



Modelos de predicción de precios mayoristas de frutas y hortalizas

Diciembre 2022

Línea de trabajo incluida en el proyecto **“Diseño y aplicación de un modelo predictivo de precios de frutas y hortalizas frescas para el comercio mayorista de Uruguay”**

FSDA_1_2018_1_154836 Fondo sectorial de investigación a partir de datos, ANII.

Equipo

Responsable técnico-científico: *Juliana Bruzzone*

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

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

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

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

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

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

Resumen ejecutivo

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. En esta publicación se presentan algunos de 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*. En el marco del proyecto se emplearon modelos de espacio estado para la predicción de los precios de un conjunto acotado de rubros, llegando 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 según las características del rubro. 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.

Introducción

La presente publicación presenta algunos de 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* (FSDA_1_2018_1_154836). En particular, los referentes a la línea de trabajo de la modelización de los precios de frutas y hortalizas para su predicción.

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. Las instituciones involucradas en el proyecto (UAM¹, DIGEGRA², IICA³, OPYPA⁴, FAGRO⁵, entre otras) tienen interés en contar con dicho modelo predictivo, que representa 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.

En el caso de Uruguay, las frutas y hortalizas representan un sector de la economía que depende fuertemente de la producción nacional, donde una proporción considerable de la misma (60% de la producción nacional aproximadamente⁶) era volcada al Mercado Modelo y en la actualidad es canalizada por 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 marco del Observatorio Granjero⁷ 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.

Además, se hace uso de la información que proveen los y las especialistas involucrados en las distintas etapas del proyecto, en especial a la hora de incorporar información en la modelización. A modo de ejemplo, se logra caracterizar los mercados de los rubros trabajados, estudiar la estacionalidad de los distintos rubros y la identificación y eventual incorporación de variables auxiliares.

El objetivo principal busca la predicción del precio con el menor error, siendo definido este último en base a un conjunto de métricas delineadas en *Hyndman y Koehler (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)*. La inclusión de variables auxiliares también es evaluada, pero muchas de las series de las variables propuestas no cuentan con registros históricos completos por lo que no es posible la incorporación en el cálculo o en algunos casos su efecto no fue claro sobre la predicción.

Marco teórico

¹ Unidad Agroalimentaria Metropolitana.

² Dirección General de la Granja.

³ Instituto Internacional de Coordinación Agrícola.

⁴ Oficina de Programación de Política Agropecuaria.

⁵ Facultad de Agronomía, Universidad de la República.

⁶ Estimación de los técnicos del Observatorio Granjero.

⁷ El Observatorio Granjero es una iniciativa de trabajo conjunto entre la UAM y el MGAP que constituye el Sistema de Información de Mercados nacional de frutas, hortalizas, huevos y animales de granja.

Rubros elegidos

Los rubros seleccionados para el análisis son: manzana, papa rosada, tomate (redondo), cebolla blanca (incluyendo la cebolla temprana) y naranja.

Para su selección se tomó en cuenta un conjunto de 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, perecibilidad y características de comercialización. La finalidad de esto es captar la diversidad de situaciones que pueden plantear distintos escenarios de modelado. Como tercer criterio para la selección de los rubros se consideró que la evolución de los precios históricos mayorista del producto pueda tener cierta respuesta a factores que afectan la producción, como por ejemplo factores ambientales, el potencial de conservación del producto, entre otros. Por último, se tomó en cuenta la volatilidad a corto plazo de la evolución de precios en el mercado, dado que este último factor 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 cortos periodos de tiempo, 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*).

En la presente publicación se presentan a modo ilustrativo los resultados obtenidos para los rubros manzana y tomate. El periodo de tiempo considerado para el estudio va de enero de 2013 a junio de 2022.

Caracterización de los rubros

Manzana

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 a corto plazo en el volumen de la producción están más asociados a factores que afectan la productividad de una zafra, que a cambios en su superficie. De todas formas, analizando el periodo 2001/02 al 2015/16, si bien se constata que la superficie total pasó de 3.716 a 2.677 hectáreas (*Bregante, 2002; Bregante, 2016*), el aumento en el número de plantas por hectárea ha permitido que la producción posiblemente no haya bajado en forma acentuada, si bien la interrupción en la realización de relevamientos de producción desde el 2016 impide analizar este comportamiento con rigurosidad.

Regularmente, el periodo de cosecha de manzana se concentra de enero a abril, meses en que predomina en el mercado el producto recién cosechado. Tradicionalmente, hacia los meses de invierno, el producto presente en la oferta fue conservado principalmente en cámaras de frío convencionales y entrada la primavera el producto principalmente proviene de cámaras de atmósfera controlada. Esta última tecnología es cada vez más utilizada por lo que en los últimos años ya en el comienzo del invierno hay productos provenientes de cámaras de atmósfera controlada. Hacia fin de año se incorpora (generalmente sobre noviembre y diciembre) producto importado, que complementa la oferta en casos de escasa oferta o relativamente baja proporción de frutas con calidad comercial.

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 de precios altos ya que son rubros de cosecha anual. En este sentido, la relativa concentración de la producción y el comercio mayorista en pocos actores determina el desarrollo de estrategias en las que según los niveles de producción se regule la oferta de manera de disponer de

producto durante todo el período de comercialización hasta la siguiente cosecha. 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 (Millán y Romero, 2019). La manzana es un cultivo que depende particularmente de los factores climáticos anteriormente mencionados, cuya inclusión es evaluada a la hora de especificar los modelos de predicción de precios. Para incluir estas variables climáticas se debe tener en cuenta que si bien se espera que estos fenómenos climáticos afecten los precios, no lo harán en el momento 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 a su vez 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 (costo condicionado 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 también puede llegar a incidir en el precio es la cotización de manzana a nivel regional. Una desventaja de incluir los precios en la región como variable auxiliar es la dificultad para la identificación del momento temporal de medición de la variable para incluirla, debido a que el estudio de los precios de la región se considera en el momento puntual de la importación y no en el momento en que el producto importado estará a la venta.

Para este rubro fue posible encontrar bases de datos de variables auxiliares como ser: la producción, las horas de frío acumuladas y las precipitaciones, además del ingreso de mercadería al mercado mayorista.

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 el año 2021 según estimaciones del Observatorio Granjero. Hay dos zonas de producción principales en el país, siendo la primera el litoral norte (principalmente de Salto, Constitución y Bella Unión); que acumula aproximadamente 60% de la producción nacional y donde casi la totalidad se produce bajo invernáculo. La segunda zona de producción es la zona sur (Canelones, Montevideo, San José), donde se produce prácticamente el 40% restante bajo las modalidades invernáculo y "a campo". El cultivo en invernáculo va ganando año a año superficie, en la modalidad a campo se producen también 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 interno 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 junto a otras particularidades del proceso de importación, hacen que normalmente demore varios días hasta que finalmente llega la partida a la venta, aumentando el riesgo de pérdidas por afectación de la calidad.

El tomate es un artículo que presenta una particular sensibilidad a los cambios de clima (temperatura, precipitaciones e intensidad de la 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 generalmente debidos a 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, marcos temporales y valores exactos para definir la incidencia en las condiciones de cuajado, por lo que no son incluidas en el desarrollo del modelo de predicción. También se identificó como variable de interés para incluir la que refiere a la concentración de la producción, pero no hay datos históricos disponibles sobre este factor para poder incluirla en el modelo.

Modelado y predicción de series de tiempo

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 y Harrison (1997).

La estructura de un modelo de espacio-estado lineal puede describirse en dos procesos vinculados. El primero es un proceso de estado θ_t , el cual se supone que se comporta como una cadena de Markov. Luego, se tiene el proceso observado y_t , que depende solamente del proceso de estado y cuyos elementos son condicionalmente independientes dado el mismo. Esta estructura se representa en la Figura 1.

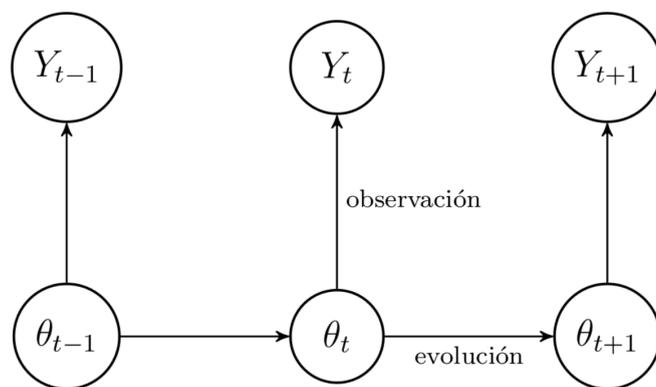


Figura 1: Estructura de dependencia de los DLM. Extraído de Hernández-Banadik et al. (2021).

Uno de los atractivos de esta clase de modelos es su flexibilidad, ya 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.

Para modelar la estacionalidad de los rubros 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 se emplean

términos de Fourier. Esto permite enfrentar las complicaciones que genera la frecuencia semanal de las series trabajadas, por ejemplo, al trabajar una estacionalidad anual (un año tiene 52,18 semanas aproximadamente, el periodo no entero dificulta el uso de las herramientas de modelado de series de tiempo).

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 plantean. 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.

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” (datos conocidos previamente registrados que simulan ser datos futuros a los efectos de la “prueba”).

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

Para el análisis presentado se hicieron uso de algunas de las métricas descritas en *Hyndman & Koehler (2006)*, en particular:

- RMSE: Raíz del error cuadrático medio.
- MAE: Error absoluto medio.
- MAPE: Error porcentual absoluto medio.

Dado que cuentan con características distintas, resulta de utilidad emplearlas en conjunto para la evaluación, para no realizar conclusión a partir de solo una de ellas.

Resultados

Descripción de los datos

Se cuenta con series digitalizadas de precios mayoristas por kilo e ingresos al mercado mayorista Mercado Modelo/UAM (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 de datos con la información que se releve en adelante e incorporarla dentro de los modelos. Las series de precios mayoristas originales tienen una frecuencia bisemanal, el relevamiento de precios es realizado por el Observatorio Granjero los lunes y los jueves,

días de mayor actividad en el Mercado Modelo/UAM (aproximadamente el 40% de los ingresos semanales se acumulan en esos días). Para dicho relevamiento de datos se considera la metodología detallada en el “Manual de procedimientos y referencias técnicas para la tipificación de la calidad” (CAMM, 2021) .

El traslado de la operativa del Mercado Modelo a la UAM en febrero del 2021, el cual a su vez fue coetáneo con la pandemia de COVID-19, pudo generar cambios en la captación de la información de volúmenes totales y de precios, si bien la metodología de relevamiento no cambió fundamentalmente. Ante este hecho deberá evaluarse si la situación amerita un cambio estructural en los modelos.

Se cuentan con series de precios para los distintos productos y sus variedades, distinguiéndose en cada caso tres calidades (en orden descendente): Categoría Extra, Categoría I y Categoría II. Para contar con una noción general del comportamiento de los precios mayoristas, los técnicos del Observatorio Granjero recurren a entrevistar un conjunto selecto de informantes calificados, que son vendedores mayoristas (operadores) del mercado, cuya información se considera fiable y representativa. Esta información se releva cuando 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 los mismos es deliberada y se toma en cuenta a aquellos considerados informantes calificados y con la complejidad agregada de considerar las variables cualitativas que diferencian a las diferentes partidas de frutas y hortalizas presentes en la oferta, las cuales también determinan cotizaciones diferentes.

Esta metodología de muestreo no probabilístico se enmarca dentro del 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, pp. 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. Estas características 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. Observando las series de precios de los rubros separando por calidad se puede apreciar que 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 a 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, ante lo cual se emplea la modelización de la estacionalidad mediante términos de Fourier.

Otras series disponibles para el periodo considerado son las de ingresos de los distintos rubros al Mercado Modelo y la UAM, es decir, el volumen 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 sistematizados, o si los había, eran pocas las observaciones en el tiempo lo que imposibilitaba la construcción de una serie. Además, fue un aspecto de suma dificultad definir con la exactitud necesaria los rezagos con los que las variables auxiliares afectan efectivamente a los precios (por ejemplo: la lluvia de qué momento exacto del año afecta los precios relevados).

Predicciones

Para los distintos rubros se ajustan los siguientes métodos y modelos: promedio simple, caminatas al azar simples (*Random walks*) y con regresores, modelo *SUTSE* (un modelo de caminata al azar bivariado, considerando precios e ingresos), *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. Adicionalmente, se realizó la descomposición de los precios en sus componentes (estacionalidad, tendencia/ciclo e irregular) haciendo uso de la metodología STR robusta, desarrollada en *Dokumentov & Hyndman (2022)*.

Para el procesamiento de datos, la estimación de los modelos y las visualizaciones se hizo uso del software *R* (*R Core Team, 2022*). En particular, para el modelado se hizo uso los paquetes *forecast* (*Hyndman et al., 2022*), *statespacer* (*Beijers, 2022*), y *dIrm* (*Petris, 2010*).

Manzana

La visualización de la descomposición STR robusta de los precios de manzana se presenta en la Figura 2. Destaca el elemento estacional (segundo gráfico de la figura), que refleja que en los años analizados 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.

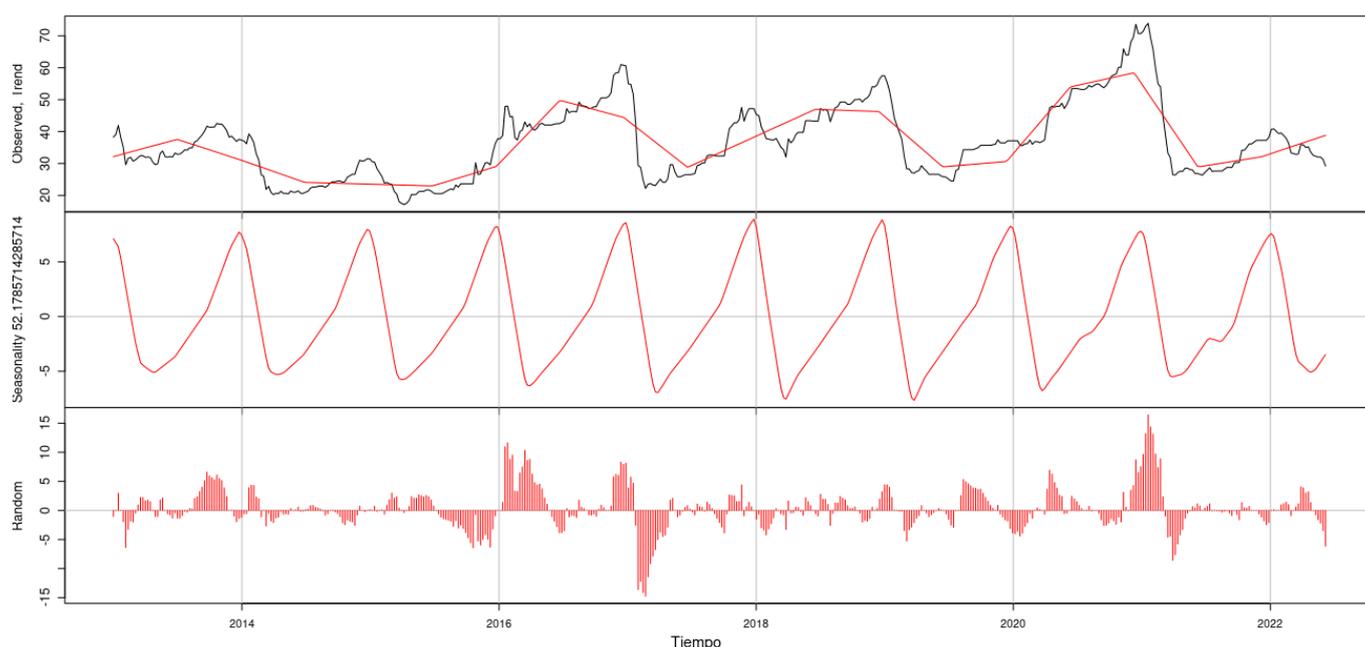


Figura 2: Descomposición STR robusta de la serie de precios de manzana.

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", esto se da en algunos cultivos frutales que tienen años con altos niveles de producción seguidos de años de bajo niveles de producción, debido al esfuerzo fisiológico del cultivo en el año de alta producción. 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 (tercer gráfico de la figura) como una perturbación al comportamiento sistemático de los precios, el efecto no resulta tan evidente como en el caso de otros rubros estudiados.

Para este rubro, el modelo por el cual se opta es el *TBATS*, debido a su mejor rendimiento a mediano plazo (12 semanas) y largo plazo (52 semanas) por sobre el resto de los modelos en términos de métricas de error. El modelo estimado presenta un elemento estacional de periodo anual, una transformación de *Box-Cox* de la variable estudiada y un componente de tendencia.

En la Figura 3 se presentan las predicciones realizadas con el modelo desarrollado, además se grafican los precios reales para comparar cómo se ajustan los valores predichos. Se puede apreciar que el modelo que utiliza para la

estimación la información de precios de enero de 2013 a junio de 2021, y que predice los precios desde junio de 2021 a junio de 2022, capta en general el comportamiento estacional de los precios. No obstante, en torno a marzo de 2022 ya comienza a no ser tan preciso, debido al amplio horizonte que se está trabajando y la incertidumbre que esto implica, aunque es importante destacar que las últimas dos zafas (2020/2021 y 2021/2022) parecerían haber mostrado comportamientos especialmente atípicos en cuanto a precios bajos, dificultando el ajuste del modelo.

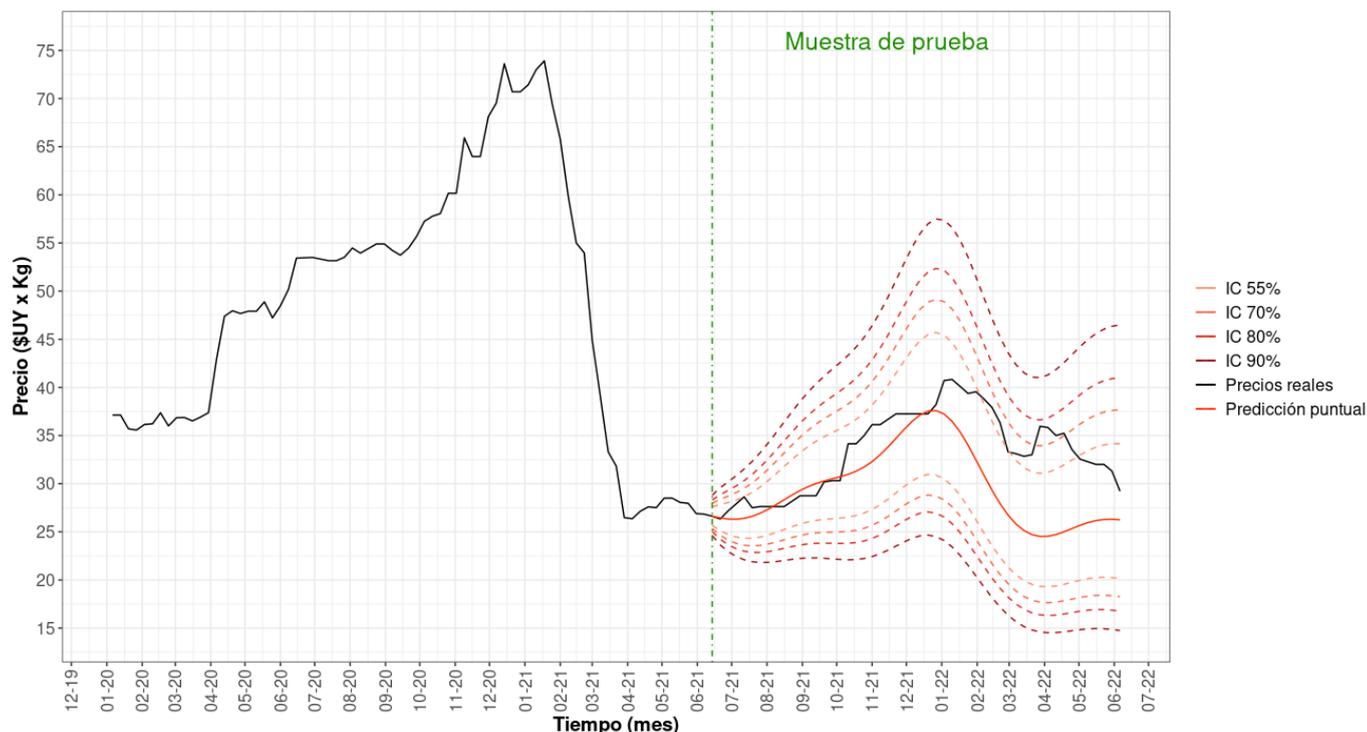


Figura 3: Predicción a 52 pasos. Manzana. TBATS.

Tomate

En la Figura 4 se puede observar la descomposición STR robusta de la evolución del precio del tomate. No hay tendencias globales que se puedan apreciar en primera instancia (parte superior de la figura). Además, la estacionalidad presenta un mayor grado de irregularidad que en el caso de la manzana (segunda parte de la figura). Cabe destacar que el año 2022 presenta precios irregularmente altos, lo cual se ve reflejado en el gráfico de residuos (tercera sección de la figura).

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, por lo que también se presentan sus resultados (Figura 4). Cabe destacar que este tipo de modelo capta el comportamiento estacional del rubro y permite el cálculo de intervalos de predicción.

El tomate es un rubro cuyos precios presentan una alta variabilidad dadas sus características productivas, lo cual se ve reflejado en la amplitud de los intervalos del modelo ARIMA + Fourier. El elevado pico de precios en marzo de 2022 afectó la precisión de las predicciones del modelo, lo cual sumado a que al comienzo del periodo predictivo hubo precios que se enmarcaron en el extremo inferior del intervalo de confianza son las posibles razones de que el promedio simple tenga un mejor desempeño que el modelo estocástico.

No obstante, el modelo ARIMA + Fourier capta la estacionalidad subyacente de la serie de precios, aunque de forma más rígida que un TBATS (dado que en el primero la estacionalidad es determinística, es decir, se considera que no cambia a lo largo de los años).

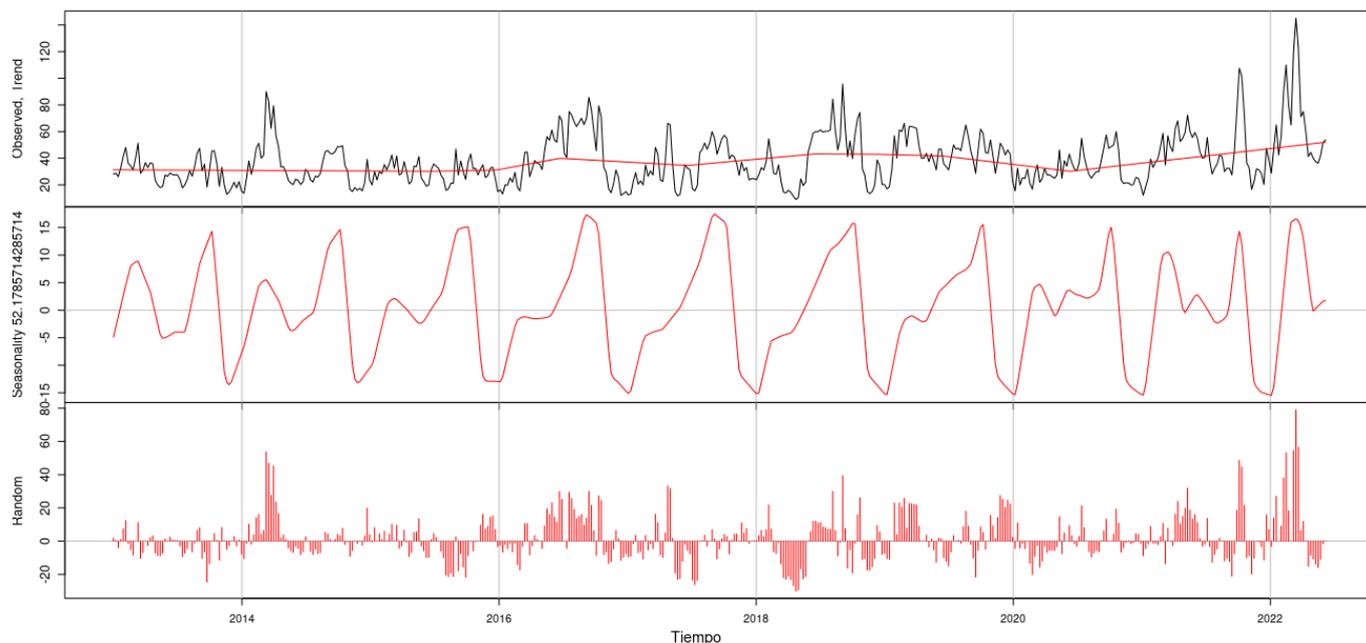


Figura 2: Descomposición STR robusta de la serie de precios de tomate.

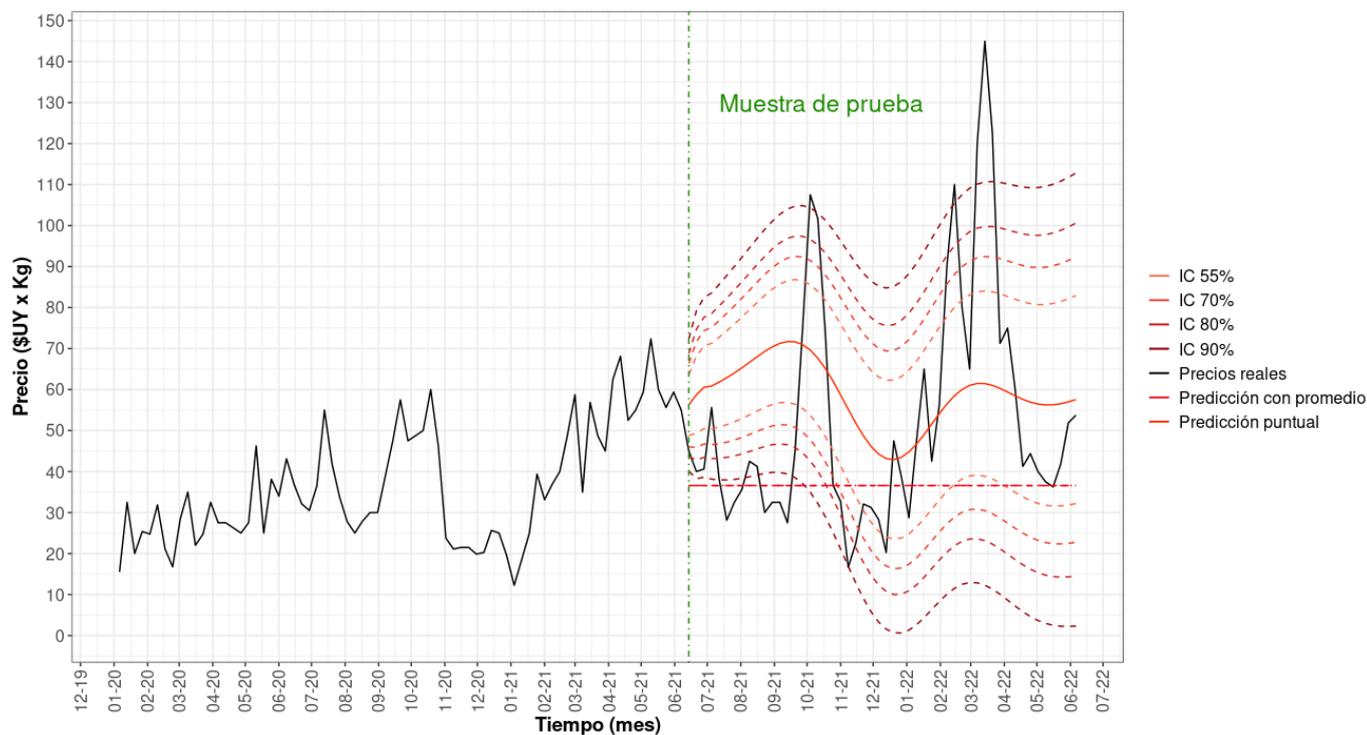


Figura 5: Predicción a 52 pasos. Tomate. ARIMA + Fourier y promedio.

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.

Para el caso de los rubros que se presentaron en esta publicación, la manzana es donde los modelos de espacio-estado estacionales planteados tienen en general el mejor desempeño a través de los distintos horizontes. En el caso del tomate, la predicción resulta de mayor dificultad debido al elevado grado de variabilidad que presentan sus precios.

La inclusión de variables auxiliares en las especificaciones de los modelos fue problemática. Incluir en los modelos el volumen de ingreso de mercadería al mercado de los distintos rubros resulta relativamente directo, aunque para la mayoría de los rubros no implicó una ganancia en precisión predictiva respecto a otros modelos planteados. 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, estas no determinaron una ganancia en poder predictivo.

Dado que los modelos de predicción pueden ser utilizados como herramienta para la toma de decisiones políticas es importante mencionar que estos requieren ser interpretado por personal técnico con conocimiento del funcionamiento del mercado hortifrutícola, y de las limitaciones de las predicciones para no realizar una sobreinterpretación de los resultados.

Además el uso de los resultados de las predicciones debe ser utilizada de forma reservada ya que la difusión de esta información puede afectar la formación de precios del mercado.

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

Bibliografía

Beijers, D. (2022). statespacer: State Space Modelling in R. R package version 0.4.1.

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

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

Dokumentov, A., & Hyndman, R. J. (2022). STR: Seasonal-Trend Decomposition Using Regression. *INFORMS Journal on Data Science*, 1(1), 50-62. <https://doi.org/10.1287/ijds.2021.0004>

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.
- Harvey, A. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.
- Hyndman, R., & Athanasopoulos, G. (2021). *Forecasting: principles and practice*. Otexts.
- Hyndman, R., & Koehler, Anne B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22, 679-688.
- Hyndman R, Athanasopoulos G, Bergmeir C, Caceres G, Chhay L, O'Hara-Wild M, Petropoulos F, Razbash S, Wang E, Yasmeeen F (2022). *forecast: Forecasting functions for time series and linear models*. R package version 8.18, <https://pkg.robjhyndman.com/forecast/>.
- 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.
- Observatorio Granjero. (2019). Anuario estadístico 2018. Mercado Modelo.
- Observatorio Granjero. (2022). Anuario estadístico 2021. Unidad Agroalimentaria Metropolitana.
- Petris, G. (2010). An R Package for Dynamic Linear Models. *Journal of Statistical Software*, 36(12), 1-16. URL <http://www.jstatsoft.org/v36/i12/>.
- Petris, G., Petrone, S., & Campagnoli, P. (2009). *Dynamic Linear Models With R*. Springer Science + Business Media.
- R Core Team (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>.
- 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.