¿Anticipa el satélite al precio? Nowcasting semanal de precios del tomate en Chile

Documento de trabajo — resultados preliminares. Proyecto Barlovento

Verónica Velásquez

Agosto 2026

Resumen

Este documento evalúa si un modelo estacional de series de tiempo puede anticipar el precio semanal del tomate en Chile con utilidad práctica, y presenta evidencia preliminar sobre si una señal satelital de vigor vegetal (NDVI) sobre las zonas productoras aporta información adicional. Sobre 2.504 observaciones semanales de precios ODEPA al consumidor (2008–2026) y un esquema de validación walk-forward estricto, un SARIMAX(1,1,1)(1,1,1,52) reduce el MAPE en 36,8% respecto de un paseo aleatorio en la Región Metropolitana (4,37% contra 6,91%; Diebold-Mariano p = 0,026), con una ventaja que crece con el horizonte: a cuatro semanas el error del modelo es menos de la mitad del benchmark. El resultado no es uniforme: en Maule la mejora es de 11,0% y no alcanza significancia convencional, y en Valparaíso el modelo estacional es batido por el paseo aleatorio. Un ejercicio exploratorio de señal satelital sobre precios mayoristas encuentra una correlación de anomalías de $-0{,}40$ a quince semanas de rezago ($R^2 = 0{,}19$), con el signo que predice el mecanismo de oferta. Esa señal aún **no** ha sido integrada al modelo de pronóstico; el criterio de incorporación queda pre-registrado en la sección final.

Índice general

1 Motivación **3**

2 Datos **3**

3 Estrategia empírica **5**
3.1 Modelo 5
3.2 Validación 5
3.3 Benchmarks 5
3.4 Contraste estadístico 6

4 Resultados **6**
4.1 El modelo bate a los benchmarks, pero no en todas partes 6
4.2 La ventaja crece con el horizonte 7
4.3 Diagnóstico: por qué el modelo falla donde falla 8

5 La capa satelital: evidencia preliminar **9**
5.1 Identificación del cultivo 9
5.2 Contenido informativo sobre el precio 10
5.3 Consecuencia operativa 11

6 Limitaciones **11**

7 Agenda y criterio pre-registrado **12**

8 Anexo: reproducibilidad **12**

1 Motivación

El componente alimentario pesa cerca de un quinto de la canasta del IPC chileno y es, sistemáticamente, su fuente principal de volatilidad de corto plazo. Sin embargo, la información de precios se publica con rezago: cuando el dato oficial está disponible, la decisión que dependía de él —de política monetaria, de compra, de cobertura, de planificación agroindustrial— ya se tomó a ciegas.

Existe una asimetría explotable. El precio de un producto agrícola fresco es, en buena medida, el resultado de una oferta que se estuvo formando físicamente en el campo durante los meses previos, y esa formación es observable desde el espacio antes de llegar al mercado. La pregunta que organiza este proyecto es si esa observación tiene contenido informativo suficiente para mejorar un pronóstico, y no solo para ilustrarlo.

Este documento reporta el estado de dos preguntas separadas y secuenciales:

1. **¿Sirve el modelo base?** Un modelo estacional de series de tiempo sobre precios, ¿mejora de forma medible y estadísticamente distinguible a los benchmarks ingenuos que un pronosticador razonable usaría sin esfuerzo? Esta pregunta está respondida.
2. **¿Aporta el satélite?** Una vez fijada esa vara, ¿la señal satelital de las zonas productoras la supera? Esta pregunta tiene evidencia preliminar favorable pero **no está respondida**, y la sección 7 fija el criterio con el que se responderá, antes de conocer el resultado.

El producto elegido como caso de estudio es el tomate. La razón es metodológica, no comercial: el tomate tiene un ciclo fenológico anual marcado —siembra, crecimiento, fructificación, cosecha— y por lo tanto una firma temporal distinguible desde el satélite. Un cultivo perenne y siempre verde, como el cítrico o el palto, presenta una curva plana y es mucho más difícil de aislar de la vegetación circundante.

2 Datos

Precios. Serie semanal de precios al consumidor de ODEPA (Oficina de Estudios y Políticas Agrarias, Ministerio de Agricultura), variedad “Tomate larga vida, primera calidad”, 2008–2026. El panel consolidado tiene 2.504 registros. Se modelan las tres series regionales con historia suficiente: Región Metropolitana, Valparaíso y Maule. La serie se indexa a frecuencia semanal con lunes como inicio (`W-MON`); las semanas sin registro se mantienen como valores faltantes, que el filtro de Kalman maneja de forma nativa, en lugar de imputarse.

Cuadro 1: Series de precios utilizadas. El porcentaje faltante corresponde a semanas del calendario sin registro publicado.

Serie	Semanas	Observadas	% faltante	Media (CLP/kg)	CV	Inicio
R. Metropolitana	956	854	10,7%	1.042,5	0,343	2008-03-17
Valparaíso	627	532	15,2%	1.243,2	0,240	2014-07-07
Maule	680	581	14,6%	1.152,6	0,273	2013-07-01

Señal satelital. Sentinel-2 (colección `COPERNICUS/S2_SR_HARMONIZED`) procesada en Google Earth Engine, con máscara de nubes por píxel vía la capa de clasificación de escena. El índice base es el NDVI, calculado como $(NIR - RED)/(NIR + RED)$ y agregado a promedio zonal semanal sobre la misma grilla temporal de los precios.

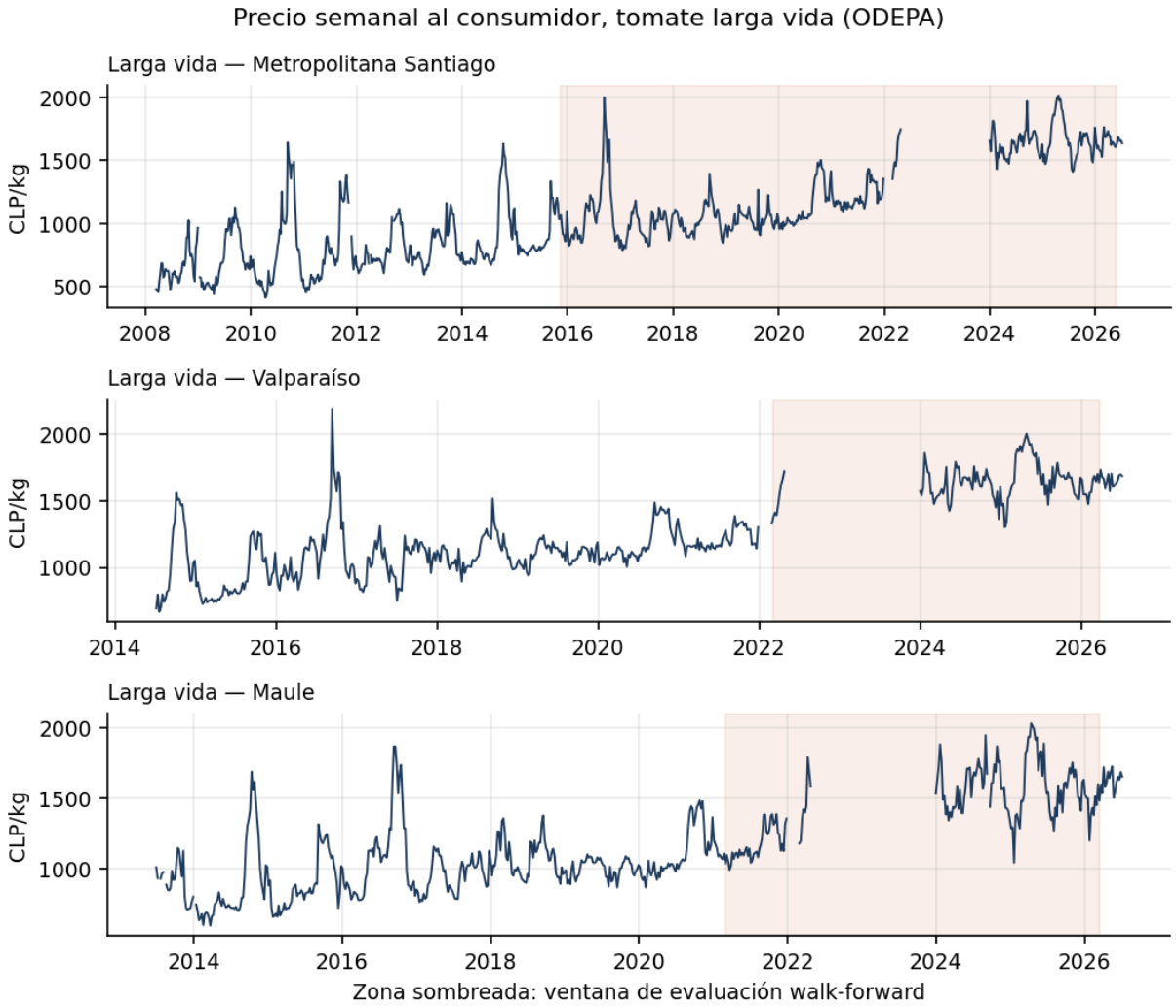


Figura 1: Series semanales de precio al consumidor. La zona sombreada marca la ventana de evaluación walk-forward.

3 Estrategia empírica

3.1 Modelo

El modelo base es un SARIMAX$(1,1,1)(1,1,1)_{52}$ con constante, estimado por máxima verosimilitud en representación de espacio de estados. La periodicidad estacional de 52 recoge el ciclo anual de la producción hortícola de campo abierto en el hemisferio sur. No se incluyen regresores exógenos en esta etapa: el modelo del que se reporta evidencia aquí es deliberadamente el más simple que puede capturar la estacionalidad, para que cualquier mejora posterior atribuida al satélite sea atribuible al satélite y no a un cambio simultáneo de especificación.

3.2 Validación

Todo lo reportado proviene de un backtest walk-forward estricto: en cada origen t el modelo se estima usando exclusivamente información hasta t y pronostica $t+1, \dots, t+4$. La ventana inicial de entrenamiento es de 400 semanas y los orígenes avanzan cada 26 semanas. Como el paso entre orígenes (26) excede el horizonte de pronóstico (4), las ventanas de evaluación no se solapan, lo que deja los errores de orígenes distintos aproximadamente independientes —condición que hace limpio el contraste estadístico de la sección siguiente.

El pipeline incorpora una guardia explícita contra filtración de información: cualquier matriz de variables exógenas pasa por un filtro que descarta el nivel contemporáneo del precio y cualquier columna derivada de valores futuros del objetivo. Esta guardia se introdujo tras una auditoría que detectó dos filtraciones silenciosas en una versión anterior del pipeline, y está cubierta por tests automatizados.

3.3 Benchmarks

La pregunta "¿es bueno un MAPE de 4,4%?" no tiene respuesta en el vacío. Se compara el modelo contra cinco alternativas, todas estimadas sobre exactamente los mismos orígenes y evaluadas sobre exactamente los mismos objetivos:

- **Naive (paseo aleatorio):** $\widehat{y}_{t+h} = y_t$. El benchmark que la literatura de pronóstico de precios rara vez logra batir a horizontes cortos.
- **Naive estacional:** $\widehat{y}_{t+h} = y_{t+h-52}$.
- **Drift:** paseo aleatorio con deriva estimada sobre toda la ventana de entrenamiento.
- **Climatología:** media histórica de la misma semana calendario dentro de la ventana.
- **ARIMA(1,1,1) sin componente estacional:** aísla cuánto del desempeño viene de la estacionalidad y cuánto de la dinámica de corto plazo.

El benchmark ARIMA se estima imponiendo estacionariedad e invertibilidad. Sin esa restricción el optimizador se desplazó a raíces explosivas en dos orígenes de Maule, produciendo pronósticos de orden 10^8 que habrían inflado artificialmente

la ventaja del modelo. Un benchmark debe ser competente; un hombre de paja no informa nada.

3.4 Contraste estadístico

Se aplica el test de Diebold-Mariano sobre las diferencias de pérdida, con corrección de Harvey, Leybourne y Newbold para muestras pequeñas y varianza de largo plazo tipo Newey-West. Se reportan resultados bajo pérdida cuadrática y absoluta. Un estadístico negativo indica que el SARIMAX pierde menos que el benchmark.

4 Resultados

4.1 El modelo bate a los benchmarks, pero no en todas partes

Cuadro 2: Error de pronóstico por modelo y serie, agregado sobre todos los horizontes (1 a 4 semanas). MASE escalado por el error del naive estacional dentro de la muestra de entrenamiento; valores bajo 1 indican mejora sobre esa referencia.

Serie	Modelo	MAPE (%)	RMSE	MASE
R. Metropolitana	**SARIMAX**	**4,37**	**64,3**	**0,308**
	Naive (RW)	6,91	100,7	0,462
	ARIMA(1,1,1)	7,06	102,7	0,473
	Drift	7,08	103,2	0,474
	Naive estacional	10,86	185,4	0,838
	Climatología	26,75	450,3	2,262
Maule	**SARIMAX**	**7,94**	**146,4**	**0,621**
	ARIMA(1,1,1)	8,64	163,0	0,689
	Drift	8,86	172,2	0,709
	Naive (RW)	8,92	173,1	0,714
	Naive estacional	10,74	209,3	0,824
	Climatología	25,44	420,1	2,033
Valparaíso	ARIMA(1,1,1)	5,07	111,3	0,501
	Drift	5,21	112,6	0,513
	Naive (RW)	5,26	114,8	0,519
	SARIMAX	*6,16*	*127,8*	*0,599*
	Naive estacional	12,05	220,0	1,157
	Climatología	27,98	477,3	2,710

En la Región Metropolitana —la serie más larga, la de mayor volumen de mercado y la más relevante para el IPC— el modelo reduce el MAPE en **36,8%** respecto del paseo aleatorio y el RMSE en 36,2%. El MASE de 0,308 indica que el error es menos de un tercio del que produciría repetir el valor del año anterior.

En Maule la mejora existe pero es modesta: 11,0% sobre el paseo aleatorio. En Valparaíso el resultado es negativo: el SARIMAX es **peor** que el paseo aleatorio (6,16% contra 5,26%). Reportarlo importa. Una tabla en la que el modelo propio

gana en las tres series sería más vendible y menos creíble; la heterogeneidad tiene además una explicación diagnosticable, que se desarrolla en la sección 4.3.

Cuadro 3: Test de Diebold-Mariano bajo pérdida cuadrática, SARIMAX contra cada benchmark, agregando horizontes. Estadístico negativo = SARIMAX pierde menos. p-valores entre paréntesis.

Serie	vs. Naive	vs. Drift	vs. ARIMA	vs. Naive est.	vs. Climatología
R. Metropolitana	−2,27 (0,026)	−2,39 (0,019)	−2,40 (0,019)	−3,17 (0,002)	−3,24 (0,002)
Maule	−2,02 (0,052)	−1,98 (0,057)	−1,09 (0,285)	−1,38 (0,181)	−3,81 (0,001)
Valparaíso	+0,83 (0,418)	+0,96 (0,347)	+0,97 (0,343)	−1,82 (0,089)	−5,73 (0,000)

En la Región Metropolitana la superioridad es estadísticamente significativa al 5% contra todos los benchmarks. En Maule queda al borde del 5% frente a los benchmarks ingenuos y no se distingue del ARIMA no estacional —es decir, allí la estacionalidad anual no está aportando. En Valparaíso el signo se invierte, sin significancia.

4.2 La ventaja crece con el horizonte

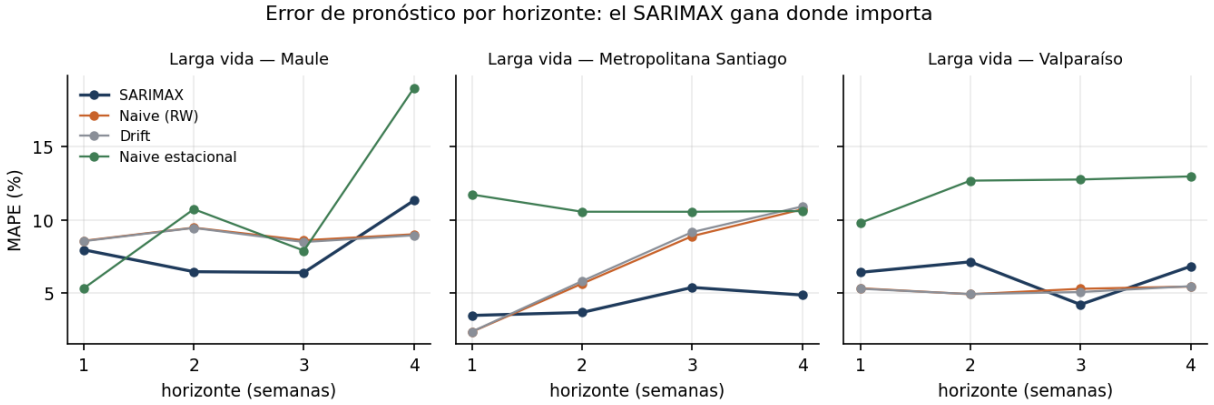


Figura 2: MAPE por horizonte de pronóstico. En Región Metropolitana el paseo aleatorio es imbatible a una semana, y la ventaja del modelo estacional aparece y se amplía al alejarse el horizonte.

La descomposición por horizonte revela la estructura del resultado y es, probablemente, el hallazgo más interesante del ejercicio. En la Región Metropolitana, a una semana de distancia el paseo aleatorio empata o supera al modelo: el precio de la próxima semana se parece demasiado al de esta semana como para que la estructura estacional agregue algo. El estadístico de Diebold-Mariano a $h = 1$ es de hecho positivo y no significativo (+0,92, p = 0,373). A cuatro semanas la situación se invierte por completo: el MAPE del modelo es 4,88% contra 10,72% del paseo aleatorio, con DM = −3,56 (p = 0,002).

Esto tiene una lectura económica directa. El valor del modelo no está en decir qué pasará la próxima semana —eso lo hace igual de bien no hacer nada—, sino en sostener la proyección cuando el horizonte se aleja lo suficiente como para que la

persistencia del nivel deje de servir. Es exactamente el rango donde una decisión de compra, cobertura o planificación tiene tiempo de ejecutarse.



Figura 3: Pronósticos walk-forward de cuatro semanas contra el precio observado, Región Metropolitana. Cada segmento corresponde a un origen distinto, con el modelo reestimado usando solo información previa.

4.3 Diagnóstico: por qué el modelo falla donde falla

Cuadro 4: Diagnóstico econométrico por serie. ADF: H_0 = raíz unitaria. KPSS: H_0 = estacionariedad. Ljung-Box sobre residuos del modelo ajustado; H_0 = ausencia de autocorrelación. p-valores entre paréntesis.

Serie	ADF nivel	KPSS nivel	ADF Δ_{52}	Ljung-Box (12)	Ljung-Box (52)
R. Metropolitana	$-3{,}44$ (0,010)	3,43 (<0,01)	$-7{,}63$ (0,000)	12,59 (0,400)	64,00 (0,123)
Valparaíso	$-2{,}89$ (0,046)	2,48 (<0,01)	$-3{,}91$ (0,002)	18,39 (0,104)	67,59 (0,072)
Maule	$-2{,}31$ (0,168)	2,52 (<0,01)	$-6{,}00$ (0,000)	32,95 (0,001)	93,59 (0,000)

El diagnóstico es coherente con el ranking de desempeño y contiene tres advertencias honestas:

Los residuos son ruido blanco solo en la Región Metropolitana. Ljung-Box no rechaza a ningún rezago (p = 0,40 a 12 semanas; p = 0,12 a 52), lo que indica que la especificación agotó la estructura lineal disponible. En Maule el test rechaza contundentemente a todos los rezagos (p < 0,001), señal inequívoca de que queda estructura sin modelar: el modelo está mal especificado para esa serie, y su ventaja marginal sobre los benchmarks debe leerse con esa reserva.

Hay indicios de sobrediferenciación. En la Región Metropolitana el parámetro de media móvil no estacional se estima en el borde del círculo unitario ($\hat{\theta}_1 = -1,00$), patrón clásico de una diferencia de más. En Valparaíso el término estacional de media móvil se estima fuera del círculo ($\hat{\Theta}_{52} = -1,23$, no invertible) y ni el componente autorregresivo ni la constante son significativos. Dicho de otro modo: el modelo estacional en Valparaíso está estimando estacionalidad que la serie no tiene con suficiente nitidez, y por eso pierde contra el paseo aleatorio. La serie de

Valparaíso es la de menor coeficiente de variación de las tres (0,240): tiene menos que explicar.

Los intervalos están mal calibrados por exceso. La cobertura empírica del intervalo de predicción nominal de 95% es de 100,0% en Región Metropolitana (72 de 72), 93,5% en Maule y 87,5% en Valparaíso. La sobrecobertura de la serie principal significa que los intervalos son demasiado anchos: el modelo es más preciso de lo que declara, lo que resulta inútil para cualquier uso que requiera cuantificar riesgo. Corregir esto es requisito previo a cualquier aplicación de cobertura o de decisión bajo incertidumbre.

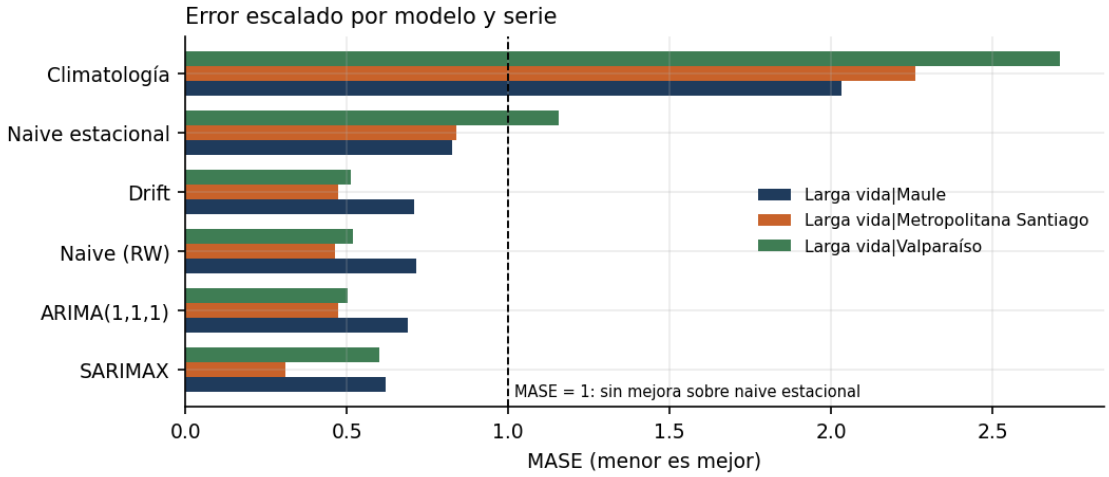


Figura 4: Error escalado (MASE) por modelo y serie.

5 La capa satelital: evidencia preliminar

Esta sección reporta resultados de un ejercicio exploratorio previo, documentado en la bitácora del proyecto (*probes* #18 y #19, julio 2026). Se presenta separado del resto por dos razones: usa una serie de precios distinta —mayorista, no al consumidor— y su resultado **no** ha pasado todavía por el criterio de evaluación del proyecto.

5.1 Identificación del cultivo

El obstáculo central no es calcular NDVI, sino saber qué píxeles son tomate. Una escena sobre Quillota contiene tomate, lechuga, paltos, viñas y maleza simultáneamente verdes; el promedio indiscriminado diluye la señal hasta hacerla inservible. El tomate, además, no existe como polígono parcelario en ninguna fuente cartográfica oficial chilena: el Censo Agropecuario 2021 llega solo a nivel comunal y no se actualizará hasta cerca de 2031.

La estrategia adoptada combina dos filtros: el censo acota la zona geográfica, y una **firma fenológica** identifica, dentro de esa zona, los píxeles cuya trayectoria anual de NDVI tiene la forma del ciclo del tomate —ascenso desde la siembra, máximo en el verano austral, caída abrupta en la cosecha. La restricción se aplica sobre la

serie temporal completa de cada píxel, no sobre una imagen aislada, que es lo que permite separar el tomate de una viña o de un palto igualmente verdes en enero.

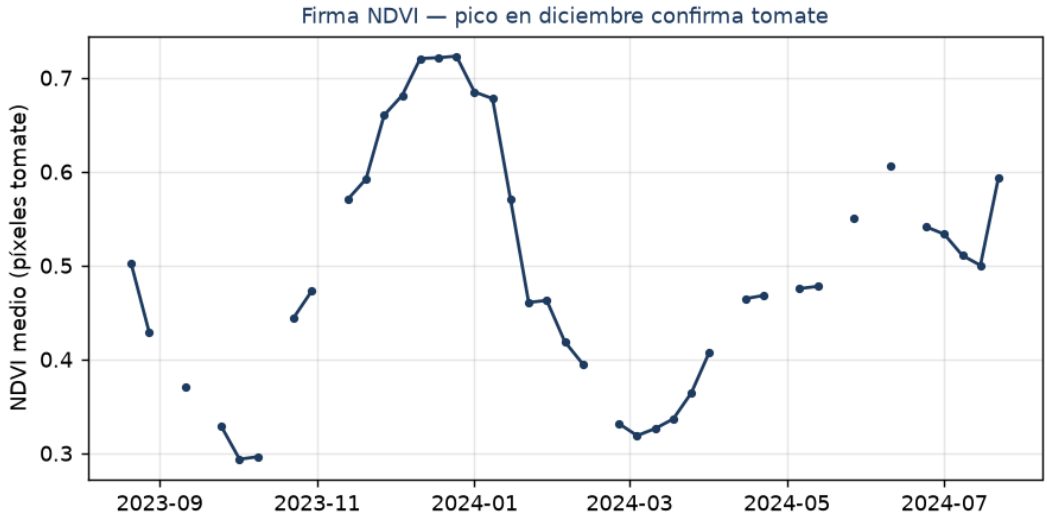


Figura 5: Curva NDVI media de los píxeles seleccionados por la firma fenológica, Quillota, temporada 2023-24. El máximo en diciembre corresponde al pico vegetativo esperado del tomate de campo abierto. El hombro de invierno (mayo-julio, NDVI ≈ 0,5) es contaminación residual identificada y no corregida.

La curva resultante (figura 5) reproduce el calendario agronómico esperado: mínimo en torno a la siembra, máximo de 0,72 en diciembre, descenso pronunciado entre febrero y marzo coincidente con la cosecha. La hipótesis de identificación se sostiene: la firma temporal distingue al cultivo.

5.2 Contenido informativo sobre el precio

Se contrastó la anomalía de NDVI —desviación respecto de la media de la misma semana calendario— contra la anomalía del precio mayorista de tomate semiduro en la Región Metropolitana, en la ventana de solape 2017–2025. Se eligió el precio mayorista y no el minorista por su mayor proximidad temporal a la cosecha.

Cuadro 5: Correlación cruzada entre anomalía de NDVI y anomalía de precio mayorista, según calidad de la máscara de cultivo.

Máscara	Zona	máx. |corr|	Rezago	R^2	n
v1 (cruda)	39.000 km², umbrales laxos	0,158	14 sem	0,042	210
v2 (pulida)	**Quillota + Aconcagua, 7.307 km²**	**0,397**	**15 sem**	**0,188**	**180**

Dos elementos hacen que este resultado merezca continuidad. Primero, **el signo es el correcto y no fue impuesto**: la correlación es negativa, es decir, un NDVI de producción por sobre lo normal precede a un precio por debajo de lo normal. Es el mecanismo de oferta que la teoría predice. Segundo, **el rezago es coherente con la agronomía**: el máximo de correlación aparece a quince

semanas, aproximadamente tres meses y medio, que es el orden de magnitud del trayecto entre vigor de campo y llegada a mercado.

El salto de R^2 entre la máscara cruda y la pulida (0,042 a 0,188) es consistente con la interpretación de que la señal es real y estaba diluida por error de clasificación, no creada por el refinamiento. Corresponde señalar, no obstante, que ambos umbrales de máscara se eligieron observando la propia señal, de modo que el R^2 de 0,188 debe leerse como una estimación optimista dentro de la muestra: la validación limpia requiere aplicar los umbrales fijados a una ventana no usada para elegirlos.

5.3 Consecuencia operativa

El rezago de quince semanas tiene una implicación concreta para la integración: la especificación de variables exógenas del pipeline construye rezagos de 1 y 4 semanas, que habrían pasado por completo al lado de la señal. La configuración se actualizó a rezagos de 4, 12 y 16 semanas antes de cualquier prueba de integración. Este es el tipo de detalle que decide entre encontrar un efecto y concluir erróneamente que no existe.

6 Limitaciones

Se enumeran sin atenuantes, porque determinan qué se puede y qué no se puede afirmar.

1. **El resultado central del proyecto no está probado.** La señal satelital tiene correlación con el precio, pero correlación no es mejora de pronóstico. Un regresor puede correlacionar y aun así no reducir el error fuera de muestra. Ese contraste está pendiente.
2. **Pocos orígenes de evaluación.** 22, 11 y 9 orígenes respectivamente. Los intervalos de confianza de los p-valores reportados son anchos; el resultado de la Región Metropolitana es robusto en magnitud, los de Maule y Valparaíso deben tratarse como indicativos.
3. **Ventana satelital corta.** Sentinel-2 opera desde aproximadamente 2017, lo que deja nueve años de solape con precios. Con estacionalidad de 52 semanas y ventana inicial de 400, el número de orígenes disponibles para un backtest del modelo con NDVI es reducido. Esto obliga a un diseño de evaluación específico para esa prueba, y a reportar el modo.
4. **Esquema de estimación.** Por restricciones de cómputo, los parámetros se reestiman completamente cada cuatro orígenes (aproximadamente cada dos años) y entre reestimaciones el estado se actualiza mediante filtro de Kalman con parámetros fijos. Una comparación puntual contra reestimación completa en cada origen arrojó diferencias de pronóstico del orden de 0,3% a 1,0% del nivel, pero la equivalencia no fue verificada de manera sistemática sobre todos los orígenes.
5. **Un solo producto y una sola variedad.** Todo lo anterior es sobre tomate larga vida. La extrapolación a la canasta alimentaria completa no está justificada por

esta evidencia.

6. **Máscara satelital sin validación de terreno.** No se dispone de coordenadas de campos verificados como tomate. La firma fenológica es consistente con el calendario agronómico, pero no se ha medido su tasa de error de clasificación.

7 Agenda y criterio pre-registrado

El siguiente experimento del proyecto es la integración del NDVI como variable exógena sobre el precio mayorista de tomate semiduro en la Región Metropolitana, con rezagos de 12 a 16 semanas, comparando el modelo base contra el modelo base más NDVI sobre la misma ventana y los mismos orígenes.

El criterio de decisión se fija aquí, antes de conocer el resultado: la señal satelital se incorpora si y solo si reduce el error fuera de muestra respecto del modelo base bajo el mismo protocolo de validación. Una correlación favorable no basta. Si el NDVI correlaciona pero no mejora el pronóstico, ese es en sí mismo un hallazgo que se reportará como tal, y la línea de investigación cambia de dirección en lugar de buscar la especificación que produzca el resultado deseado.

Trabajo posterior, en orden de prioridad:

1. Recalibración de los intervalos de predicción, cuya sobrecobertura es hoy el defecto más serio del modelo para uso en decisiones.
2. Reespecificación para Maule y Valparaíso: la evidencia de sobrediferenciación y autocorrelación residual sugiere que la especificación única para las tres regiones no es adecuada.
3. Validación fuera de muestra de los umbrales de la máscara fenológica, fijados con información de la propia señal.
4. Incorporación de shocks climáticos georreferenciados, que a diferencia del NDVI admiten una interpretación causal más limpia por su carácter plausiblemente exógeno.

8 Anexo: reproducibilidad

Todos los resultados de las secciones 4 y 5 se generan con `experiments/benchmark_suite.py` (backtest y contrastes), `experiments/refit_arima.py` (benchmark ARIMA con raíces restringidas), `experiments/diagnosticos.py` (raíz unitaria y residuos) y `experiments/figuras.py`. Las salidas quedan en `experiments/results/`: predicciones por origen y horizonte, tabla de métricas, tabla de Diebold-Mariano y diagnósticos en formato JSON. La configuración del modelo, la ventana de entrenamiento, el paso entre orígenes y los umbrales de la máscara satelital residen en `config.yaml`, sin parámetros incrustados en el código.

Los resultados de la sección 6 provienen de la bitácora de experimentos del proyecto y no fueron regenerados para este documento, por requerir acceso autenticado a Google Earth Engine.