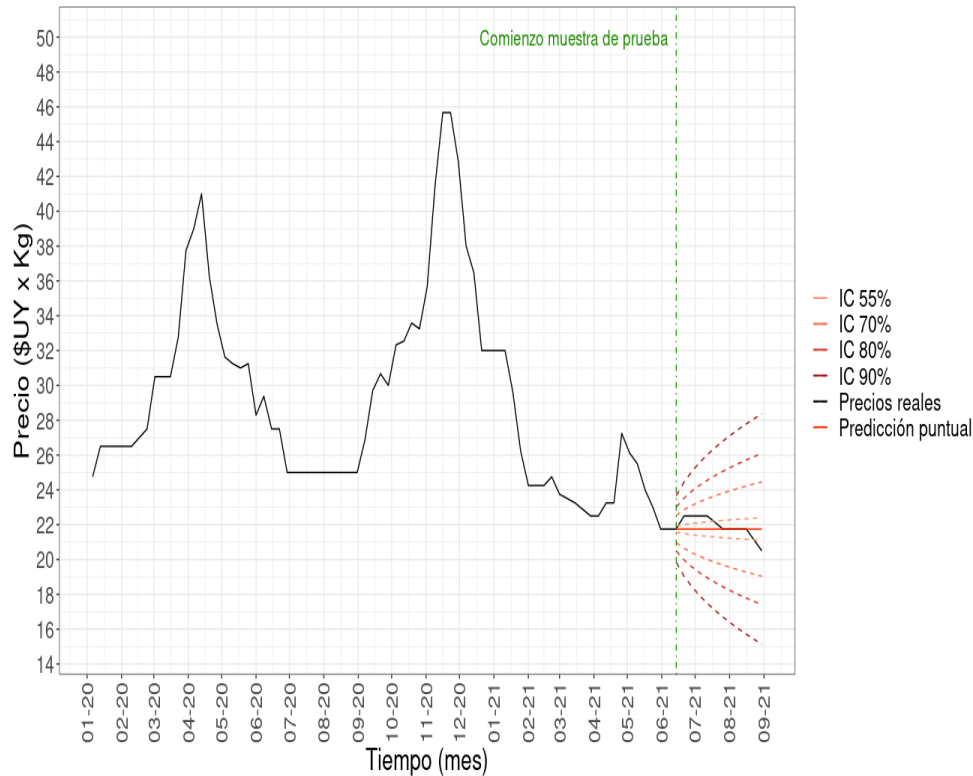


Figura 5.2: Predicción a 12 pasos. Papa. Random Walk.

El modelo estimado es una caminata al azar, donde las varianzas de los procesos observados y de estado son 0,000011 y 2,22. La predicción del precio promedio es la misma, es decir, 21,75. Se da un aumento de la amplitud de los intervalos de confianza a medida que se vuelve más lejano el horizonte predictivo, lo cual refleja la incertidumbre creciente asociada.

4.5 Resultados para el rubro manzana

a) Descomposición de la serie de precios

En contraste claro con la papa, en los precios de la manzana ya a primera vista se aprecia un comportamiento sistemático, en la forma de una estacionalidad aparentemente anual.

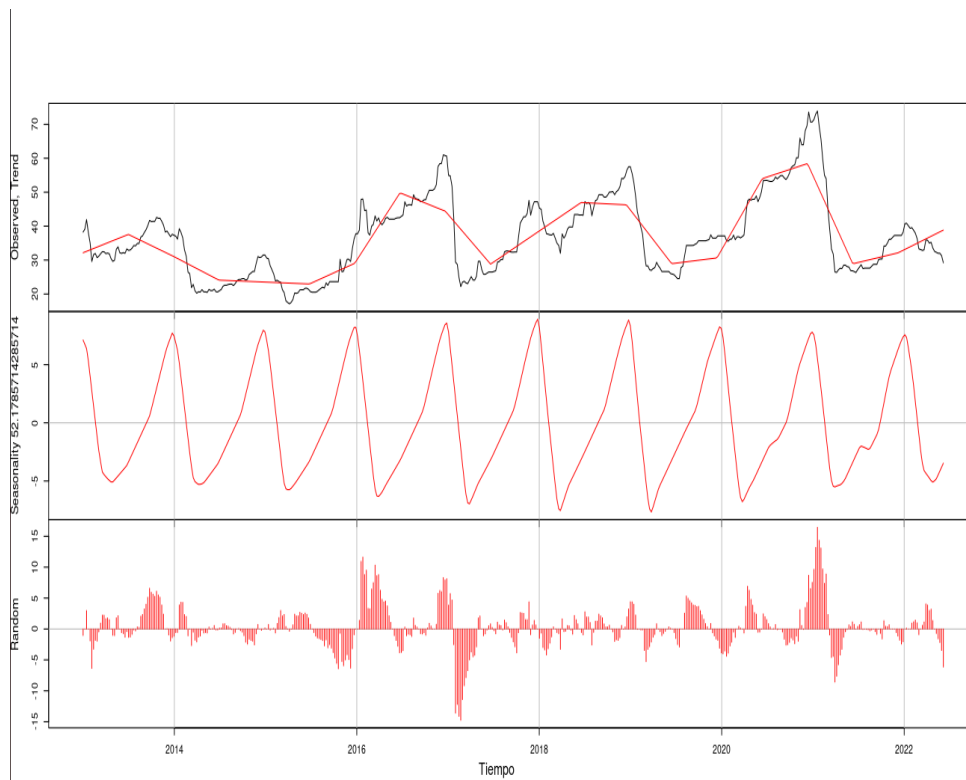


Figura 5.3: Descomposición STR robusta de la serie de precios de manzana.

La visualización de la descomposición STR robusta se presenta en la [Figura 5.3](#). Destaca el elemento estacional, que refleja que en los años el precio comienza siendo elevado, descendiendo hasta la mitad del año, momento a partir del cual vuelve a aumentar. Este componente presenta poca variación en el tiempo.

Por otro lado, cabe destacar que a partir de 2016 es posible identificar un ciclo corto (parte superior de la figura) donde cada dos años el precio vuelve a su nivel anterior. Esto es descrito por los especialistas del Observatorio Granjero como “añerismo”, quienes tienen como hipótesis que es el resultado de la observación de los precios del rubro del año pasado por parte de los productores. Cuando observan que los precios fueron altos, aumentan la siembra, lo cual tiene como resultado una baja de los precios. Al otro año, al observar esta baja disminuyen la siembra.

Se pueden distinguir dos periodos donde el componente irregular toma más fuerza, que son en torno a los años 2016 y 2017 y luego a fines del año 2020. Si bien se podría considerar que la pandemia se manifiesta en el componente irregular como una perturbación al comportamiento sistemático de los precios, el efecto no resulta tan evidente como en el caso de la papa.

b) Ajuste de modelos, selección y diagnóstico

Se ajustan a los precios de manzana las mismas clases de modelos anteriormente descritos cuando se trataron los precios de papa, cambiándose la especificación en algunos casos para incluir, por ejemplo, el componente cíclico y el estacional en los modelos de espacio-estado, y teniéndose en cuenta las variables auxiliares que atañen al rubro.

Al ponerse a prueba los supuestos hechos al momento de la especificación, se obtiene como resultado que solo los modelos de caminata al azar con componente tendencia/ciclo, ARIMA con componente estacional de Fourier determinístico y el TBATS cumplen la hipótesis de autocorrelación conjunta nula de los residuos, lo cual es evidencia a favor de que estos modelos están captando la dinámica de dependencia temporal de la serie. Ninguno de ellos cumple el supuesto de normalidad de sus residuos. En este caso se podría pensar que el proceso generador de datos de los residuos en general no sigue una distribución normal, aunque también puede ser el resultado de residuos atípicos. Ante esta cuestión, el problema puede ser enfrentado mediante la aplicación de técnicas de *bootstrap* a la hora de calcular los intervalos de confianza de las predicciones.

Tabla 5.3: Tabla de errores de predicción a 12 pasos. Manzana.

Método	RMSE	MAE	MAPE
Mean	9.4152	9.3905	34.0622
Random Walk	1.0564	0.877	3.139
SUTSE, Random walk	1.0498	0.918	3.2823
Random walk + trend	3.7644	3.2781	11.7263
ARIMA(3,1,0) + Fourier	1.4687	1.1884	4.2493
TBATS	0.9531	0.73	2.6105

Tabla 5.4: Tabla de errores de predicción a 12 pasos. Validación cruzada. Manzana.

Método	RMSE	MAE	MAPE
Mean	9.3007	9.2718	33.6368
Random walk-dlm	0.5661	0.4324	1.5578
Random walk + trend	0.6425	0.4428	1.5973
SUTSE, Random	0.5654	0.4326	1.5584

Método	RMSE	MAE	MAPE
walk			
ARIMA(3,1,0) + Fourier	0.6248	0.4554	1.6513
TBATS	0.5624	0.4449	1.612

Se calculan las predicciones a 12 pasos condicionando a la muestra de entrenamiento y la predicción de validación cruzada para evaluar el desempeño predictivo de los modelos estimados. Los valores de las métricas de error de predicción a 12 pasos para los distintos modelos se presentan en la [Tabla 5.3](#). En este caso el modelo TBATS es el de mejor desempeño predictivo en todas las métricas. Por otro lado, para la validación cruzada el resultado no es tan concluyente, a muy corto plazo no hay ningún modelo que supere ampliamente a la predicción *naive*.

Tabla 5.5: Tabla de errores de predicción a 52 pasos. Manzana.

Método	RMSE	MAE	MAPE
Seasonal Naive	23.212	20.744	63.254
ARIMA(3,1,0) + Fourier	5.0969	3.8163	11.2617
TBATS	5.1411	3.8614	11.15

Dada las características de comportamiento del rubro se valora apropiada la predicción a 52 pasos, mediante el uso de los modelos con componentes estacionales. En la [Tabla 5.5](#) se presentan las métricas de error predictivo de los dos modelos estacionales y adicionalmente el de la predicción *naive* estacional a modo de *benchmark*. El modelo ARIMA con términos de Fourier es el de mejor desempeño en términos de RMSE y MAE, pero no en MAPE; aunque las diferencias con el modelo TBATS no son substanciales. Ambos tienen un desempeño considerablemente mejor a la predicción *naive* estacional.

c) Modelo de mejor desempeño predictivo

Se presentan los resultados del modelo TBATS, debido a su mejor rendimiento a 12 pasos y que para la predicción a 52 pasos la diferencia con el modelo ARIMA es mínima.

El modelo estimado es un TBATS donde el parámetro de la transformación Box-Cox es $\omega = 0,015$, los parámetros del modelo ARMA(p,q) de los errores son ambos 0 (es un ruido blanco), su desvío estándar estimado es $\sigma = 0,0504$, el parámetro de amortiguación del componente de tendencia es $\phi = 0.8$ y son empleados 3 armónicos para representar el componente estacional anual $m = 52,18$. Los parámetros de suavizado de los errores ARMA para el componente de nivel local, la tendencia de corto plazo y de los componentes estacionales son $\alpha = 1,0093$, $\beta = 0,113$, $\gamma_1 = 0,00045$ y $\gamma_2 = -0,00042$.

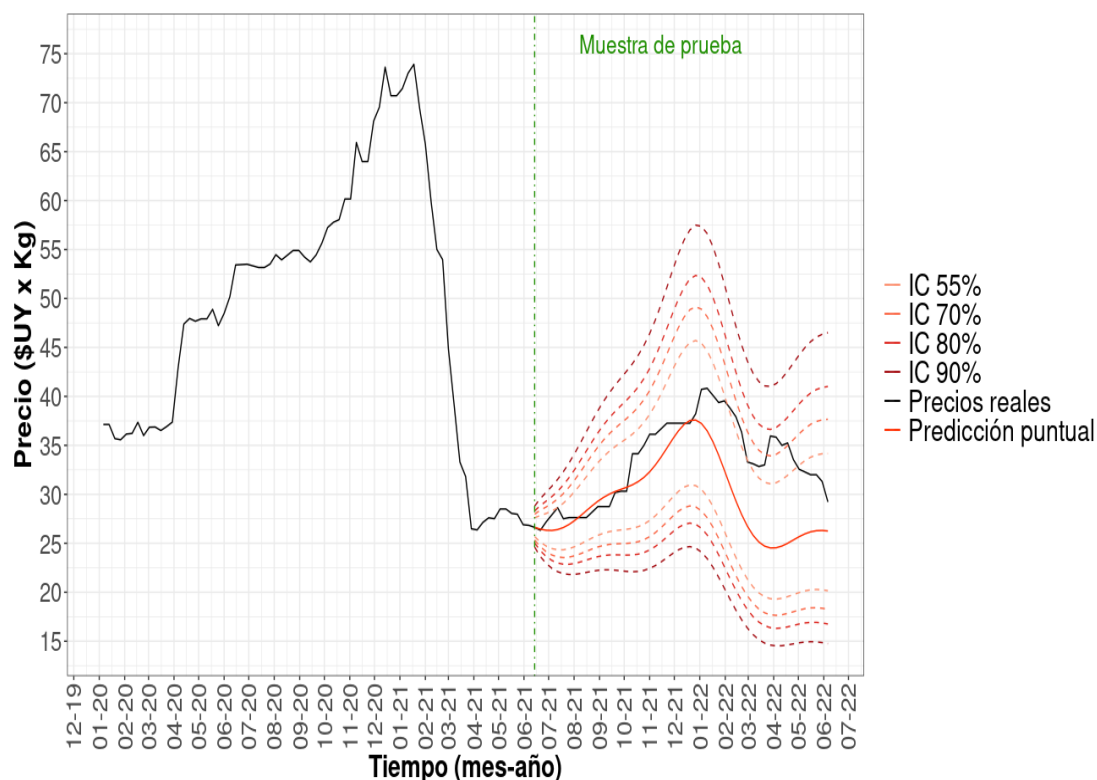


Figura 5.4: Predicción a 52 pasos. Manzana. TBATS

En la Figura 5.4 se presentan las predicciones a 52 semanas para los precios de la manzana mediante el modelo TBATS. Estas van desde la semana que comienza el 14/05/2021 a la que comienza el 06/06/2022 y adicionalmente se presentan los verdaderos valores de los precios observados. Se destaca que el comportamiento de los precios es en general predicho hasta comienzos de 2022, pero ya para mediados de este año el modelo comienza a flaquear, prediciendo un crecimiento de los precios mientras que los reales descienden. No obstante, se debe tener en cuenta que esto ocurre al final del horizonte predictivo, donde la incertidumbre es mayor para el modelo, lo que se ve reflejado en la amplitud de los intervalos de confianza.

4.6 Resultados para el rubro tomate

a) Descomposición de la serie de precios

Dadas las características productivas del rubro, los precios del tomate presentan una mayor volatilidad que los de los rubros anteriormente tratados. Por este motivo, resulta de gran utilidad la descomposición de la serie en sus elementos inobservables para comprender en un mayor grado su comportamiento.

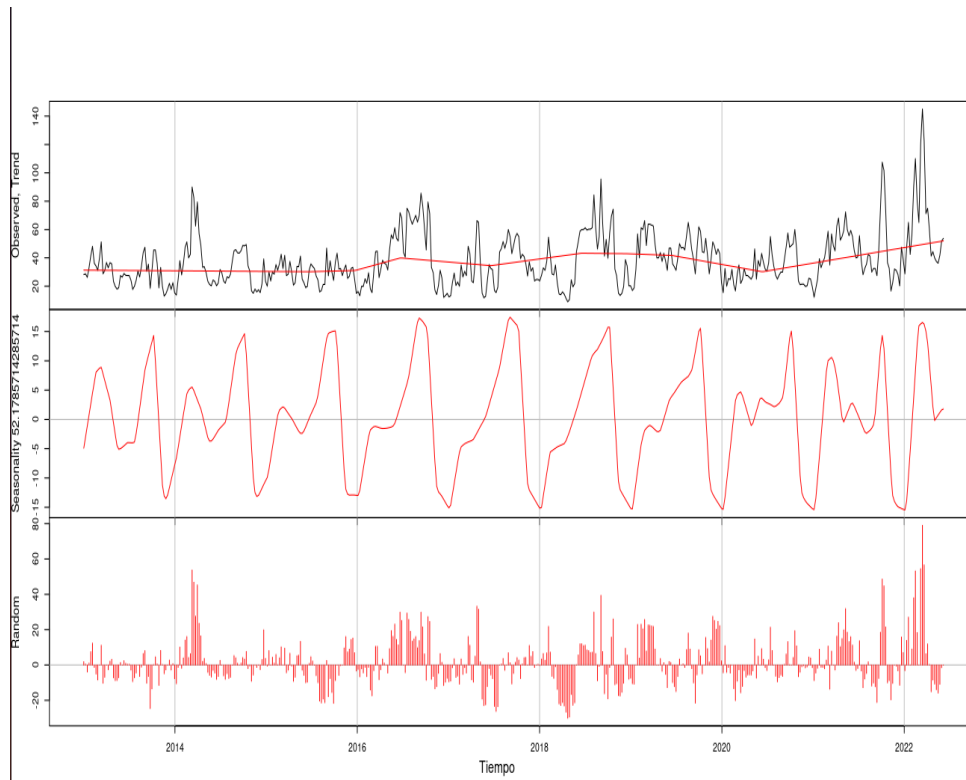


Figura 5.5: Descomposición STR robusta de la serie de precios de tomate.

La visualización de la descomposición se presenta en la Figura 5.5. En primera instancia, destaca que aparentemente la serie presenta una estacionalidad anual subyacente, donde el pico de precios se da a fines de cada año. Esto indica que será necesario incorporar un elemento estacional a la especificación del modelo. Luego, no parece haber una tendencia global al alza de precios, aunque el máximo histórico de precios está en el último año de la muestra, 2022.

b) Ajuste de modelos, selección y diagnóstico

Al igual que en el caso de los rubros anteriores, se aplica un conjunto de modelos, en este caso compuesto por un modelo de *Random Walk*, un *SUTSE* bivariado que considera los precios del tomate y sus ingresos, un *ARIMA(1,1,2)*, un *ARIMA(0,1,4)* con estacionalidad representada por 3 términos de Fourier y un modelo *TBATS*, además del promedio simple como referencia.

Entre los modelos, el que cumple el supuesto de autocorrelación conjunta es el modelo *ARIMA + Fourier*. Para este modelo la normalidad de los residuos no se cumple.

Tabla 5.6: Tabla de errores de predicción a 12 pasos. Tomate.

Método	RMSE	MAE	MAPE
Promedio simple	7.4674	5.9319	15.0445

Método	RMSE	MAE	MAPE
Random walk-stsp	19.2008	17.9318	51.2834
SUTSE random walk-dlm	25.0919	24.0296	67.9235
ARIMA(1,1,2)	19.555	18.1329	52.1317
ARIMA(0,1,4) + 3 términos de Fourier	27.1256	25.2451	72.1469
TBATS	27.464	25.6146	72.9585

En la [Tabla 5.6](#) se presentan las métricas de error para las predicciones a 12 pasos de los precios del tomate. Como se puede apreciar, en este caso la mejor predicción resulta del promedio simple. Esta deficiencia de los modelos aplicados se puede deber a la volatilidad de la serie, ante la cual las especificaciones planteadas no pueden captar toda la dinámica y también a la muestra de prueba tomada.

Lo mismo ocurre con la predicción a 1 paso de 12 semanas con validación cruzada ([Tabla 5.7](#)).

Tabla 5.7: Tabla de errores de predicción a 12 pasos. Validación cruzada. Tomate.

Método	RMSE	MAE	MAPE
Promedio simple	7.4695	5.9365	15.0782
Random walk-stsp	8.8114	7.3956	19.3881
SUTSE random walk-dlm	13.9736	12.1439	34.0733
ARIMA(1,1,2)	12.8287	10.1086	31.1391
ARIMA(0,1,4) + 3 términos de Fourier	9.4619	7.8833	21.2862
TBATS	8.9052	7.1668	18.837

Tabla 5.8: Tabla de errores de predicción a 52 pasos. Tomate.

Método	RMSE	MAE	MAPE
Promedio	32.9171	20.9557	32.2062
ARIMA(0,1,4) + Fourier	28.3181	22.8459	53.3701
TBATS	30.209	22.6721	46.3226

Considerándose una predicción a largo plazo (52 semanas), la situación cambia. El promedio simple resulta superior en las métricas del MAE y MAPE, mientras que el modelo ARIMA con estacionalidad de Fourier funciona mejor en RMSE. Esto se puede justificar en cierta medida por la incertidumbre asociada a la serie. Por ejemplo, considerando el modelo ARIMA +

estacionalidad de Fourier (Figura 5.6) se observa la gran amplitud de los intervalos de confianza, una medida de la incertidumbre asociada a la predicción puntual.

c) Modelo de mejor desempeño predictivo

Si bien a corto y mediano plazo la predicción considerando el promedio simple es la de mejor desempeño, a largo plazo el modelo ARIMA con estacionalidad determinística con términos de Fourier resulta competitivo. Cabe destacar que este tipo de modelo capta el comportamiento estacional del rubro y permite el cálculo de intervalos de predicción. Por este motivo, se presentan sus resultados además de la predicción mediante el promedio simple.

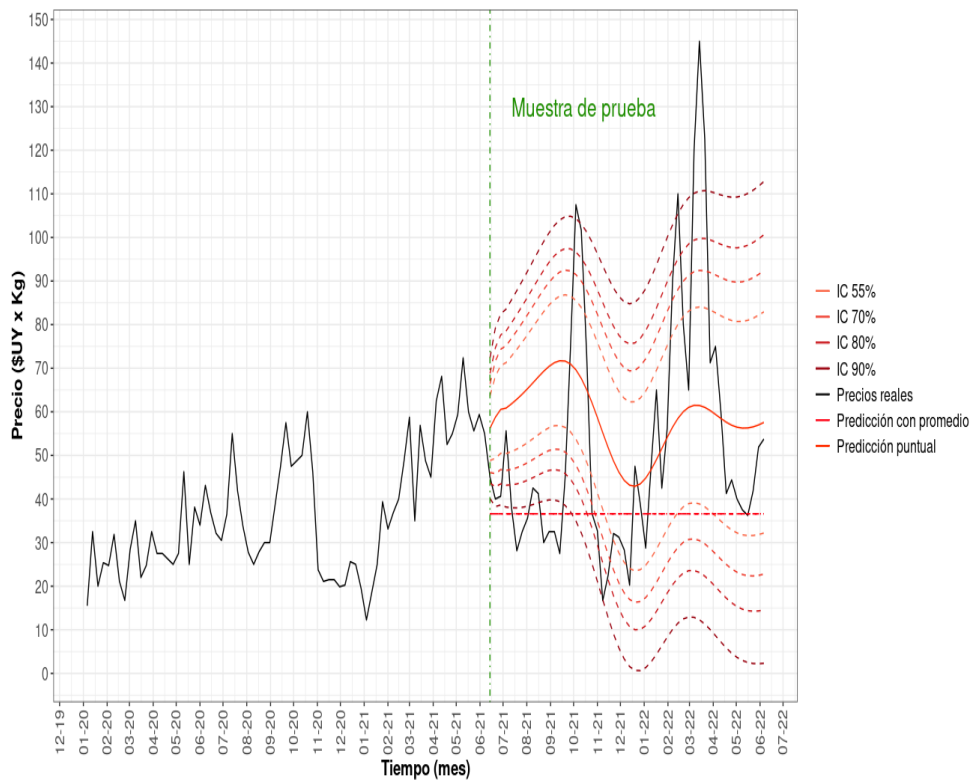


Figura 5.6: Predicción a 52 pasos, tomate.

El modelo ARIMA en cuestión es un $ARIMA(0,1,4)$, donde los parámetros de medias móviles son: $\theta_1 = -0.2153$, $\theta_2 = -0.3019$, $\theta_3 = -0.1145$ y $\theta_4 = 0.0827$. La estacionalidad se representa con 3 términos de Fourier y donde el desvío de los términos de error es $\sigma = 9.8$.

La predicción empleando el promedio es 36.6 para el periodo de prueba.

Como se describió anteriormente en la sección [Capítulo 4](#), el tomate es un rubro cuyo volumen de producción se ve afectado rápidamente por un conjunto de factores rápidamente cambiantes (como lo son las condiciones climáticas que afectan el cuajado), lo cual dificulta su predicción.

Incluso contándose con las variables que cuantifican estos factores, resulta necesaria su predicción para ser incorporada al modelo, lo cual es una tarea que presenta gran dificultad y excede los límites de este trabajo.

4.7 Resultados para el rubro cebolla

a) Descomposición de la serie de precios

La descomposición STR robusta se presenta gráficamente en la [Figura 5.7](#). En la parte superior de la figura se puede observar el componente de tendencia-ciclo, pudiéndose apreciar cierto ciclo corto bianual, el cual puede ser una manifestación de un añerismo similar al encontrado para el rubro manzana. Por otro lado, el componente estacionalidad presenta un patrón anual, donde a principios del año comienza bajo para ir aumentando hasta fines del año. El rango de variación de este patrón se va volviendo menor año a año. El componente irregular tiene una presencia fuerte en torno a los años 2015/2016 y el periodo de la pandemia.

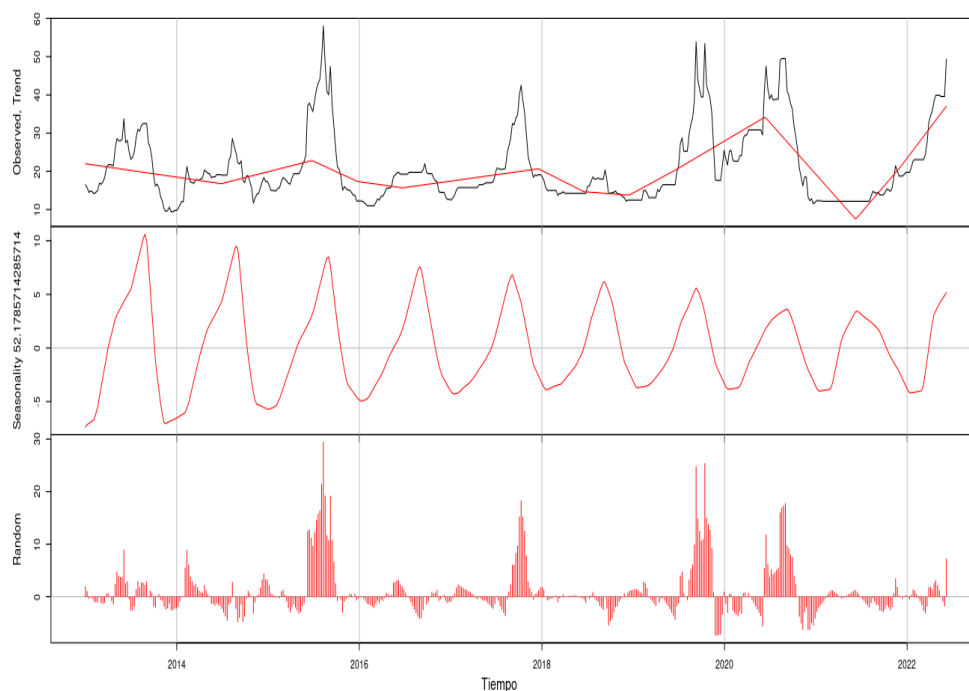


Figura 5.7: Descomposición STR robusta de la serie de precios de cebolla.

a) Ajuste de modelos, selección y diagnóstico

Se realiza la estimación del conjunto de modelos anteriormente empleados, pero para los precios de cebolla. En particular estos son: *Random Walk*, un *SUTSE* bivariado que considera los precios

de la cebolla y sus ingresos, un ARIMA(3,0,2), un ARIMA(1,1,1) con estacionalidad representada por 2 términos de Fourier y un modelo TBATS; además de las predicción mediante el promedio simple como referencia.

En lo que refiere al cumplimiento de los supuestos de los modelos estocásticos, la hipótesis de autocorrelación conjunta nula de los residuos no se rechaza para el modelo ARIMA, el ARIMA con términos de Fourier y el TBATS. Para ninguno de los modelos se cumple el supuesto de normalidad de los errores.

Tabla 5.9: Tabla de errores de predicción a 12 pasos. Cebolla.

Método	RMSE	MAE	MAPE
Promedio simple	7.9485	7.8833	62.2589
Local level-stsp	3.8098	3.5833	22.2451
ARIMA(2,0,1)	4.0923	3.7665	29.0028
SUTSE random walk-dlm	1.2208	0.6959	4.8651
ARIMA + Fourier	5.3774	4.9749	38.4265
TBATS	2.3264	2.0947	16.2121

Tabla 5.10: Tabla de errores de predicción a 12 pasos. Validación cruzada. Cebolla.

Método	RMSE	MAE	MAPE
Promedio simple	7.8531	7.7803	61.4821
Local level-stsp	3.1782	3.066	19.2917
ARIMA(2,0,1)	0.5108	0.4729	3.6193
SUTSE random walk-dlm	0.5328	0.2649	1.8852
ARIMA + Fourier	1.1301	1.0024	8.0221
TBATS	0.5532	0.4269	3.2012

Las métricas de precisión de predicción a 12 pasos y de validación cruzada se presentan en la [Tabla 5.9](#) y la [Tabla 5.10](#). Como se puede apreciar, para la predicción a corto plazo el modelo bivariado de caminata al azar (SUTSE) es el de mejor desempeño, mientras que en la validación cruzada este modelo es el mejor en dos de las tres métricas empleadas (MAE y MAPE), el ARIMA se desempeña mejor en RMSE.

Considerando la predicción a largo plazo (Tabla 5.11), los modelos estacionales no superan al promedio simple de las observaciones. Todo esto indica que no resultan apropiadas para la predicción a largo plazo las especificaciones planteadas.

Tabla 5.11: Tabla de errores de predicción a 52 pasos. Cebolla.

Método	RMSE	MAE	MAPE
Promedio	9.8854	7.5363	35.0371
ARIMA(1,1,1) + Fourier	16.9684	13.877	56.449
TBATS	15.3789	11.7225	44.3316

b) Modelo de mejor desempeño predictivo

Se presentan las predicciones a corto plazo del modelo de mejor desempeño (SUTSE) para el rubro cebolla. Como se puede apreciar, el comportamiento atípico de los precios a lo largo de 2021, donde se mantuvieron fijos, puede ser una causa del pobre desempeño de los modelos.

El modelo estimado es una caminata al azar bivariada donde:

$$V = \begin{pmatrix} 0.00043 & -0.0027 \\ -0.0027 & 9.2 \times 10^{10} \end{pmatrix}$$

$$W = \begin{pmatrix} 6.7 & -0.0001 \\ -0.0001 & 0.031 \end{pmatrix}$$

La predicción puntual del modelo SUTSE es 19.75 para todo el periodo, mientras que los precios reales por las primeras 8 semanas de la muestra de prueba es de 12.17.

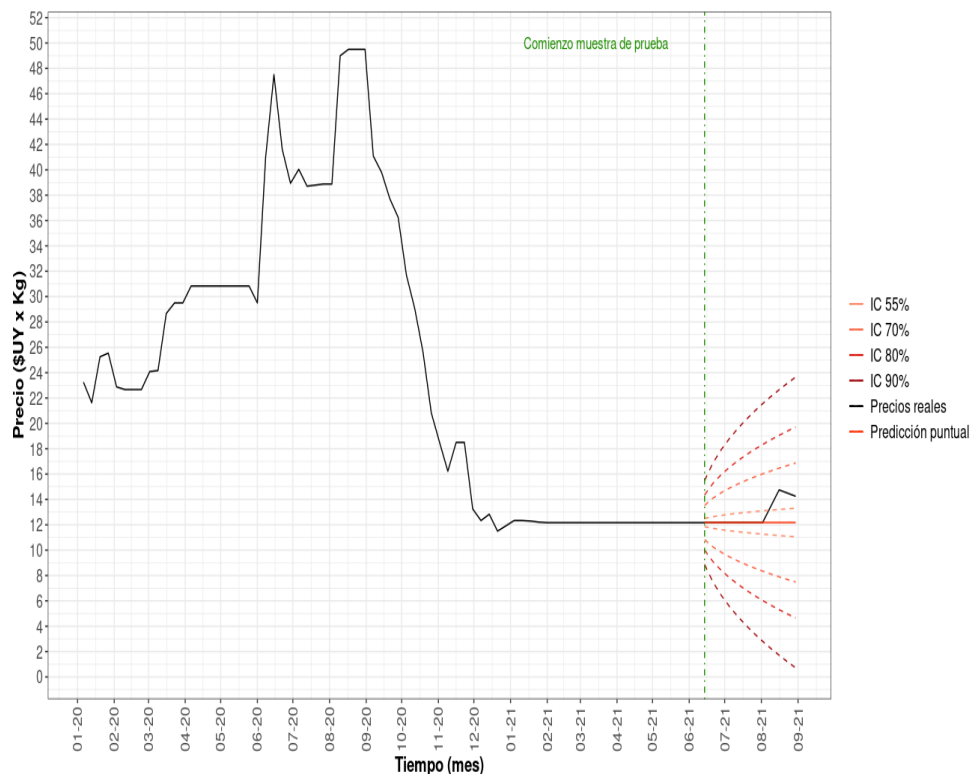


Figura 5.8: Predicción a 12 pasos. Cebolla. SUTSE Random Walk.

Quiebres estructurales de este tipo son difíciles de predecir para cualquier tipo de modelo y posiblemente se requiera de una intervención en el modelado una vez que se toma constancia de que ocurre.

4.8 Resultados para el rubro naranja

a) Descomposición de la serie de precios

La descomposición STR robusta (Figura 5.9) permite apreciar cierto ciclo corto (de 2 años) y una estacionalidad con un periodo anual, donde el rango de valores de este elemento va en aumento año a año. Se destaca el comportamiento del elemento irregular en torno al año 2016 y 2021.

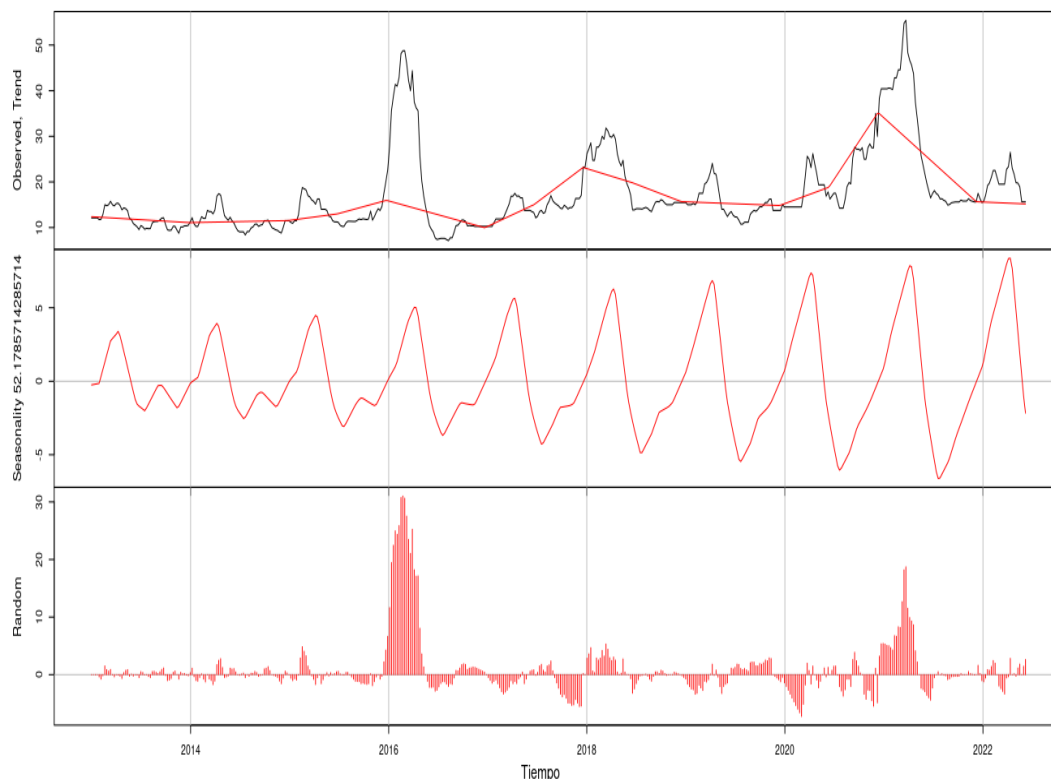


Figura 5.9: Descomposición STR robusta de la serie de precios de naranja.

b) Ajuste de modelos, selección y diagnóstico

Se ajustan un modelo *Random Walk*, un $ARIMA(4, 1, 0)$, un SUTSE bivariado donde se consideran precios e ingresos del rubro naranja, un $ARIMA(1,1,1)$ con dos componentes de Fourier estacionales y un TBATS; además de contarse con la predicción del método de *benchmark* del promedio de precios.

El supuesto de autocorrelación nula de los residuos se cumple para los modelos $ARIMA$, $ARIMA$ con términos de Fourier y TBATS. Para ninguno de los modelos se da el cumplimiento del supuesto de normalidad.

Tabla 5.12: Tabla de errores de predicción a 12 pasos. Naranja.

Método	RMSE	MAE	MAPE
Promedio simple	1.2293	1.0202	6.3912
Local level-stsp	3.3256	3.1611	19.5211
Arima(4, 1, 0)	1.5902	1.4781	8.93
SUTSE random	3.3259	3.1614	19.5229

Método	RMSE	MAE	MAPE
walk-dlm			
ARIMA + fourier	1.3752	1.1599	7.0928
TBATS	2.1207	1.5594	9.9116

Tabla 5.13: Tabla de errores de predicción a 12 pasos. Validación cruzada. Naranja.

Método	RMSE	MAE	MAPE
Promedio simple	1.2259	1.0184	6.3792
Local level-stsp	0.8472	0.6753	4.0019
Arima	3.9863	3.2617	26.5
SUTSE random	0.8476	0.6758	4.005
walk-dlm			
ARIMA + Fourier	0.8022	0.6452	3.858
TBATS	0.8585	0.7679	4.6583

El rendimiento predictivo de los modelos a 12 pasos y en la validación cruzada, expresado en las métricas de error, se presenta en la [Tabla 5.12](#) y la [Tabla 5.13](#). Para el primer tipo de predicción, el promedio simple es el que presenta el menor error. Por otro lado, en la predicción a un paso de validación cruzada el modelo ARIMA con términos de Fourier es el de mejor desempeño.

Tabla 5.14: Tabla de errores de predicción a 52 pasos. Naranja.

Método	RMSE	MAE	MAPE
Promedio simple	2.7703	2.1533	11.2931
ARIMA(1,1,1) + Fourier	4.3938	3.6894	20.421
TBATS	6.8568	5.8894	32.3605

Evaluando las predicciones de largo plazo, se puede observar como el error predictivo de los modelos estacionales planteados resulta mayor que realizar la predicción mediante el promedio simple de las observaciones de la muestra de entrenamiento. El modelo más cercano en desempeño a esta predicción es el modelo ARIMA con términos de Fourier. Esto puede ser una señal de que el tipo de modelo y las especificaciones planteadas no están siendo capaces de captar la dinámica de largo plazo de la serie.

c) Modelo de mejor desempeño predictivo

Se presenta la predicción del promedio, que es la de mejor desempeño, junto a la del modelo ARIMA con estacionalidad representada por términos de Fourier, que es el que le seguía en desempeño a largo plazo. Este último es un modelo $ARIMA(1,1,1)$, donde los parámetros de la parte autorregresiva son $\phi = 0.7373$, de la parte de medias móviles $\theta = -0.591$ y el desvío estándar es $\sigma = 2.347$. La estacionalidad se considera en este contexto como determinística y se modela mediante dos componentes de Fourier.

La predicción mediante el promedio es de 17.26 para todo el periodo de evaluación. En la figura se puede ver como el modelo ARIMA + Fourier capta el comportamiento estacional, a pesar de su peor desempeño en las métricas de error.

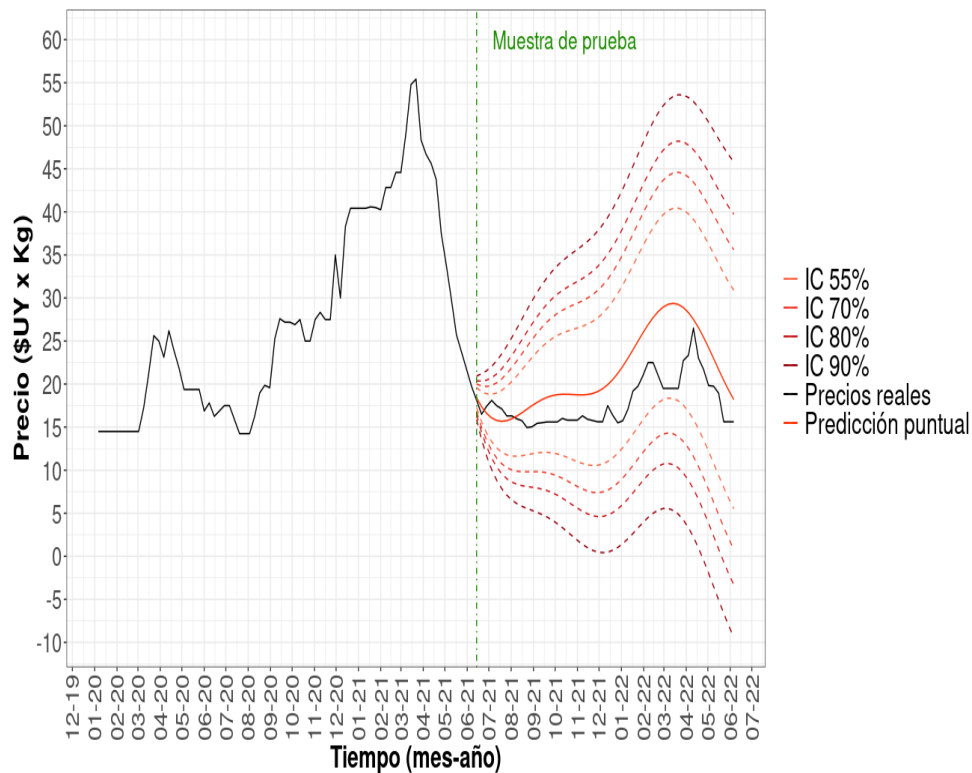


Figura 5.10: Predicción a 52 pasos. Naranja. ARIMA + Fourier.

4.9 Conclusiones y pasos futuros

Al modelar los precios del conjunto de frutas y hortalizas consideradas mediante metodologías de espacio-estado, en distintos horizontes, los resultados fueron variados. No se puede concluir que hay un tipo de modelo particular que tenga desempeño superior para todos los horizontes predictivos y todos los rubros, sino que para cada caso hay un método de predicción particular

que se destaca por sobre los otros. Esto en parte responde a la diversidad de las dinámicas de formación de precios y su susceptibilidad a poder ser captadas por modelos del tipo planteado.

Tabla 6.1: Modelo o método de mejor desempeño para cada rubro y horizonte predictivo.

	Corto plazo	Mediano Plazo	Largo plazo
Papa	Random Walk	Random Walk/TBATS	-
Manzana	Random Walk	TBATS	TBATS/ARIMA+Fourier
Tomate	Promedio	Promedio	Promedio/ARIMA+Fourier
Cebolla	SUTSE Random Walk/ARIMA	SUTSE Random Walk	Promedio
Naranja	ARIMA+Fourier	Promedio	Promedio

Para la manzana es donde los modelos de espacio estado estacionales planteados tienen, en general, el mejor desempeño a través de los distintos horizontes.

Los precios de rubros como la papa y el tomate resultan difíciles de predecir a cualquier horizonte predictivo. El primero debido a la concentración de la oferta, que hace que los precios dependan de las decisiones de pocos actores. En el caso del segundo se debe al elevado grado de variabilidad que presentan sus precios.

La cebolla y la naranja comparten que cuentan con cierto componente estacional pero afectado por periodos con picos elevados de precios. Esto último es un factor que dificulta la predicción de sus precios.

La inclusión de variables auxiliares en las especificaciones de los modelos fue una problemática. Incluir en los modelos el volumen de ingreso de mercadería de los distintos rubros resulta relativamente directo, y para el caso de la cebolla aporta al poder predictivo. Para el resto de los rubros, no implicó una ganancia en precisión. Para el resto de las variables que se discutieron al comienzo del proyecto, resultó un obstáculo definir los periodos en los que afectan a los precios dichas variables y en muchos casos los datos que se pudieron encontrar estaban incompletos o no tenían una frecuencia suficiente para aportar al modelado. Cuando sí se pudo incluir alguna de estas variables, no implicaron una ganancia en poder predictivo.

Como posibles líneas de investigación futura están la aplicación de otras especificaciones de modelos de espacio-estado y a otros rubros. Adicionalmente, también queda abierta la posibilidad de emplearse otro tipo de modelos, como Redes Neuronales y otro tipo de modelos de *Deep Learning*. Queda sobre la mesa la necesidad de sistematización del registro de variables agrícolas, para poder ser empleadas en el modelado de series de tiempo vinculadas (en el caso presente, los precios de frutas y hortalizas).

5. Objetivo ii: Desarrollo de plataforma unificada de información de mercado de Frutas y hortalizas

Este objetivo acompañó lo que fue la mudanza del Mercado Modelo a la UAM. Previendo una mejora en los sistemas de información y de declaración de información por parte de los usuarios se identificó como imprescindible la mejora del sistema de almacenamiento y análisis de información de ingreso de mercadería que tenía el Mercado Modelo. Un año previo a la mudanza se comenzó a diseñar esta nueva plataforma en el marco de este proyecto.

Esta plataforma pretendió ser una herramienta que permitiera la carga de guías online por parte del operador de forma ágil.

Esta plataforma permitió la carga de la información histórica de precios e ingresos.

5.1 Descripción general de la plataforma e ingresos a la UAM y precios mayoristas

La aplicación posee una arquitectura simple y la lógica necesaria para el relevamiento de precios e ingresos de guías. A su vez posee varias funcionalidades para gestionar la configuración de todas las entidades requeridas para su funcionamiento tales como: especies, variedades, países, etc. Por último, posee un componente de métricas que se encarga de hacer todos los cálculos y reportes del sistema.

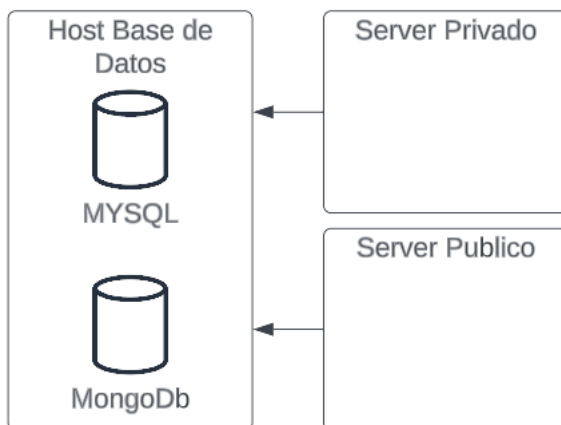
a) Tecnologías:

1. LDAP: (Lightweight Directory Access Protocol) es un protocolo de red estándar que se utiliza para acceder y mantener información de directorios en un sistema operativo o aplicación de servidor. El uso más común de LDAP es en la autenticación y autorización de usuarios en una red.
2. PHP: Es un lenguaje de programación de código abierto que se utiliza principalmente para desarrollar aplicaciones web y de servidor. PHP se ejecuta en el servidor y se utiliza para generar contenido dinámico y trabajar con bases de datos.
3. Laravel: Laravel es un popular marco de trabajo de código abierto de PHP que se utiliza para desarrollar aplicaciones web. Laravel facilita el desarrollo de aplicaciones web de alta calidad y eficiencia mediante el uso de patrones de diseño y prácticas de programación modernas.
4. Laravel-Backpack: Es un paquete de complementos para el marco de trabajo Laravel que se utiliza para acelerar el desarrollo de aplicaciones web. Laravel-Backpack incluye características preconstruidas y configuraciones preestablecidas que facilitan el desarrollo de aplicaciones web.

5. MongoDB: Es una base de datos de documentos de código abierto que utiliza un modelo de datos NoSQL. MongoDB se utiliza para almacenar datos no estructurados y semiestructurados y es altamente escalable y flexible.

6. MySQL: Es un sistema de gestión de bases de datos relacionales (RDBMS) de código abierto que se utiliza para almacenar y organizar grandes cantidades de datos. MySQL es ampliamente utilizado en aplicaciones web y de servidor debido a su capacidad de manejar múltiples conexiones simultáneas y su escalabilidad.

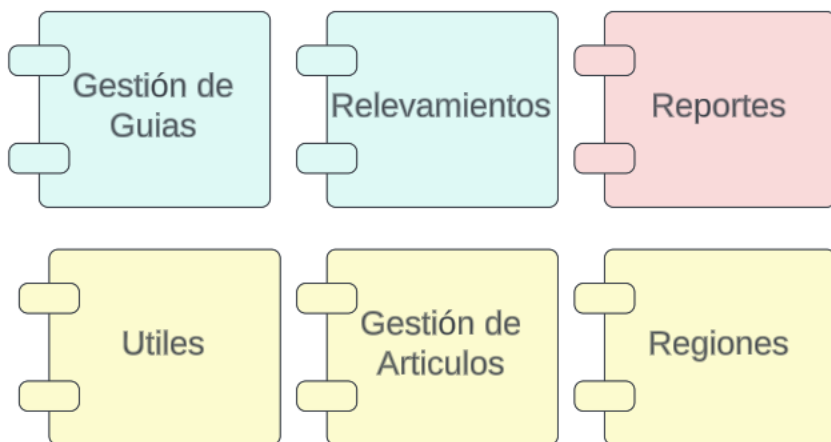
b) Arquitectura:



El sistema contiene 3 elementos principales en cuanto a los servicios que están operando. Por un lado la UAM provee una instancia privada que contiene las bases de datos y un server privado con acceso a las bases para los usuarios dentro de la UAM. Por otro lado, FlowLabs se encarga de hostear el servidor público el cual accede a las bases de datos mediante un proxy.

Para profundizar en las bases de datos, el sistema contiene dos bases de datos por un lado una base de datos relacional (MYSQL) para todo lo que es manejo de configuraciones y datos transaccionales y por otro lado (MongoDB) Una base de datos no relacional que se utiliza para guardar los ingresos tanto de precios como de volúmenes.

c) Resumen de componentes:



El sistema puede descomponerse en 6 grandes componentes que se separan en 3 categorías:

- Ingresos de datos:
 - Gestión de Guías
 - Relevamientos de precios
- Reportes:
 - Métricas y Reportes
- Gestión de configuraciones:
 - Útiles
 - Gestión de artículos
 - Regiones.

i. **Gestión de guías**

Este componente permite la carga de las guías al sistema, además de permitir visualizar y editar cada una de las guías ya ingresadas (figura 1).

Es el único componente que está abierto al público, aunque con ciertas limitaciones, la versión publica solo permite ver las guías creadas por el usuario que está utilizando la herramienta y deben pasar por la aprobación de personal autorizado para confirmar sus ingresos.

La mayor complejidad de este componente se encuentra en la creación de nuevas guías. Para esto, el sistema ofrece una interfaz visual separada en tres grandes partes: Metadatos, Plantilla para líneas y Líneas.

EL metadato posee información general del ingreso (cabezal): Fecha, Numero de guía, Matricula, Destinatario, Transportista y tipo de vehículo⁸ (figura2).

Luego de ingresada esta información se despliega el siguiente apartado para la carga de la información de los productos (plantilla para líneas).

Por cada producto ingresado se debe indicar: País y departamento de origen. Sector dentro de la UAM al que va. Presentación, número de unidades y el peso por unidad (figura 3). El producto se puede buscar mediante código o por el nombre de su especie o variedad. Las presentaciones pueden seleccionarse por su número o desplegando el cuadro de selección. Luego se ingresa la cantidad de unidades y se puede cambiar el peso por unidad si no se está de acuerdo con aquel que provee la base de datos para ese producto. Debajo de la plantilla se puede ver el peso total del ingreso hasta ahora.

Para optimizar el tiempo de digitación, se utiliza un predictor de la información del cabezal, en base a las matrículas previamente cargadas. Además, permite fijar País y departamento si todos los productos de la guía tienen mismo origen.

El listado de guías cargadas se puede visualizar y filtrar según varios parámetros, lo que nos permite buscar errores o detectar incumplimiento de operadores por no entrega de la guía. Además, se cuenta, con alertas según peso, esto surge como herramienta para detectar errores en lo declarado, se cuenta con rangos estimados de peso por categoría de vehículo si este es superior se muestra una alerta, que es revisada para su corrección.

⁸ Categoría 1: Auto o camioneta; camión rueda simple; categoría 2: camión rueda doble categoría 3: camión doble eje, categoría 4: camión y zorra.

Panel | Gestión de Guías | Relevamientos | Gestión de Artículos | Regiones | Útiles | Extras | Reportes

Admin / Guías / Listar

Guías Reiniciar

+ Añadir guía

Buscar...

▼ Date range ▼ Zonas UAM ▼ Número de guía ▼ Destinatario ▼ Matrícula ▼ Confirmación ▼ Hecha por operador Pasada de Peso ⇄ Remover filtros

Número de Guía	Matrícula	Fecha de ingreso	Destinatario	Peso	Usuario	Confirmación	Acciones
2025699	ATM1373	2 feb. 2022	Da silva 30	3,900	amalaquina	<input type="text"/>	Editar Eliminar
2025699	ATM1373	1 feb. 2022	Da silva 30	3,954	amalaquina	<input type="text"/>	Editar Eliminar
2025698	ATM1373	31 ene. 2022	Da silva	3,912	amalaquina	<input type="text"/>	Editar Eliminar
2025697	ATM1373	31 ene. 2022	Da silva 30	2,910	amalaquina	<input type="text"/>	Editar Eliminar
2025696	atm1373	28 ene. 2022	el 30	3,590	Imarveggio	<input type="text"/>	Editar Eliminar
2025695	ATM1373	27 ene. 2022	Da silva	3,140	amalaquina	<input type="text"/>	Editar Eliminar
2025694	atm1373	26 ene. 2022	el 30	3,110	Imarveggio	<input type="text"/>	Editar Eliminar
2025693	ATM1373	25 ene. 2022	Da silva	3,234	amalaquina	<input type="text"/>	Editar Eliminar

Figura 1

Panel | Gestión de Guías | Relevamientos | Gestión de Artículos | Regiones | Útiles | Extras | Reportes

Guía

02/05/2023
Fecha de Ingreso de la Mercadería

Matrícula

Transportista

Número

Número de Guía

Destinatario

Destinatario

Tipo de Vehículo

Tipo de Vehículo

Crear

Ingresar producto mediante código

Peso total: 0

Realizado por Tech Lab.

Figura 2

Hortalizas de Hoja, Brote e Inflorescencias	Región	Sector	Unidad	Peso	
Acelga (Docena)	UY CAN	MF	80	15	🔍 🗑️
Apio - Hoja (Docena)	UY CAN	MF	10	12	🔍 🗑️
Cebolla Verde (Docena)	UY CAN	MF	40	11	🔍 🗑️
Lechuga	UY CAN	MF	60	5	🔍 🗑️
Lechuga - Crespa (Docena)	UY CAN	MF	100	5	🔍 🗑️
Perejil (Docena)	UY CAN	MF	30	7	🔍 🗑️
Puerro (Docena)	UY CAN	MF	15	8	🔍 🗑️

Hierbas Aromáticas	Región	Sector	Unidad	Peso	
Menta (Docena)	UY CAN	MF	10	2	🔍 🗑️

Peso total: 3.900

Figura 3

ii. Relevamiento de precios

iii. En esta pestaña se cargan los Excel del relevamiento de precios mayoristas de frutas y hortalizas de la UAM.

Fecha de Relevamiento	Acciones
25 ago. 2003	📄 Exportar 🗑️ Eliminar
29 ago. 2003	📄 Exportar 🗑️ Eliminar
1 sep. 2003	📄 Exportar 🗑️ Eliminar
4 sep. 2003	📄 Exportar 🗑️ Eliminar
8 sep. 2003	📄 Exportar 🗑️ Eliminar
11 sep. 2003	📄 Exportar 🗑️ Eliminar
15 sep. 2003	📄 Exportar 🗑️ Eliminar
18 sep. 2003	📄 Exportar 🗑️ Eliminar

La plataforma permite la carga de toda la información del boletín de precios de forma rápida, esto fue diseñado en base al proceso que se llevaba a cabo en el proceso de elaboración del boletín de precios frutihortícola. El proceso de relevamiento culmina con una reunión técnica en que se comparten los datos relevados por cada relevador, y se conforma una planilla de Excel donde se escriben los rangos de precio por producto, calibre y categoría de calidad. La plataforma

permite cargar directamente este Excel, eliminando el tiempo de digitación requerido en la plataforma anterior.

Grupo	Especie	Variedad	País	Unidad	Cultivar	Categoría	Min/Un	Max/Un	Ref	Diff	Coste	Min/Kg	Max/Kg	Min/Un Art	Max/Un
1	Hortalizas de Hoja, Brote e Inflorescencias	Aceituna	UY	DOC	GRANDE	I	35	40	0		15	2,333333333	2,666666667		35
2	Hortalizas de Hoja, Brote e Inflorescencias	Aceituna	UY	DOC	GRANDE	II	30	35	0		15	2	2,333333333		30
3	Hortalizas de Hoja, Brote e Inflorescencias	Ajo	CN	KG	GRANDE	I	30	33	0		1	30	33		30
4	Hortalizas Secas o Pesadas	Ajo	AR	KG	GRANDE	I	36	40	0		1	36	40		36
5	Hortalizas Secas o Pesadas	Ajo	UY	KG	GRANDE	I	40	45	0		1	40	45		40
6	Hortalizas Secas o Pesadas	Ajo	UY	KG	MEDIANO	I	35	40	0		1	35	40		35
7	Hortalizas Secas o Pesadas	Banana	Brasilera	BR	KG	GRANDE	I	8,5	9,1	0		1	8,5	9,1	8,3
8	Exóticos/Importados	Banana	Brasilera	BR	KG	MEDIANO	II	7,9	8,5	0		1	7,9	8,5	7,7
9	Exóticos/Importados	Berenjena	UY	KG	GRANDE	E	26	27	0		1	26	27		28
10	Hortalizas de Fruto	Berenjena	UY	KG	GRANDE	I	25	26	0		1	25	26		26
11	Hortalizas de Fruto	Berenjena	UY	KG	MEDIANO	I	24	25	0		1	24	25		25
12	Hortalizas de Fruto	Boniato	Arapey/Morada	UY	KG	MEDIANO	E	3	3,5	0		1	3	3,5	3
13	Hortalizas Secas o Pesadas	Boniato	Arapey/Morada	UY	KG	MEDIANO	II	2,3	2,8	0		1	2,3	2,8	2,3
14	Hortalizas Secas o Pesadas	Boniato	Arapey/Morada	UY	KG	MEDIANO	E	1,8	2,3	0		1	1,8	2,3	1,8
15	Hortalizas Secas o Pesadas	Cebolla	AR	KG	MEDIANO	E	9,5	10,5	0		1	9,5	10,5		10,5
16	Hortalizas Secas o Pesadas	Cebolla	UY	KG	MEDIANO	E	8,5	9	0		1	8,5	9		9
17	Hortalizas Secas o Pesadas	Cebolla	UY	KG	MEDIANO	I	8	8,5	0		1	8	8,5		8,5
18	Hortalizas Secas o Pesadas	Cebolla	UY	KG	MEDIANO	II	7,5	8	0		1	7,5	8		7,5
19	Hortalizas de Fruto	Chaucha	Bomboneta Chata	UY	KG	MEDIANO	I	37,5	45	0		1	37,5	45	37,5
20	Hortalizas de Fruto	Chaucha	Hibrida Chata	UY	KG	GRANDE	I	32,5	40	0		1	32,5	40	32,5
21	Hortalizas de Hoja, Brote e Inflorescencias	Choclo	Dulce	BR	DOC	GRANDE	I	50	55	0		5	10	11	55
22	Hortalizas de Hoja, Brote e Inflorescencias	Coliflor	UY	DOC	GRANDE	I	80	100	0		11	7,272727273	9,090909091		90
23	Hortalizas de Hoja, Brote e Inflorescencias	Coliflor	UY	DOC	MEDIANO	II	60	80	0		11	5,454545455	7,272727273		70
24	Hortalizas de Hoja, Brote e Inflorescencias	Espinaca	UY	DOC	GRANDE	II	25	30	0		6	4,166666667	5		25
25	Hortalizas de Hoja, Brote e Inflorescencias	Espinaca	UY	DOC	GRANDE	III	20	25	0		6	3,333333333	4,166666667		20
26	Hortalizas de Hoja, Brote e Inflorescencias	Frutilla	UY	KG	GRANDE	E	37,5	45,5	0		1	37,5	45,5		37,5
27	Frutos de Huerta	Frutilla	UY	KG	GRANDE	I	30	37,5	0		1	30	37,5		30
28	Frutos de Huerta	Lechuga	UY	DOC	12 PLANTAS	E	50	60	0		5	10	12		50
29	Hortalizas de Hoja, Brote e Inflorescencias	Lechuga	UY	DOC	12 PLANTAS	II	35	40	0		5	7	8		35

El Excel que se importa es una planilla inteligente con valores por defectos, y valores que el sistema va a verificar o agregar utilizando la base de datos de la UAM en caso de que algunos valores no hayan sido ingresados al Excel. A su vez lanza errores claros si es que hay un problema con la importación del relevamiento.

En el caso que se desee actualizar un relevamiento simplemente se debe sobrescribir en el sistema el Excel para la fecha dada.

iv. Módulos de gestión

Los 3 módulos que quedan por cubrir son aquellos que gestionan datos necesarios para la lógica de negocio. Son aquellos datos que se almacenan de forma relacional en la base de datos MySQL que luego se utilizan para generar los ingresos de guías y los relevamientos.

Están separados en 3 categorías:

1. Gestión de Artículos:
Son los datos relacionados con el producto: Su grupo, especie, variedad y presentación.
2. Regiones:
Son los datos relacionados con la región de donde proviene el producto, se admite el ingreso y modificación de Países y Departamentos
3. Útiles:
Son los datos que se requieren para el correcto funcionamiento de la lógica de negocio pero que no son estrictamente representativos de un producto tales como:

Unidades de Medida, Categorías, Calibres, Tipos de Vehículos, IPC, Localidades y plataforma para consolidar.

Este módulo permite modificar la información existente de agrupación y/o descripción de variables. Por ejemplo, la definición de Grupos de productos (Frutas Cítricas, Frutas de Huerta, Frutas de hoja caduca etc.).

Además, fácilmente se puede agregar o editar productos de la base de datos, tanto especie, variedad y presentación.

Species [Reiniciar](#)

[+ Añadir especies](#)

Grupo	Nombre	es activo	Acciones
Exóticos/Importados	Pitahaya	<input checked="" type="radio"/>	Vista previa Editar Eliminar
Exóticos/Importados	Echalote	<input checked="" type="radio"/>	Vista previa Editar Eliminar
Exóticos/Importados	Maracuyá	<input checked="" type="radio"/>	Vista previa Editar Eliminar
Frutas de Hoja Caduca/Nativos	Zarzamora	<input checked="" type="radio"/>	Vista previa Editar Eliminar
Hierbas Aromáticas	Ñame	<input checked="" type="radio"/>	Vista previa Editar Eliminar
Exóticos/Importados	Cúrcuma	<input checked="" type="radio"/>	Vista previa Editar Eliminar
Frutas de Hoja Caduca/Nativos	Mora	<input checked="" type="radio"/>	Vista previa Editar Eliminar
Exóticos/Importados	Plátano	<input checked="" type="radio"/>	Vista previa Editar Eliminar

Varieties [Reiniciar](#)

[+ Añadir variety](#)

Especie	Nombre de la variedad	código	es activo	Acciones
Mandarina	Moria	40,288	<input checked="" type="radio"/>	Vista previa Editar Eliminar
Ñame	-	50,980	<input checked="" type="radio"/>	Vista previa Editar Eliminar
Pitahaya	-	50,960	<input checked="" type="radio"/>	Vista previa Editar Eliminar
Echalote	-	50,350	<input checked="" type="radio"/>	Vista previa Editar Eliminar
Maracuyá	-	50,950	<input checked="" type="radio"/>	Vista previa Editar Eliminar
Uva	Prima	31,004	<input type="radio"/>	Vista previa Editar Eliminar
Ñame	-		<input checked="" type="radio"/>	Vista previa Editar Eliminar
Zarzamora	-	32,100	<input checked="" type="radio"/>	Vista previa Editar Eliminar

Descripción	Variedad Especie	Peso/Unidad por defecto	es activo	por defecto	Acciones
Kilo	- Choclo	1.000	<input checked="" type="radio"/>	<input type="radio"/>	Vista previa Editar Eliminar
Caja	- Choclo	17.000	<input checked="" type="radio"/>	<input type="radio"/>	Vista previa Editar Eliminar
Kilo	Dulce Choclo	1.000	<input checked="" type="radio"/>	<input type="radio"/>	Vista previa Editar Eliminar
Caja	Dulce Choclo	17.000	<input checked="" type="radio"/>	<input type="radio"/>	Vista previa Editar Eliminar
Petaca	- Mora	0.125	<input checked="" type="radio"/>	<input type="radio"/>	Vista previa Editar Eliminar
Petaca	- Frambuesa	0.125	<input checked="" type="radio"/>	<input type="radio"/>	Vista previa Editar Eliminar
Petaca	- Zarzamora	0.125	<input checked="" type="radio"/>	<input type="radio"/>	Vista previa Editar Eliminar
Cajon	Otras Manzana	20.000	<input checked="" type="radio"/>	<input type="radio"/>	Vista previa Editar Eliminar

Además, permite modificar las Unidades de Medida por producto (variable necesaria porque para algunos productos se suelen relevar precio por docena o atado), categorías de calidad, categorías de calibres, tipos de vehículos y su carga máxima, y sectores dentro de la UAM.

v. REPORTES

El componente de reportes es la que nos permite obtener toda la información cargada en el sistema referente a ingresos y precios.

Tanto como para gráficas de precios o ingresos el formato visual es el mismo:

1. Plantilla para línea de búsqueda
2. Líneas de búsqueda
3. Parámetros de búsqueda
4. Gráficas y exportación

Lo primero es la plantilla para las líneas de búsqueda. En un principio se esperaba hacer una consulta sola que recopile todos los datos, pero dado a las dificultades que presentaba esta solución y por la necesidad de procesamiento de los datos a nivel de línea, el sistema recopila los datos línea por línea, por lo tanto, si hay 5 líneas de búsqueda se van a hacer 5 llamadas a la base de datos de MongoDB.

Dependiendo el tipo de reporte los campos que tiene la plantilla. Un ejemplo es la plantilla para ingresos:

Ingresar Línea de Búsqueda ▼

Todas las Especies ▼

Todas las Variedades ▼

Todos los Países ▼

Todos los Departamentos ▼

Todos las Localidades ▼

Todas las presentaciones ▼

Identificador (Opcional)

Ingresar

Forzar Validaciones

Cada uno de estos campos permite modificar la granularidad de la línea de búsqueda, esto requiere cierta complejidad a nivel de la api interna del sistema para hacer las llamadas a la base.

A su vez las líneas pueden tener diferente nivel de granularidad permitiendo en una misma búsqueda ingresar una línea que sea “Pera – todas las variedades” y otra línea que sea “Manzana – Granny Smith – Uruguay”.

Esto se va a ver en la siguiente sección visual que es bloque de “Líneas de Búsqueda”, el mismo permite visualizar las líneas a buscar, eliminar las líneas que no son necesarias y a su vez permite guardar y cargar búsquedas para hacerlas luego nuevamente.

Especie	Variedad	País	Departamento	Presentacion	Localidad	Identificador	
Pera	Todas las Variedades	Todos los Países	Todos los Departamentos	Todas las Presentaciones	Todas las Localidades	Pera	✖
Manzana	Granny Smith	URUGUAY	Todos los Departamentos	Todas las Presentaciones	Todas las Localidades	Manzana Granny Smith	✖

📄
↓
🔍

La siguiente sección es el módulo de parámetros de búsqueda, en este se define los datos para la búsqueda que no tienen que ver con las líneas.

Parámetros de Búsqueda ▼

Fecha Inicio

Fecha Fin

Rango de Búsqueda

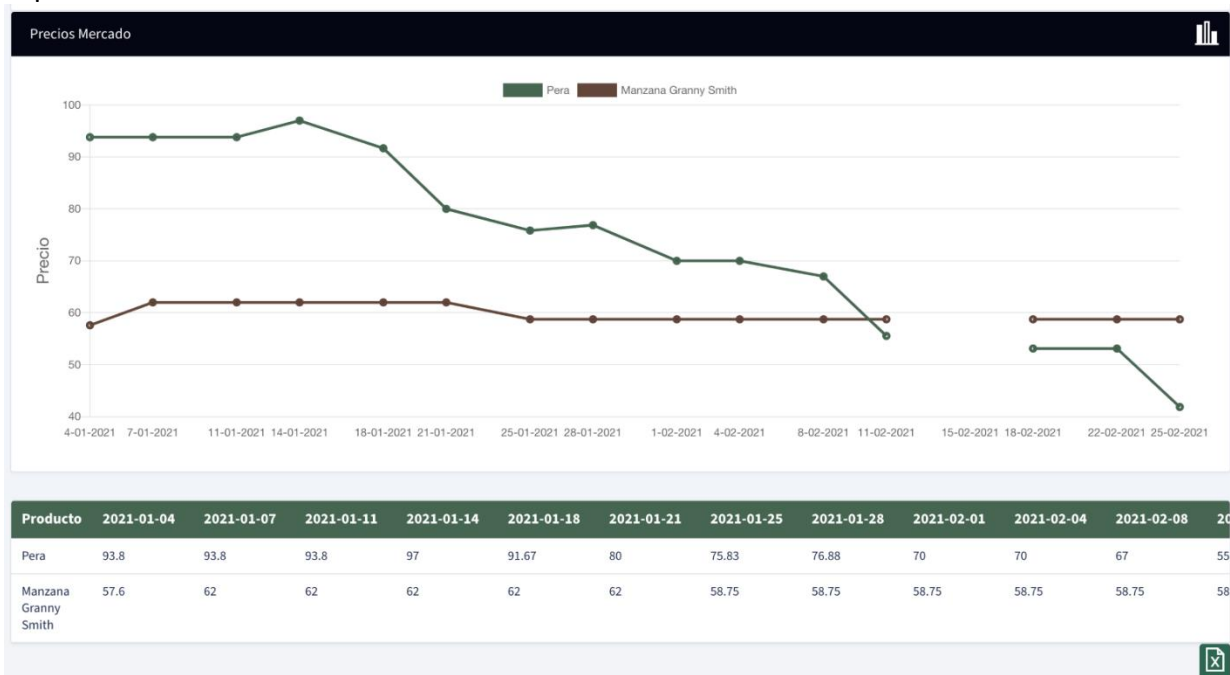
Valores Unitarios

Realizar Búsqueda

El caso más complejo es el Rango de Búsqueda que muestra el lapso por el que se agrupan los datos, si hago un periodo de 2 meses con rango diario voy a tener los ingresos separados por día, si pongo quincenal o mensual voy a tener los ingresos agrupados en cuatro o dos puntos respectivamente.

En el caso de relevamientos de precios, los datos son promediados usando el rango de búsqueda. Y el rango “Diario” solo toma en cuenta los relevamientos en los días destinados a relevamientos teniendo en cuenta que a veces los relevamientos no se hacen o se corren unos días.

La última sección muestra las gráficas que pueden ser modificadas dinámicamente para que se muestren en formato de barras o tabla y a su vez tiene una tabla con los datos que puede ser exportada a Excel.



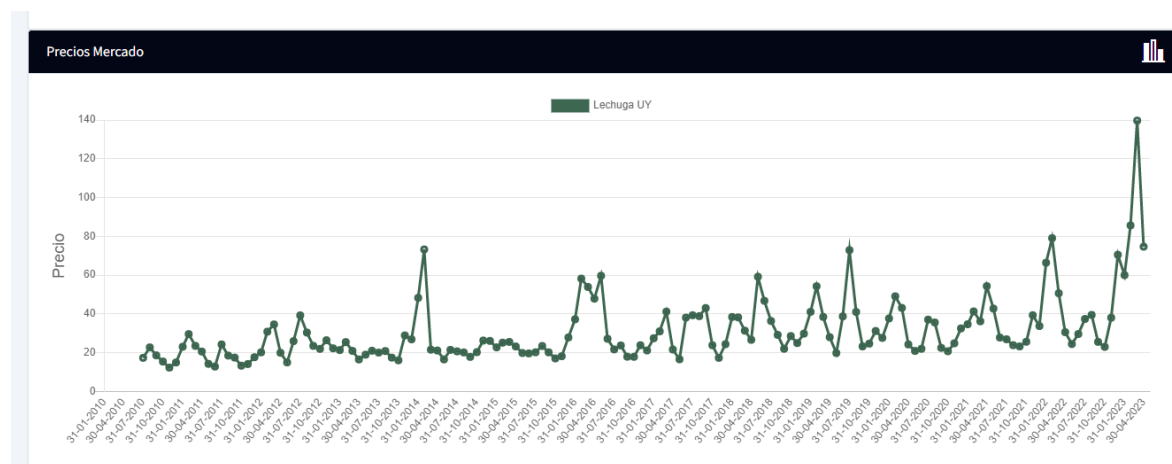
i. Precios

Permite obtener la serie de precios para cada Especie y variedad, filtrando por origen calibre y categoría de calidad, para un periodo dado.

La información de precios solicitada puede ser de forma anual, mensual, quincenal, semanal, diario o en un periodo particular. Además, el valor brindado puede ser el promedio de los máximos por kg o los mínimos por kg,

El reporte permite sacar el grafico directamente, o descargar el Excel con la información del mismo.

La consulta puede ser por grupo de especies o por cada especie por separado.

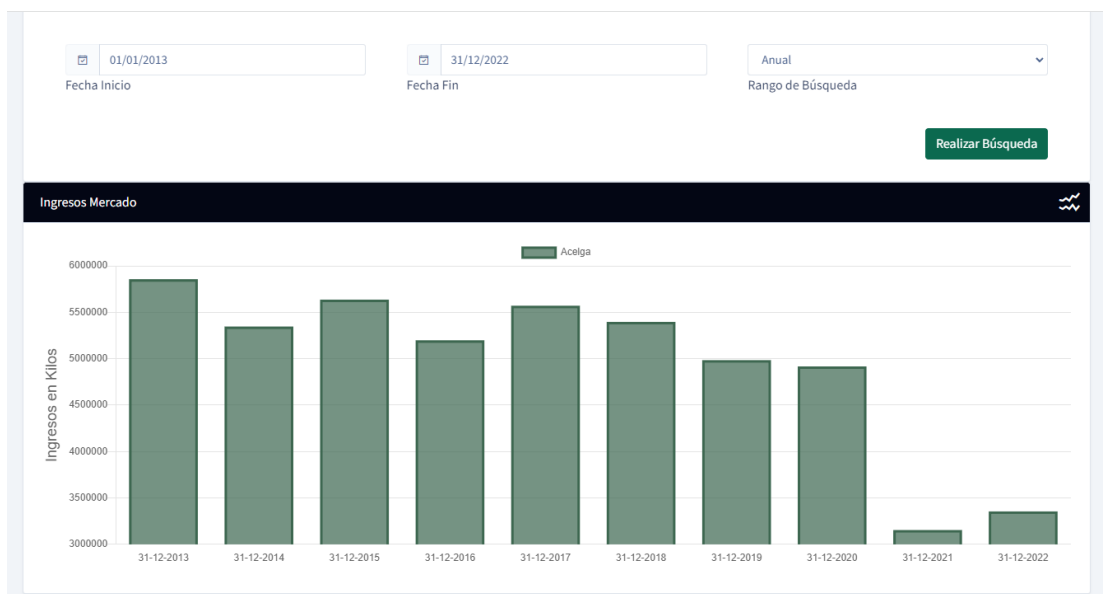


Una funcionalidad muy útil que fue desarrollada especialmente, es la de poder realizar una selección de especies, variedad, presentación, calibre y categoría, y guardarla para futuras búsquedas.

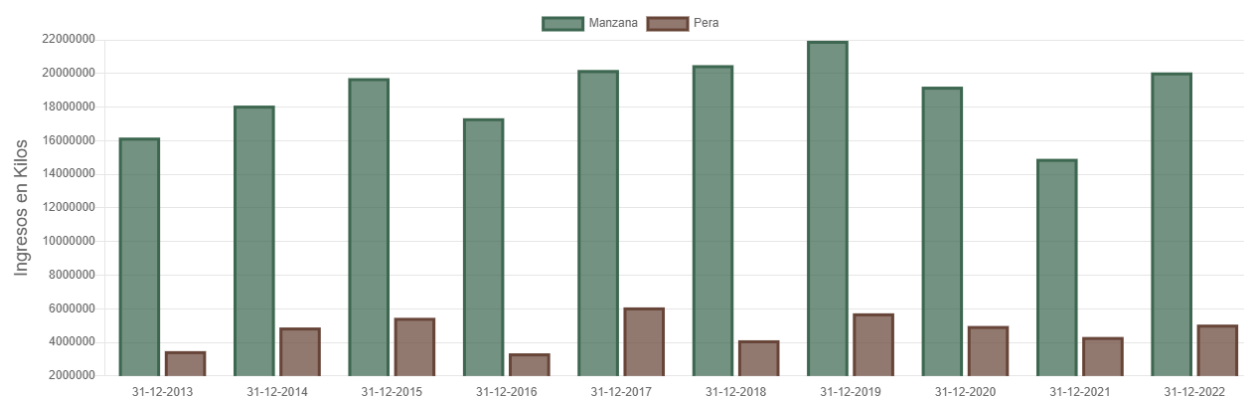
ii. Ingresos de mercadería

Para la consulta de ingresos de mercadería, se puede consultar para un periodo determinado el volumen ingresado por Especie. Además, se puede seleccionar específicamente una variedad, tipo de presentación, País o departamento de origen y/o sector de destino dentro de la UAM.

Al igual que en precios se puede seleccionar que el dato sea diario, semanal, quincenal, mensual, o anual.



También el sistema permite consultar varias especies juntas como se muestra en la siguiente imagen.

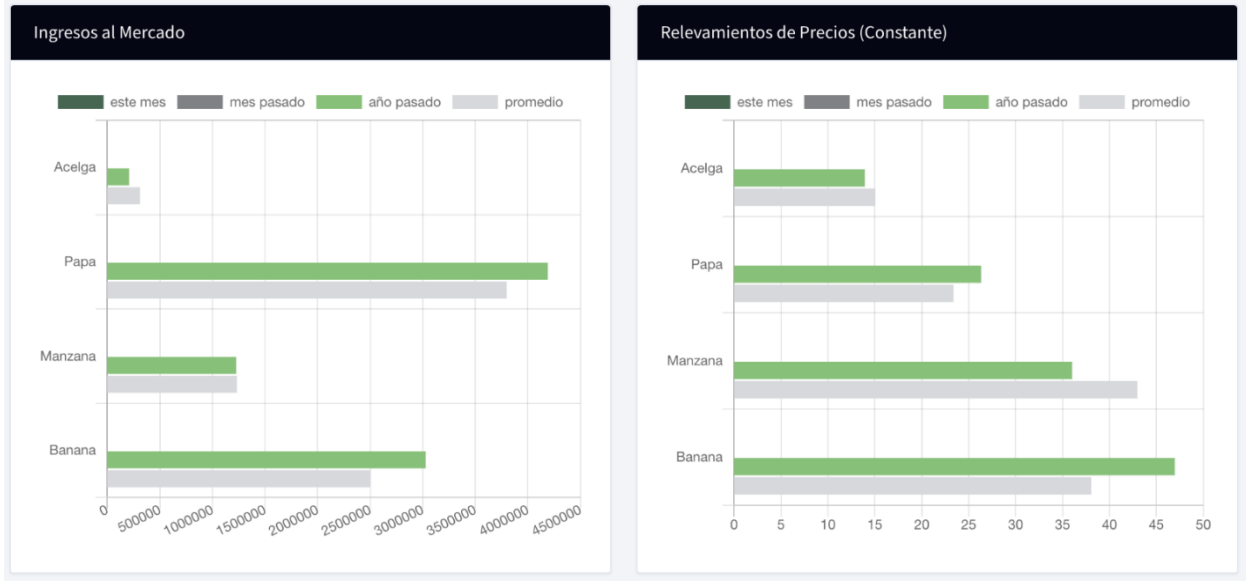


- vi. Además de estos reportes se desarrolló el cálculo automatizado de dos indicadores: el valor kilo canasta (VKC) y el índice de anomalía de precios (ICAP) que se desarrollarán en el último objetivo de este proyecto.

Uno de los últimos componentes en agregarse al sistema luego de varias iteraciones en su lógica fue el reporte de valor kilo canasta. Para el mismo se reestructuró la interfaz interna que hace los cálculos de ingresos y precios para que pueda ser reutilizable en módulos más complejos como el VKC y el ICAP.

Este módulo se encarga de calcular el Vkc a partir de los ingresos y relevamientos en MongoDB y datos que pueden ser ingresados en el momento por el usuario acorde con la lógica provista por la UAM.

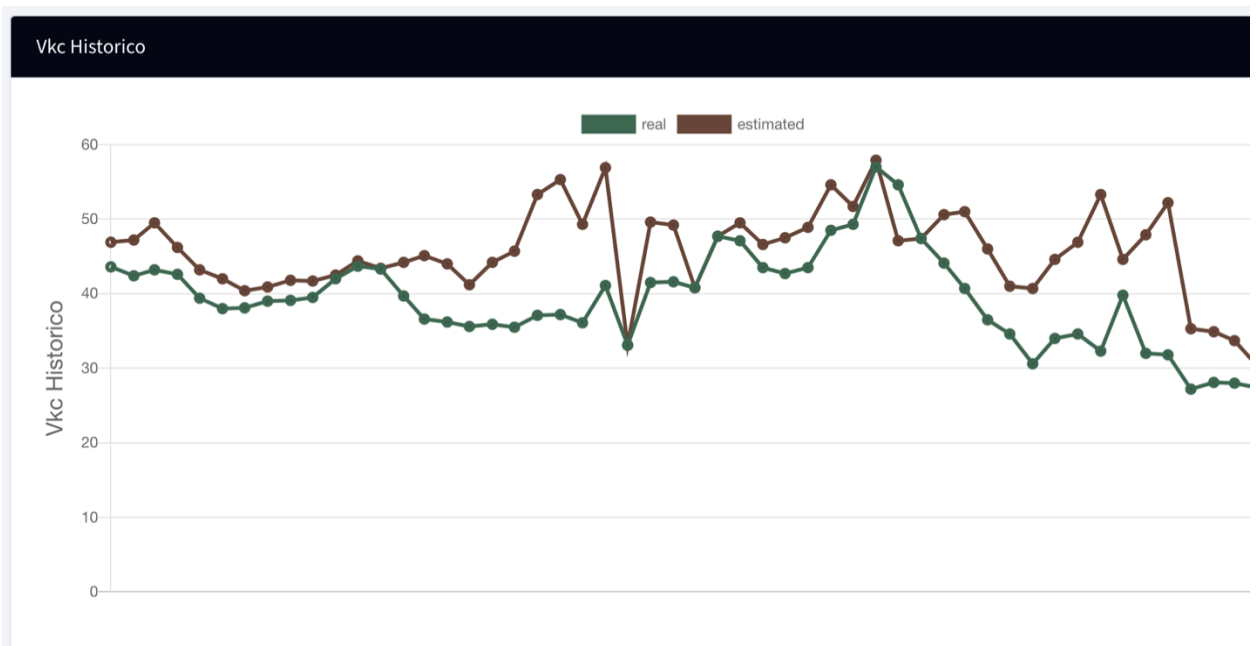
Muestra ingresos al mercado y precios para el mes correspondiente, el mes anterior, el año anterior y un promedio de los últimos 5 años:



Luego muestra el vkc a lo largo del periodo de un año



Y por último se agregó un VCK histórico:



Icap:

Es otro de los últimos componentes agregados al sistema, a partir de los precios ingresados en el sistema se calcula anomalías y probabilidades para los próximos meses teniendo en cuenta la lógica de negocio de la UAM.

Ingresar Línea de Búsqueda
▼

Banana

Todas las Variedades

70

Predicción

0,4

0,4

0.2

Ponderación - Mes
Ponderación - trimestral
Ponderación - anual

Identificador

Identificador (Opcional)

Ingresar

Especie	Variedad	Ponderación - mensual	Ponderación - trimestral	Ponderación - anual	Identificador
Papa	Todas las Variedades	0.5	0.5	0	Papa ✖
Limón	Todas las Variedades	0.3	0.5	0.2	Limón ✖
Banana	Todas las Variedades	0.4	0.4	0.2	Banana ✖

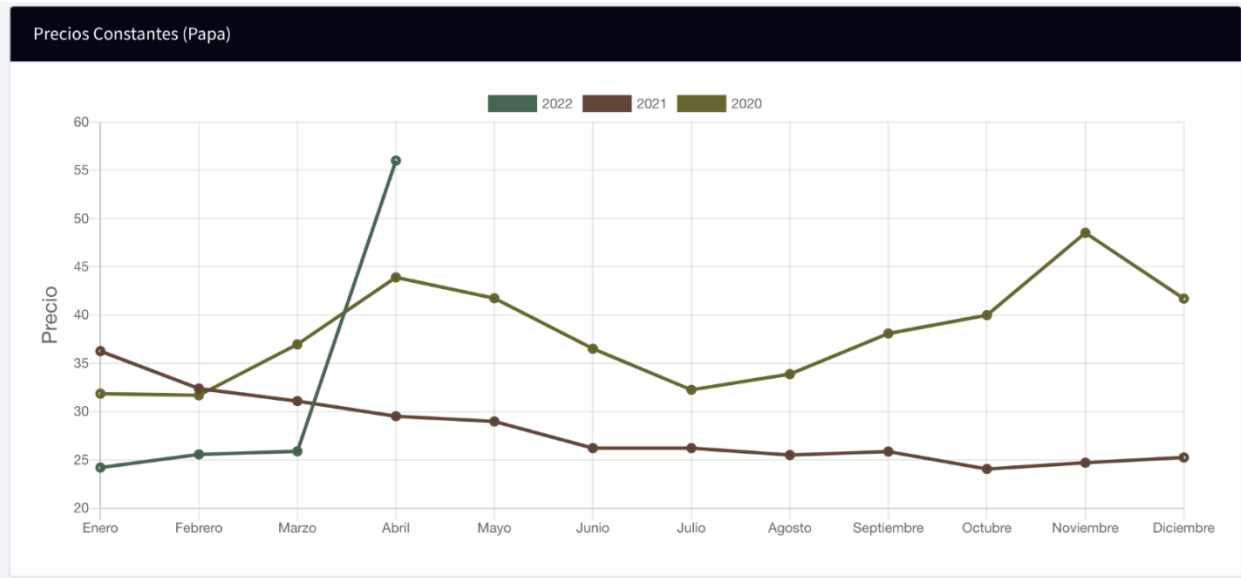
Nombre de Búsqueda

📄
⬇️
🗑️

Seleccionar Mes:
Realizar Búsqueda
Limpiar

En este componente se agregan las líneas de búsqueda y sus ponderaciones para los valores calculados en cuanto a mes, trimestre y año. (La suma de las ponderaciones debe dar 1)

Y luego el sistema devuelve los indicadores de anomalía y precios constantes para cada línea ingresada.

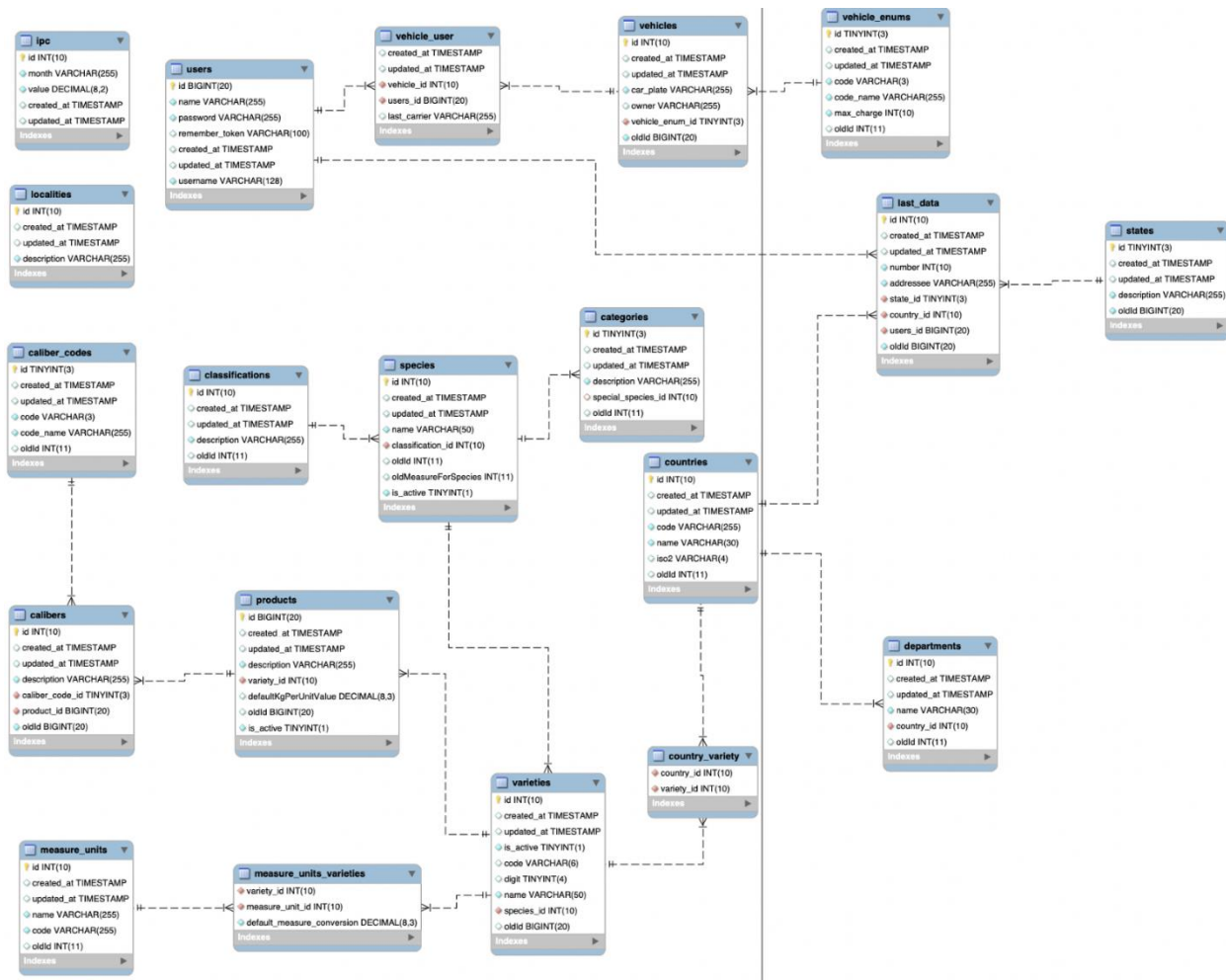


Los precios y anomalías van hasta el mes definido en la búsqueda y utilizando el último punto como el mes a predecir. Hay que tener en cuenta que si no existen valores de IPC para los meses anteriores el ICAP no va a ser posible de calcular por lo que se requiere ingresar el IPC a la plataforma todos los meses.

d) Diagramas de bases de datos utilizadas

Base de Datos MySQL:

A continuación, se presenta el diagrama relacional de la base de datos en MySQL:



Base de datos MongoDB

En MongoDB guardamos varias entidades:

Guías consolidadas: es una limpieza diaria de las guías ingresadas para facilitar la velocidad de cálculo para las métricas

consolidated_guide_lines

Storage size:	Documents:	Avg. document size:	Indexes:	Total index size:
37.99 MB	783 K	190.00 B	1	8.63 MB

Contadores: mantiene un registro de varios contadores utilizados en el proyecto

counters

Storage size:	Documents:	Avg. document size:	Indexes:	Total index size:
20.48 kB	1	65.00 B	1	20.48 kB

Guías: mantiene todas las guías ingresadas hasta el momento.

guides

Storage size:	Documents:	Avg. document size:	Indexes:	Total index size:
412.36 MB	1 M	2.57 kB	6	71.35 MB

Relevamientos: mantiene todos los relevamientos ingresados hasta el momento

surveys

Storage size:	Documents:	Avg. document size:	Indexes:	Total index size:
29.87 MB	1.3 K	141.43 kB	3	159.74 kB

Consultas de métricas: mantiene todas las consultas guardadas previamente para ser reutilizadas.

metrics_queries

Storage size:	Documents:	Avg. document size:	Indexes:	Total index size:
614.40 kB	19	276.26 kB	1	36.86 kB

5.2 Portal de visualización de modelo de predicción

El portal permite el fácil acceso a las salidas provenientes del código planteado para la predicción, en la forma de tablas y gráficos, para su uso en el proceso de información comercial. El portal debe permitir además que a medida que se vayan actualizando las bases de datos referentes a los precios, las predicciones al horizonte predictivo deseado a su vez se vayan actualizando.

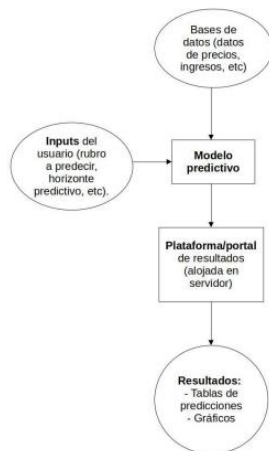


Figura 4: Esquema del portal

a) Estructura de la solución y sistemas

La solución consiste de una capa de visualización, una de datos y otra de integración. Cada una de ellas cumple una función y está interconectada con el resto, formando el todo que representa el producto final del Portal de Predicción de Precios.

Capa de integración: A través de una *API* desarrollada por Flow se obtienen los datos sin procesar (en formato *JSON*), referentes a los precios y volúmenes de mercadería de la UAM. Dicha información es procesada mediante *Pentaho Data Integration*, para llevarla a un formato a ser guardado en la base de datos que constituye la capa de datos y que es el insumo de las predicciones. Este software se emplea dentro de un contenedor *Docker* con la finalidad de garantizar el correcto funcionamiento de la solución, que a su vez se encuentra alojado en un servidor *AWS EC2* empleado sólo para este propósito.

Capa de datos: Consiste en un *datawarehouse*, compuesto por una base de datos *Postgresql* donde se almacenan los datos procesados, para ser empleados en el modelado predictivo y el dashboard del portal. Esta capa está en contacto con la de integración, la cual facilita la actualización de los datos.

Capa de visualización: La parte visible del Portal. Está programada en *R*, haciendo uso de múltiples paquetes, en particular, resulta central el uso de *Shiny*. Este paquete facilita la construcción de aplicaciones web interactivas en conexión directa con código *R*, lo cual resulta especialmente apropiado ya que fue el lenguaje empleado para la construcción de los modelos predictivos.

En esta capa se puede elegir el tipo de modelo a emplearse, el rubro a predecir, el horizonte predictivo y qué tanto hacia atrás se emplea la serie de tiempo. La salida es un gráfico de serie de tiempo que permite observar la serie original, la predicción puntual en el horizonte indicado y los intervalos de confianza. Además, también se presenta la tabla de dichos resultados en formato descargable.

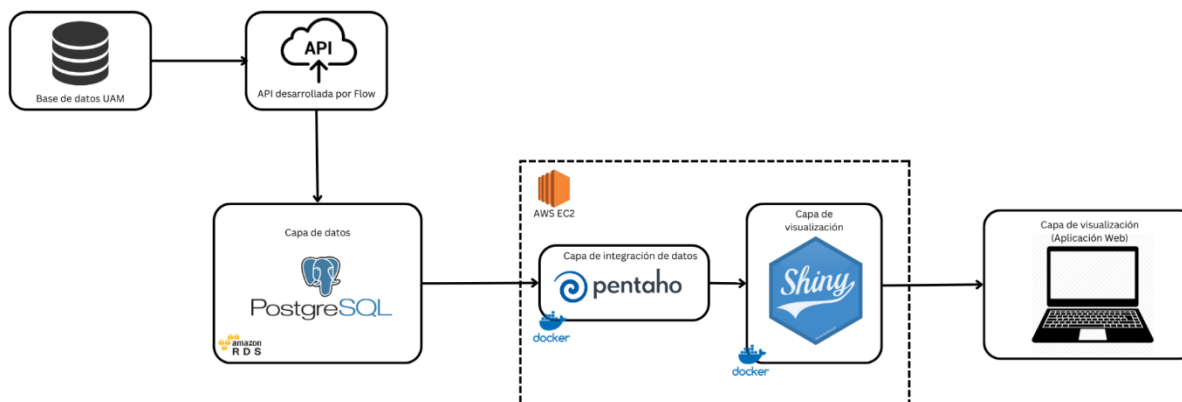


Diagrama de la estructura del Portal.

a) Tipo de datos

Los datos empleados son series de tiempo de precios y volumen de ingresos de mercadería, relevados en el marco del Observatorio Granjero. Dichas series van desde enero de 2013 hasta junio de 2022 y tienen una frecuencia semanal (obtenida mediante el promedio de los dos relevamientos que se hacen los lunes y jueves de todas las semanas). Se cuentan con 493 observaciones para cada rubro estudiado correspondientes al periodo.

Si bien es posible obtener datos previos a 2013, no son considerados del todo fiables, por lo que no se tomarán en cuenta. Además, se acota el problema a los rubros de origen nacional, dado que para los productos de origen extranjero no se cuenta con los ingresos semanales, sino con un agregado mensual.

Las series empleadas corresponden a los rubros papa, manzana, tomate, naranja y cebolla.

Luego de ser parseado el formato *JSON* en el que son obtenidos los datos desde la *API*, son volcados a una tabla de hecho con sus respectivas dimensiones en la base de datos *PostgreSQL* mencionada anteriormente. El “hecho” aquí es el relevamiento que ocurre en un día determinado, mientras que las dimensiones son la variedad, país de origen y categoría. Esta información es luego resumida semanalmente, empleando la información relevante (en este caso, nos interesa el origen local y solo las categorías I y II), llevándose a un formato consumible por la plataforma y los modelos contenidos en ella.

b) Especificaciones técnicas de los modelos de predicción

Los modelos que son el foco de atención en este trabajo son los modelos de espacio-estado, una metodología flexible y general que permite tratar una variedad de problemas de análisis de series de tiempo.

Como se mencionó previamente, bajo este marco se considera que el sistema o proceso que se estudia es determinado por una serie de vectores no observables $\theta_1, \dots, \theta_n$ (proceso de estado), que se asocian a una serie de vectores que sí es observada y_1, \dots, y_n (proceso observado). La estructura del modelo es definida en base a estos conjuntos de vectores, matrices que definen la especificación y vectores de términos de error; todos ellos pudiendo depender de parámetros desconocidos a ser estimados.

La estructura de un modelo de espacio-estado lineal se puede describir de la siguiente manera:

$$Y_t = F_t \theta_t + v_t, v_t \sim N(0, V_t)$$
$$\theta_t = G_t \theta_{t-1} + w_t, w_t \sim N(0, W_t)$$

Donde:

- Y_t es el vector m -dimensional de los valores de las m series de las variables consideradas en el momento t .
- θ_t es un vector p -dimensional de procesos no observables.
- F_t y G_t son matrices de dimensiones $m \times p$ y $p \times p$ respectivamente que permiten caracterizar la especificación del modelo.
- v_t y w_t son perturbaciones de los procesos observados y de estado respectivamente.

La flexibilidad de estos modelos ha dado lugar a una miríada de especificaciones. Entre ellas se destacan los modelos *DLM*, *TBATS*, *SUTSE* y los modelos *ARIMA* con estacionalidad expresada en términos de Fourier; todas ellas aplicadas en la plataforma.

5.3 Comentarios finales

Se logro consolidar una herramienta que permitió mejorar la sistematización de la información de volúmenes y precios, y asegura una mejorara en la calidad de ésta, además de modernizar la herramienta de declaración permitiendo sea online. El diseño de esta nueva plataforma, y su versatilidad permite obtener información específica y el cálculo de indicadores complejos.

Se logro consolidar una herramienta para el cálculo de los modelos de predicción, que será evaluada en meses futuros.

6. Objetivo iii: Indicador de comportamiento anómalo de precios en frutas y hortalizas (ICAP)

Juliana Bruzzone, Juan Millán, Andrés Díaz y María Noel Ackermann

En esta oportunidad, y considerando como base el avance realizado en 2016, se busca actualizar el cálculo del Indicador de Comportamiento atípico de precios (ICAP) para los rubros que se tomaron en cuenta para elaborar el modelo de predicción de precios previstos en el proyecto de ANII. Además de incorporar en la plataforma de información de la UAM el cálculo automático del indicador para todos los rubros que cuenten con series de precios históricas.

Como objetivo secundario, se diseñará un mecanismo para generar predicciones del ICAP a uno o dos meses hacia adelantes utilizando los valores predichos por los modelos de predicción. Con esto se podrá detectar anticipadamente valores de precios atípicos.

Consecuentemente, el resultado de dicho proceso es la obtención de una herramienta que conjuga los distintos puntos del proyecto, la que permitirá robustecer la toma de decisiones de diversos actores vinculados a la producción, comercialización y hacedores de políticas. Por ejemplo, junto con otra información disponible, la herramienta contribuirá a detectar si en los meses venideros se registrarán precios que justifiquen la necesidad de importación de productos para moderar la escalada de precios. También se podrá identificar productos a promocionar, si la baja de precios es relevante, mediante el desarrollo de campañas de consumo llevadas adelantes por distintas instituciones para impulsar su comercialización.

6.1 Antecedentes

En situaciones de crecimiento sostenido, y en particular en el caso de rubros alimentarios, los precios suelen ser objeto de análisis debido a su incidencia sobre los registros inflacionarios y la seguridad alimentaria de la población. Por otro lado, bajas sustanciales y persistentes de los precios pueden comprometer la rentabilidad de quienes producen alimentos. Por lo tanto, en productos que en general presentan una estacionalidad marcada en la producción, que se traslada

a su vez a precios, cobra relevancia analizar si los vaivenes observados responden a los fundamentos propios de los mercados o a hechos excepcionales y la persistencia de dichos eventos.

Entre las metas establecidas para el Objetivo de Desarrollo Sostenible (ODS) número 2, anima a los países a adoptar medidas para asegurar el funcionamiento apropiado de los mercados de alimentos básicos y sus derivados y facilitar el acceso oportuno a la información de mercado (incluyendo los datos sobre reservas de alimentos) de manera de contribuir a limitar la alta volatilidad de sus precios. Concretamente la meta 2.c sugiere la adopción de un indicador que permita capturar variaciones anómalas en el precio de los alimentos.

World Food Programme (2014) destaca que el uso de este tipo de indicadores constituye una herramienta de monitoreo para detectar fácilmente los productos que presentan un comportamiento “atípico” en sus precios, así como conocer la severidad y persistencia del fenómeno. Estos indicadores se nutren de información histórica, lo que podría contribuir a generar alertas tempranas y posibles medidas de política de contingencia.

De acuerdo a Baquedano (2015), el indicador para detectar comportamientos atípicos de precios se basa en el cálculo de una tasa de variación compuesta ponderada que toma en cuenta tanto los cambios de precios dentro del año como entre años. Una vez calculadas estas tasas, se definen los umbrales que determinan los comportamientos atípicos, tanto en variaciones al alza como a la baja. La ponderación de medias y desvíos se realiza de manera tal que el pasado reciente tiene un peso mayor respecto al lejano, para disminuir los falsos negativos (no detectar datos atípicos cuando en realidad existen).

En Uruguay, y con el objetivo de monitorear cambios atípicos en los precios mayoristas de frutas y hortalizas, la Unidad Agroalimentaria Metropolitana (UAM - ex Mercado Modelo) y OPYPA-MGAP realizaron en 2016 un primer esfuerzo en construir un indicador para detectar anomalías de precios en cuatro productos (papa, morrón, manzana, lechuga) siguiendo la metodología de Baquedano (2015).

6.2 Método

En este estudio se toma como base la metodología desarrollada por Baquedano (2015) y presentada en Anuario OPYPA (2016), la cual plantea el desarrollo de un indicador que permite medir la volatilidad de los precios y de esa manera detectar variaciones anómalas de los precios de alimentos. Esto permitiría también generar alertas tempranas ante cambios atípicos de tales precios.

Así, el ICAP para un determinado rubro y en el mes t es la suma ponderada de tres tasas de crecimiento compuestas (TCC), las cuales se definen de acuerdo a la ecuación [1].

$$TCCX_T = \left(\frac{P_{t_n}}{P_{t_0}} \right)^{\frac{1}{t_n - t_0}} - 1 \quad [1]$$

Donde P_{t_n} es el precio al final del periodo, P_{t_0} es el precio al inicio del período y $(t_n - t_0)$ son los meses transcurridos entre t_0 y t_n , y X indica el período en el cual se calcula la variación. A modo de ejemplo, la TCC anual para el mes de diciembre de 2021 se calcula de la siguiente manera:

$$TCCA_{dic-2021} = \left(\frac{P_{dic-2021}}{P_{dic-2020}} \right)^{\frac{1}{12}} - 1 \quad [1']$$

Cabe mencionar que los mercados de frutas y hortalizas presentan una estacionalidad dentro del año debido a los procesos productivos. Asimismo, este hecho puede verse acentuado por la presencia de factores intra-anales (climáticos, sanitarios, entre otros) que pueden afectar no sólo a la producción sino también a la comercialización. A tales efectos, para contemplar en un único indicador las variaciones dentro de un mismo año y las registradas entre años se modificó el TCCX. Para ello, se definió una tasa de crecimiento compuesta mensual (TCCM), una trimestral (TCCT) y una anual (TCCA). Estas tasas se calcularon como una variación promedio móvil respecto a 1, 3 o 12 meses precedentes, respectivamente.

Posteriormente, se define que los precios presentan variaciones atípicas cuando la diferencia absoluta entre el TCCX y su media histórica para el mismo mes de referencia es mayor a un desvío estándar, tal como lo muestra la ecuación [2]:

$$\left(\frac{TCCX_{yt} - \overline{TCCX_t}}{\hat{\sigma}_t} \right) = ICAPX_t \{ ICAPX_t \geq 1, \text{variación atípica } (X_ICAP_t^A) \text{ en otro caso, variación normal } (X_ICAP_t^N) \} \quad [2]$$

Donde $TCCX_{yt}$ es la tasa de crecimiento compuesta mensual, trimestral o anual en el mes t para el año y . Por su parte, $\overline{TCCX_t}$ es el promedio ponderado de la tasa de crecimiento compuesta mensual, trimestral o anual en el mes t a lo largo de la serie histórica, hasta el año y . Asimismo, $\hat{\sigma}_t$ es el desvío estándar ponderado de la tasa de crecimiento compuesta mensual, trimestral o anual para el mes t a lo largo de la serie histórica, hasta el año y . Finalmente, $ICAPX_t$ es el indicador de comportamiento atípico de precios mensual, trimestral o anual en el mes t .

La ponderación de medias y desvíos se realiza de forma tal que el pasado reciente tiene un peso mayor en relación al lejano, para disminuir los falsos negativos (no detectar datos atípicos cuando en realidad existen). Por ejemplo, si todos los períodos de tiempo considerados tuvieran el mismo peso, uno de precios altos y volátiles como otro de precios bajos y estables tendrían la misma ponderación en el cálculo de media y desvío histórico de las tasas de crecimiento. Por lo tanto, si se registran precios altos y volátiles al inicio de la serie de datos mientras se constatan precios bajos y menos volátiles en la actualidad, podría no detectarse un comportamiento atípico ya que la comparación se realiza con un umbral mayor (Baquedano, 2015).

El indicador síntesis de comportamiento atípico de precios (ICAP) es la suma ponderada de los indicadores de comportamientos atípicos mensuales, trimestrales y anuales según se puede apreciar en la Ecuación [3]:

$$ICAP_t = \gamma.ICAPM_t + \delta.ICAPT_t + (1 - \gamma - \delta).ICAPA_t \quad [3]$$

Un aspecto relevante a definir del cálculo del ICAP es el valor de las ponderaciones γ y δ , dado que son parámetros que se fijan de manera arbitraria, de acuerdo al juicio de experto.

6.3 Resultados

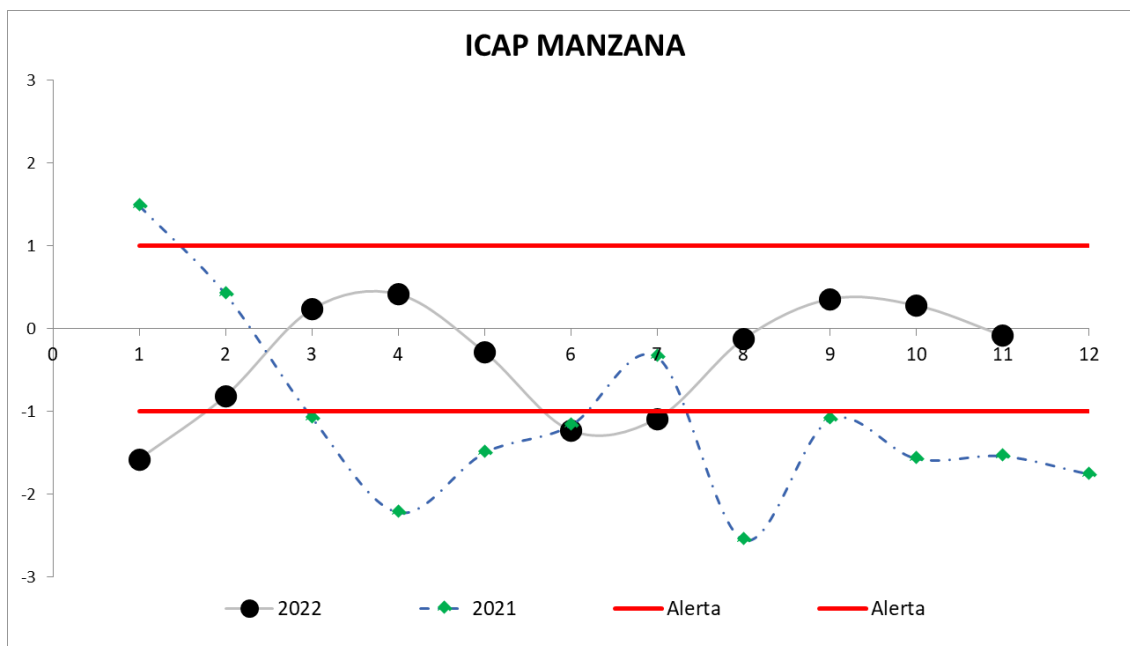
Bajo la metodología anteriormente descrita, se consideran datos mensuales de precios mayoristas de los rubros a analizar, expresados a precios constantes, en el período enero de 2012 al último dato disponible 2022. Si bien se cuenta con información de precios de años anteriores, se entiende que un periodo de diez años es razonable para detectar la historia más reciente y contemplar así formas de producción y comercialización actuales, que distan de lo ocurrido más allá de diez años (por ejemplo, avances han permitido reducir la estacionalidad de la oferta).

En cuanto a los ponderadores adoptados para el cálculo del ICAP sintético en cada producto se establecieron en 50%, 25% y 25% para la variación interanual, trimestral y mensual, respectivamente.

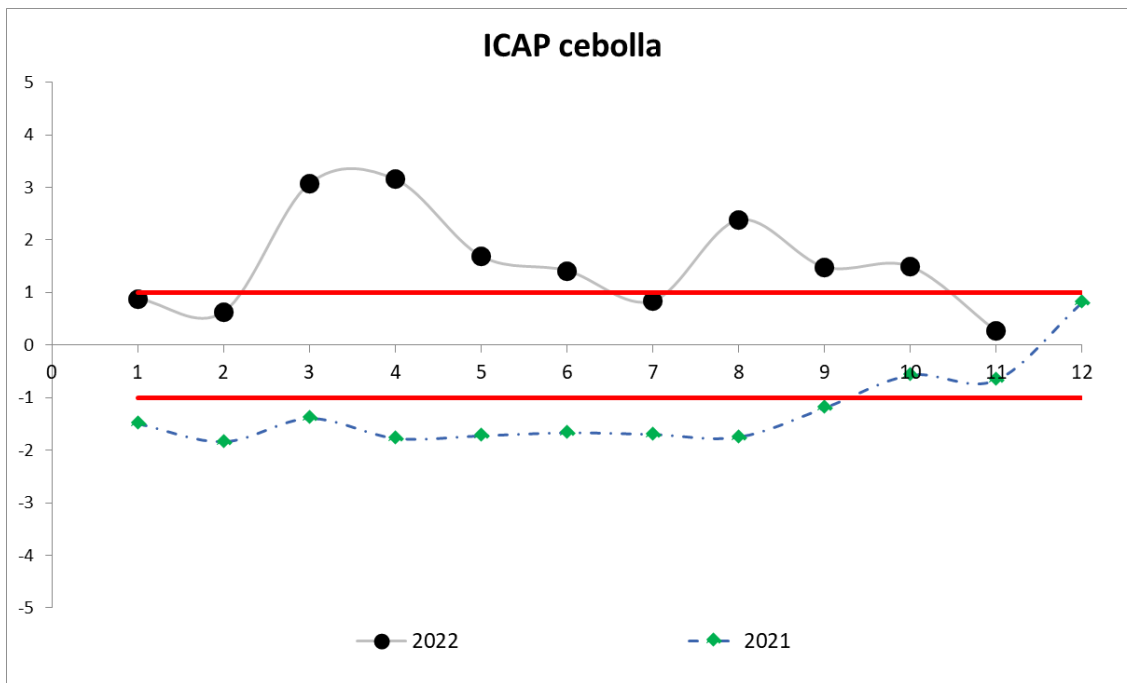
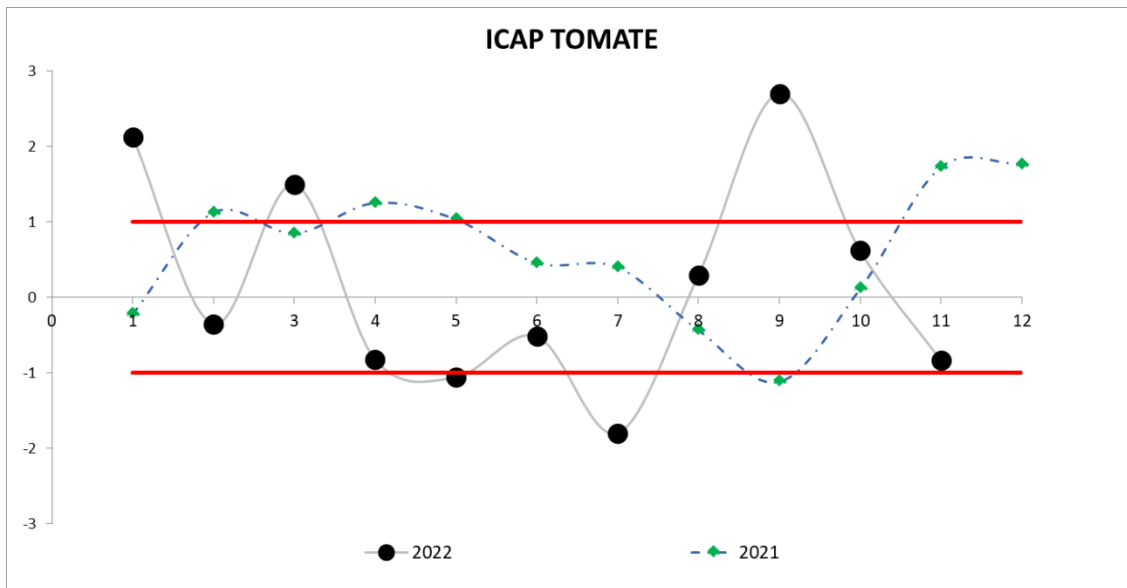
A continuación, se presentan los gráficos del indicador para Manzana y Tomate, en los mismos la curva roja indica el valor límite de normalidad igual a la unidad.

En el caso de manzana se aprecian valores dentro del rango de normalidad o inferiores a este para los años 2020, 2021 y 2022. El año 2021 se caracterizó por ser un año de alta oferta, que junto con la mejora en las técnicas de conservación han generado un escenario de precios bajos,

para el año 2022, si bien los valores fueron bajos, al ser comparados con varios años de precios bajos, están dentro del margen esperable como es observado



Para el caso de Tomate se identifica para el 2022 entre febrero y abril, y entre agosto y octubre se registraron periodos de precios altos. Estos picos de precios del 2022 se debieron, en primer lugar, a un periodo de intensas lluvias en febrero que hicieron descender tanto la cantidad como la calidad de la oferta en el corto plazo. Y en segundo lugar en el mes de setiembre se registró un periodo de precios altos como consecuencia de las malas condiciones para la formación y crecimiento de fruta de los meses anteriores (días nublados y fríos intensos) que generaron poca fruta de valor comercial superior, que junto con una demanda creciente actuaron incrementando los precios



Como se observa en cebolla, prácticamente todo el año se observaron valores anormalmente altos o por al menos en los márgenes superiores de lo esperable, esto viene dado por las malas condiciones de crecimiento para la cebolla, que por un lado bajaron el stock a conservar y por el otro disminuyeron notablemente su calidad y potencial de conservación, dando como resultado menor oferta de un producto de baja elasticidad de precios y por ende plausible de alcanzar precios altos a muy altos.

6.4 Automatización del cálculo

Uno de los objetivos de trabajar este indicador en el marco de este proyecto, además de su optimización y ajuste fue desarrollar la herramienta para facilitar su cálculo y automatizarlo.

Es por esto por lo que se desarrolló dentro de la plataforma de ingreso y precios mayoristas de frutas y hortalizas de la UAM, un reporte que permite el cálculo del ICAP para todos los productos.

Actualmente este indicador se utiliza únicamente internamente, pero será incorporado como parte de los informes mensuales y especiales.

6.5 Comentarios finales:

Este indicador permite tener una visión general del comportamiento de precios histórico, no es una herramienta como tal para adelantarse a un comportamiento atípico. Asimismo, la incorporación del precio predicho en la serie de precios podría anticipar la alerta, pero esto debe ser evaluado.

Este indicador es una herramienta que debe ser utilizada por conocedores del mercado, ya que su interpretación en caso de tiempo prolongado de precios atípicos puede llevar a una interpretación de normalidad, siendo esto un error.

El Observatorio Granjero integra desde 2015 la Organización de Información de Mercados de las Américas como representante de Uruguay. Esta Organización coordina y articula las oficinas de información de mercados agroalimentarios de 33 países de América del Sur, América del Norte, Centro América y el Caribe. Esta participación se implementa a través de la realización de reuniones anuales plenarias, en que se realizan una serie de actividades de capacitación e intercambio y a través de coordinaciones regionales, dónde Uruguay integra el grupo Cono Sur, junto a Chile, Argentina, Brasil y Paraguay. Esta participación permitió la difusión de lo desarrollado en este proyecto en varias instancias de intercambio, siendo un tema de interés para los referentes.

7. Comentarios finales del proyecto

Este proyecto permitió y propició el desarrollo de nuevas herramientas para el análisis de la información del mercado frutihortícola, con un avance importante en la tecnología utilizada para el almacenamiento y manejo de la información, cálculo de indicadores y perspectivas de precios a corto plazo.

En el análisis para la generación de los modelos de predicción para los distintos rubros, se detecta la falta de información de volumen de producción real a nivel de campo, que podría contribuir al fortalecimiento del modelo de predicción. Los modelos de predicción desarrollados serán utilizados en indicadores para anticiparnos a situaciones particulares.

Las plataformas generadas, permiten el análisis de la información y aseguran la confiabilidad de los datos.

Además de la generación de estas herramientas de análisis de mercado, el proyecto permitió ponderar la importancia del análisis estadístico de la información, como un recurso imprescindible para la toma de decisiones.

8. Bibliografía

Bregante, A. (2002). *Encuesta Frutícola. Producción, variedades, comercialización y endeudamiento. Zafra 2001/02*. Ministerio de Ganadería Agricultura y Pesca, Dirección de Estadísticas Agropecuarias.

Bregante, A. (2016). *Encuesta Frutícola de Hoja Caduca. Zafra 2016*. Ministerio de Ganadería Agricultura y Pesca, Dirección de Estadísticas Agropecuarias.

Casares, I. (2017). *Encuesta Citrícola, primavera 2016*. Ministerio de Ganadería Agricultura y Pesca, Dirección de Estadísticas Agropecuarias.

Casares, I. (2020). *Encuesta de papa: Primavera 2020*. Ministerio de Ganadería, Agricultura y Pesca, DIEA.

Cleveland, W. P., Evans, T., & Scott, S. (2014). *Weekly Seasonal Adjustment - A Locally-weighted Regression Approach* (N.º 473). U.S. Bureau of Labor Statistics.

Comisión Administrativa del Mercado Modelo. (2021). *Manual de procedimientos y referencias técnicas para la tipificación de la calidad de frutas y hortalizas frescas*. Unidad Agroalimétrica Metropolitana.

De Livera, A. M., Hyndman, R. J., & Snyder, R. D. (2011). Forecasting Time Series With Complex Seasonal Patterns Using Exponential Smoothing. *Journal of the American Statistical Association*, 106, 1513-1527.

- Díaz, A. (2018). *Consumo de frutas: Estimación y análisis de sus determinantes para lograr una ingesta acorde a la recomendación internacional*. Universidad de la República, Facultad de Ciencias Económicas y Administración, IESTA.
- Dieng, A. (2008). Alternative Forecasting Techniques for Vegetable Prices in Senegal. En *Revue sénégalaise des recherches agricoles et agroalimentaires* (N.º 3; Vol. 1).
- Durbin, J., & Koopman, S. J. (2012). *Time Series Analysis by State Space Methods*. Oxford University Press.
- Freiría, G. (2020). *Anuario estadístico agropecuario 2020*. Ministerio de Ganadería, Agricultura y Pesca, Dirección de Estadísticas Agropecuarias.
- Godahewa, R., Bergmeir, C., Webb, G. I., & Montero Manso, P. (2020). *A Strong Baseline for Weekly Time Series Forecasting*. <https://arxiv.org/abs/2010.08158#:~:text=Many%20businesses%20and%20industries%20require,approaches%20dedicated%20to%20this%20task>.
- Güenaga, M. (2014). *Componentes variables del IPC: Frutas y Verduras*. Universidad de la República, Facultad de Ciencias Económicas y Administración.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.
- Harvey, A. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.
- Harvey, A., Koopman, S. J., & Riani, M. (1997). The Modeling and Seasonal Adjustment of Weekly Observations. *Journal of Business & Economic Statistics*, 15, 354-368.
- Hyndman, R. J., & Anne B. Koehler. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22, 679-688.
- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: principles and practice*. OTexts.
- Li, G., Xu, S., & Li, Z. (2010). Short-Term Price Forecasting For Agro-products Using Artificial Neural Networks. *Agriculture and Agricultural Science Procedia* 1, 278-287.
- Luo, C., Wei, Q., Zhou, L., Zhang, J., & Sun, S. (2011). Prediction of Vegetable Price Based on Neural Network and Genetic Algorithm. *Computer and Computing Technologies in Agriculture IV*, 672-681.
- Millán, J., & Romero, D. (2019). *Aportes para la construcción de un modelo de predicción de precios mayoristas de frutas y hortalizas en el Uruguay*. Universidad de la República, Facultad de Agronomía.

Mitra, D., & Paul, R. K. (2017). Hybrid time-series models for forecasting agricultural commodity prices. *Model Assisted Statistics and Applications*, 12, 255-264.

Observatorio Granjero. (2019). *Anuario estadístico 2018*. Mercado Modelo.

Observatorio Granjero. (2022). *Anuario estadístico 2021*. Unidad Agroalimentaria Metropolitana.

Peng, Y.-H., Hsu, C.-S., & Huang, P.-C. (2015). *Developing Crop Price Forecasting Service Using Open Data from Taiwan Markets* (pp. 172-175). IEEE.

Petris, G., Petrone, S., & Campagnoli, P. (2009). *Dynamic Linear Models With R*. Springer Science + Business Media.

West, M., & Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*. Springer New York.

Wolf, C., Joye, D., Smith, T. W., & Fu, Y. (2016). *The SAGE Handbook of Survey Methodology*. SAGE Publications Ltd.

Young, P. C., Pedregal, D. J., & Tych, W. (1999). Dynamic Harmonic Regression. *Journal of Forecasting*, 18, 679-688.