

# QHAWARINA

## Documento Metodológico

### Nowcasting Económico para el Perú

Carlos César Chávez Padilla

Center for the Economics of Human Development

University of Chicago

qhawarina.pe · Versión 1.0 — Marzo 2026

#### Resumen

QHAWARINA es una plataforma de datos abiertos que produce estimaciones diarias de indicadores macroeconómicos para el Perú: un índice de precios de supermercados (estilo Billion Prices Project), un índice de riesgo político basado en medios de comunicación, un nowcast del PBI trimestral, y una proyección de pobreza monetaria. Este documento detalla la metodología de cada indicador, incluyendo las fórmulas de cálculo, las fuentes de datos, los supuestos, la validación y las limitaciones conocidas. Todo el código fuente y los datos están disponibles bajo licencia abierta.

**Palabras clave:** nowcasting, inflación, riesgo político, PBI, pobreza, Perú, datos abiertos

**Clasificación JEL:** C53, C82, E31, E37, I32

---

Datos bajo licencia CC BY 4.0. Código disponible en <https://github.com/cesarchavezp29/qhawarina>

## Índice

<b>1. Índice de Precios Diarios</b>	<b>3</b>
1.1. Descripción general	3
1.2. Fuentes de datos	3
1.3. Construcción del índice	3
1.3.1. Emparejamiento de productos	3
1.3.2. Filtro de ratios extremos	3
1.3.3. Índice Jevons por categoría	4
1.3.4. Encadenamiento	4
1.3.5. Agregación ponderada por IPC	4
1.3.6. Tasa de inflación interanual	4
1.4. Cobertura y limitaciones	5
1.5. Referencia metodológica	5

<b>2. Índice de Riesgo Político</b>	<b>5</b>
2.1. Descripción general . . . . .	5
2.2. Fuentes de datos . . . . .	5
2.3. Clasificación de artículos . . . . .	6
2.4. Construcción del índice (Versión 2 — EPU) . . . . .	6
2.4.1. Paso 1: Estandarización por fuente . . . . .	6
2.4.2. Paso 2: Estandarización y agregación ponderada por volumen . . . . .	6
2.4.3. Paso 3: Normalización EPU . . . . .	7
2.4.4. Paso 4: Suavizado . . . . .	7
2.4.5. Escala del indicador en la plataforma . . . . .	7
2.4.6. Clasificación de niveles . . . . .	7
2.5. Factor principal . . . . .	7
2.6. Limitaciones . . . . .	8
<b>3. Nowcast de PBI</b>	<b>8</b>
3.1. Descripción general . . . . .	8
3.2. Modelo . . . . .	8
3.2.1. Paso 1: Extracción de factores . . . . .	8
3.2.2. Paso 2: Ecuación puente . . . . .	8
3.2.3. Ventana de estimación . . . . .	9
3.2.4. Manejo del borde irregular (ragged edge) . . . . .	9
3.3. Indicadores utilizados . . . . .	9
3.4. Validación . . . . .	9
3.5. Limitaciones . . . . .	9
<b>4. Nowcast de Inflación</b>	<b>10</b>
4.1. Descripción general . . . . .	10
4.2. Modelo . . . . .	10
4.3. Desempeño . . . . .	10
4.4. Limitaciones . . . . .	10
<b>5. Proyección de Pobreza Monetaria</b>	<b>10</b>
5.1. Descripción general . . . . .	11
5.2. Modelo . . . . .	11
5.3. Desagregación temporal: Chow-Lin . . . . .	11
5.4. Intervalos de confianza . . . . .	11
5.5. Pobreza distrital . . . . .	12
5.6. Limitaciones . . . . .	12
<b>6. Infraestructura de Datos</b>	<b>12</b>
6.1. Pipeline de datos . . . . .	12
6.2. Degradación elegante . . . . .	13
6.3. Trazabilidad de vintages . . . . .	13
<b>7. Datos Abiertos y Reproducibilidad</b>	<b>13</b>

# 1. Índice de Precios Diarios

## 1.1 Descripción general

El Índice de Precios Diarios de QHAWARINA mide la variación diaria de precios de bienes de consumo vendidos en supermercados de Lima. Está basado en la metodología del Billion Prices Project (Cavallo & Rigobon, 2016) y la experiencia de Inflación Verdadera en Argentina (Cavallo, 2013).

El INEI publica el Índice de Precios al Consumidor (IPC) de Lima con frecuencia mensual. Nuestro índice ofrece una señal complementaria de alta frecuencia, actualizada diariamente, basada en precios observados directamente en puntos de venta.

## 1.2 Fuentes de datos

Recolectamos precios individuales de tres cadenas de supermercados en Lima metropolitana:

Cadena	Grupo	Productos aprox.
Plaza Vea	InRetail	~15,000
Metro	Cencosud	~15,000
Wong	Cencosud	~12,000

Cuadro 1: Fuentes del índice de precios

La recolección se realiza diariamente mediante scraping automatizado. Cada snapshot captura aproximadamente 42,000 productos con nombre, SKU, categoría, precio y tienda de origen.

## 1.3 Construcción del índice

### 1.3.1 Emparejamiento de productos

Para cada día  $t$ , identificamos los productos observados tanto en  $t$  como en  $t - 1$  usando el par (tienda, SKU). Solo los productos emparejados entran al cálculo. Los productos nuevos o discontinuados se excluyen de ese día, siguiendo el enfoque de “matched-model index” de Cavallo (2013).

### 1.3.2 Filtro de ratios extremos

Excluimos cualquier producto cuyo ratio de precios esté fuera del rango  $[0,5, 2,0]$ :

$$0,5 < \frac{p_{i,t}}{p_{i,t-1}} < 2,0$$

Esto elimina errores de scraping, cambios de presentación y promociones extremas que no reflejan movimientos genuinos de precios.

### 1.3.3 Índice Jevons por categoría

Para cada categoría  $c$  (carne, arroz y cereales, aceites, lácteos, etc.), calculamos un índice Jevons bilateral — la media geométrica de los ratios de precios de todos los productos emparejados:

$$J_{c,t} = \left( \prod_{i \in \mathcal{M}_{c,t}} \frac{p_{i,t}}{p_{i,t-1}} \right)^{1/N_{c,t}}$$

donde  $\mathcal{M}_{c,t}$  es el conjunto de productos emparejados en la categoría  $c$  en el día  $t$ , y  $N_{c,t} = |\mathcal{M}_{c,t}|$ .

### 1.3.4 Encadenamiento

El índice de cada categoría se encadena multiplicativamente:

$$I_{c,t} = I_{c,t-1} \times J_{c,t}, \quad I_{c,0} = 100$$

Para días sin observación (fines de semana, feriados), aplicamos interpolación lineal.

### 1.3.5 Agregación ponderada por IPC

El índice agregado se calcula como una media geométrica ponderada de los índices por categoría, utilizando las ponderaciones de gasto del IPC de Lima publicadas por INEI (base diciembre 2021):

$$I_t = \prod_c (I_{c,t})^{w_c / \sum_c w_c}$$

donde  $w_c$  es la ponderación del IPC para la categoría  $c$ . Esto asegura que el índice refleje la estructura de consumo real: un alza en arroz pesa más que un alza idéntica en productos de limpieza.

Categoría	Ponderación IPC (%)
Carnes y preparados de carne	5.8
Arroz y cereales	4.2
Lácteos y huevos	4.1
Verduras y legumbres	2.3
Frutas	2.1
Bebidas no alcohólicas	2.3
Cuidado personal	2.1
Productos de limpieza	1.8
Aceites y grasas	1.4

Cuadro 2: Ponderaciones IPC aplicadas (selección). Fuente: INEI, estructura IPC Lima dic-2021. Los valores corresponden a las categorías mapeadas desde los productos de supermercados.

### 1.3.6 Tasa de inflación interanual

Siguiendo la metodología estándar del BPP (Cavallo & Rigobon, 2016), calculamos la inflación interanual como:

$$\pi_t^{YoY} = \left( \frac{\bar{I}_t}{\bar{I}_{t-365}} - 1 \right) \times 100$$

donde  $\bar{I}_t$  es el promedio móvil de 30 días del índice:

$$\bar{I}_t = \frac{1}{30} \sum_{s=0}^{29} I_{t-s}$$

Este suavizado elimina el ruido diario y produce una serie directamente comparable con el IPC mensual del INEI.

## 1.4 Cobertura y limitaciones

- **Incluye:** Bienes de consumo vendidos en supermercados formales de Lima — alimentos, bebidas, productos de limpieza, cuidado personal.
- **No incluye:** Servicios (salud, educación, transporte, alquileres), mercados tradicionales, bodegas, comercio informal. El IPC de Lima del INEI cubre todos estos rubros.
- **Cobertura geográfica:** Lima metropolitana únicamente.
- **Sesgo de cobertura:** Los supermercados formales representan aproximadamente el 30 % del gasto en alimentos de los hogares limeños. Los precios en mercados tradicionales pueden comportarse de manera diferente, especialmente para productos frescos.
- **Fragilidad del scraping:** Si un supermercado rediseña su sitio web, la recolección puede interrumpirse temporalmente. El sistema de validación detecta estas fallas y mantiene el último índice válido hasta la recuperación.

## 1.5 Referencia metodológica

La metodología sigue explícitamente a [Cavallo \(2013\)](#) y [Cavallo & Rigobon \(2016\)](#). La principal diferencia es que nuestro índice aplica ponderaciones del IPC para la agregación, mientras que el BPP original usaba ponderaciones iguales en la mayoría de los países.

# 2. Índice de Riesgo Político

## 2.1 Descripción general

El Índice de Riesgo Político mide la intensidad diaria de la inestabilidad política en el Perú, construido a partir del análisis de contenido de medios de comunicación. El diseño se inspira en el Economic Policy Uncertainty Index de [Baker, Bloom & Davis \(2016\)](#), adaptado al contexto político peruano.

## 2.2 Fuentes de datos

Monitoreamos 6 fuentes RSS de 3 medios de comunicación peruanos:

Medio	Secciones	Tipo
El Comercio	Política, Economía	Diario nacional
Gestión	Economía, General	Diario económico
La República	Política, Economía	Diario nacional

Cuadro 3: Fuentes RSS del índice de riesgo político

Se recopilan aproximadamente 60–100 artículos diarios. Se eliminan duplicados por hash de URL y similitud de título.

### 2.3 Clasificación de artículos

Cada artículo se clasifica mediante Claude Haiku (Anthropic, modelo claude-haiku-4-5-20251001) en dos dimensiones:

1. **Categoría:** político, económico, ambos, o irrelevante.
2. **Severidad:** escala 1–5 (mínimo, bajo, moderado, alto, crítico).

Los artículos se clasifican en lotes de 10 para eficiencia. Los resultados se almacenan en caché para evitar reclasificación.

### 2.4 Construcción del índice (Versión 2 — EPU)

La versión actual del índice (v2) sigue un enfoque estilo EPU con estandarización por fuente y ponderación por volumen:

#### 2.4.1 Paso 1: Estandarización por fuente

Para cada fuente  $s$ , calculamos la proporción diaria de artículos políticos ponderados por severidad (SWP):

$$SWP_{s,t} = \frac{\sum_{i \in \mathcal{P}_{s,t}} \text{severity}_i / 3}{N_{s,t}}$$

donde  $\mathcal{P}_{s,t}$  son los artículos clasificados como políticos o “ambos” de la fuente  $s$  en el día  $t$ , y  $N_{s,t}$  es el total de artículos de esa fuente.

#### 2.4.2 Paso 2: Estandarización y agregación ponderada por volumen

Cada fuente se estandariza dividiendo por su desviación estándar histórica:

$$Y_{s,t} = \frac{SWP_{s,t}}{\sigma_s}$$

donde  $\sigma_s$  es la desviación estándar de  $SWP_{s,\cdot}$  sobre todo el período disponible.

Se agregan las fuentes mediante una media ponderada por volumen de artículos:

$$Z_t = \sum_s w_{s,t} Y_{s,t}, \quad w_{s,t} = \frac{N_{s,t}}{\sum_{s'} N_{s',t}}$$

### 2.4.3 Paso 3: Normalización EPU

El índice agregado se normaliza para tener media 100 sobre el período de referencia:

$$EPU_t = \frac{Z_t}{\bar{Z}} \times 100$$

donde  $\bar{Z}$  es la media de  $Z_t$  sobre todo el período disponible. Un valor de 150 indica un nivel de riesgo 50 % superior al promedio histórico. El índice no tiene techo fijo — puede superar 100.

### 2.4.4 Paso 4: Suavizado

Se aplica un promedio móvil centrado de 7 días:

$$EPU_t^{smooth} = \frac{1}{7} \sum_{s=-3}^3 EPU_{t+s}$$

### 2.4.5 Escala del indicador en la plataforma

Para la visualización en el sitio web, se construye un índice compuesto de inestabilidad que combina los componentes político y económico:

$$C_t = 0,6 \times S_t^{pol} + 0,4 \times S_t^{econ}$$

donde  $S_t^{pol}$  y  $S_t^{econ}$  son las puntuaciones SWP suavizadas (escala 0–1) de los artículos clasificados como políticos y económicos, respectivamente. La ponderación 60/40 refleja el mayor peso relativo de la inestabilidad política en el contexto peruano.

El compuesto  $C_t$  está acotado entre 0 y 1. Se reporta multiplicado por 100 en el sitio web (escala 0–100). La clasificación de niveles (Mínimo a Crítico) se aplica sobre  $C_t \times 100$ , no sobre el índice EPU directamente.

### 2.4.6 Clasificación de niveles

Nivel	Rango	Interpretación
Mínimo	[0, 15)	Estabilidad política alta
Bajo	[15, 35)	Tensiones menores
Medio	[35, 60)	Tensiones moderadas
Alto	[60, 80)	Inestabilidad significativa
Crítico	[80, 100]	Crisis política activa

Cuadro 4: Niveles del índice de riesgo político

## 2.5 Factor principal

Para días con un compuesto SWP suavizado superior a 0.6 en la escala 0–1 (correspondiente a los niveles Alto o Crítico), se identifican los 3 artículos de mayor severidad y se genera una

frase resumen de 5–10 palabras mediante Claude Haiku, describiendo el factor principal de inestabilidad. Esta frase se reporta como “Factor principal hoy” en la plataforma.

## 2.6 Limitaciones

- **Cobertura limitada:** Solo 3 medios de comunicación. Una cobertura más amplia (10–15 medios) mejoraría la robustez del índice.
- **Sesgo de medios:** Los medios seleccionados pueden tener inclinaciones editoriales que afectan la detección de riesgo.
- **Dependencia de NLP:** La clasificación por Claude Haiku es un modelo de caja negra. Aunque fijamos la versión del modelo (claude-haiku-4-5-20251001) para reproducibilidad, la clasificación puede variar si se actualiza a versiones futuras. No se ha realizado una evaluación formal de precisión (e.g., F1-score) del clasificador — esta es una prioridad futura que requiere un conjunto de datos etiquetados manualmente.
- **Sin validación externa consolidada:** El índice aún no se ha validado sistemáticamente contra medidas institucionales como el ICRG o el EMBI+ spread. Esta validación es una prioridad futura.

## 3. Nowcast de PBI

### 3.1 Descripción general

Estimamos el crecimiento interanual del PBI trimestral del Perú en tiempo real, combinando un Modelo de Factores Dinámicos (DFM) con ecuaciones puente tipo Ridge. La estimación se actualiza semanalmente conforme se publican nuevos indicadores mensuales.

### 3.2 Modelo

#### 3.2.1 Paso 1: Extracción de factores

A partir de un panel de  $K$  indicadores mensuales  $\mathbf{x}_t = (x_{1,t}, \dots, x_{K,t})'$ , extraemos  $r = 3$  factores comunes mediante Análisis de Componentes Principales:

$$x_{i,t} = \lambda_i' f_t + e_{i,t}$$

donde  $f_t \in \mathbb{R}^3$  son los factores latentes,  $\lambda_i$  son las cargas factoriales, y  $e_{i,t}$  es el componente idiosincrásico.

#### 3.2.2 Paso 2: Ecuación puente

Los factores mensuales se agregan a frecuencia trimestral y se estima una regresión Ridge ( $\alpha = 1,0$ ) mapeando factores a crecimiento del PBI:

$$\Delta \ln Y_t^q = \beta_0 + \beta_1 \bar{f}_{1,t}^q + \beta_2 \bar{f}_{2,t}^q + \beta_3 \bar{f}_{3,t}^q + \varepsilon_t$$

donde  $\bar{f}_{j,t}^q$  es el promedio trimestral del factor  $j$  y  $\Delta \ln Y_t^q$  es el crecimiento interanual del PBI.

### 3.2.3 Ventana de estimación

Se utiliza una ventana rodante de 7 años para la estimación, lo cual permite capturar cambios estructurales — en particular, el quiebre asociado al COVID-19.

### 3.2.4 Manejo del borde irregular (ragged edge)

Los meses con cobertura inferior al 50% de los indicadores se excluyen del cálculo. Los factores se estiman solo con los meses que tienen cobertura suficiente.

## 3.3 Indicadores utilizados

El panel incluye aproximadamente 30–34 series mensuales del BCRP, INEI, SUNAT y otras fuentes:

Categoría	Series	Fuente
Producción	Producción industrial, electricidad, cemento	BCRP
Comercio	Exportaciones, importaciones (valor, volumen)	BCRP
Fiscal	Recaudación tributaria (IGV, renta)	SUNAT
Financiero	Crédito al sector privado, tipo de cambio	BCRP
Laboral	Empleo urbano (EPE)	INEI
Confianza	Índice de confianza del consumidor	BCRP
Alta frecuencia	Índice de precios de supermercados (QHAWARINA)	Propio
Satelital	Luminosidad nocturna (NTL)	VIIRS

Cuadro 5: Panel de indicadores para el nowcast de PBI

## 3.4 Validación

El modelo se evalúa mediante un ejercicio pseudo-tiempo-real: para cada trimestre histórico, se reconstruye el conjunto de información que habría estado disponible en la fecha de publicación, se estima el modelo con la ventana rodante, y se compara la predicción con el dato oficial del INEI.

Los resultados del backtest (excluyendo los trimestres afectados por COVID-19, 2020-Q1 a 2021-Q4) se reportan en la plataforma junto con el nowcast actual.

## 3.5 Limitaciones

- **Modelo lineal:** El DFM asume una estructura factorial lineal. En períodos de crisis (COVID, conflictos sociales), las relaciones pueden ser no lineales.
- **Muestra corta:** Con ~80–100 trimestres desde el año 2000, los grados de libertad son limitados para la ecuación puente.
- **Rezagos de publicación:** Algunos indicadores (empleo, producción industrial) se publican con 2–3 meses de retraso, limitando la información disponible para el trimestre en curso.

- **Sin descomposición sectorial completa:** La ecuación puente no produce una descomposición por sector económico comparable a la del INEI. Las contribuciones sectoriales reportadas son aproximaciones.

## 4. Nowcast de Inflación

### 4.1 Descripción general

Además del índice de precios diarios de supermercados (Sección 1), estimamos la variación mensual del IPC oficial del INEI mediante un modelo de factores dinámicos aplicado a indicadores mensuales. Esta estimación complementa el índice BPP: mientras el BPP mide precios de supermercados directamente, el nowcast de inflación predice el IPC oficial — que incluye servicios, alquileres y otros rubros no cubiertos por el scraping.

### 4.2 Modelo

Utilizamos un DFM similar al del nowcast de PBI (Sección 3), pero con la variación mensual del IPC como variable objetivo. El modelo extrae factores de un panel de indicadores mensuales y estima una ecuación puente:

$$\pi_t^{IPC} = \beta_0 + \beta' \bar{f}_t + \varepsilon_t$$

donde  $\pi_t^{IPC}$  es la variación mensual del IPC y  $\bar{f}_t$  son los factores extraídos del mismo panel de indicadores del BCRP.

La inflación reportada es el promedio móvil de 3 meses de la variación mensual predicha.

### 4.3 Desempeño

El RMSE del backtest es de aproximadamente 0.32 puntos porcentuales sobre la variación mensual. Esto implica que si el modelo predice una inflación mensual de 0.5 %, el intervalo de confianza al 95 % es aproximadamente [−0.13 %, 1.13 %].

### 4.4 Limitaciones

- **Alcance limitado vs. el índice BPP:** El nowcast de inflación predice el IPC oficial, que se publica mensualmente. El índice BPP (Sección 1) ofrece señales diarias pero solo cubre supermercados. Ambos indicadores son complementarios.
- **Sensibilidad a shocks de oferta:** El modelo puede subestimar la inflación en meses con shocks de oferta (e.g., fenómeno El Niño) que afectan precios de alimentos frescos no capturados en los indicadores mensuales del panel.

## 5. Proyección de Pobreza Monetaria

## 5.1 Descripción general

Proyectamos la tasa de pobreza monetaria a nivel nacional y departamental, utilizando la relación histórica entre pobreza (medida por ENAHO) y un conjunto de indicadores macroeconómicos. Actualizamos la proyección conforme se actualizan los indicadores subyacentes.

**Nota metodológica importante:** La pobreza monetaria es una estadística distribucional que depende de la distribución completa del consumo per cápita, observada solamente en la ENAHO anual. Cualquier actualización entre encuestas es una *proyección* basada en correlaciones históricas, no una medición directa. Usamos el término “proyección” deliberadamente.

## 5.2 Modelo

Utilizamos un *Gradient Boosting Regressor* (GBR) entrenado en un panel departamental:

$$\Delta\text{Pobreza}_{d,t} = f(\Delta\text{PBI}_{d,t}, \pi_{d,t}^{\text{food}}, \text{Agri}_{d,t}, \text{Empleo}_{d,t}, \text{NTL}_{d,t}) + \varepsilon_{d,t}$$

donde  $\Delta\text{Pobreza}_{d,t}$  es el cambio interanual en la tasa de pobreza del departamento  $d$ , y las variables explicativas son indicadores departamentales de crecimiento económico, inflación alimentaria, producción agrícola, empleo e intensidad de luminosidad nocturna satelital.

La tasa de pobreza proyectada se obtiene sumando el cambio predicho a la tasa del año anterior:

$$\widehat{\text{Pobreza}}_{d,t} = \text{Pobreza}_{d,t-1} + \Delta\widehat{\text{Pobreza}}_{d,t}$$

## 5.3 Desagregación temporal: Chow-Lin

La ENAHO produce estimaciones anuales de pobreza. Para obtener señales trimestrales, aplicamos el método de desagregación temporal de [Chow & Lin \(1971\)](#). Este procedimiento distribuye el valor anual proyectado en cuatro trimestres, utilizando indicadores trimestrales relacionados (PBI trimestral, empleo trimestral) como variables auxiliares. El resultado respeta la restricción de agregación temporal: la suma de los cuatro trimestres es consistente con la estimación anual.

Es importante notar que la desagregación temporal (Chow-Lin, anual  $\rightarrow$  trimestral) y la desagregación espacial (NTL, departamento  $\rightarrow$  distrito) son operaciones independientes que se aplican en diferentes dimensiones.

## 5.4 Intervalos de confianza

Reportamos intervalos de confianza al 95 % basados en el RMSE del backtest:

$$\text{IC}_{95\%} = \widehat{\text{Pobreza}}_{d,t} \pm 1,96 \times \text{RMSE}$$

donde el RMSE del backtest es aproximadamente 2.5 puntos porcentuales. Este intervalo es amplio — lo reportamos para reflejar honestamente la incertidumbre inherente al ejercicio.

## 5.5 Pobreza distrital

Las estimaciones distritales utilizan interpolación dasimétrica con luminosidad nocturna. La base es la tasa de pobreza del departamento correspondiente, no la nacional:

$$\widehat{\text{Pobreza}}_{\text{distrito}} \approx \widehat{\text{Pobreza}}_{\text{depto}} + \omega \times \left( \frac{\text{NTL}_{\text{distrito}}}{\text{NTL}_{\text{depto}}} \right)$$

donde NTL es la intensidad de luminosidad nocturna del VIIRS y  $\widehat{\text{Pobreza}}_{\text{depto}}$  es la proyección departamental del modelo GBR. Así, un distrito en Huancavelica se desagrega a partir de la tasa de Huancavelica, no del promedio nacional. Esta es una aproximación espacial, no una estimación directa — los resultados distritales deben interpretarse como indicativos, no como mediciones precisas.

## 5.6 Limitaciones

- **Muestra pequeña:** Con ~20 años de ENAHO (2004–2024), los grados de libertad a nivel nacional son muy limitados. El panel departamental (~500 observaciones) alivia parcialmente esta restricción.
- **Intervalos amplios:** El RMSE de ~2.5pp implica que una proyección de 24 % tiene un intervalo real de aproximadamente [19 %, 29 %]. Esto es inherente al ejercicio, no un defecto del modelo.
- **Sin microsimulación:** No simulamos el efecto de cambios de precios sobre cada hogar individualmente. Un enfoque de microsimulación (aplicando cambios de precios del BPP a los microdatos de ENAHO) es un objetivo futuro.
- **Pobreza distrital es aproximada:** La interpolación por NTL no reemplaza una estimación de áreas pequeñas (SAE) como la de [Elbers, Lanjouw & Lanjouw \(2003\)](#).

## 6. Infraestructura de Datos

### 6.1 Pipeline de datos

La plataforma opera un pipeline diario automatizado:

1. **Ingesta:** Scraping de supermercados (~35K productos), recolección de RSS (~80 artículos), descarga de series BCRP/INEI.
2. **Validación:** Antes de cada paso de procesamiento, se verifican: conteo de productos ( $\geq 70\%$  del día anterior), cobertura de fuentes ( $\geq 2/3$  supermercados,  $\geq 3/6$  feeds), distribución de precios (mediana de cambio  $< 10\%$ ), y frescura de datos.
3. **Procesamiento:** Cálculo de índices, modelos de nowcasting, clasificación NLP.
4. **Exportación:** Generación de archivos JSON para el frontend.
5. **Despliegue:** Push a repositorio Git → Vercel reconstruye el sitio automáticamente.
6. **Alertas:** Reporte diario por correo electrónico con el estado de cada componente.

## 6.2 Degradación elegante

Si un componente falla (e.g., un supermercado cambia su sitio web), el sistema:

- Mantiene los datos del último día válido
- Registra la falla en `pipeline_status.json`
- Muestra un aviso en el sitio: “Datos actualizados hasta el [fecha]. Actualización en proceso.”
- Envía una alerta por correo con los detalles de la falla

## 6.3 Trazabilidad de vintages

Los datos oficiales (BCRP, INEI) se almacenan con la fecha de descarga, permitiendo reconstruir el conjunto de información disponible en cualquier fecha pasada. Esto es esencial para evaluar honestamente la precisión del nowcast en tiempo real.

## 7. Datos Abiertos y Reproducibilidad

- **Licencia de datos:** CC BY 4.0
- **Código fuente:** <https://github.com/cesarchavezp29/qhawarina> (pipeline de datos, modelos y scripts de procesamiento)
- **API:** Documentación en [qhawarina.pe/api/docs](https://qhawarina.pe/api/docs)
- **Descarga directa:** Archivos JSON disponibles en [qhawarina.pe/datos](https://qhawarina.pe/datos)
- **Cita sugerida:** Chávez Padilla, C.C. (2026). “Qhawarina: Nowcasting Económico para el Perú.” Disponible en <https://qhawarina.pe>.

## Referencias

- Baker, S.R., Bloom, N. & Davis, S.J. (2016). Measuring Economic Policy Uncertainty. *Quarterly Journal of Economics*, 131(4), 1593–1636.
- Bañbura, M. & Modugno, M. (2014). Maximum Likelihood Estimation of Factor Models on Datasets with Arbitrary Pattern of Missing Data. *Journal of Applied Econometrics*, 29(1), 133–160.
- Bañbura, M., Giannone, D. & Reichlin, L. (2010). Large Bayesian Vector Auto Regressions. *Journal of Applied Econometrics*, 25(1), 71–92.
- Cavallo, A. (2013). Online and Official Price Indexes: Measuring Argentina's Inflation. *Journal of Monetary Economics*, 60(2), 152–165.
- Chow, G.C. & Lin, A. (1971). Best Linear Unbiased Interpolation, Distribution, and Extrapolation of Time Series by Related Series. *The Review of Economics and Statistics*, 53(4), 372–375.
- Cavallo, A. & Rigobon, R. (2016). The Billion Prices Project: Using Online Prices for Measurement and Research. *Journal of Economic Perspectives*, 30(2), 151–178.
- Elbers, C., Lanjouw, J.O. & Lanjouw, P. (2003). Micro-Level Estimation of Poverty and Inequality. *Econometrica*, 71(1), 355–364.
- Ghysels, E., Santa-Clara, P. & Valkanov, R. (2004). The MIDAS Touch: Mixed Data Sampling Regression Models. *CIRANO Working Paper*.
- Giannone, D., Reichlin, L. & Small, D. (2008). Nowcasting: The Real-Time Informational Content of Macroeconomic Data. *Journal of Monetary Economics*, 55(4), 665–676.