

# ÍNDICE DE PRECIOS AGROPECUARIOS IPAP-BMC

Julio 2022

Metodología

La Gerencia Corporativa de Análítica y Estudios Económicos de la Bolsa Mercantil de Colombia presenta en este documento la metodología utilizada en la generación de un nuevo índice de precios agropecuarios. Considerando que los insumos de este indicador usan fuentes de información inéditas, este trabajo representa un complemento indispensable a los indicadores de precios agropecuarios en Colombia.

## Índice de Precios Agropecuarios IPAP-BMC

### EL ÍNDICE DE PRECIOS DE LA BOLSA MERCANTIL DE COLOMBIA

La BMC, como parte de un proyecto para posicionarse como líder en información y análisis de tendencias en el sector agropecuario, ha preparado un nuevo Índice de Precios Agropecuarios (IPAP-BMC). Este indicador presenta una nueva perspectiva para considerar variaciones de precio en los precios agropecuarios a nivel nacional, aprovechando la información disponible en el BMC. Estos datos ofrecen, entonces, una perspectiva única sobre las variaciones de los precios.

El índice se calcula usando la metodología de Laspeyres y tomando como base para las ponderaciones la estructura de la producción nacional de un año base de acuerdo con la Encuesta Nacional Agropecuaria de 2014, lo que es equivalente a usar las ponderaciones del IPP agropecuario del DANE. Un buen indicador de la robustez del IPAP-BMC es que está altamente correlacionado con otros indicadores de precios, como el IPC de alimentos, el IPC de bienes no durables y el IPP agropecuario.

La fuerte relación entre el IPAP-BMC y otros indicadores de actividad económica en el sector reflejan un resultado técnicamente sólido y acorde con las tendencias generales. Los resultados están altamente correlacionados con el fenómeno de El Niño, la tasa de cambio y el ISE agropecuario. De la misma forma, encontramos una correlación alta entre el IPAP-BMC y otros indicadores de precios como el IPP agropecuario, el IPC de alimentos, y el IPC de bienes no durables.

El principal valor agregado del índice proviene de su fuente de información: el registro de facturas. A diferencia de otros indicadores disponibles para el sector agropecuario, que tienen mecanismos de recolección de información con base en entrevistas y encuestas en plazas de mercado de diferentes ciudades del país, este índice presenta como gran ventaja la posibilidad de acceder a los datos con mayor inmediatez y menor rezago en su registro.

### ¿QUÉ ES EL REGISTRO DE FACTURAS?

El registro de facturas es el sistema que incluye las transacciones de productos agropecuarios y agroindustriales realizadas fuera del escenario de negociación de la BMC y cuyo mecanismo contribuye a formalizar el sector, construir precios de referencia y proporcionar estadísticas del mercado. Estos registros permiten al vendedor de bienes o productos de origen agrícola, pecuario y/o pesquero, sin procesamiento industrial o con transformación industrial primaria, optimizar su flujo

de caja mediante un beneficio tributario establecido por la ley<sup>1</sup>. El registro es realizado por Firmas Comisionistas a través del Sistema de Información Bursátil de la Bolsa Mercantil de Colombia.

El mecanismo para registrar facturas es el siguiente: el productor o empresa vende sus productos generando una factura. Luego, a través de una sociedad comisionista de bolsa, la factura se registra en la BMC. La BMC emite un comprobante de registro para el comprador y vendedor. Al momento del pago, según el caso, se da aplicación al decreto 1555 de 2017 (postergación del pago de retención en la fuente). Finalmente, toda la información de la factura se almacena en el Sistema de Información Bursátil de la BMC.

El principal beneficio del registro de facturas para el vendedor es la postergación del pago de retención en la fuente. De esta forma, el vendedor puede tener mayor liquidez en el corto plazo, un alivio tributario y la liberación de su capital de trabajo. La idea principal de este registro es generar mayor formalidad en el sector agropecuario, y simultáneamente generar una mayor información en el sector.

La base de datos es una valiosa fuente de información: cuenta con 50 millones de facturas registradas entre 2006 y 2022, con más de 2.000 productos padres y 10.000 productos caracterizados por diferencias de empaque y calidad. De estos productos, 98% han sido codificados a la nomenclatura internacional CPC.

## DESAFÍOS DEL REGISTRO DE FACTURAS

Dada la particularidad de la data del registro de facturas de la Bolsa Mercantil de Colombia, es natural encontrar algunos desafíos a la hora de procesar los datos. A diferencia de otras bases de datos, estos desafíos no afectan exclusivamente la data histórica, sino que también pueden afectar la información registrada en un futuro y, por lo tanto, podrían afectar al índice como lo presentamos actualmente. Para enfrentar estos desafíos, proponemos aquí usar técnicas econométricas y de ciencia de datos que explicamos con más detalle más adelante.

Al procesar los datos del registro de facturas se resaltan las diferencias en la forma en que se recolectan los datos frente a aquella utilizada por otros indicadores de precios. En primer lugar, la disponibilidad de datos depende de que las empresas decidan o no registrar sus facturas. Hay diversos factores exógenos que determinan esta decisión, pero no los describiremos en este documento por tratarse de un tema complejo en sí mismo y que requiere una explicación detallada. Lo importante es resaltar que puede que en un mes ciertos productos agropecuarios no tengan registros, o que productos que se habían registrado en el pasado dejen de

---

<sup>1</sup> Para mayores detalles, visite <https://www.bolsamercantil.com.co/Productos/RegistrodeFacturas.aspx>

registrarse en un momento dado (y viceversa, que productos de los que no teníamos registros empiecen a registrarse). Adicionalmente, el registro de algunos productos podría disminuir y generar vacíos de información, limitar la trazabilidad y moderar la robustez de los precios que son el insumo fundamental del índice. Lo esencial es, entonces, escoger las series de precios históricamente más sólidas, y prever un modelo que permita identificar este valor si se trata de un único mes sin datos.

Otras complicaciones vienen de la heterogeneidad de los productos registrados en la Bolsa Mercantil de Colombia. En algunos casos, el tipo de empaque y las cantidades registradas del producto inciden directamente en los precios (por dar un ejemplo, una botella de gaseosa pequeña no tiene el mismo precio por mililitro que una botella de gaseosa tamaño familiar). De forma similar, hay productos con diferentes niveles de calidad difíciles de agrupar y también productos que están registrados en diferentes niveles de la cadena de producción. Esta variedad de precios es un desafío a la hora de agrupar los precios, e implicó un estudio minucioso de los productos que se incluyen en este índice.

## **METODOLOGÍA**

Esencialmente, el proceso para la elaboración del índice consta de tres pasos:

- i. Selección de productos incluidos en el índice
- ii. Procesamiento de datos
- iii. Construcción del índice

### **I. SELECCIÓN DE PRODUCTOS INCLUIDOS EN EL ÍNDICE**

El IPAP-BMC se calcula con información de precios del registro de facturas de la BMC, teniendo en cuenta las subclases del CPC con las que se calcula el IPP agropecuario. Para estimar el índice, tomamos como base las ponderaciones del IPP agropecuario del DANE, las cuales provienen de la Encuesta Nacional Agropecuaria.

Para poder emparejar cada una de las ponderaciones de subclase del DANE se agrupan los productos más importantes del registro de facturas en categorías similares. A diferencia del DANE, los productos del registro de facturas no están catalogados usando códigos CPC, sino que tienen su propia codificación. Por esto, se usa una correlativa entre los productos de la Bolsa Mercantil y su equivalente del DANE a la hora de clasificar un producto en una subclase del índice.

Al revisar las series por mes y año desde el 2010, se seleccionaron los productos de cada subclase para los cuales la BMC tiene registros durante más del 80% del período analizado. Entre las series de tiempo, la única subclase del IPP del DANE que no se consideró en el cálculo fue la caña de azúcar, por no contar con registros frecuentes y no cumplir con la anterior condición.

## **II. PROCESAMIENTO DE DATOS**

El registro de facturas de la Bolsa Mercantil de Colombia recibe en promedio 20.000 facturas al día. Por lo tanto, el proceso de agregación requiere de un proceso de depuración y limpieza en dos etapas: por un lado, se corre un proceso para detectar valores atípicos en el total de facturas individuales; posteriormente, se realiza una limpieza de valores atípicos en las series de tiempo agregadas. Adicionalmente, se realiza una imputación de valores en las series incompletas (aquellas que cumplan con la condición de 80% o más tiempo con registros válidos).

El proceso es el siguiente:

- A. Limpieza de outliers en facturas individuales (bandas IQR)
- B. Generación de serie de tiempo (agregación por promedio ponderado) y detección de outliers.
- C. Imputación de valores usando Dynamic Time Warping – ARIMA.

A continuación, describimos con detalle cada uno de estos procesos.

### **A. Limpieza de outliers en facturas individuales (bandas IQR)**

Considerando que el registro de facturas consiste en varias facturas para diferentes productos observadas en diferentes periodos de tiempo y cada una por una cantidad y un valor individual, es indispensable trabajar estos datos antes de agregarlos. De esta forma, evitamos la presencia de valores atípicos al agregar los valores de precio en cada fecha.

Un procedimiento estándar a la hora de limpiar outliers son las llamadas bandas de valores extremos. Este, consiste en estimar un valor superior e inferior para cada producto, y considera que cualquier valor por fuera de estos límites es, por definición, un valor atípico. Como hay que considerar que, además, los precios de estas facturas tienen tendencia -normalmente tienden a aumentar en el largo plazo- algo que podría ser un valor fuera de lo normal hace algunos años podría ser un valor normal en el año en curso. Por esto, es importante que estas bandas de valores

extremos también puedan cambiar con cierta regularidad; en nuestro caso, estas bandas se ajustan mensualmente con ventanas móviles de 360 días.

Considerando lo anterior, construimos las bandas de precios para cada producto y para cada mes, tomando como límite inferior y límite superior la distribución intercuartílica de los datos, para cada producto  $i$  y cada mes  $t$ :

$$\text{Límite Inferior}_{it} = Q_{1it} - (3 * RIQ_{it})$$

$$\text{Límite Superior}_{it} = Q_{3t} + (3 * RIQ_t)$$

$$RIQ_{it} = Q_{3t} - Q_{1t}$$

Donde  $Q_1$  corresponde al cuartil 1 (percentil 25) y  $Q_3$  corresponde al cuartil 3 (percentil 75), en ambos casos, de la distribución del producto  $i$  en el mes  $t$ .

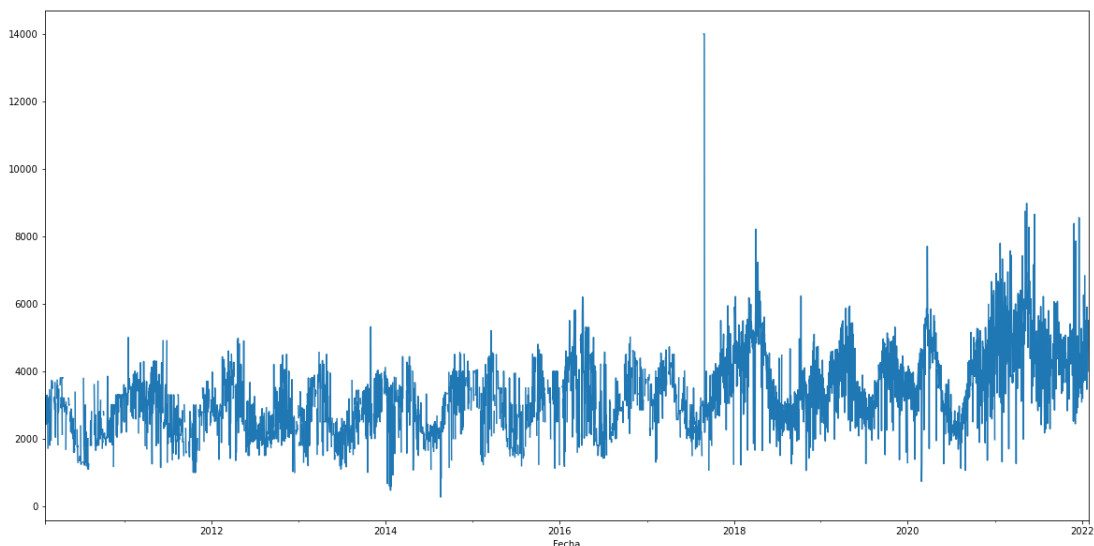
## B. Generar serie de tiempo (agregar datos con promedio ponderado) y detectar outliers en la serie de tiempo.

A continuación, para cada fecha se promedian las  $N$  facturas  $r_{it}$  para generar las series de precio  $y_t$ :

$$y_t = \frac{1}{N} \sum_{i=1}^N r_{it}$$

Al agregarse, estas series siguen exhibiendo valores atípicos (ver figura 1), por lo que hay que hacer una segunda limpieza de outliers en las series de tiempo.

### 1. Serie de tiempo obtenida tras limpieza de outliers en facturas individuales.



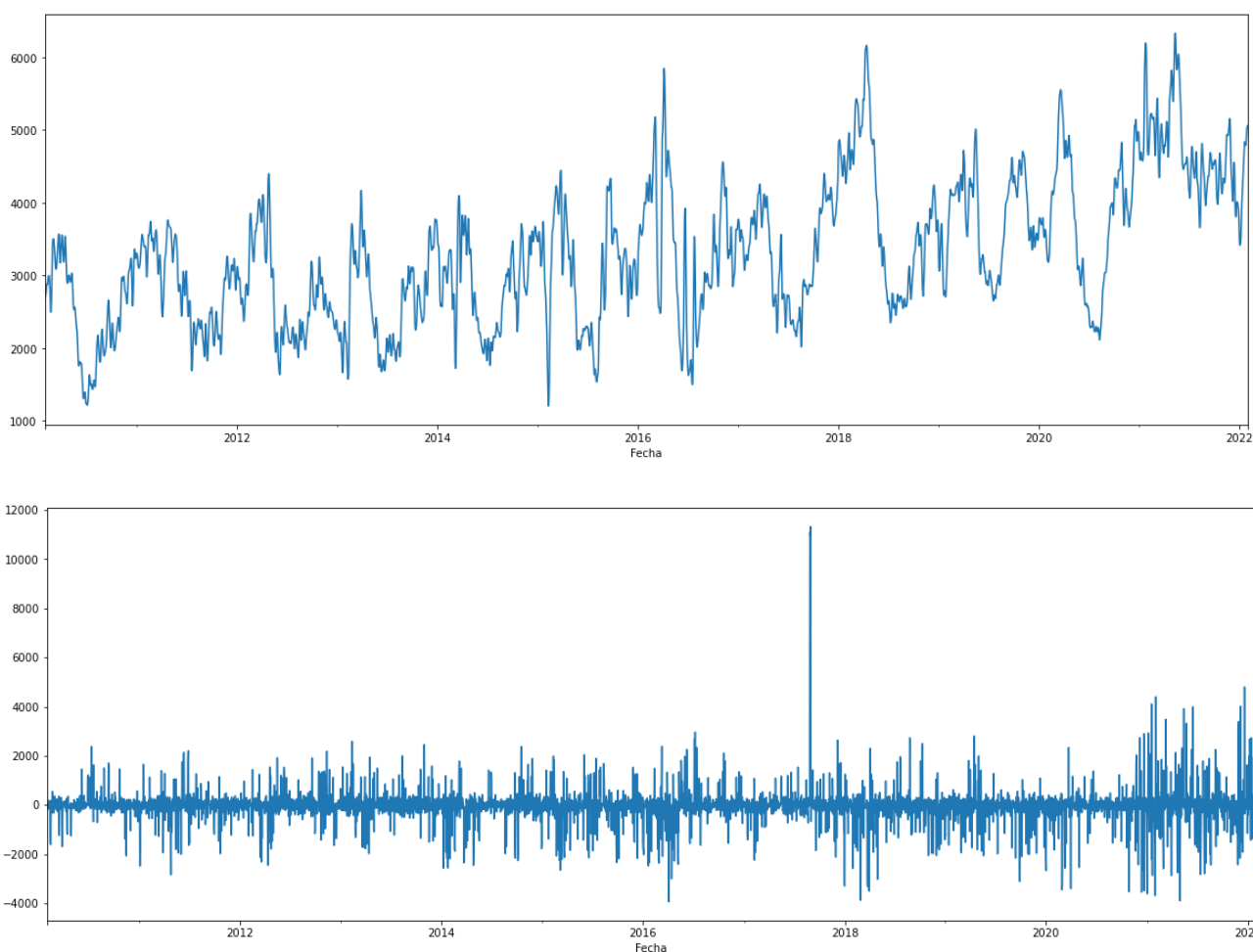
Para detectar valores extremos, empleamos un filtro de outliers con el método de descomposición STL con LOESS (*Seasonal and Trend decomposition using locally*

*weighted regression and scatterplot smoothing*)<sup>2</sup>. Este proceso busca valores atípicos usando el término de error que resulte de la descomposición de la misma, tomando como punto de partida el método de descomposición para series de tiempo. Este método es relativamente flexible, ya que permite utilizar diferentes medidas de tiempo y, además, permite encontrar relaciones no lineales entre los componentes de la serie. Esto puede describirse matemáticamente en las siguientes ecuaciones:

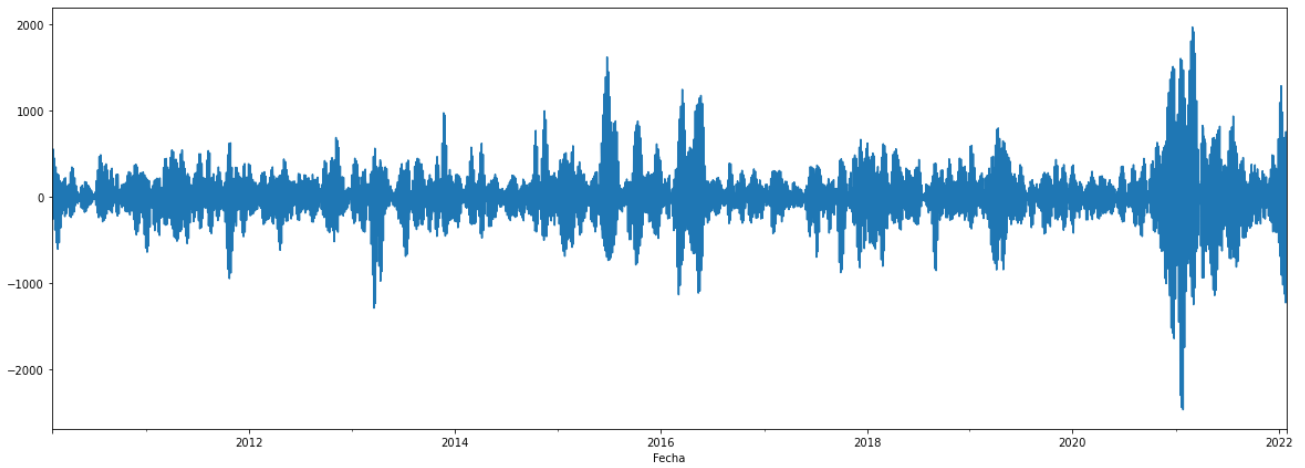
$$y_t = s_t + d_t + r_t$$

donde  $y_t$  es el valor de la serie de tiempo en el período  $t$ ,  $s_t$  es el valor periódico de la serie,  $d_t$  es el componente de tendencia de la serie, y  $r_t$  es el valor residual de la descomposición de la serie. Gráficamente, la descomposición de la serie de la gráfica 1 puede verse en la gráfica 2.

## 2. Descomposición de series de tiempo usando STL con LOESS.

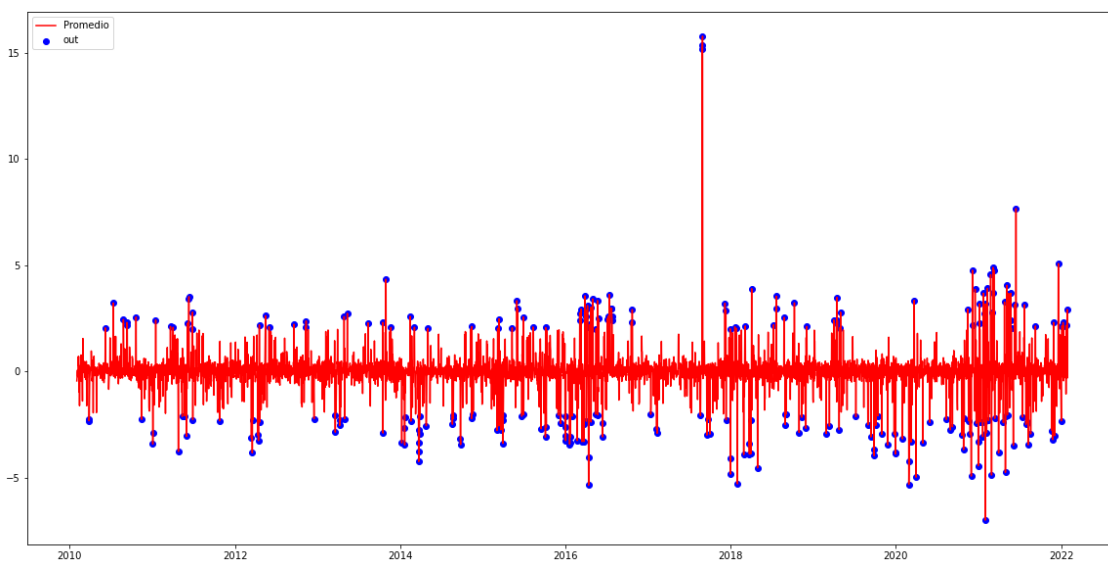


<sup>2</sup> Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. J. (1990). STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6(1), 3–33.

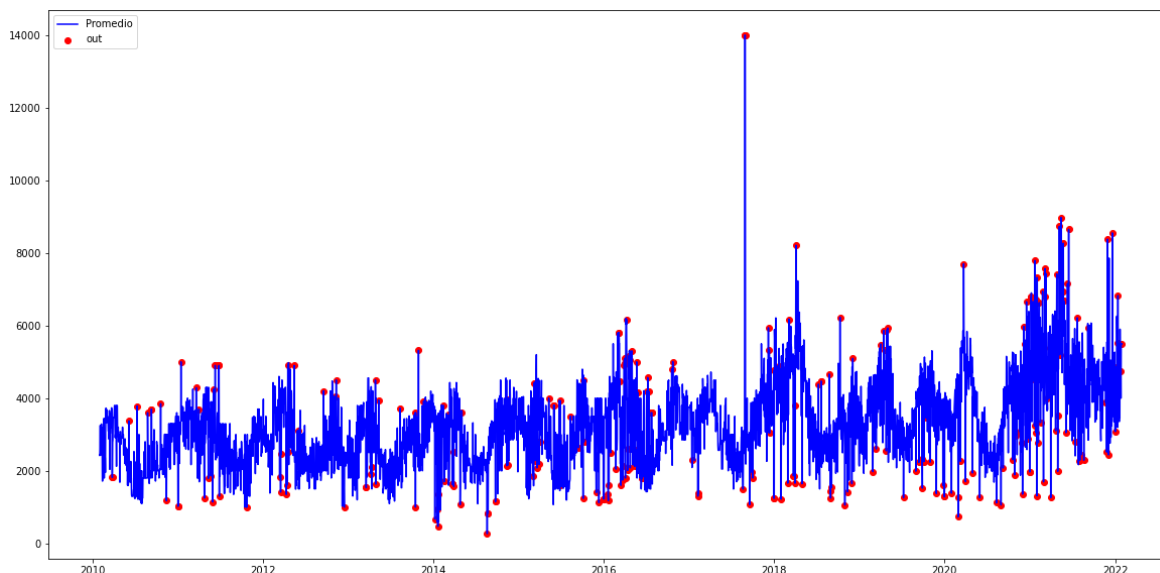


Para determinar si un valor de la serie es un valor atípico, tomamos valores del término residual  $r_t$  entre dos desviaciones estándar alrededor de la media. Al excluir los valores atípicos, se reemplaza el valor de éste por el valor de los términos de tendencia y la periodicidad obtenidos de la descomposición de la serie. Gráficamente, el resultado de este proceso puede verse en las gráficas 3 y 4.

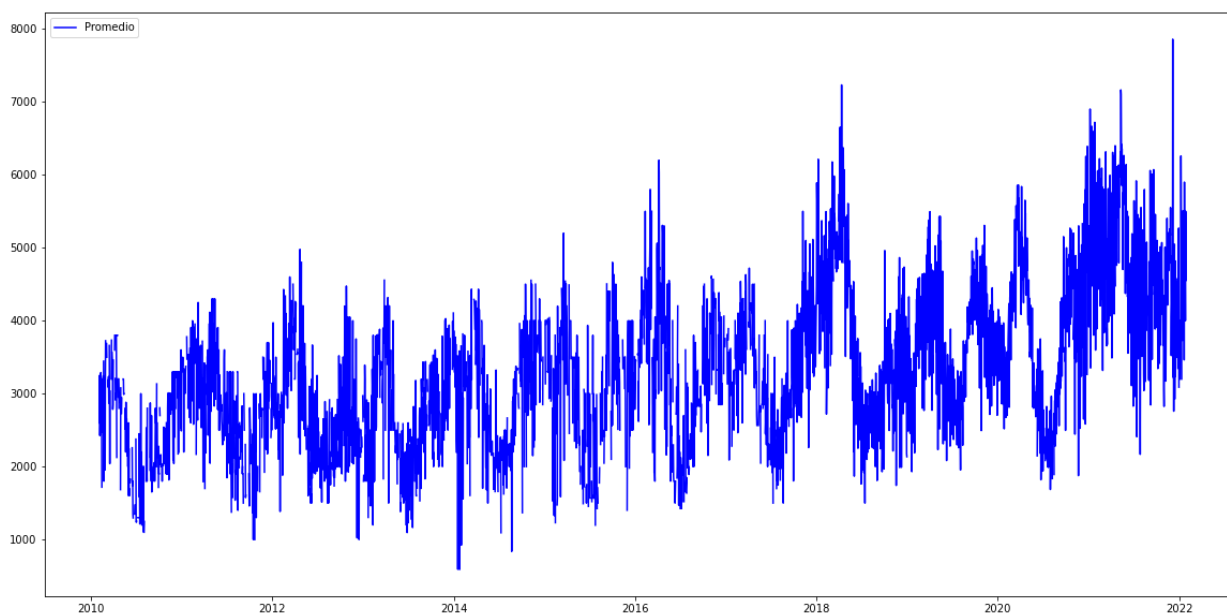
### 3. Identificación de valores atípicos usando el término residual de la descomposición.







4. Resultado de la limpieza de outliers.



### C. Imputación de valores usando Dynamic Time Warping – ARIMA.

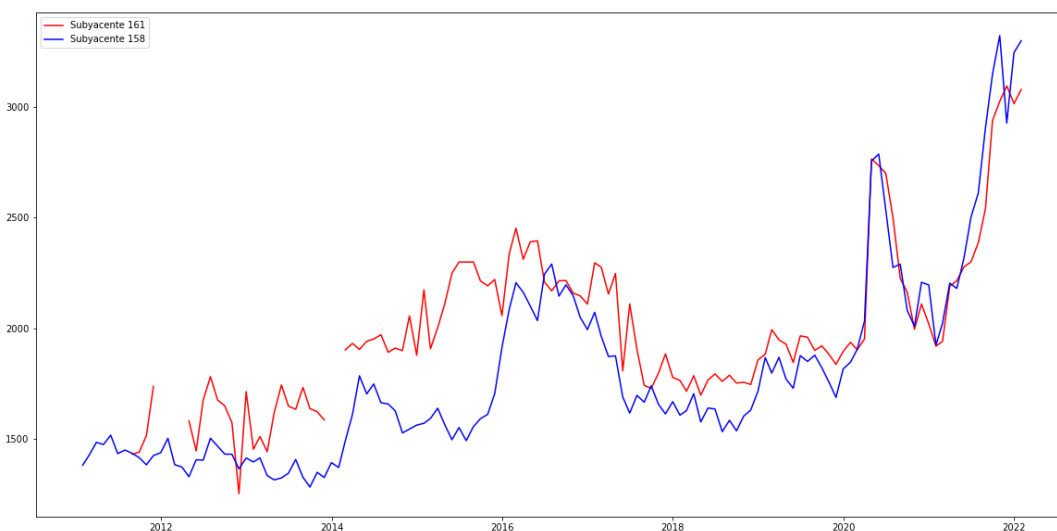
Una vez tenemos las series de tiempo procesadas, imputamos valores para las series de tiempo que tengan más del 80% de las observaciones disponibles en el período seleccionado. Para imputar valores, una aproximación útil es buscar series de tiempo similares, que tengan oscilaciones similares aún si los niveles no son idénticos, las series no son contemporáneas, o estas ocurren en ciclos de escalas con una amplitud diferente. Una metodología ideal para encontrar series de tiempo que sigan estos patrones es el Dynamic Time Warping (DTW), una medida de distancia que toma en cuenta la posibilidad de que dos series tengan patrones similares aún si tienen variaciones como las que mencionamos.

El algoritmo para imputar valores en una serie incompleta contempla el siguiente proceso:

- A. Encontramos la distancia DTW entre la serie incompleta y el conjunto de todas las series que estén completas.
- B. Tomamos la serie con el menor DTW de la base de datos.
- C. A continuación, usamos la serie obtenida en el paso anterior para imputar valores, ajustándola en niveles y tendencia a la serie incompleta.
- D. Esta serie se incorpora ahora al conjunto de series completas usado en el paso A.

Como las series que se completan terminan por incluirse en el conjunto de las series completas, el algoritmo es más robusto cuando se comienza por imputar valores en las series menos incompletas y se termina por las series más incompletas. Un ejemplo de los resultados de este algoritmo para una serie y su serie más cercana en términos de DTW puede observarse en la gráfica 5.

#### *5. Resultado del algoritmo DTW para encontrar series con tendencias similares.*



### III. CONSTRUCCIÓN DEL ÍNDICE

Una vez se terminan de procesar los datos, donde se garantiza que las series de precios sean robustas frente a la presencia de valores atípicos y que se imputen valores en series incompletas, estamos en condiciones de agregar las series obtenidas y estimar el índice de precios agropecuarios.

Se calculan los índices bajo la metodología de Laspeyres<sup>3</sup> usando las ponderaciones del IPP del sector agropecuario del DANE en su componente de oferta interna. Conservamos estas ponderaciones al considerarlas más robustas y representativas del sector agropecuario nacional que otros pesos que se pudieran obtener de las cantidades tranzadas en el registro de facturas de la Bolsa Mercantil de Colombia.

Debido a la baja cantidad de registros, como ya se mencionó, se excluyó la caña de azúcar del índice, y se hicieron las debidas reponderaciones. Con esto, el índice está clasificado por CPC, compuesto por 16 grupos, 30 clases, 33 subclases y 142 productos.

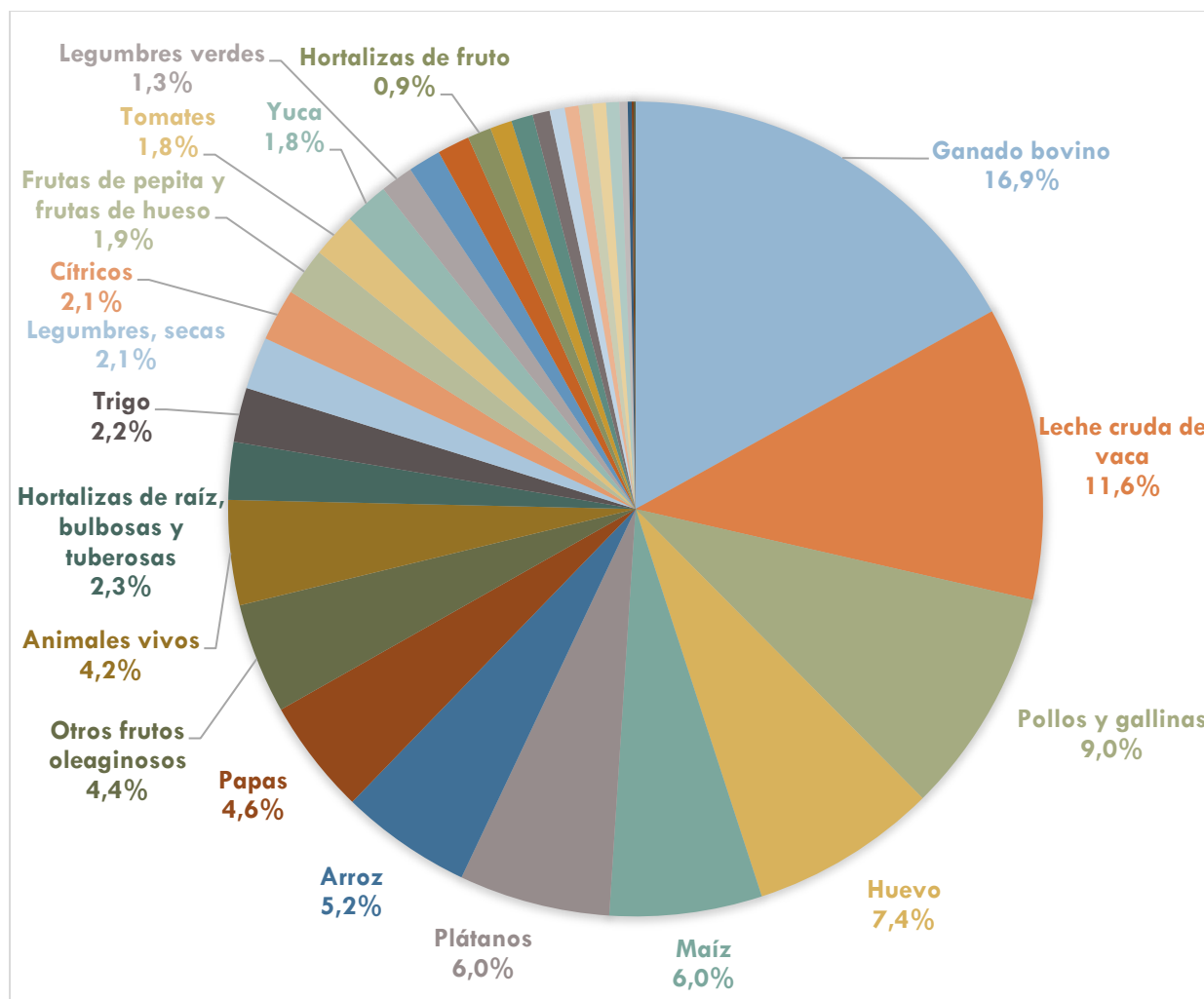
Como se ve en la tabla 1, los productos más importantes del índice son Ganado Bovino, Leche Cruda de Vaca, Pollos y Gallinas, Huevo, y Maíz, que representan casi la mitad del índice. Gráficamente, el total de las ponderaciones se presentan en el cuadro 1, y una tabla completa se presenta en el apéndice. Se toma como mes base del índice Diciembre de 2014, lo que lo hace completamente comparable con el IPP.

### 1. Ponderaciones del índice por producto.

<b>Ponderaciones de los 15 productos más importantes del Índice de Precios Agropecuarios IPAP-BMC</b>	
Ganado bovino	16.94%
Leche cruda de vaca	11.65%
Pollos y gallinas	8.97%
Huevo	7.43%
Maíz	6.04%
Plátanos	5.99%
Arroz	5.21%
Papas	4.57%
Otros frutos oleaginosos	4.39%
Animales vivos	4.16%
Hortalizas de raíz, bulbosas y	2.29%
Trigo	2.16%
Legumbres, secas	2.07%
Cítricos	2.07%
Demás productos	16.05%

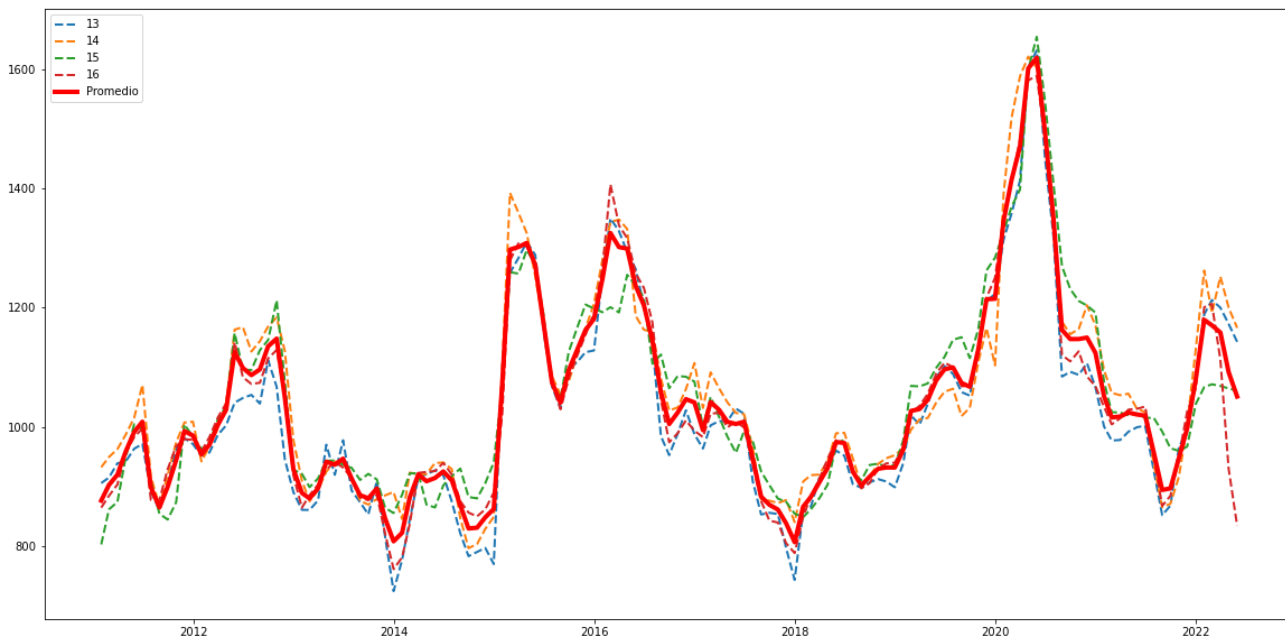
<sup>3</sup> Este método parte de un índice basado en una canasta fija de productos en un período base, en la que se sustituyen los precios período a período. Es decir, la variabilidad del índice proviene completamente de la variabilidad de los precios. Ver Laspeyres, E. (1871), "Die Berechnung einer mittleren Waarenpreissteigerung", *Jahrbücher für Nationalökonomie und Statistik*, vol. 16, Lucius & Lucius Verlagsgesellschaft mbH.

## 6. Ponderaciones del índice por subclase CPC.

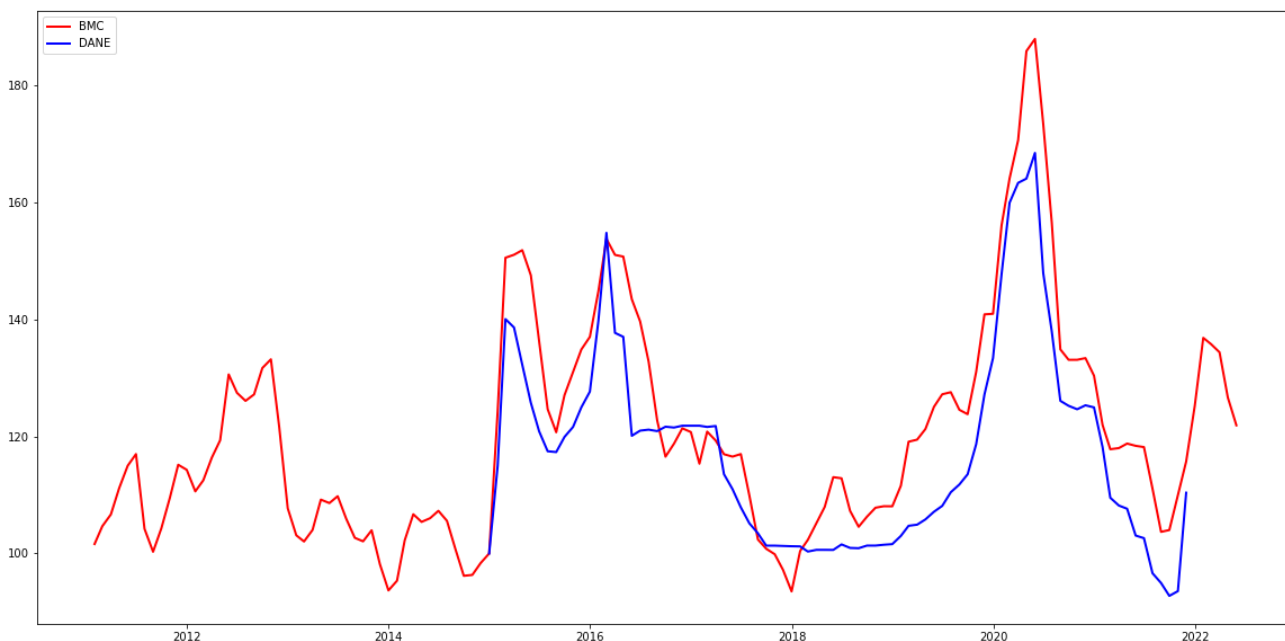


Para cada una de las 33 subclases, agregamos los productos clasificados en el paso I, usando un promedio simple entre los productos. Como guía, se tomaron los componentes del IPP del DANE para compararlos con los resultados de la BMC, y, en general, ambas series siguen una tendencia similar (ver gráficas 7 y 8).

7. Productos agregados por subclase.



8. Comparación tendencia para una subclase BMC contra la subclase del DANE.

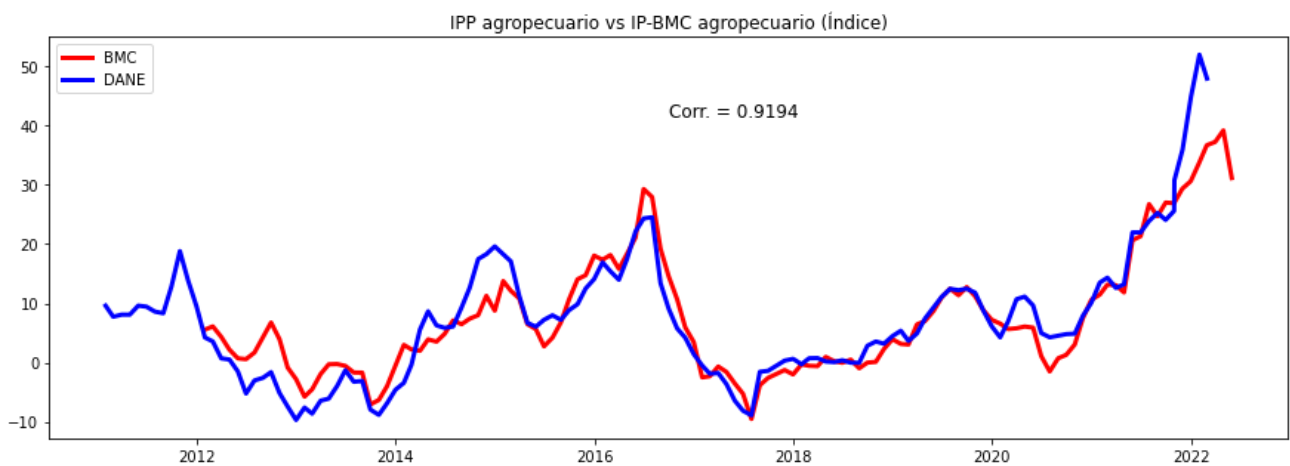
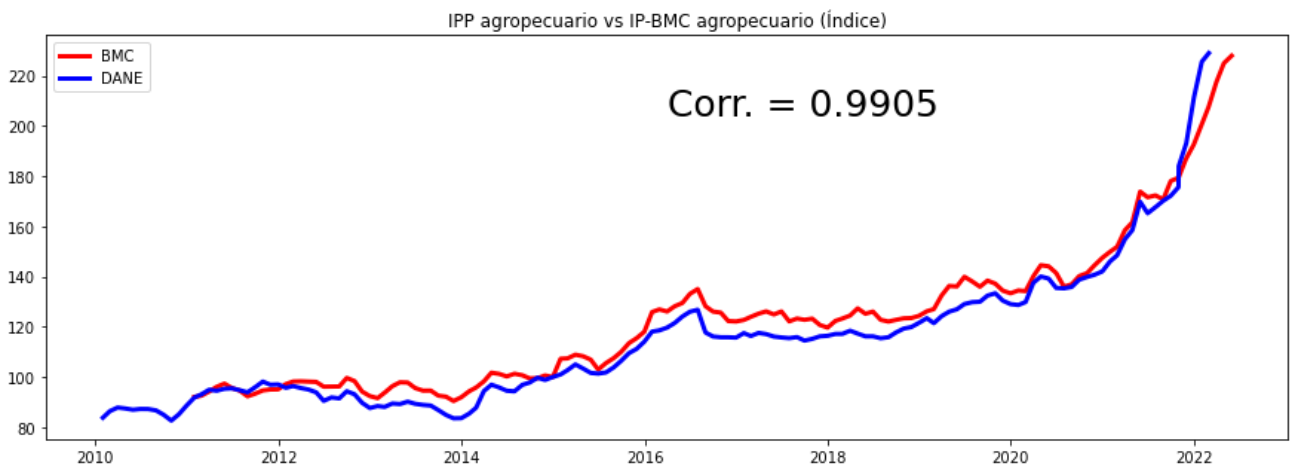


#### IV. COMPARACIÓN CON INDICADORES ECONÓMICOS.

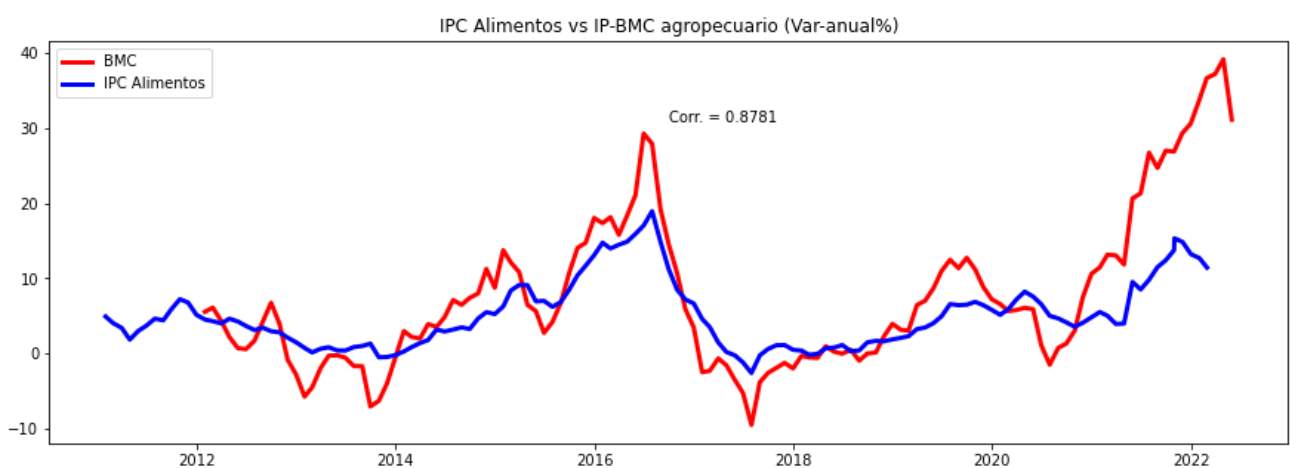
Resulta interesante comparar el IPAP-BMC con otros indicadores de interés a través del tiempo. Nuestros resultados muestran que, efectivamente, el índice refleja un comportamiento esperado en un indicador coyuntural de precios agropecuarios: existe una alta correlación con indicadores similares, como el IPP agropecuario y el

IPC de alimentos del DANE tanto en términos de niveles como de crecimiento anual. (Ver gráficas 9, 10 y 11).

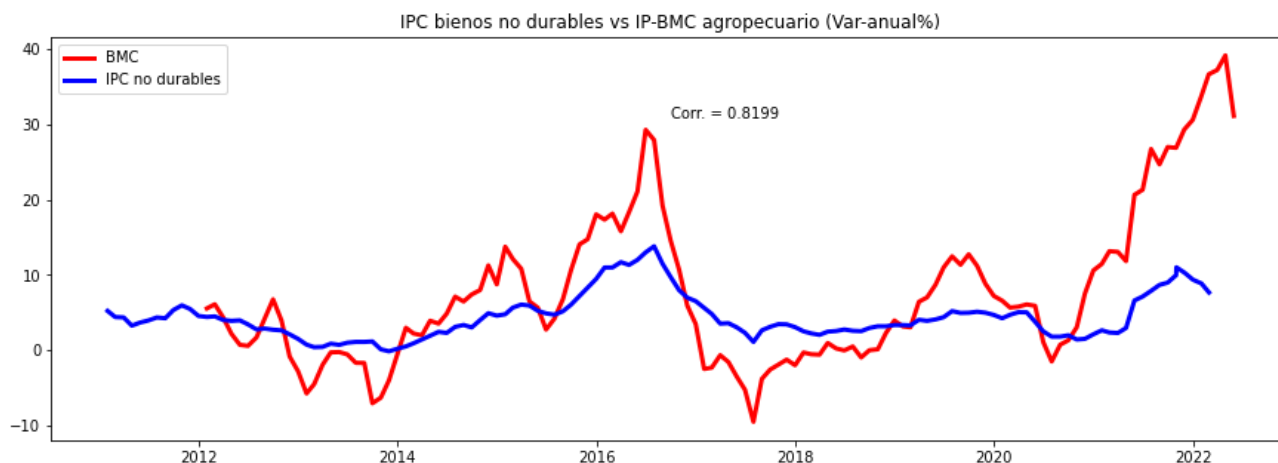
9. IPP Agropecuario (DANE) vs IPAP-BMC



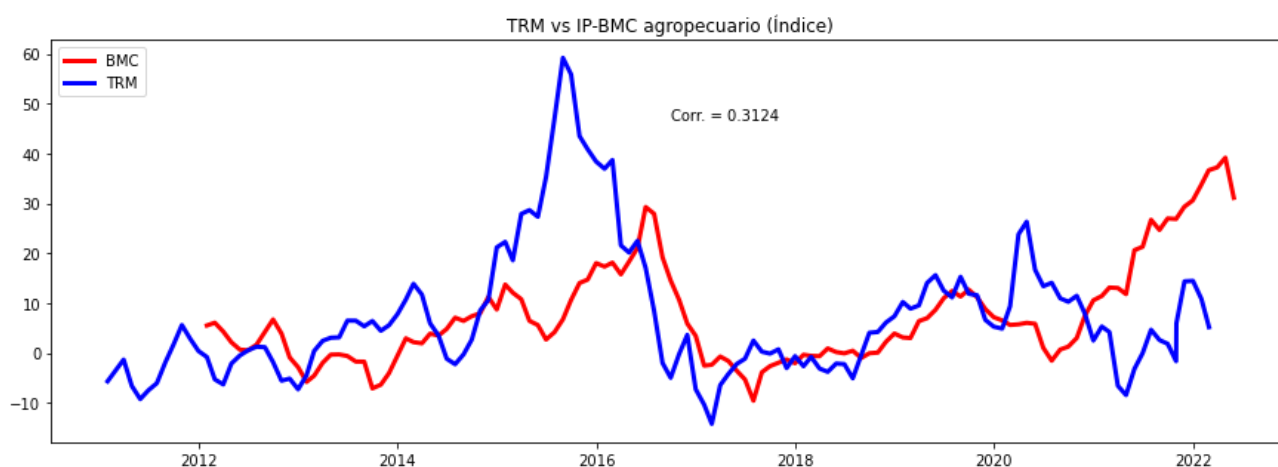
10. IPC de Alimentos (DANE) vs IPAP-BMC.



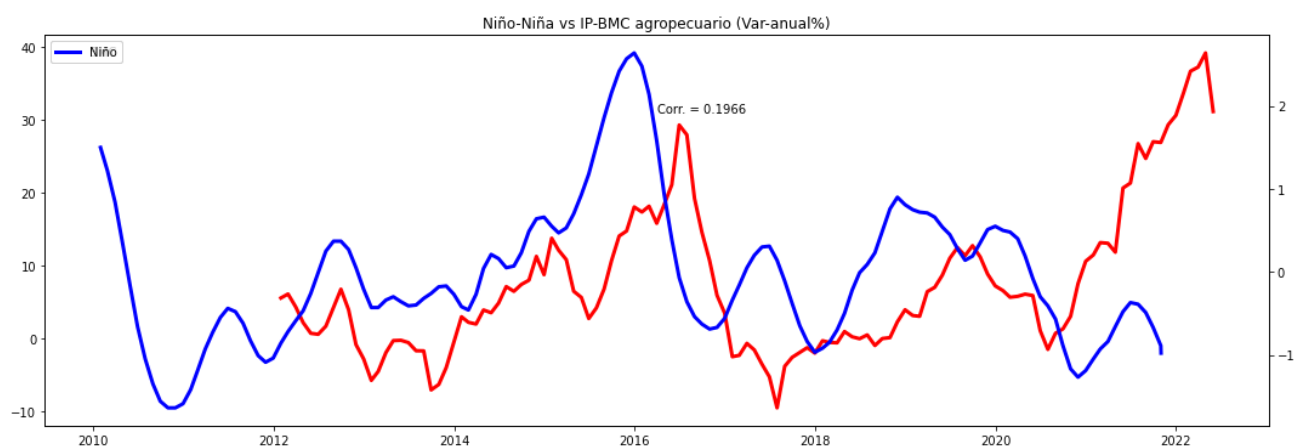
## 11. IPC bienes no durables (DANE) y IPAP-BMC



También se observa una correlación no contemporánea fuerte entre la tasa de cambio y el índice de precios: movimientos fuertes en la tasa de cambio afectan la tasa de crecimiento del índice. Uno de los motivos de este comportamiento es que varios productos incluidos en el índice, como el maíz y el trigo, son en su mayoría importados, al igual que algunos agroinsumos. Oscilaciones en la tasa de cambio pueden no afectar el precio inmediatamente, pero sí con algún tiempo por efecto de inventarios, transporte y comercialización.



Otro indicador fuertemente correlacionado con nuestro índice es el fenómeno del Niño/Niña. Al tratarse de un indicador de precios agropecuarios, es claro que puede haber una fuerte influencia de períodos de clima extremo sobre precios al afectar directamente la cosecha de la mayoría de los productos del índice. De forma similar a la tasa de cambio, el efecto del Niño/Niña parece no ser contemporáneo, sino afectar a los precios con cierto rezago.



## V. COMENTARIOS FINALES

El índice de precios agropecuarios de la Bolsa Mercantil de Colombia (IPAP-BMC) es un indicador esencial a la hora de ver tendencias generales en el sector agropecuario. Dada la frecuencia y el detalle de la fuente de datos, probablemente el índice aquí presentado sea el punto de partida de indicadores con frecuencias temporales más altas (semanal o quincenal).

A juzgar por el comportamiento del índice agropecuario contra otros indicadores semanales, el indicador que presenta la Bolsa Mercantil es un indicador sólido y en línea con otras variables similares en el país, a pesar de que los insumos para construirlos tienen un origen muy distinto. Aunque el registro de facturas sea una base de datos con particularidades, complejidades y desafíos únicos, la metodología aquí presentada hace que el índice de precios sea un indicador robusto.

### APÉNDICE

#### 2. Ponderaciones completas del índice.

Alimentos	Ponderación original	Reponderación
<b>Ganado bovino</b>	0,019058684	0,169370782
<b>Leche cruda de vaca</b>	0,013108282	0,116490732
<b>Pollos y gallinas</b>	0,010098471	0,089743128
<b>Huevo</b>	0,008357986	0,074275778
<b>Maíz</b>	0,006801856	0,060446759
<b>Plátanos</b>	0,00674265	0,059920606
<b>Arroz</b>	0,005862363	0,052097672
<b>Papas</b>	0,005145137	0,045723827
<b>Otros frutos oleaginosos</b>	0,004940834	0,043908222
<b>Animales vivos</b>	0,004680515	0,041594821
<b>Hortalizas de raíz, bulbosas</b>	0,002571502	0,022852431



<b>Trigo</b>	0,002429394	0,021589547
<b>Legumbres, secas</b>	0,002333026	0,020733145
<b>Cítricos</b>	0,002329731	0,020703863
<b>Frutas de pepita y frutas de</b>	0,002137234	0,018993176
<b>Tomates</b>	0,001998469	0,017760004
<b>Yuca</b>	0,001979389	0,017590443
<b>Legumbres verdes</b>	0,001461507	0,012988126
<b>Hortalizas de hoja o de tallo</b>	0,001461462	0,012987726
<b>Frutas tropicales y</b>	0,001444123	0,012833636
<b>Hortalizas de fruto</b>	0,001060739	0,009426576
<b>Cultivos de fibra</b>	0,00097735	0,008685522
<b>Pescado</b>	0,000966262	0,00858698
<b>Nueces</b>	0,000766947	0,006815703
<b>Sorgo, otros</b>	0,000676046	0,006007886
<b>Soja, otros</b>	0,000620327	0,005512721
<b>Granos de cacao</b>	0,000610961	0,005429492
<b>Flores</b>	0,000604265	0,005369985
<b>Banano</b>	0,000587467	0,005220701
<b>Cebada, otros</b>	0,000364494	0,003239188
<b>Uvas</b>	0,000178471	0,001586038
<b>Semillas de remolacha</b>	8,83258E-05	0,000784934
<b>Crustáceos</b>	8,21273E-05	0,000729849