

Documento de trabajo 1225: Metodología para la predicción de los precios minoristas de los alimentos localmente disponibles en las principales ciudades de Colombia

Pontificia Universidad Javeriana, Cali

14 de diciembre de 2025

1. Presentación

El documento de trabajo presenta una metodología para la predicción de los precios minoristas de alimentos en las principales ciudades de Colombia, utilizando exclusivamente información oficial del DANE. El análisis se apoya en dos fuentes de datos: los precios minoristas empleados en la construcción del Índice de Precios al Consumidor (IPC), con cobertura mensual entre 1999 y 2018 para trece ciudades, y los precios mayoristas provenientes del Sistema de Información de Precios y Abastecimiento del Sector Agropecuario (SIPSA), con información mensual para Cali, Bogotá y Medellín. Un elemento central del estudio es el mapeo entre los productos del IPC y los reportados por SIPSA, el cual no es uno a uno, sino de tipo muchos a uno, y permite vincular coherentemente ambas bases de datos para el análisis empírico.

A partir de esta información, el documento evalúa seis metodologías para estimar y predecir precios minoristas: una aproximación basada en la variación del IPC, la estimación mediante márgenes medianos de comercialización, modelos de regresión en niveles y en primeras diferencias con control estacional, un modelo de corrección de error (ECM) y una versión asimétrica de este modelo estimada por ciudad. Las metodologías se comparan dividiendo la muestra en conjuntos de entrenamiento y validación, y se evalúan mediante métricas estándar de error. Además, algunas especificaciones permiten derivar márgenes de comercialización implícitos, los cuales son analizados como parte de la discusión y validación de los resultados presentados en el documento

2. Datos

2.1. Insumo del IPC (DANE)

La primera fuente corresponde a los precios minoristas reportados por el DANE, utilizados como insumo para el cálculo del Índice de Precios al Consumidor (IPC) (DANE, 2024). El periodo de análisis abarca desde enero de 1999 hasta marzo de 2018 y cubre las 13 principales ciudades del país:

- Bogotá D.C.
- Medellín A.M.
- Cali A.M.
- Barranquilla A.M.
- Bucaramanga A.M.
- Manizales A.M.
- Pereira A.M.
- Cúcuta A.M.
- Pasto

- Ibagué
- Montería
- Cartagena
- Villavicencio

Una característica relevante de estos datos es su estructura de clasificación, basada en la canasta de seguimiento del IPC 2008. Esta clasificación contempla distintos niveles de agregación: Grupo, Subgrupo, Clase, Gasto Básico y Artículo. A continuación, se presenta la estructura de esta clasificación:

Subclase	IPC	Gasto_basico	DANE
011101	Arroz	11101	Arroz
011102	Pastas alimenticias	11103	Pastas alimenticias
011103	Avena y sus derivados	11105	Otros cereales
011105	Trigo y sus derivados	11102	Harina de maíz y otras harinas
		11105	Otros cereales
011106	Maíz y sus derivados	11102	Harina de maíz y otras harinas
		11105	Otros cereales
011108	Pan	11201	Pan
011198	Otras harinas, cereales y almidones	11102	Harina de maíz y otras harinas
		11104	Cereales preparados
		11202	Otros productos de panadería
011199	Otros productos de panadería	11202	Otros productos de panadería
011201	Carne de res y derivados	15101	Res
011202	Carne de Cerdo y Derivados	15102	Cerdo

A continuación, se muestra la estructura de la base de datos de precios proporcionada por el DANE:

```
## # A tibble: 15 x 8
##   ano  ciudad nombre_ciudad codigo_articulo articulo      unidad mes    precio
##   <chr> <chr> <chr>          <chr>        <chr>      <chr> <chr> <chr>
## 1 1999 05 MEDELLÍN 1110101 ARROZ PARA SE~ 500Gr~ enero 588.81
## 2 1999 05 MEDELLÍN 1110101 ARROZ PARA SE~ 500Gr~ febr~ 588.37
## 3 1999 05 MEDELLÍN 1110101 ARROZ PARA SE~ 500Gr~ marzo 587.72
## 4 1999 05 MEDELLÍN 1110101 ARROZ PARA SE~ 500Gr~ abril 587.71
## 5 1999 05 MEDELLÍN 1110101 ARROZ PARA SE~ 500Gr~ mayo  590.53
## 6 1999 05 MEDELLÍN 1110101 ARROZ PARA SE~ 500Gr~ junio 592.17
## 7 1999 05 MEDELLÍN 1110101 ARROZ PARA SE~ 500Gr~ julio 590.42
## 8 1999 05 MEDELLÍN 1110101 ARROZ PARA SE~ 500Gr~ agos~ 587.68
## 9 1999 05 MEDELLÍN 1110101 ARROZ PARA SE~ 500Gr~ sept~ 586.29
## 10 1999 05 MEDELLÍN 1110101 ARROZ PARA SE~ 500Gr~ octu~ 586.14
## 11 1999 05 MEDELLÍN 1110101 ARROZ PARA SE~ 500Gr~ novi~ 581.46
## 12 1999 05 MEDELLÍN 1110101 ARROZ PARA SE~ 500Gr~ dici~ 578.25
## 13 1999 05 MEDELLÍN 1110102 ARROZ PARA SO~ 500Gr~ enero 0
## 14 1999 05 MEDELLÍN 1110102 ARROZ PARA SO~ 500Gr~ febr~ 0
## 15 1999 05 MEDELLÍN 1110102 ARROZ PARA SO~ 500Gr~ marzo 0
```

2.2. Datos de SIPSA

La información sobre los precios mayoristas corresponde a los datos del Sistema de Información de Precios y Abastecimiento del Sector Agropecuario (SIPSA), publicados por el DANE. El SIPSA no sólo informa, con frecuencia diaria, sobre los precios mayoristas de los productos agroalimentarios que se comercializan en el

país; sino que, además, proporciona información, con frecuencia quincenal, sobre el nivel de abastecimiento de los alimentos en las ciudades. En este estudio se utilizarán datos de SIPS A para las tres principales ciudades de Colombia (Cali, Bogotá y Medellín) durante el período 2013:1 – 2024:1, con frecuencia mensual.

A continuación, se presenta la estructura de los datos:

```
## # A tibble: 15 x 8
##   cod_mun alimento_sipsa Year Month precio_sipsa articulo_ipc codigo_articulo
##   <chr>     <chr>      <dbl> <dbl>      <dbl> <chr>      <chr>
## 1 76001 Aguacate papel~ 2013     1        1468. AGUACATE    1410503
## 2 76001 Ahuyama          2013     1        265. AHUYAMA    1310507
## 3 76001 Arracacha amar~ 2013     1        503. ARRACACHA  1210303
## 4 76001 Arroz de prime~ 2013     1       1044. ARROZ PARA ~ 1110101
## 5 76001 Arveja verde e~ 2013     1        837. ARVEJA FRES~ 1320201
## 6 76001 Arveja verde s~ 2013     1       1166. ARVEJA SECA 1320202
## 7 76001 Avena en hojue~ 2013     1       3058. AVENA HOJUE~ 1110501
## 8 76001 Avena molida     2013     1       3080. AVENA MOLIDA 1110502
## 9 76001 Azúcar morena    2013     1        879. AZÚCAR NATU~ 1810202
## 10 76001 Azúcar sulfita~ 2013     1        707. AZÚCAR REFI~ 1810201
## 11 76001 Banano criollo  2013     1        293. BANANOS    1410201
## 12 76001 Bocachico impo~ 2013     1       3256. PESCADO DE ~ 1610102
## 13 76001 Café molido     2013     1       7201. CAFÉ MOLIDO 1820101
## 14 76001 Carne de cerdo~ 2013     1       4360. CARNE DE CE~ 1510202
## 15 76001 Carne de cerdo~ 2013     1       3135. CARNE DE CE~ 1510202
## # i 1 more variable: precio_ipc <dbl>
```

3. Metodologías

Con base en las fuentes de información descritas anteriormente, en este estudio se evaluarán tres metodologías para estimar los precios minoristas a partir de la información de precios mayoristas. Primero, se propone el cálculo del margen mediano para cada uno de los productos reportados en SIPS A, así como el margen en el percentil 25 y 75. Segundo, se propone la estimación de un modelo SARIMAX, que permita capturar la evolución temporal de los precios minoristas, considerando los precios mayoristas como variable exógena y la variación estacional. En tercer lugar, se propone un modelo dinámico de corrección de errores (ECM), en el cual la dinámica de corto plazo de las variables está determinada por su variación respecto del equilibrio. Para validar el ajuste de las metodologías para la predicción de los precios minoristas, se propone dividir la muestra en dos: un conjunto de datos de entrenamiento, que corresponde al 70% de la muestra total; y un conjunto de datos de validación, que corresponde al 30%. Los métodos son evaluados a partir de las métricas comúnmente empleadas: error cuadrático medio y error absoluto medio porcentual.

3.1. Mapeo: IPC-SIPS A

La implementación de las metodologías I-III presupone un mapeo previo entre los alimentos reportados por el DANE en la construcción del IPC y los alimentos reportados por SIPS A (en adelante, mapeo IPC-SIPS A). Es decir, para cada alimento reportado en SIPS A, existe un alimento equivalente en la canasta del IPC. A continuación, se presenta el mapeo:

Table 2: Mapeo IPC – SIPS A

codigo_tcac	retail	mapeo_sipsa	sipsa
26	FECULA DE MAÍZ	26	Fécula de maíz
A010	ARROZ PARA SECO	A010	Arroz blanco importado

codigo_tcac	retail	mapeo_sipsa	sipsa
A010	ARROZ PARA SECO	A010	Arroz de primera
A010	ARROZ PARA SECO	A010	Arroz de segunda
A010	ARROZ PARA SECO	A010	Arroz excelso
A012	AVENA HOJUELAS	A012	Avena en hojuelas
A029	AVENA MOLIDA	A029	Avena molida
A034	HARINA PRECOCIDA	A034	Harina precocida de maíz
A040	HARINA DE TRIGO	A040	Harina de trigo
A044	CUCHUCO DE MAÍZ	A044	Cuchuco de maíz
A044	CUCHUCO DE MAÍZ	A044	Maíz amarillo trillado
A072	PASTA PARA SECO	A072	Pastas alimenticias
B006	AHUYAMA	B006	Ahuyama
B006	AHUYAMA	B006	Ahuyamín (Sakata)
B009	AJO	B009	Ajo
B015	ARRACACHA	B015	Arracacha amarilla
B019	ARVEJA FRESCA	B019	Arveja amarilla seca importada
B019	ARVEJA FRESCA	B019	Arveja verde en vaina
B019	ARVEJA FRESCA	B019	Arveja verde en vaina pastusa
B027	CEBOLLA	B027	Cebolla cabezona blanca
	CABEZONA		

Nótese que el mapeo no es 1:1, sino n: 1 (muchos: 1). Esto quiere decir que un mismo alimento de la base de datos del IPC puede tener varios alimentos equivalentes en la base de datos de SIPSAs. (Un ejemplo: el arroz para seco, que es un alimento del IPC, corresponde tanto al arroz de primera como al arroz de segunda y el arroz excelsio en la base de datos de SIPSAs).

3.2. Metodologías

La siguiente tabla resume las metodologías implementadas en este documento de trabajo:

No.	Metodología
I	Ajuste sobre los precios minoristas a partir del IPC
II	Estimación del margen mediano (Q2 [Q1, Q3])
III	Regresión lineal en niveles, controlando por el efecto estacional.
IV	Regresión en primeras diferencias, controlando por el efecto estacional.
V	Modelo de corrección de error
VI	Modelo de corrección de error asimétrico

4. Descripción de las metodologías y resultados

4.1. Metodología I: Aproximación a partir del IPC

Para estimar los precios minoristas incompletos, se implementa una estrategia basada en la variación mensual del IPC a nivel de subclase. Para cada artículo en cada ciudad, se separaron los datos en dos subconjuntos:

- Conjunto de entrenamiento hasta enero de 2015 con observaciones sobre precios minoristas.
- Conjunto de validación a partir de enero de 2015, a partir del cual se estimarán los precios usando las variaciones del IPC.

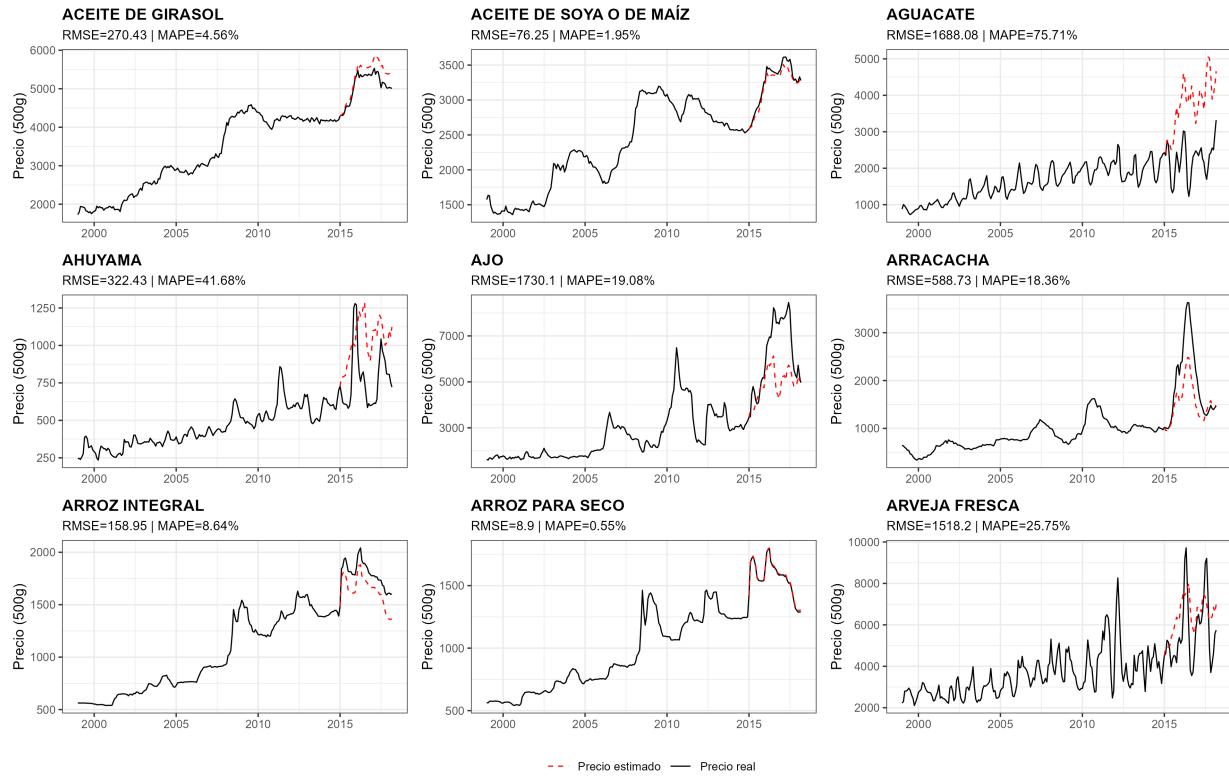
Se utilizó el último precio observado del conjunto de entrenamiento (enero de 2015) y se estimaron los precios hacia adelante multiplicando iterativamente por la variación mensual del IPC:

$$\hat{p}_t = \hat{p}_{t-1} \times \frac{IPC_t}{IPC_{t-1}}$$

Esta operación se repite mes a mes hasta llegar al final del conjunto de validación (marzo de 2018). Esta metodología proporciona una estimación directa de los precios minoristas y, por tanto, evita el tratamiento explícito del problema de la transmisión entre precios mayoristas y minoristas.

Resultados:

CALI — Precio real vs estimado (IPC subclase)



4.2. Metodología II: Estimación a partir del margen mediano (Q1 – Q3)

La primera metodología propuesta consiste en el cálculo del margen de comercialización para cada uno de los productos reportados por SIPSA. A partir del mapeo IPC-SIPSA, para el alimento k en t, el margen de comercialización se calcula simplemente como:

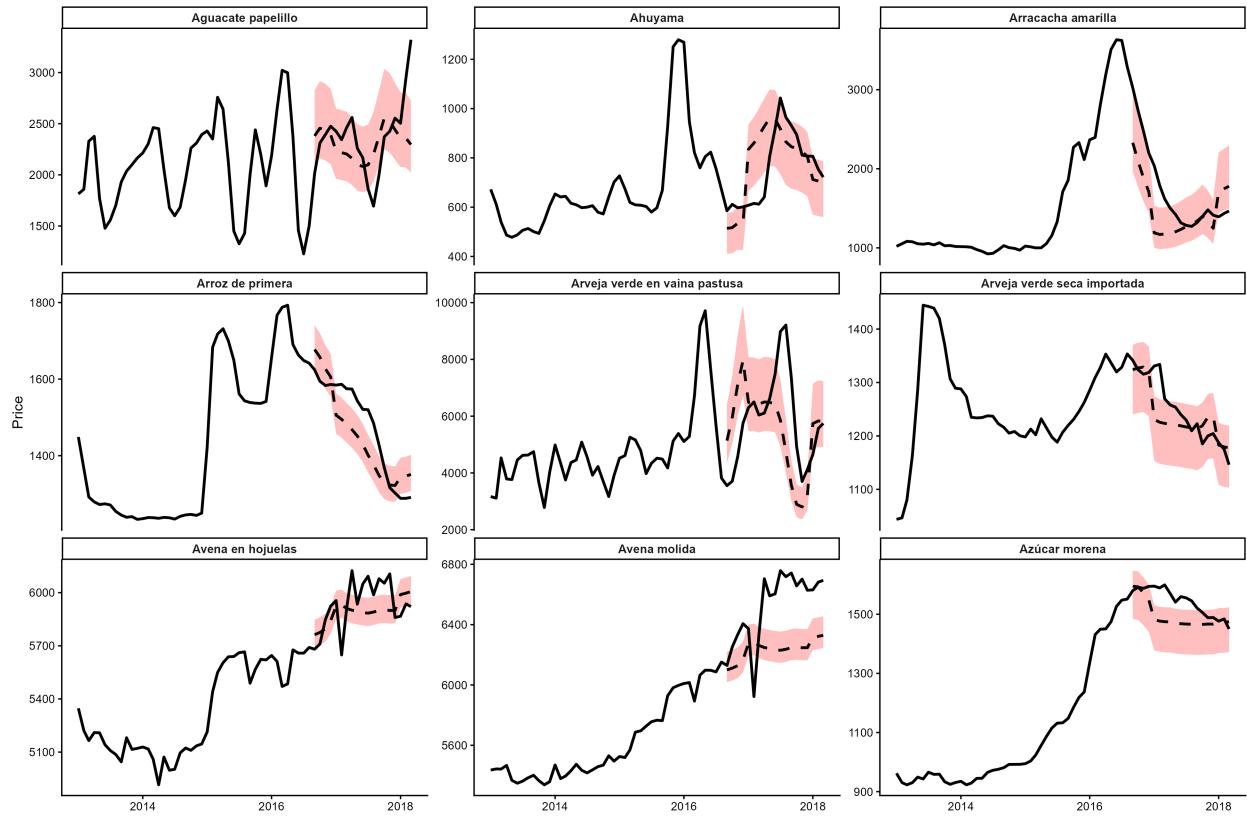
$$\gamma_{k,t} = \frac{P_{min} - P_{may}}{P_{may}} \times 100$$

Con base en los márgenes de comercialización estimados, se obtiene la distribución para cada uno de los alimentos reportados por SIPSA. A partir de la distribución, se calcula el margen mediano (Q2), así como el margen en el percentil 25 (Q1) y 75 (Q3).

Resultados:

Observed vs Predicted Retail Prices — Methodology 1

Q1–Q3 margin band | Page 1 of 9



4.3. Metodología III: Regresión lineal en niveles y primeras diferencias

Sea $\ln(P_{it}^{min})$ el logaritmo natural del precio minorista, y sea $\ln(P_{it}^{may})$ el logaritmo natural del precio mayorista del alimento i en el período t . Para cada alimento i , se estima el siguiente modelo de regresión lineal en niveles logarítmicos:

$$\ln(P_{it}^{min}) = \alpha_i + \beta_i \ln(P_{it}^{may}) + \sum_{m=1}^{11} \delta_{im} S_{mt} + e_{it}$$

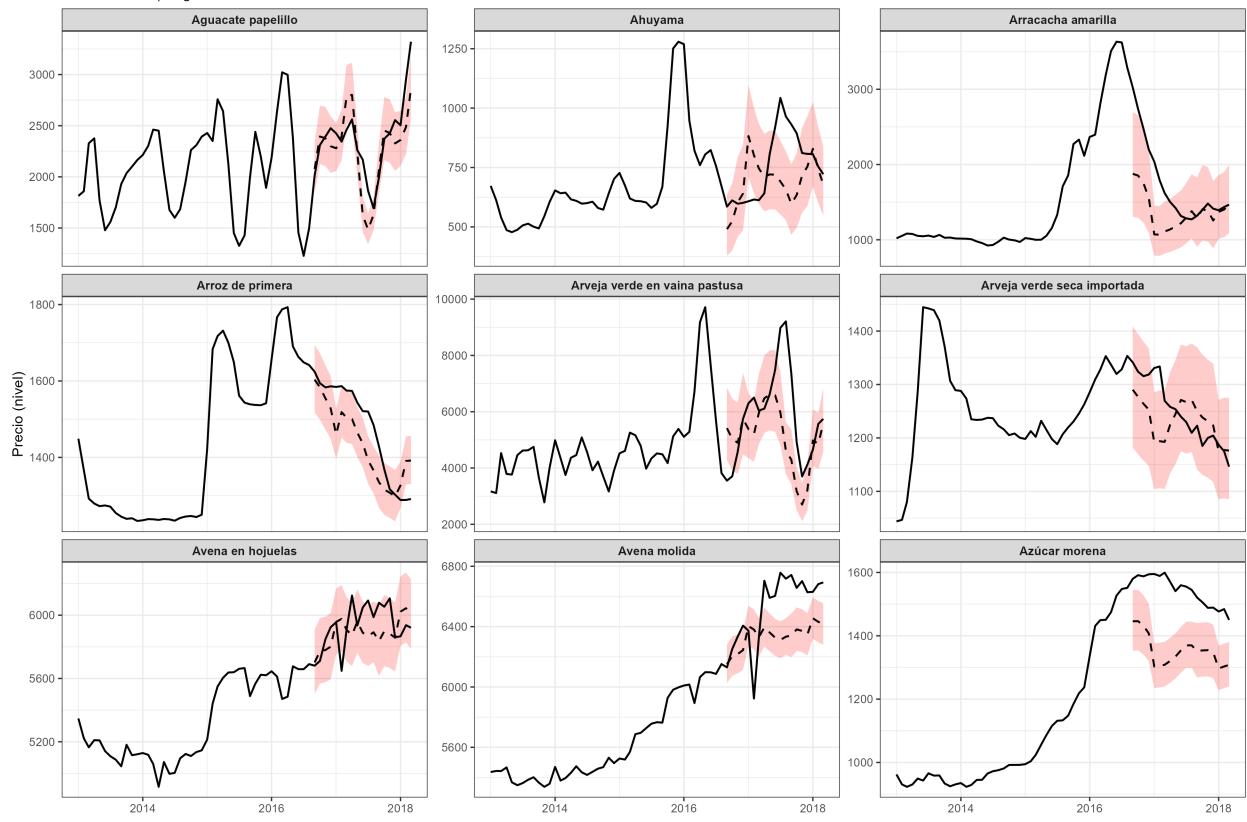
donde

$$S_{mt} = \begin{cases} 1, & \text{si } t \text{ está en el mes } m \\ 0, & \text{en caso contrario} \end{cases}$$

Nótese que, para cada alimento i , β_i mide la elasticidad (o *pass-through* contemporáneo) del precio minorista respecto del precio mayorista. Esto implica que β_i captura el margen de comercialización implícito.

Resultados:

M6: Observed vs Predicted Retail Price (levels) — Month dummies
95% CI in test | Page 1 of 9



4.4. Metodología IV: Regresión lineal en niveles y primeras diferencias

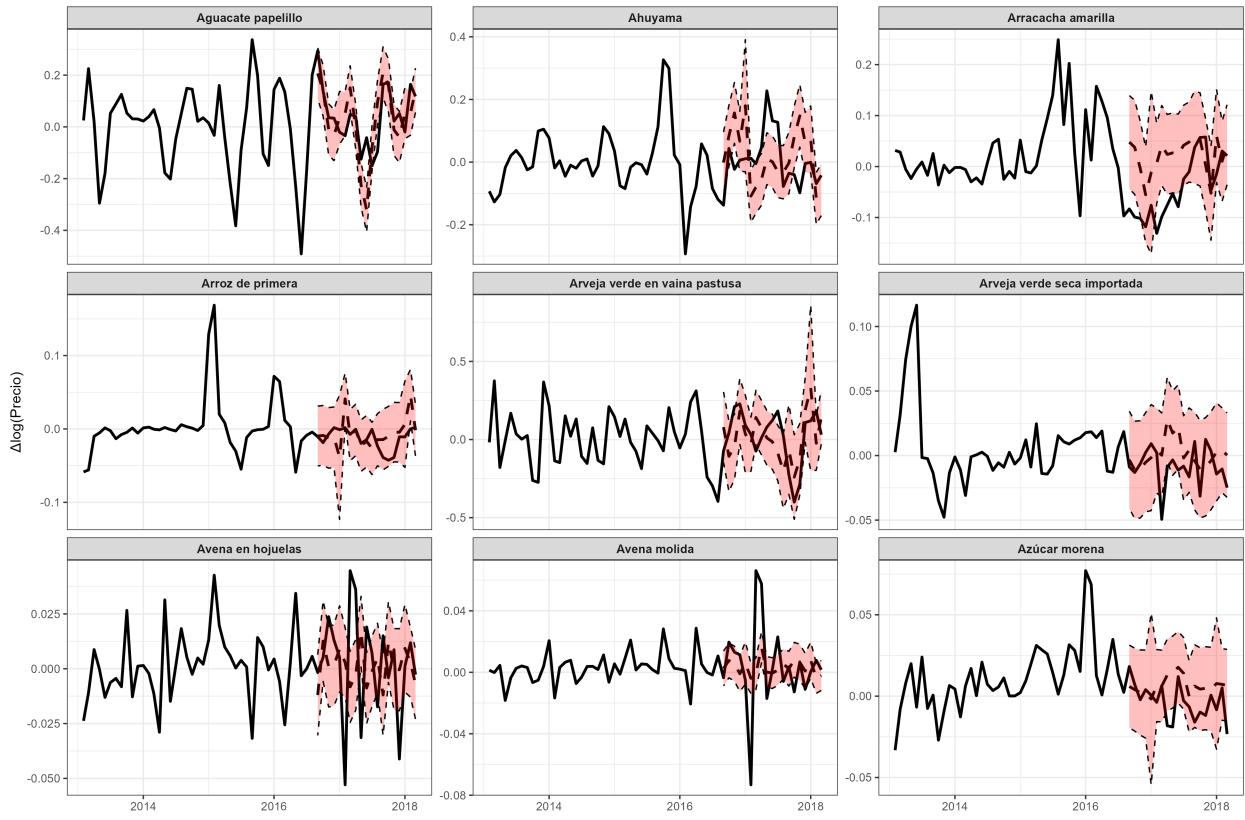
A partir de pruebas ADF, se verificó que los precios de los alimentos corresponden a series I(1). La estimación en niveles puede verse afectado por problemas de no estacionariedad y correlaciones espurias. Para garantizar series covarianza-estacionarias, se estima la siguiente especificación:

$$\Delta \ln(P_{it}^{min}) = \alpha_i + \beta_i \Delta \ln(P_{it}^{may}) + \sum_{m=1}^{11} \delta_{im} S_{mt} + e_{it}$$

En este caso, β_i captura la respuesta inmediata del crecimiento del precio minorista ante cambios en el precio mayorista, es decir, captura el *pass-through* de corto plazo.

Resultados:

M7: $\Delta \log(P_{\min}) \sim \Delta \log(P_{\text{may}}) + \text{month dummies}$
 Observed vs Predicted ($\Delta \log$) | Page 1 of 8



4.5. Metodología V: Modelo de Corrección de Error (ECM)

Como señala Enders (2014), una característica de las variables cointegradas es que su trayectoria temporal está determinada por la desviación respecto del equilibrio de largo plazo. Esto implica que las dinámicas de corto plazo deben estar influenciadas por su desviación respecto de la relación de largo plazo. El modelo dinámico utilizado para capturar simultáneamente la relación de largo plazo y los ajustes dinámicos de corto plazo es el modelo de corrección de error (ECM).

La relación de equilibrio de largo plazo entre ambos precios viene dada por:

$$\ln(P_{it}^{\min}) = \alpha_i + \beta_i \ln(P_{it}^{\text{may}}) + e_{it}$$

¿Se incorpora el componente estacional?

Así, el término de corrección de error corresponde al residuo rezagado de la relación de cointegración:

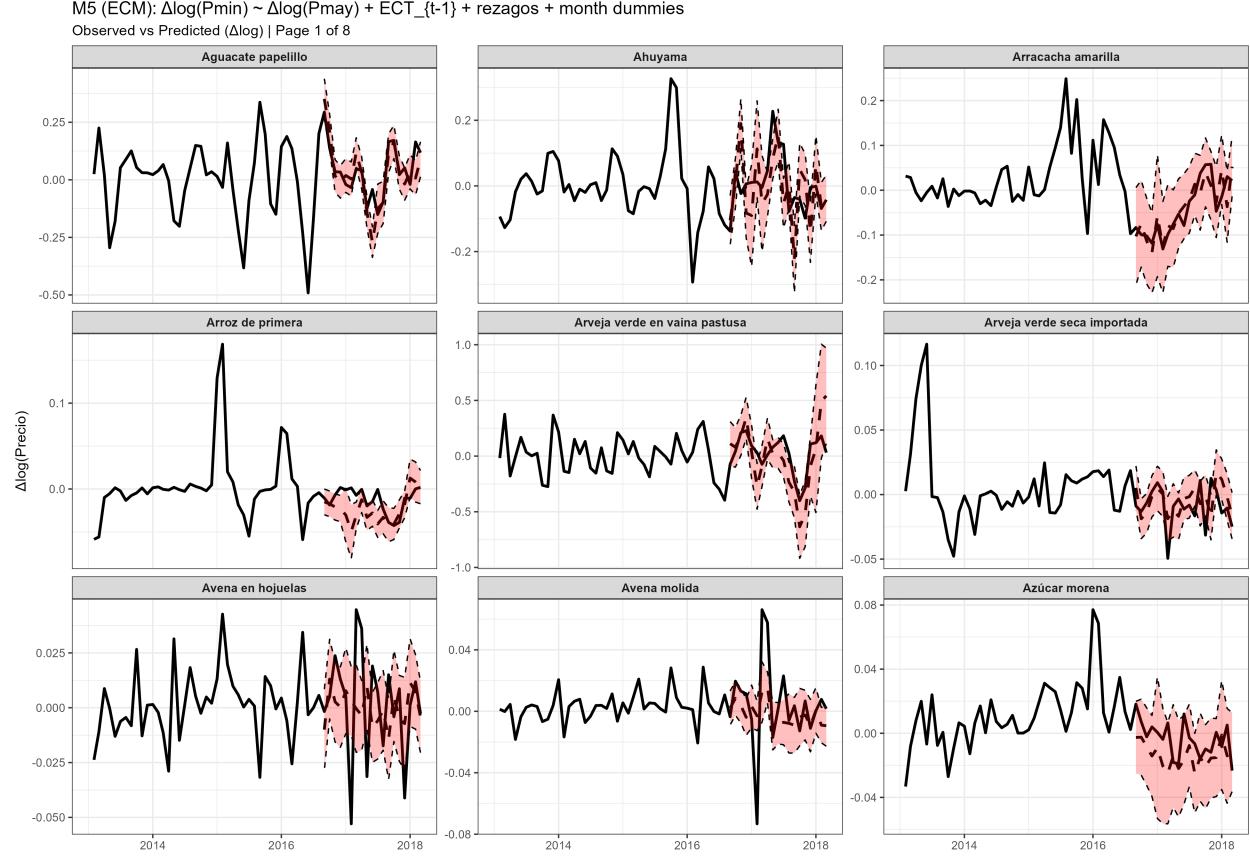
$$e_{i,t-1} = \ln(P_{i,t-1}^{\min}) - \alpha_i - \beta_i \ln(P_{i,t-1}^{\text{may}})$$

Note que la expresión anterior captura el desequilibrio en $t-1$. Si $e_{i,t-1} > 0$, el precio minorista está por encima del nivel consistente con el equilibrio de largo plazo (dado el precio mayorista). En consecuencia, la dinámica de corto plazo se captura mediante la siguiente ecuación ECM:

$$\Delta \ln(P_{it}^{\min}) = c_{i0} + \sum_{p=1}^P \beta_{ip} \Delta \ln(P_{i,t-p}^{\min}) + \sum_{q=1}^Q \gamma_{iq} \Delta \ln(P_{i,t-q}^{\text{may}}) + \theta_i e_{i,t-1} + u_t$$

Naturalmente, β_{ip} captura la dinámica autorregresiva de corto plazo del precio minorista, y γ_{iq} captura el impacto de corto plazo de variaciones en el precio mayorista. En general, θ_i es una medida de la velocidad de ajuste hacia el equilibrio de largo plazo. (Se espera que $\theta_i < 0$, es decir, que desviaciones positivas sean corregidas mediante reducciones en el crecimiento del precio minorista).

Resultados:



4.6. Metodología VI: Modelo de Corrección de Error (A-ECM) Asimétrico

En tercer lugar, se estima una versión del modelo ECM que permite especificar un modelo individual para cada ciudad principal. Siguiendo el estudio de Chesnes (2010), se propone la implementación de un modelo ECM para capturar las posibles asimetrías en la velocidad y el patrón de ajuste ante aumento y reducciones en el precio mayorista.

Sea P_t^R el precio minorista observado en la ciudad de análisis en el período t , y sea P_t^W el precio mayorista correspondiente. El modelo sigue la forma:

$$\begin{aligned} \Delta P_t^{min} = & \sum_{i=0}^{L_1^+} \beta_{1i}^+ \Delta^+ P_{t-i}^{may} + \sum_{i=0}^{L_1^-} \beta_{1i}^- \Delta^- P_{t-i}^{may} + \sum_{i=1}^{L_2^+} \beta_{2i}^+ \Delta^+ P_{t-i}^{min} + \sum_{i=1}^{L_2^-} \beta_{2i}^- \Delta^- P_{t-i}^{min} \\ & + \beta_3^+ (P_{t-1}^{min} - \gamma_0 - \gamma_1 P_{t-1}^{may}) + \beta_3^- (P_{t-1}^{min} - \gamma_0 - \gamma_1 P_{t-1}^{may}) + \varepsilon_t \end{aligned}$$

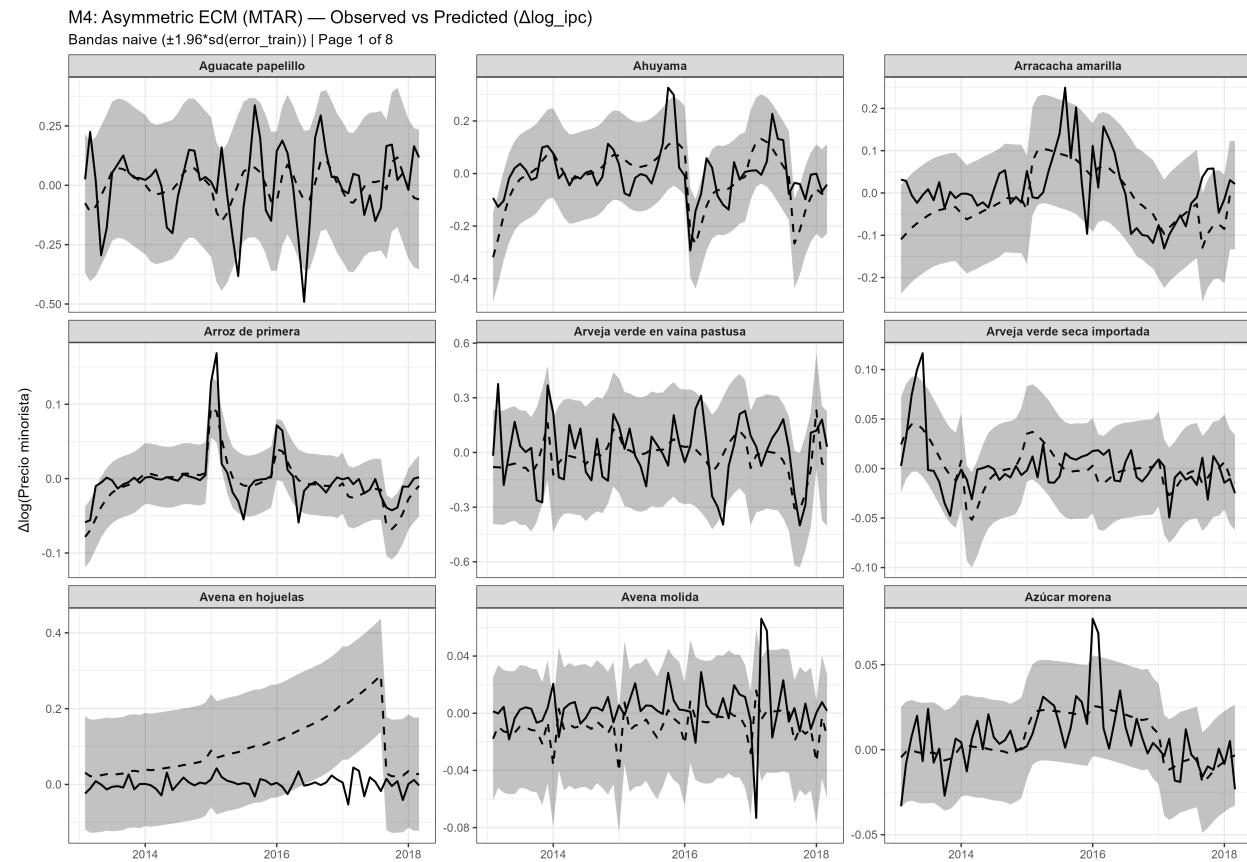
La expresión z_{t-1} corresponde al término de corrección de error y captura la relación de largo plazo entre el precio minorista y el precio mayorista. Así, se espera que ambos β_3 sean negativos: si el precio minorista

está por encima $z_{t-1} > 0$ (por debajo $z_{t-1} < 0$ del equilibrio, los precios minoristas deberían disminuir (aumentar)).

Siguiendo la metodología en dos etapas de Engle & Granger (1987), se estima γ_0 y γ_1 a partir de la siguiente ecuación de largo plazo:

$$P_{t-1}^{min} = \gamma_0 + \gamma_1 P_{t-1}^{may} + z_{t-1}.$$

Resultados:

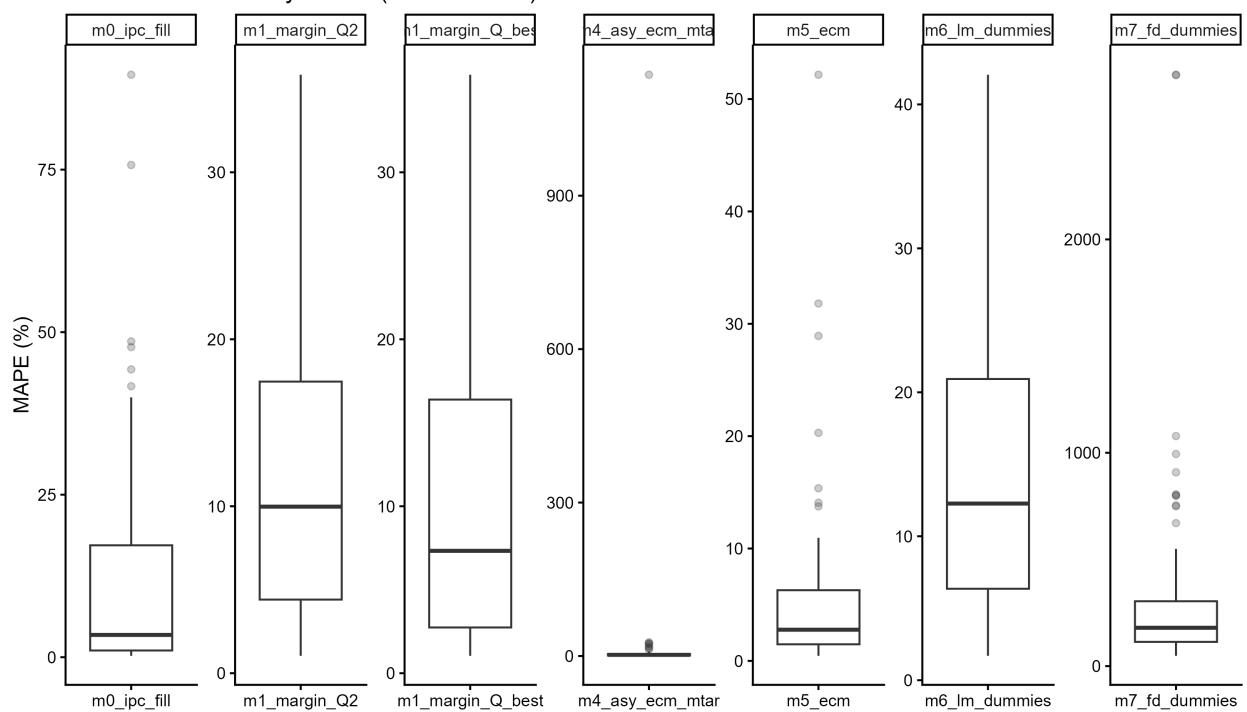


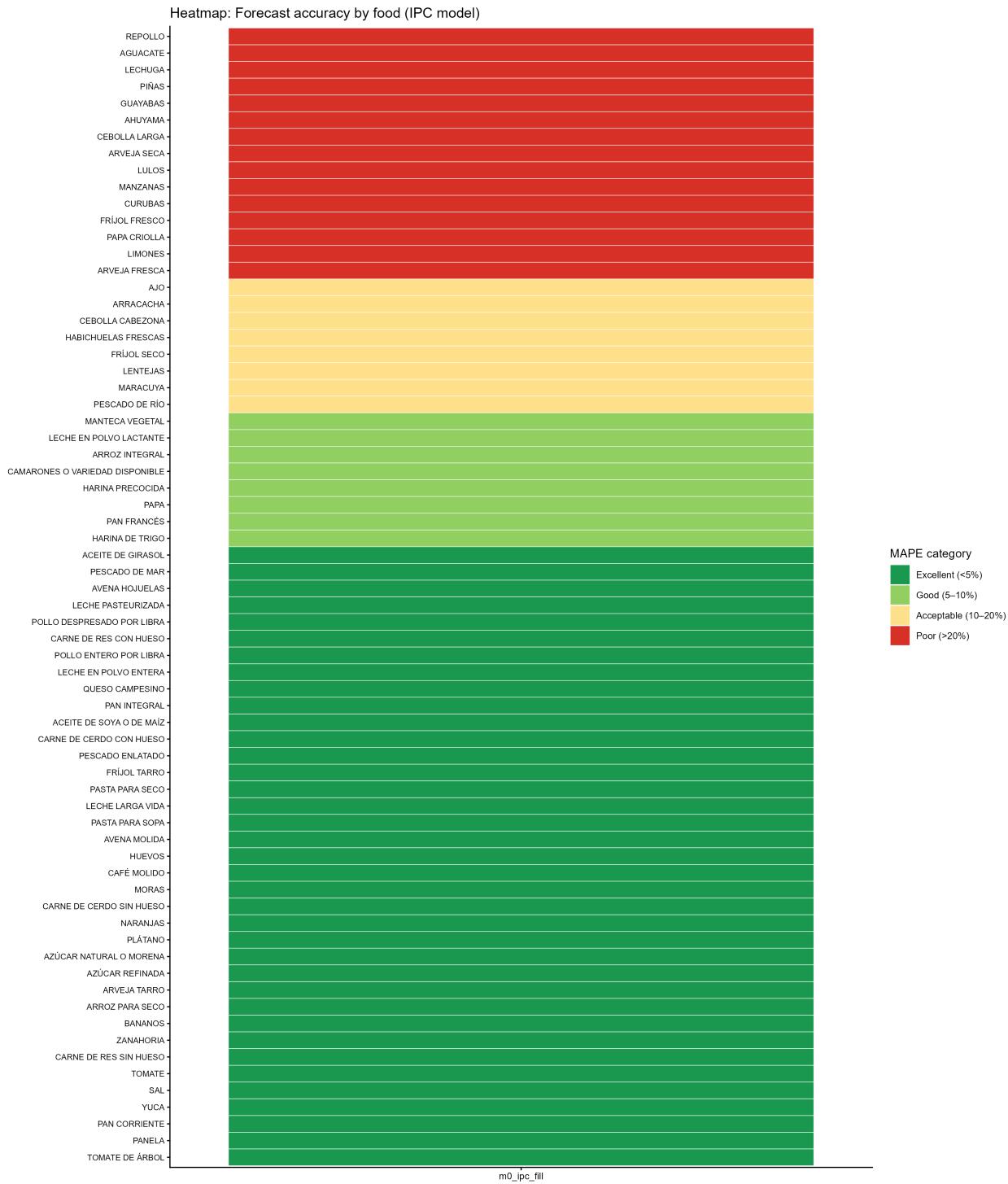
5. Discusión

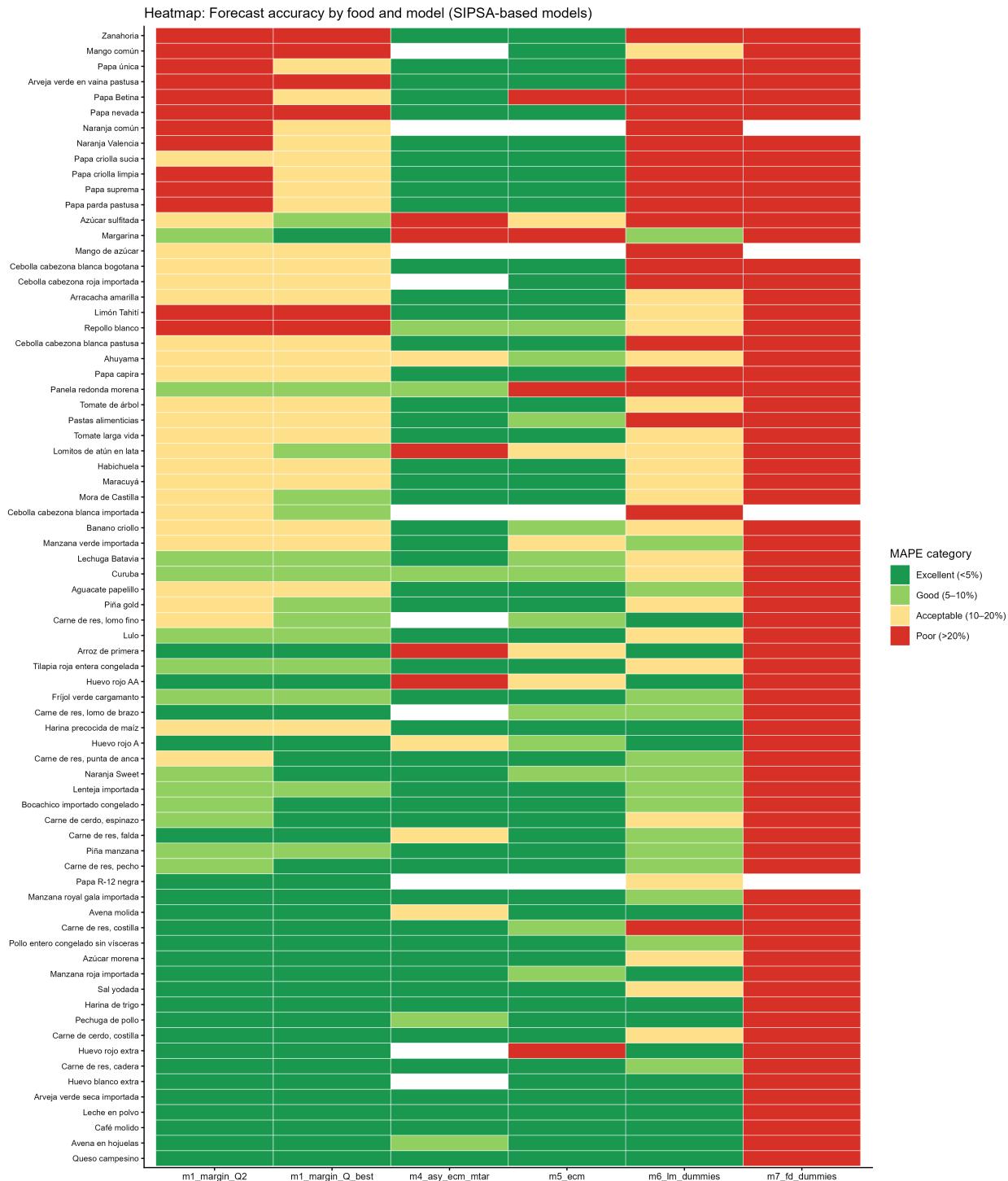
5.1. Métricas de validación

La siguiente tabla muestra las métricas obtenidas a partir de cada metodología:

MAPE distribution by model (across foods)







5.2. Márgenes de comercialización implícitos

5.3. Validación con márgenes implícitos

6. Conclusiones