Graduate Institute of International and Development Studies

International Economics Department

Working Paper Series

Working Paper No. HEIDWP20-2026

Construction and Forecasting of the Imported Food Price Index in Azerbaijan

Tamkin Nuriyev
Central Bank of the Republic of Azerbaijan

Aygun Garayeva
Central Bank of the Republic of Azerbaijan

Gulzar Tahirova
Central Bank of the Republic of Azerbaijan

Chemin Eugène-Rigot 2
P.O. Box 136
CH - 1211 Geneva 21
Switzerland

Construction and Forecasting of the Imported Food Price Index in Azerbaijan

Tamkin Nuriyev, Aygun Garayeva, Gulzar Tahirova

Central Bank of the Republic of Azerbaijan

Abstract

Using 800,000 transaction-level customs records from January 2018 to February 2026, the paper constructs a trade-weighted Imported Food Price Index (IFPI), covering 34 items from the consumer basket with significant import dependence. The index is developed using the Fisher ideal methodology to provide a timely measure of external food price pressures. The results indicate that the IFPI leads official food Consumer Price Index (CPI) by approximately two months, with a maximum correlation of 0.81, highlighting its potential usefulness as an early indicator of domestic food inflation. Building on this, the paper develops a forecasting framework for the IFPI by combining non-parametric Binary Segmentation and Hidden Markov Models with a regularized machine-learning ensemble. The model employs an ensemble approach that combines Histogram-based Gradient Boosting Regression Tree, Random Forest, and Extreme Gradient Boosting, alongside rigorous time-series cross-validation. The optimized ensemble achieves a 58% out-of-sample R^2 relative to a random walk benchmark, vastly outperforming traditional linear Autoregressive Distributed Lag (ARDL) (13.60%) and Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX) (0.18%) baselines. The forecast results are intended to be incorporated into broader inflation forecasting models to improve short-term projections.

Keywords: Import price index; Fisher Ideal index; Food price inflation; Machine learning forecasting; Hidden Markov models.

JEL: C43, C53, C55, E31, F14

The authors thank Elliot Beck from the Swiss National Bank for the academic supervision of this paper. This research took place through the coaching program under the Bilateral Assistance and Capacity Building for Central Banks (BCC), financed by SECO, and the Graduate Institute in Geneva.

1. INTRODUCTION

Accurately forecasting imported food prices constitutes a fundamental challenge for monetary policy authorities in emerging economies, where food expenditures represent a substantial proportion of household consumption and import dependence is high, making domestic food prices particularly susceptible to external price shocks. In Azerbaijan, food items comprise approximately 43% of the consumption basket. A significant share of domestic food consumption is imported, making domestic prices highly sensitive to international commodity price movements and global supply shocks. Given the substantial contribution of imported inflation to domestic price dynamics, deficiencies in the monitoring and forecast of the external price pressures can undermine the accuracy of domestic inflation forecasts. Therefore, accurate estimates of import price dynamics are crucial for anticipating inflationary pressures and calibrating monetary policy responses.

The challenge extends beyond the availability of international price information. It also includes constructing measures that accurately reflect how external price shocks transmit into domestic inflation. Conventional import price indices are often compiled using broad product classifications that may not correspond to the composition of household consumption or the structure of a country's import basket. Consequently, they can mask substantial heterogeneity in price dynamics across products, reducing their ability to capture the inflationary pressures faced by consumers. For economies with a high dependence on imported food products, timely and disaggregated measures of imported inflation are therefore essential for improving inflation forecasting performance and identifying the channels through which external shocks propagate to domestic prices.

In Azerbaijan, the official import price indices published monthly by the State Statistics Committee (SSC) are subject to aggregation bias, which may dampen their effectiveness in capturing external price pressures. This limits their usefulness for forecasting domestic food inflation and underscores the need for alternative measures to monitor and predict imported inflationary pressures.

This paper provides two primary contributions to empirical trade and forecasting literature. First, we establish a highly disaggregated micro-data framework for tracking imported food inflation. We construct a trade-weighted Imported Food Price Index (IFPI) using approximately 800,000 transaction-level customs records derived from a broader 16-million raw transactions covering monthly period from January 2018 to February 2026. Our newly constructed index is aligned with the consumption basket by mapping transaction-level customs records for a wide range of imported brands and product varieties to 34 product categories that are represented in the consumer basket and are highly import-dependent. This step eliminates traditional aggregation bias by mapping micro customs filings directly to the actual national household consumption basket (Cavallo and Rigobon, 2016). Second, we develop a forecasting framework for the IFPI by combining non-parametric Binary Segmentation and Hidden Markov Models with a regularized machine-learning (ML) ensemble. We implement a hybrid forecasting framework designed to handle small-sample constraints. The architecture combines non-parametric change-point detection (Fryzlewicz, 2014) and Hidden Markov Model regime sorting (Hamilton, 1989) with histogram-stabilized gradient boosting. This approach stabilizes prediction variance when working with short time-series windows where standard ML models overfit.

Our findings demonstrate that the proposed IFPI possesses substantial leading-indicator properties for domestic food price inflation, with correlation coefficient peaking at 0.81 at a lag of

2 months. It is important to note that customs data become available after an average delay of 20–25 days following the reference month, whereas the official food Consumer Price Index (CPI) data is typically released within 10–15 days after the month-end. Consequently, once the IFPI for month *t* becomes available, it provides an information advantage of approximately 6-7 weeks over the official food CPI for the month *t+2*, enabling inflationary pressures to be identified in advance of the corresponding CPI data release.

Furthermore, the proposed ML forecasting framework exhibits substantially stronger predictive performance than conventional linear forecasting approaches in projecting the dynamics of imported food prices. The index and its projections can be readily incorporated into disaggregated food inflation forecasting models at the Central Bank of Azerbaijan Republic (CBAR), enhancing the projection of food price inflation and, consequently, improving the accuracy of broader inflation forecasts.

The paper is organized into seven sections. Section 2 reviews literature on global price transmission. Section 3 explains the employed data and the mathematical construction of the price index and discusses its properties. Section 4 presents the empirical forecast methodology, including feature selection and ensemble weight optimization. Section 5 describes the properties of the newly constructed IFPI, along with the results of the diagnostic analysis. Section 6 reviews the results of the forecasting model and gives comparative analysis of the model's performance. Section 7 briefly covers the limitations of the study and Section 8 concludes**.**

2. LITERATURE REVIEW

2.1 Global Price Transmission and Exogenous Climate Shocks

In trade-dependent small open economies consumer food inflation is primarily supply-driven and tied to international trade balances rather than domestic demand conditions (Galí & Monacelli, 2005). Upstream supply shocks propagate across borders through direct energy, transport, and fertilizer linkages, quickly reshaping downstream food value chains (Hamilton, 2003; Headey & Fan, 2008; Abbott et al., 2008). These issues are amplified by physical cross-border trade costs, which restrict market adjustments (Anderson & van Wincoop, 2004).

Classical trade theory posits that spatial arbitrage drives asset and commodity price convergence via the law of one price. However, real-world international market integration remains highly imperfect due to persistent structural frictions (Rogoff, 1996). Agricultural markets compound these frictions. Because of biological production lags, storage constraints, and weather dependencies, agricultural commodities experience long periods of price stability punctuated by sudden, sharp spikes (Deaton & Laroque, 1992; Pindyck, 2001).

Large-scale climatic anomalies—most notably the El Niño–Southern Oscillation (ENSO)—systematically alter global food markets by depressing crop yields in primary exporting regions (Brunner, 2002; Cashin et al., 2017). For import-dependent small open economies, these international climate shocks act as purely exogenous inflationary forces. They are transmitted directly through higher border values rather than local domestic harvest failures.

A major gap in empirical literature centers on how imported inflation is operationalized in forecasting models. Most predictive studies rely on highly aggregated national import price indices (IPIs) or broad international benchmarks like the FAO Food Price Index. However, axiomatic index number theory shows that an index is only as valid as its basket is representative (Diewert, 1976). Official national IPIs typically blend capital machinery and industrial inputs with primary consumer goods, which introduces significant measurement errors and dilutes the unique statistical signals generated by food supply shocks. Modern measurement theory addresses this data dilution by shifting focus away from macro-aggregates toward administrative, transaction-level micro-data (Cavallo & Rigobon, 2016). Constructing customized, trade-weighted price metrics directly from administrative customs data filters out irrelevant macroeconomic noise, allowing researchers to align border supply shocks directly with equivalent retail consumer baskets.

Another challenge relates to formulating robust forecasting frameworks for these indices. To map pass-through vectors, empirical econometrics has traditionally relied on linear frameworks, such as Autoregressive Distributed Lag (ARDL) bounds models (Pesaran et al., 2001) and Autoregressive Integrated Moving Average with Exogenous Regressors (ARIMAX) architectures (Stock & Watson, 2016). While mathematically tractable, these models assume that structural parameters remain constant over time. When confronted with volatile commodity regimes or sudden policy shifts, linear parameters suffer from structural instability, which leads to predictive breakdown in small samples. To address these limitations without requiring vast sample horizons, recent literature uses non-parametric signal segmentation. The Binary Segmentation approach (Fryzlewicz, 2014) isolates sudden mean and variance shifts without restrictive functional assumptions, bypassing the strict asymptotic distribution requirements of classical structural break tests (Bai & Perron, 2003).

To connect structural shifts with predictive accuracy, modern macroeconomic forecasting combines state-space formulations with regularized ML models. Hidden Markov Models capture non-linear transmission patterns by transitioning across discrete, unobserved latent states (Hamilton, 1989). When combined with tree-based ensembles like Random Forests (Breiman, 2001) or Gradient Boosting (Friedman, 2001), these frameworks map high-dimensional interaction effects without linear constraints.

To prevent overfitting under tight data constraints, recent applications implement histogram-stabilized gradient boosting (Ke et al., 2017). This structural binning technique serves as an algorithmic regularization layer that dampens transaction noise, controls prediction variance, and outperforms static linear benchmarks in high-volatility, short-horizon environments.

3. DATA AND INDEX ARCHITECTURE

3.1 Data Coverage

The IFPI tracks monthly trade prices using transaction-level administrative customs data for food products entering the Republic of Azerbaijan. By measuring prices directly at the point of entry, the IFPI isolates border price developments from domestic retail markups, internal transport costs, and commercial distribution margins.

Data are sourced from the international trade relations database compiled by the State Customs Committee of the Republic of Azerbaijan[1]. The raw administrative archive contains approximately 16 million individual import transaction records spanning from January 2017 through February 2026. The 12-month period in 2017 serves strictly as an initialization window

[1] The underlying data are confidential and not publicly available. Access is restricted to the CBAR for balance of payments compilation and authorized analytical and research purposes.

to establish base weights for the rolling index framework. The operational time-series covers the period from January 2018 through February 2026 (98 monthly observations).

Each transactional record contains five core attributes: 1) Harmonized Commodity Description and Coding System (HS) classification codes at 6-digit international levels (HS6) and 10-digit national tariff lines (HS10); 2) Cost-Insurance-Freight (CIF) import value, logged in both US dollars and its equivalent Azerbaijani manat (AZN); 3) the net physical mass of imports (in kilograms); 4) the country of origin; 5) the statutory date of customs clearance.

Primary data auditing is conducted at the granular HS10 level to isolate specific product variants, whereas index aggregation uses the 6-digit (HS6) tier to maintain longitudinal consistency. Filtering the entries via the food category of the international Classification of Individual Consumption by Purpose (COICOP3=101) yields an active micro-dataset of roughly 800K commercial transactions. The unit import price for each product is calculated by dividing the import value (in AZN) by the net import quantity (in kilograms), after extensive data treatment and processing.

3.2 Data Treatment and Processing

The raw dataset requires systematic processing to handle clerical noise, isolate commercial trade flows, and manage seasonal gaps before index estimation.

- *Nomenclature and metric standardization*: Product description fields at the HS10 level—which specify brand names and product characteristics—frequently contain typos, language inconsistencies across customs checkpoints, and omissions. These entries are extracted using regular expressions and validated via manual audits.
- *Outlier truncation*: The administrative registry includes small-scale, personal-use imports alongside bulk commercial trade. These non-commercial transactions often carry highly inflated retail premiums that do not reflect macroeconomic food market realities. To filter out true commercial trade baselines the data was asymmetrically trimmed by removing observations below the 2nd percentile and above the 95th percentile of each commodity's monthly cross-sectional unit-value distribution. This asymmetric cutoff addresses the pronounced right-skewness typical to trade unit values, where reporting errors and extreme price spikes predominantly distort the upper tail of the distribution. This truncation filters out personal-use premium distortions and eliminates extreme data-entry errors, stabilizing variance for the ML models.
- *Imputation of seasonal gaps*: Missing monthly observations at the individual product level, primarily driven by seasonal trade pauses, were resolved via linear interpolation. For extended suspensions where an entire product category recorded zero volume, the price component was held constant at the last market-clearing unit value, and its structural index weight was temporarily redistributed proportionally among the remaining active basket components.

3.3 Product Selection and Basket Construction

To satisfy axiomatic index requirements of representativeness and economic relevance, candidate food product entries from the preprocessed data were filtered through a four-stage screening framework:

- ✓ Alignment with the consumption basket: Candidate import products were mapped against the food consumption categories within the official retail CPI basket. Items not represented in the consumption basket were omitted.
- ✓ Import dependence: Products were selected to retain only those with substantial and persistent import dependence. In the resulting sample, the lowest observed import dependency ratio was 37%, while 14 products exhibited structural import dependency exceeding 90% throughout the sample period[2].
- ✓ Market exposure: Commodity lines with irregular trading activity—specifically those with observation gaps exceeding four consecutive months—were excluded to guarantee a continuous stream of monthly market data.
- ✓ Statistical continuity: To maintain a consistent time series, we systematically excluded products that experienced structural data gaps. Such discontinuities generally occur due to market-driven shifts, for instance, when importers transition to new substitute products or cease importing specific items altogether in response to changing consumer demand, thereby disrupting the long-term continuity of those commodity lines.

Following this filtering process, the final analytical basket comprises 34 major food-product categories represented by 44 individual product varieties, ensuring complete consistency of the basket throughout the entire empirical period.

3.4 Index Construction and Weighting Framework

To mitigate substitution bias arising from substantial changes in agricultural trade volumes, the IFPI is constructed using the Fisher Ideal Index. Unlike the Laspeyres and Paasche indices, which rely exclusively on base-period or current-period expenditure weights, the Fisher index combines information from both weighting schemes through their geometric mean. As a result, it generally provides a more balanced measure of price change when expenditure patterns evolve over time and is widely regarded in the index-number literature as a "superlative" index that closely approximates the underlying cost-of-living index under flexible consumer preferences (Diewert, 1976; Balk, 2008). Moreover, the Fisher index satisfies important axiomatic properties, including the time-reversal and factor-reversal tests, which contribute to its desirable theoretical and practical characteristics (Fisher, 1922).

The calculation begins at the transaction level by deriving monthly unit prices ($UP_{j,t}$) for each import line *j* in month *t*:

$$UP_{j,t} = \frac{V_{j,t}}{Q_{j,t}}$$

where $V_{j,t}$ represents the nominal import value (CIF) in AZN, and $Q_{j,t}$ represents the net physical mass of imports in kilograms. These granular records are then aggregated into value-weighted average prices ($P_{i,t}$) at the operational 6-digit (HS6) level:

$$P_{i,t} = \frac{\sum_{j\in\Omega_i} V_{j,t}}{\sum_{j\in\Omega_i} Q_{j,t}}$$

where $\Omega_i$represents the complete set of transactional records nesting within HS6 category *i*.

[2] Product-level import dependence ratios were estimated by the authors using officially reported average, import, and domestic prices for consumer basket food products published by the SSC.

To accommodate shifting trade contracts and structural seasonality, each reporting month *t* is evaluated against a trailing 12-month moving reference window (τ) immediately preceding the current observation period. The base-period-weighted Laspeyres index and the current-period-weighted Paasche index are unified into a single geometric index expression:

$$F_t = \sqrt{\left(\frac{\sum_i P_{i,t} Q_{i,\tau}}{\sum_i P_{i,\tau} Q_{i,\tau}}\right) \times \left(\frac{\sum_i P_{i,t} Q_{i,t}}{\sum_i P_{i,\tau} Q_{i,t}}\right)}$$

where $P_{i,t}$ and $Q_{i,t}$ represent the value-weighted price and cumulative quantity of category *i* over the moving reference horizon.

Finally, the 34 distinct HS6 commodity trajectories are consolidated into the aggregate time series using lagging annual nominal trade shares as dynamic weights $w_{i,t}$:

$$w_{i,t} = \frac{V_{i,\tau}}{\sum_k V_{k,\tau}}$$

where $V_{i,\tau}$ is the total nominal import value of category *i* over the rolling reference window.

4. EMPIRICAL METHODOLOGY

The empirical framework matches variables across different scales to model the newly constructed IFPI. The pipeline follows four steps: 1) data transformation; 2) latent regime extraction; 3) feature filtering; 4) an optimized ML ensemble.

4.1 Data Transformation and Structural Breaks

Nominal series enter the model as month-on-month ($\Delta \ln X_t$) and year-on-year ($\Delta_{12} \ln X_t$) log differences. Some of the global leading indicators contain structural drift but retain useful baseline signals. We track these indicators using raw levels, first differences, and sequential Z-scores:

$$Z_t = \frac{X_t - \mu_{t-1}}{\sigma_{t-1}}$$

The rolling mean μ_{t-1} and standard deviation σ_{t-1} are calculated over past windows $\tau \in [1, t-1]$. Lagging these parameters by one month prevents future data from leaking into the training step.

Structural instability can distort parameter estimates. We handle this via a dual-track break detection approach. First, we apply a parametric rolling Chow (1960) test to the target series ($y_t = \Delta \ln \text{IFPI}_t \times 100$). An autoregressive specification with a time trend is split at sequential candidate months (T_β). We evaluate parameter constancy using an F-distribution:

$$F = \frac{\left(RSS_{\text{pooled}} - (RSS_1 + RSS_2)\right)/k}{(RSS_1 + RSS_2)/(N_1 + N_2 - 2k)}$$

where RSS is the residual sum of squares, *k* is the number of parameters, and *N* is the sample size. Second, we use a non-parametric Binary Segmentation algorithm (Scott and Knott, 1974) to find splits by minimizing an ℓ_2-norm quadratic error loss function.

These break dates are converted into time-varying proximity variables that measure the months elapsed since the last break $(t - T_b)$ the distance to an anticipated break $(T_{B+1} - t)$ and localized interval dummies $D_t \in \{0,1\}$.

4.2 Hidden Markov Model Regime Identification

We capture unobserved shift dynamics by fitting a 3-state Gaussian Hidden Markov Model (Hamilton, 1989) over an empirical observation vector:

$$O_t = \left[\Delta Y_t, |\Delta Y_t|, \sigma_t^{(3)}, \Delta^2 Y_t\right]^T$$

This vector includes the target index's first difference, its absolute value, a 3-month rolling standard deviation, and inflationary acceleration. The hidden state $S_t \in \{0, 1, 2\}$ follows a first-order Markov chain governed by a transition matrix P. The emission probabilities are modeled as a multivariate Gaussian density:

$$b_r(O_t) = \frac{1}{(2\pi)^{d/2}|\Sigma_r|^{1/2}} \exp\left(-\frac{1}{2}(O_t - \mu_r)^T \Sigma_r^{-1}(O_t - \mu_r)\right)$$

We sort the hidden states by their conditional variance and absolute returns. This step groups the data into three explicit regimes: low volatility (S_t = 0), normal transmission (S_t = 1), and high-volatility shocks (S_t = 2). The model yields monthly forward state probabilities and a state duration variable, both of which serve as predictive features in the subsequent forecasting specification.

4.3 Feature Selection and Filtering

Large numbers of correlated macroeconomic inputs cause tree-based models to overfit. We filter our feature space using a joint ranking metric that balances linear and non-linear associations:

$$\Phi(X_k) = \lambda\rho(X_k, y) + (1 - \lambda)I(X_k; y)$$

where $\rho(X_k, y)$ is the Pearson correlation coefficient, $I(X_k; y)$ is the non-parametric Mutual Information score (Kraskov et al., 2004), and $\lambda = 0.5$ balances the two metrics.

4.4 Ensemble Optimization and Forecasting Mechanics

The forecasting framework combines four base estimators:

- Random Forest Regressor: A bagging model that averages de-correlated decision trees (Breiman, 2001).
- Extreme Gradient Boosting: An additive gradient boosting framework that minimizes a regularized loss function (Chen and Guestrin, 2016).
- Histogram-based Gradient Boosting Regression Tree: A histogram-based boosting engine that bins continuous inputs to reduce split variance and smooth out small-sample noise.
- Unobserved Components Model: A linear state-space benchmark that decomposes the series into stochastic trends and cycles ($y_t = \mu_t + \gamma_t + \varepsilon_t$), where μ_t is the trend and β_t handles time-varying drift (Harvey, 1989).

We implement a strict 60-month sliding rolling window (N = 60), driven by our constrained sample of 98 monthly observations. An expanding window is unsuitable as it incorporates outdated structural regimes, diluting current trade dynamics. Conversely, shorter windows (e.g., 36–42 months) fail to provide sufficient degrees of freedom; specifically, fitting a 3-state multivariate Gaussian Hidden Markov Model requires enough regime-specific observations to prevent the covariance matrix Σ_r from becoming singular. The 60-month window strikes an optimal balance,

dropping outdated data while maintaining the numerical stability required for Hidden Markov Model state-sorting and ensemble calibration.
Base model predictions are blended using a constrained optimization routine solved via Sequential Quadratic Programming (SQP). The weight vector $w = [w_{\text{RF}}, w_{\text{XGB}}, w_{\text{UC}}]^T$ minimizes historical squared calibration errors:

$$\min_{w} \sum_{t\in\text{Calib}} \left(y_t - \left(w_{\text{RF}} y_{t,\text{RF}} + w_{\text{XGB}} y_{t,\text{XGB}} + w_{\text{UC}} y_{t,\text{UC}}\right)\right)^2$$

subject to the following boundary constraints:

$$\sum w_m = 1, \quad w_m \geq 0, \quad w_{\text{XGB}} \leq 0.45, \quad w_{\text{UC}} \geq 0.15$$

Because tree models overfit on small samples, we limit the high-variance Extreme Gradient Boosting algorithm to a maximum weight of 45%. Concurrently, the structural linear baseline (UC) is assigned a minimum weight of 15%. These boundaries prevent the framework from over-relying on volatile tree splits during sudden data shifts, maintaining baseline stability.

We track the average relative feature importance across our tree-based base models. For each tree model, a variable's importance is determined by measuring its total reduction in residual variance during the node-splitting process:

$$I_i = \sum_{t\in T} w_t \cdot \Delta\text{MSE}_t(i)$$

where *T* represents the internal nodes split by feature *i*, w_t is the fraction of samples passing through node *t*, and Δ $\text{MSE}_t(i)$ is the resulting reduction in mean squared error.

To create a single metric for the complete framework, we calculate a weighted average of these individual feature scores using the active blending weights (w_m) optimized by the SQP step. This aligns the final importance rankings with the model carrying the most weight in the ensemble at that specific historical point. Tracking these normalized percentages over our fixed rolling windows reveals how the primary drivers of imported food inflation shift during trade disruptions.

A regime-specific bias adjustment corrects systematic errors within each Hidden Markov Model state during calibration. Residuals are tracked separately within each active state, and the final prediction is adjusted by the corresponding state modifier, bounded by a conservative clipping threshold.

Finally, empirical magnitude clipping locks out-of-sample predictions within the boundaries of the active training distribution:

$$y_{t,\text{final}}^{\text{clipped}} = \max\left(Q_{0.01}(\Delta y_{\text{train}}), \min\left(y_{t,\text{final}}, Q_{0.99}(\Delta y_{\text{train}})\right)\right)$$

Restricting forecasts within the 1st and 99th percentiles of historical changes prevents erratic forecast spikes during major global supply shocks, stabilizing the model's out-of-sample properties.

5. IFPI Properties and Diagnostic Analysis

This section reviews the data properties of newly constructed IFPI, the official food CPI, and 13 covariates (N = 98 monthly observations) structured into four empirical blocks: 1) Global commodity prices block captures global input and agricultural price pressures using global natural gas prices, fertilizer prices and FAO global food price index; 2) Global cost-push block tracks external supply and regional pricing via Global Supply Chain Pressure Index (GSCPI), as well as official CPI inflation in Russia and Eurozone, and producer price index (PPI) in Türkiye; 3) Domestic macro block controls for local monetary and economic conditions through demand deposits and GDP deflator; 4) Exogenous shock block accounts for structural and climate anomalies using ENSO index, alongside dummy variables indicating periods of high inflation volatility (Inflation_Regime), Russia-Ukraine war (War_Dummy), and the COVID-19 pandemic (COVID_Dummy). Table 1 summarizes their descriptive statistics and unit root properties.

Table 1. Descriptive statistics and unit root properties

Variable	Data source	Mean	Median	Std Dev	ADF Level p	ADF Diff p	KPSS Level p	KPSS Diff p	Order of integration
Constructed IFPI	Authors' calculations	104.5	104.1	5.17	0.36	0.01	0.09	0.10	I(1)
Food CPI	SSC	107.3	105.3	6.22	0.48	0.00	0.10	0.10	I(1)
Natural gas prices	World Bank	112.8	92.6	76.86	0.23	0.00	0.10	0.10	I(1)
Global food prices index	FAO	116.8	121.3	17.73	0.25	0.00	0.01	0.10	I(1)
Fertilizer prices	World Bank	129.8	119.5	54.29	0.10	0.20	0.03	0.10	Unclear
Demand deposits	CBAR	0.0	0.0	0.15	0.15	0.00	0.10	0.10	I(1)
GDP deflator	SSC	106.7	101.1	16.21	0.27	0.00	0.10	0.10	I(1)
ENSO	NOAA National Centers for Environmental Information	0.1	0.0	0.69	0.14	0.00	0.10	0.10	I(1)
GSCPI	Federal Reserve Bank of New York	0.8	0.4	1.41	0.21	0.00	0.10	0.10	I(1)
CPI in Russia	Official figures	0.1	0.1	0.04	0.15	0.01	0.02	0.10	I(1)
CPI in Eurozone		0.0	0.0	0.03	0.38	0.02	0.09	0.10	I(1)
PPI in Turkiye		0.4	0.3	0.36	0.65	0.03	0.10	0.10	I(1)

The IFPI displays stable central tendency and muted volatility relative to the official food CPI, suggesting that the aggregation of transaction-level data produces a stable measure of imported price dynamics while retaining the underlying macroeconomic signal. Upstream inputs show significantly greater volatility, led by natural gas prices and fertilizer prices.

ADF and KPSS diagnostics confirm that the primary price series are non-stationary in levels. Level p-values for the IFPI and food CPI fail to reject the unit root null, but both series reject it at the 1% significance level in first differences, establishing I(1) integration. Because unit-root test for fertilizer prices returns an ambiguous integration result, classical linear VAR specifications become invalid, necessitating the augmented Toda–Yamamoto causal framework (see Table 3 below).

Figure 1. Official food CPI and IFPI, Δyoy %

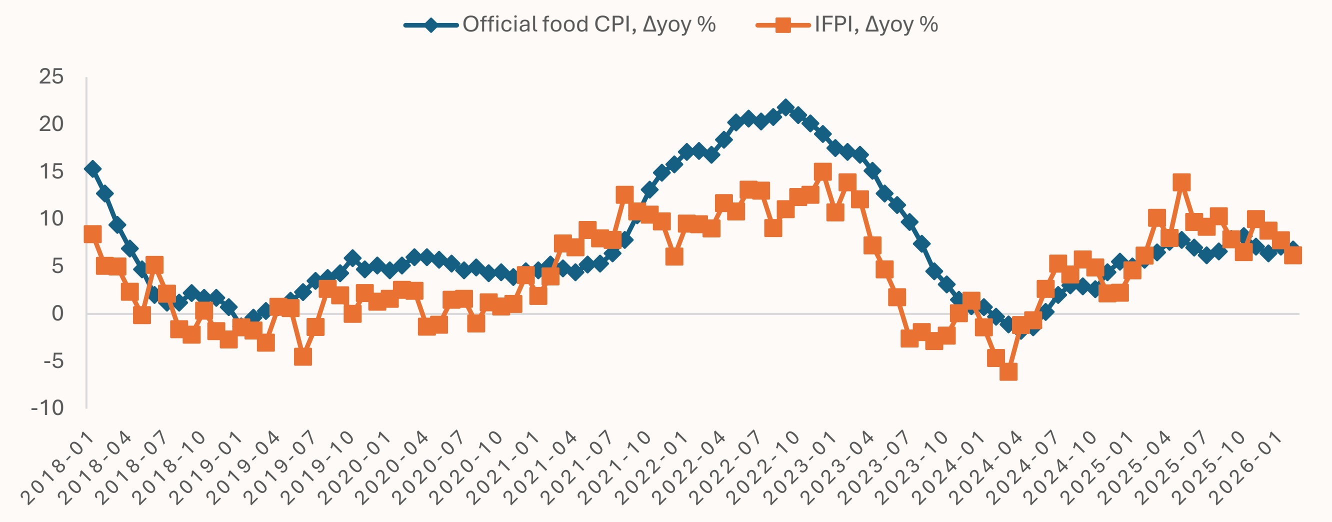
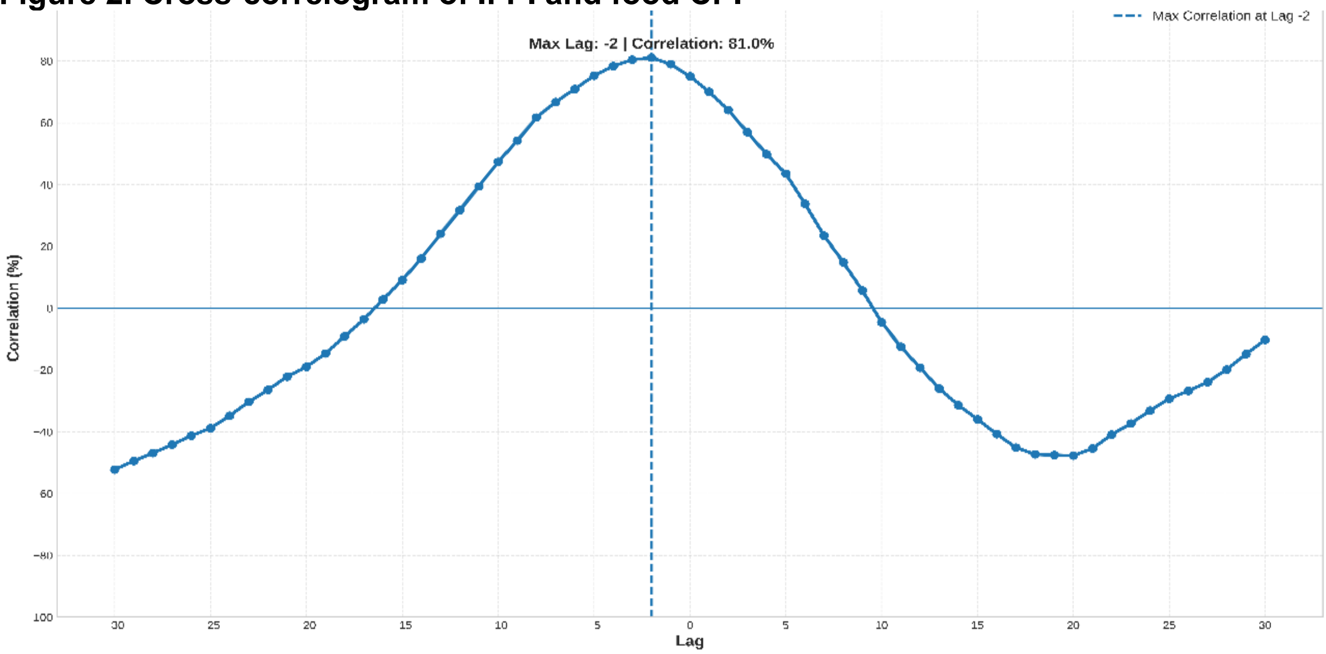


Figure 1 presents the year-on-year growth rates of the newly constructed IFPI alongside the official food CPI. While both track the same long-term inflationary cycles, the transaction-level IFPI exhibits greater high-frequency volatility. The IFPI consistently precedes changes in the official food CPI, suggesting that imported price movements are transmitted to retail food prices with a measurable lag.

Figure 2. Cross-correlogram of IFPI and food CPI



To quantify this relationship, we examine the contemporaneous and lagged correlations between the IFPI and official food CPI inflation (Figure 2). Cross-correlation functions map the lead-lag structure of these logistical transmission delays. The results indicate that the contemporaneous correlation is r = 0.7506, but the correlation coefficient peaks at Lag 2 (r = 0.8102). The border index thus operates as a two-month leading indicator, capturing wholesale supply shocks before they filter down to retail markets. This pattern is expected, as import prices measured at the customs border respond immediately to fluctuations in international commodity prices, exchange rates, freight costs, and global supply chain disruptions. By contrast, retail food prices adjust more gradually because imported goods pass through multiple stages of the domestic supply chain, including transportation, wholesale distribution, processing, and retailing. In addition, pricing strategies, inventory management, menu costs and other factors smooth the transmission of external price shocks to consumer prices over time. The results provide evidence that the index contains valuable forward-looking information on domestic food price dynamics. This finding supports the use of the IFPI as a leading indicator of food inflation.

Table 2. Bivariate Correlation Matrix with Constructed IFPI

Variable type	Variable	Correlation with IFPI (contemporaneous)
Domestic macro block	Food CPI	0.75
	Demand deposits	0.68
	GDP deflator	0.61
Exogenous shock block	ENSO	-0.73
Global commodity prices block	Natural gas prices	0.62
	FAO global food price index	0.70
	Fertilizer prices	0.66
Global cost-push block	GSCPI	0.39
	CPI in Russia	0.59
	CPI in Eurozone	0.57
	PPI in Türkiye	0.52

Next we look at the bivariate correlation dynamics between the IFPI and the 13 variables structured into 4 blocks (Table 2):

- Domestic macro block displays the strongest positive co-movement, led by official food CPI (r=0.75). This confirms that micro-level import pricing captures the underlying retail food inflation signal, supported by domestic liquidity (0.68) and output pricing (0.61).
- Global commodity prices block: Shows uniformly high positive passthrough across upstream inputs, led by the FAO global food price index (0.70), fertilizer prices (0.66), and natural gas prices (0.62).
- Global cost-push block: Reflects steady positive cost transmission from major trade partners—Russia CPI (0.59), Eurozone CPI (0.57), and Türkiye PPI (0.52)—alongside supply chain bottlenecks (GSCPI, 0.39).
- Exogenous shock block: Climate dynamics captured by ENSO display a strong inverse relationship (-0.73), co-moving inversely with global agricultural import pricing cycles.

Table 3. Predictive Causality Frameworks

Criterion	Selected Lag p	dmax	Estimated Lag p plus_dmax	Test_Statistic	P value
AIC	12	1	13	26.62	0.01
BIC	2	1	3	13.88	0.00
HQIC	2	1	3	13.88	0.00

Baseline Granger testing rejects the non-causality null ($F = 18.88$, $p < 0.001$), showing that historical IFPI records help project retail food CPI. Because the mixed integration horizons—I(0) and I(1)—identified by unit-root tests (Table 1) invalidate standard asymptotic distributions, we apply an augmented Toda–Yamamoto Wald test overfitted by $d_{max} = 1$ lag.

Toda–Yamamoto tests confirm predictive causality across all information criteria. Tight BIC/HQIC lags $(p + d_{max} = 3)$ yield a significant Wald statistic of 13.880 (p = 0.0031), while extended AIC specifications $(p + d_{max} = 13)$ retain significance (p = 0.0140). This stability across different lag lengths proves the index's leading capacity is an inherent property of the data, not a tuning artifact.

Table 4. Subsample Robustness and Structural Stability

Subsample	Start Date	End Date	Observations	IFPI and food CPI Contemporaneous Correlation Coefficient	Peak Correlation Coefficient (Lag)
Pre_COVID	1/31/2018	2/29/2020	26	0.7590	0.7590 (0)
COVID_Period	3/31/2020	12/31/2021	22	0.5293	0.9199 (-7)
Post_COVID	1/31/2022	2/28/2026	50	0.7413	0.8014 (-2)

A 36-month rolling window correlation analysis yields a stable mean coefficient of 0.72 (σ = 0.11, min = 0.42), demonstrating long-term structural stability. Subsample tracking across major structural fractures reveals distinct regime behaviors:

- Pre-COVID era (N = 26): Price transmission was near-instantaneous, peaking sharply at Lag 0 (r = 0.76) under frictionless global supply conditions.
- COVID-19 disruption (N = 22): Contemporaneous correlation fell to 0.53, but the lead-lag profile stretched to Lag -7, peaking at 0.92. This lag reflects an operational split: pandemic logistical bottlenecks delayed domestic retail surveys, while customs registries captured supply shocks in real time.
- Post-COVID recovery (N = 50): System dynamics normalized. Static correlation recovered to 0.74, and the predictive peak returned to a stable 2-month horizon (Lag -2, r = 0.80).

This structural reversion confirms that the 2-month early warning capacity of the IFPI is an enduring economic feature.

6. Comparative Forecast Performance

This section presents empirical results, statistical tests, and out-of-sample evaluations for our forecasting model of the newly constructed trade-weighted IFPI. We break the validation process down into distinct stages: residual autocorrelation tests, structural break tracking, out-of-sample rolling estimation windows, benchmark model comparisons, multi-model scenario projections, and an information-theoretic analysis of feature dynamics.

Table 5. Residual Autocorrelation Diagnostics

Fold	Ljung–Box Q(3) p-value	Empirical Conclusion
Fold 1	0.2486	Statistically insignificant
Fold 2	0.1564	Statistically insignificant
Fold 3	0.1414	Statistically insignificant

To ensure the multi-tiered forecasting framework adequately captures the underlying temporal dynamics of the target series—defined $as\, y_t = \Delta \ln \mathrm{IFPI}_t$ 100—residual diagnostics were executed across all evaluation partitions. Testing for remaining serial correlation involved applying the Ljung–Box Q(3) test sequentially to each of the three rolling validation folds. Because the p-values across all validation folds are uniformly above 0.05, the null hypothesis of no serial correlation cannot be rejected.

Table 6. Structural Break Analysis

Estimation Method	Documented Empirical Breakpoints / Horizon
Chow Test (Parametric)	Systematic cluster from 2023-02 to 2023-06
Strongest Single Break	2023-04-30 (F = 8.60; p < 0.001)
Ruptures BinSeg L2	2018-12, 2023-02, 2023-07, 2024-05, 2026-01

Global logistical realignments and domestic market adjustments frequently trigger severe parameter instability. To intercept and track these shifts within the target index, this analysis relies on a dual-track testing mechanism. First, a rolling Chow framework evaluates parametric constancy over time. Second, non-parametric regime transitions are captured using the Ruptures Binary Segmentation (L2-norm) approach.

A clear alignment emerges between the parametric and non-parametric estimators, pinpointing a window of profound regime volatility throughout the 2023 trade horizon. The single sharpest break occurs on April 30, 2023 (F = 8.60, p < 0.001). Because this structural instability is so pronounced, executing full-sample static estimations would yield invalid conclusions. This finding provides the direct mathematical justification for shifting to a strict rolling training window framework.

Table 7. Fixed Rolling-Window Validation Results

Training Horizon	Out-of-Sample Testing Period	RMSE	OOS R^2 vs RW %	Direction Accuracy %
2018-05-31 – 2023-04-30	2023-05-31 – 2024-04-30	1.50	69.31	83.33
2019-05-31 – 2024-04-30	2024-05-31 – 2025-04-30	1.33	63.74	66.67
2020-05-31 – 2025-04-30	2025-05-31 – 2026-04-30	2.02	41.24	66.67
Average		1.62	58.10	72.22

To isolate model stability from localized sample dependencies, we evaluate the predictive architecture across sequential rolling estimation windows. Given the constrained sample size inherent to monthly macroeconomic data, the cross-validation scheme is structured into three non-overlapping 12-month test folds spanning May 2023 through April 2026. This three-fold design ensures a consistent, fixed training window across folds while maximizing out-of-sample evaluation without over-fragmenting a small sample.

The proposed model demonstrates strong generalization across distinct market regimes. Out-of-sample performance remains consistently superior to the naive random walk benchmark, achieving an average out-of-sample R-squared ($OOS\ R^2$) of 58.10% across all folds[3]. While forecast accuracy moderates during the final testing window (2025–2026) due to heightened

[3] $OOS\ R^2 = 1 - {}^{MSFE_{ensemble\ model}}/_{MSFE_{benchmark\ model}}$, where MSFE is mean squared forecast error.

macroeconomic volatility—yielding an RMSE of 2.02 and an $OOS\ R^2$ of 41.24%—directional accuracy remains robust at 66.67%. Across the full evaluation sequence, the framework maintains an average directional accuracy of 72.22% (peaking at 83.33% in Fold 1), confirming that the model captures persistent structural dynamics rather than fitting to localized noise.

Table 8. Forecasting Performance Comparison Matrix

Model Architecture	RMSE	MAE	MAPE %	OOS R^2 vs. RW %	Directional Accuracy %
Ensemble (ML)	1.61	1.29	1.23	58.09%	72.22%
ARIMAX	5.28	5.25	5.00	0.18%	56.94%
ARDL	5.28	4.43	4.22	13.60%	68.06%

We compared the out-of-sample performance of our regularized hybrid ensemble against traditional linear time-series baselines: a classical ARIMAX model and an ARDL bounds specification. The regularized hybrid ensemble beats both benchmarks across every error metric and directional test.

Table 9. Pairwise Diebold–Mariano Tests for Predictive Superiority

Model Comparison	N	DM Statistic	p-value
Ensemble model (ML) vs. random walk	34	2.434	0.0231
Ensemble model (ML) vs. ARIMAX	24	5.433	< 0.0001
Ensemble model (ML) vs. ARDL	24	4.563	0.0001

To establish whether the observed reductions in point-forecast errors are statistically meaningful rather than artifacts of sampling variation, pairwise Diebold–Mariano (1995) tests were conducted using the Harvey et al. (1997) small-sample modification under quadratic loss. Relative to linear econometric specifications, the ML framework achieves highly significant performance gains over both ARIMAX (DM = 5.433, $p < 0.001$) and ARDL (DM = 4.563, $p = 0.0001$), confirming its capacity to capture non-linear dynamics that structural linear models omit. Furthermore, the ensemble model significantly outperforms the naive random walk persistence baseline at the 5% significance level (DM = 2.434, $p = 0.0231$), verifying genuine out-of-sample predictive power.

To preserve strict observation-level alignment—a fundamental requirement of the Diebold–Mariano test statistic—pairwise comparisons are computed across the maximum synchronized evaluation window feasible for each model pair. Because the extended in-sample estimation requirements of the ARIMAX and ARDL specifications shifted their out-of-sample availability to March 2024 (N = 24), these models are evaluated against the ML framework over this common sub-sample, whereas the random walk baseline is tested against the ML model across the full 34-month out-of-sample horizon (May 2023–February 2026) to avoid discarding valid forecast information.

Table 10. Feature Importance and Economic Interpretation (Top Ten Predictive Features)

Rank	Feature Taxonomy	Relative Importance %

1	y_diff_lag12 (Annual target persistence)	12.31%
2	FAO Food Price Index (yoy_mom3 Transformation)	5.68%
3	FAO Food Price Index (yoy_lag3 Transformation)	5.47%
4	FAO Food Price Index (mom1_roll3 Transformation)	5.02%
5	y_level_lag1 (Autoregressive level lag)	4.84%
6	Natural Gas Price Index (yoy_mom3 Transformation)	4.57%
7	FAO Food Price Index (mom1_lag3 Transformation)	4.24%
8	Natural Gas Index (mom1_roll6 Transformation)	4.23%
9	y_level_lag12 (Seasonal level lag)	4.09%
10	Fertilizer Price Index (mom1_lag3 Transformation)	3.51%

The relative feature importance rankings show the main internal and external drivers within the regularized forecasting engine. Internal persistence is the leading driver (y_diff_lag12 at 12.31%), which reflects long-term structural relationships in national trade agreements. Global market factors dominate the remaining share. Transformed metrics for the FAO Food Price Index take three of the top four spots, totaling 16.17% in the top ten. This confirms that prices transmit slowly from international agricultural spot markets into border entry registries.

Upstream supply factors are also prominent. The natural gas price index (totaling 8.80% across ranks 6 and 8) and the fertilizer price index (3.51% at rank 10) show that global energy shocks flow rapidly into domestic border valuations by driving up agricultural production costs and processing metrics.

Interestingly, raw ENSO metrics do not appear in the top ten predictive split vectors. This does not indicate the failure of climate hypothesis. Rather, it illustrates proximal feature shadowing in regularized tree-based ensembles. Macroclimatic anomalies act as distal supply shocks that alter global crop yields. These effects are immediately picked up and aggregated by the proximal FAO Food Price Index series. Because tree-based split algorithms select variables with the highest immediate signal density, the multi-scale transformations of the FAO index (holding a combined 35.42% predictive weight) absorb and hide the underlying low-frequency ENSO cycles. This makes direct climate indices redundant for high-frequency monthly forecasting horizons, even though they remain vital as long-run structural anchors.

7. Limitations of the Study

Several technical and structural limitations should be acknowledged when interpreting the results of this study. First, the available dataset is limited to 98 consecutive monthly observations beginning in January 2018, reflecting the historical coverage of the customs database. These data constraints restrict the application of more data-intensive forecasting approaches, particularly unconstrained deep learning architectures that require substantially longer time series for reliable estimation. Second, although the proposed forecasting framework incorporates Hidden Markov Model regime conditioning to account for structural changes, its predictive performance may deteriorate in the presence of unprecedented geopolitical events, major supply chain disruptions, or abrupt trade corridor closures that fall entirely outside the information contained within the 60-month rolling estimation window. Finally, the construction of the IFPI relies on transaction-level unit-value proxies, calculated as the ratio of nominal CIF import values to net import quantities, rather than directly observed transaction prices. While extensive product-specific data cleaning and asymmetric trimming procedures mitigate the influence of outliers, the use of unit values may still introduce a modest degree of measurement error arising from unobserved differences in product quality and composition.

8. Conclusion

This study develops a new IFPI for Azerbaijan using highly disaggregated transaction-level customs data and demonstrates its value for monitoring and forecasting imported inflationary pressures. By aligning millions of import transactions with highly import-dependent product categories in the national consumption basket, the proposed index overcomes many of the aggregation limitations associated with conventional import price measures and provides a more timely representation of external food price dynamics. Empirical results show that the IFPI possesses strong leading-indicator properties, preceding official food CPI inflation by approximately two months when the cross-correlation coefficient peaks at 0.81. These findings suggest that the index captures the transmission of international price shocks into domestic food prices at an early stage, making it a valuable input for short-term inflation monitoring.

Building on this measurement framework, the paper develops a regime-conditioned machine learning ensemble to forecast the IFPI under conditions of limited data availability and frequent structural changes. By combining change-point detection, Hidden Markov Model regime classification, and histogram-based gradient boosting, the proposed framework effectively captures the nonlinear dynamics of imported food prices and substantially outperforms conventional linear forecasting approaches. The results demonstrate that imported food price dynamics are driven by a combination of their own persistence, international food commodity prices, and upstream energy market developments, allowing the model to generate accurate projections even in a relatively short time-series environment.

References

Abbott, P.C., Hurt, C. and Tyner, W.E. (2008) *What's driving food prices?* USDA Economic Research Service, WDP-08-01.

Anderson, J.E. and van Wincoop, E. (2004) 'Trade costs', *Journal of Economic Literature*, 42(3), pp. 691–751.

Bai, J. and Perron, P. (2003) 'Computation and analysis of multiple structural change models', *Journal of Applied Econometrics*, 18(1), pp. 1–22.

Balk, B.M. (2008) *Price and Quantity Index Numbers: Models for Measuring Aggregate Change and Theory*. Cambridge: Cambridge University Press.

Breiman, L. (2001) 'Random forests', *Machine Learning*, 45(1), pp. 5–32.

Brunner, A.D. (2002) 'El Niño and world primary commodity prices: Warm water or hot air?', *Journal of International Economics*, 58(2), pp. 351–383.

Cashin, P., Mohaddes, K. and Raissi, M. (2017) 'Fair weather or foul? The macroeconomic effects of El Niño', *Journal of International Economics*, 106, pp. 37–54.

Cavallo, A. and Rigobon, R. (2016) 'The Billion Prices Project: Using online prices for measurement and research', *Journal of Economic Perspectives*, 30(2), pp. 151–178.

Chen, T. and Guestrin, C. (2016) 'XGBoost: A scalable tree boosting system', in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM, pp. 785–794.

Chow, G.C. (1960) 'Tests of equality between sets of coefficients in two linear regressions', *Econometrica*, 28(3), pp. 591–605.

Deaton, A. and Laroque, G. (1992) 'On the behaviour of commodity prices', *Review of Economic Studies*, 59(1), pp. 1–23.

Diebold, F.X. and Mariano, R.S. (1995) 'Comparing predictive accuracy', *Journal of Business & Economic Statistics*, 13(3), pp. 253–263.

Diewert, W.E. (1976) 'Exact and superlative index numbers', *Journal of Econometrics*, 4(2), pp. 115–145.

Fisher, I. (1922) *The Making of Index Numbers: A Study of Their Varieties, Tests, and Reliability*. Boston: Houghton Mifflin.

Friedman, J.H. (2001) 'Greedy function approximation: A gradient boosting machine', *Annals of Statistics*, 29(5), pp. 1189–1232.

Fryzlewicz, P. (2014) 'Wild binary segmentation for multiple change-point detection', *Annals of Statistics*, 42(6), pp. 2243–2281.

Galí, J. and Monacelli, T. (2005) 'Monetary policy and exchange rate volatility in a small open economy', *Review of Economic Studies*, 72(3), pp. 707–734.

Hamilton, J.D. (1989) 'A new approach to the economic analysis of nonstationary time series and the business cycle', *Econometrica*, 57(2), pp. 357–384.

Hamilton, J.D. (2003) 'What is an oil shock?', *Journal of Econometrics*, 113(2), pp. 363–398.

Harvey, A.C. (1989) *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

Harvey, D., Leybourne, S. and Newbold, P. (1997) 'Testing the equality of prediction mean squared errors', *International Journal of Forecasting*, 13(2), pp. 281–291.

Headey, D. and Fan, S. (2008) 'Anatomy of a crisis: The causes and consequences of surging food prices', *Agricultural Economics*, 39(s1), pp. 375–391.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. and Liu, T.-Y. (2017) 'LightGBM: A highly efficient gradient boosting decision tree', *Advances in Neural Information Processing Systems*, 30, pp. 3146–3154.

Kraskov, A., Stögbauer, H. and Grassberger, P. (2004) 'Estimating mutual information', *Physical Review E*, 69(6), p. 066138.

Pesaran, M.H., Shin, Y. and Smith, R.J. (2001) 'Bounds testing approaches to the analysis of level relationships', *Journal of Applied Econometrics*, 16(3), pp. 289–326.

Pindyck, R.S. (2001) 'The dynamics of commodity spot and futures markets: A production, consumption, and inventory model', *Energy Journal*, 22(3), pp. 1–29.

Rogoff, K. (1996) 'The purchasing power parity puzzle', *Journal of Economic Literature*, 34(2), pp. 647–668.

Scott, A.J. and Knott, M. (1974) 'A cluster analysis method for grouping means in the analysis of variance', *Biometrics*, 30(3), pp. 507–512.

Stock, J.H. and Watson, M.W. (2016) *Introduction to Econometrics*. Updated 3rd edn. Harlow: Pearson.