

Forecasting Iraq's Broad Money Supply for 2026–2030: Evidence from Box–Jenkins Models and Monetary Policy Implications

Ehsan Jabr ASHOOR

<https://orcid.org/0009-0002-1692-370X>

Department of Economics, College of Administration and Economics
University of Baghdad, Iraq
ehssan_jabbar@coadec.uobaghdad.edu.iq

Burhan Othman HUSSEIN

<https://orcid.org/0009-0007-6560-9481>

Department of Economics, College of Administration and Economics
University of Kirkuk, Iraq
burhanohman@uokirkuk.edu.iq

Omar Mahmood AKAWEE

<https://orcid.org/0009-0006-3649-0461>

Department of Economics, College of Administration and Economics
University of Diyala, Iraq
omareco@uodiyala.edu.iq

Article's History

Received 9th of June, 2026; *Revised* 29th of July, 2026; *Accepted* 31st of August, 2026; *Available online*: 30th of September, 2026. *Published* as research article in the Volume XXI, Fall, Issue 4(94), 2026.

Copyright© 2026 The Author(s). This article is distributed under the terms of the license CC-BY 4.0., which permits any further distribution in any medium, provided the original work is properly cited.

Suggested Citation

Ashoor, E. J., Hussein, B. O., & Akawee, O. M. (2026). Forecasting Iraq's Broad Money Supply for 2026–2030: Evidence from Box–Jenkins Models and Monetary Policy Implications. *Journal of Applied Economic Sciences*, Volume XXI, Fall, 4(94), 1154 - 1170.
[https://doi.org/10.57017/jaes.v21.4\(94\).04](https://doi.org/10.57017/jaes.v21.4(94).04)

Abstract

This study forecasts Iraq's monthly broad money supply (M2) from February 2026 to December 2030 using 265 observations covering January 2004–January 2026. M2 is modelled as an observed monetary aggregate rather than a structural money-demand function. Unit-root. stationarity diagnostics indicate non-stationarity in levels. support one ordinary difference: the Phillips–Perron $Z(\alpha)$ test strongly rejects the unit-root null after differencing (-226.863 . $p < 0.01$). while KPSS does not reject stationarity (0.153 . $p > 0.10$); ADF evidence is specification-sensitive. Monthly effects remain relevant, although OCSB indicates that seasonal differencing is unnecessary ($D = 0$). A transparent ARIMA/SARIMA candidate set is evaluated using AIC/BIC, residual diagnostics, a 37-month holdout, expanding-window one-step forecasts, and simple benchmarks. SARIMA(0,1,1)(0,0,1)₁₂ without drift is retained because it combines significant MA parameters, acceptable Ljung–Box diagnostics. the lowest fixed-holdout RMSE among diagnostically adequate parsimonious candidates (4,279.3 billion dinars) and the best expanding-window RMSE (2,043.4 billion dinars). slightly better than the random-walk benchmark (2,121.5).

The central forecast remains near 169.3 trillion dinars. while uncertainty widens substantially; the December 2030 95% interval is approximately 134.5–204.2 trillion dinars. BDS rejects i.i.d. residuals, so the forecast is interpreted as a conditional linear baseline for liquidity monitoring, requiring uncertainty-aware apply and frequent re-estimation.

Keywords: broad money (M2), Iraq, Box–Jenkins, ARIMA, SARIMA, monetary policy.

JEL Classification: C22; C53; E51; E52.

Introduction

Iraq's broad money supply (M2) is an observed monetary aggregate that summarises liquid monetary balances in the economy. It needs to not be interpreted, by itself, as a structural money-demand function. A structural money-demand model relates desired real balances to determinants such as income, interest rates, inflation, exchange rates, opportunity costs, and financial innovation. The present study addresses a narrower forecasting question: how accurately might the historical monthly path of Iraq's M2 be represented. projected using Box–Jenkins time-series methods? This distinction is important because movements in M2 may reflect banking liquidity, government spending, oil-related foreign-currency flows, credit conditions. and policy operations rather than a single causal mechanism (Abdullah and Daoud, 2024; Akawee et al., 2024; International Monetary Fund, 2024).

M2 is relevant for policy monitoring given that it summarises currency, deposits, and banking liquidity. In Iraq. where public finance. external oil receipts can affect domestic liquidity conditions. forecasts of M2 can support scenario planning. early monitoring without being interpreted as causal estimates of monetary-policy effects. Evidence on Iraqi money demand. monetary transmission likewise underscores the distinction between forecasting an observed aggregate. estimating a behavioural demand equation (Ali and Ghafor, 2024).

Significance and Objectives of the Study

The study develops a transparent monthly forecasting framework for Iraq's M2 that explicitly addresses integration. 12-month seasonal dependence. ARIMA/SARIMA candidate comparison. residual diagnostics. and genuine out-of-sample validation.. The final forecasting horizon is February 2026–December 2030, comprising 59 monthly forecasts.

The objectives are to identify a parsimonious. empirically defensible Box–Jenkins model for Iraq's monthly M2. compare non-seasonal ARIMA specifications with appropriate SARIMA alternatives. evaluate predictive performance against simple benchmarks. and interpret the resulting forecasts in relation to monetary-policy. liquidity-management considerations.

1. Literature Review and Research Background

The literature relevant to this study may be organised into three strands. First, monetary-aggregate and money-demand research distinguishes observed monetary stocks from behavioural demand relationships. Work on monetary measurement. money-demand stability shows why monetary aggregates can remain informative while still responding to institutional. policy changes (Barnett and Chauvet, 2011; Lucas and Nicolini, 2015). Iraqi studies similarly examine the response of money supply to policy instruments. the behaviour of money demand. monetary velocity (Abdullah and Daoud, 2024; Ali and Ghafor, 2024; Akawee et al., 2024).

Second, the forecasting literature emphasises parsimonious ARIMA/SARIMA modelling together with evaluation on observations not used for estimation. Automatic. competition-based evidence demonstrates the importance of transparent model selection. benchmark comparison. and multiple forecast origins rather than relying exclusively on in-sample fit

(Hyndman and Khandakar, 2008; Makridakis and Hibon, 2000; Makridakis et al., 2020; Tashman, 2000). Forecast-accuracy measures. validation designs require particular care for persistent time series. and rolling or cross-validation approaches might provide a more informative assessment than a single in-sample statistic (Bergmeir et al., 2018; Hyndman and Koehler, 2006). Formal forecast-comparison procedures. proper scoring principles further reinforce the need to evaluate predictive performance directly (Diebold and Mariano, 1995; Gneiting and Raftery, 2007; Harvey et al., 1997).

Third, structural instability and nonlinear dependence are crucial sources of model risk in macroeconomic forecasting. Structural breaks may alter persistence. forecast performance. while regime-switching models provide one way to represent discrete changes in economic dynamics (Bai and Perron, 1998; Hamilton, 1989; Perron, 1989; Casini and Perron, 2019). Modern forecasting research moreover shows that machine-learning methods. forecast combinations may be useful when a single linear specification is insufficient. although additional complexity does not guarantee superior performance (Lim and Zohren, 2021; Makridakis et al., 2018; Makridakis and Winkler, 1983; Wang et al., 2023). The contribution of the present study is consequently a reproducible univariate forecasting application for Iraq's observed M2 aggregate, with explicit seasonality assessment, out-of-sample validation, benchmark comparison and uncertainty-aware monetary-policy interpretation.

2. Research Methodology

The study uses 265 monthly observations of Iraq's broad money supply (M2), in billion Iraqi dinars, from 2004M01 through 2026M01, sourced from the Central Bank of Iraq Monthly Statistical Bulletin (Central Bank of Iraq, 2026). ARIMA/SARIMA models are estimated by maximum likelihood in a state-space framework using statsmodels 0.14.6 (Seabold & Perktold, 2010), and the Phillips–Perron robustness test uses the $Z(\alpha)$ implementation in pmdarima 2.1.1 (Smith et al., 2017). Model selection combines AIC/BIC, parameter significance, residual diagnostics, parsimony, a fixed holdout (2004M01–2022M12 for estimation, 2023M01–2026M01 for validation), and an expanding-window one-step-ahead evaluation. Final forecasts are generated after re-estimating the selected model on the full sample, with 80% and 95% prediction intervals.

Box–Jenkins Modelling Framework

The Box–Jenkins framework identifies, estimates, diagnoses, and forecasts stochastic time-series models. ARIMA and seasonal ARIMA specifications are considered as monthly M2 may display both persistence and recurring annual dependence. Model choice therefore uses differencing requirements, ACF/PACF behaviour, seasonal lags, information criteria, parameter significance, residual diagnostics, parsimony, and out-of-sample forecast performance. For a covariance-stationary process, the lag- k autocorrelation might be expressed as follows:

$$\rho_k = \frac{\text{Cov}(Y_t, Y_{t-k})}{\text{Var}(Y_t)} = \frac{\gamma_k}{\gamma_0} \quad (1)$$

An autoregressive process of order p is represented by:

$$y_t = c + \sum_{i=1}^p \phi_i y_{t-i} + \varepsilon_t \quad (2)$$

where: ε_t is a white-noise innovation and the autoregressive coefficients satisfy the stationarity conditions of the selected specification. A moving-average process of order q is represented by:

$$y_t = \mu + \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j} \quad (3)$$

For monthly data, the general multiplicative seasonal ARIMA structure with period 12 is:

$$\Phi_p(L^{12})\phi_p(L)(1-L)^d(1-L^{12})^D y_t = \theta_q(L^{12})\theta_q(L)\varepsilon_t \quad (4)$$

where: L is the lag operator; p , d , q are the non-seasonal orders; p , d , q are the seasonal orders; the seasonal period is 12. The non-seasonal and seasonal orders are selected from directly estimated candidate models.

Stationarity, Seasonality and Diagnostic Testing

Integration is evaluated with the Augmented Dickey–Fuller, Phillips–Perron, and KPSS procedures. The ADF regression is:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{i=1}^k \delta_i \Delta y_{t-i} + \varepsilon_t \quad (5)$$

For the ADF-PP tests, the null hypothesis is a unit root; for KPSS, the null hypothesis is stationarity (Dickey and Fuller, 1979; Kwiatkowski et al., 1992; Phillips and Perron, 1988). The PP calculation uses a Newey–West long-run variance adjustment, and seasonal differencing is evaluated with the OCSB procedure (Newey and West, 1987; Osborn et al., 1988). Ordinary first differencing is defined by:

$$\Delta M2_t = M2_t - M2_{t-1} \quad (6)$$

Residual linear dependence is assessed with the Ljung–Box statistic (Ljung and Box, 1978):

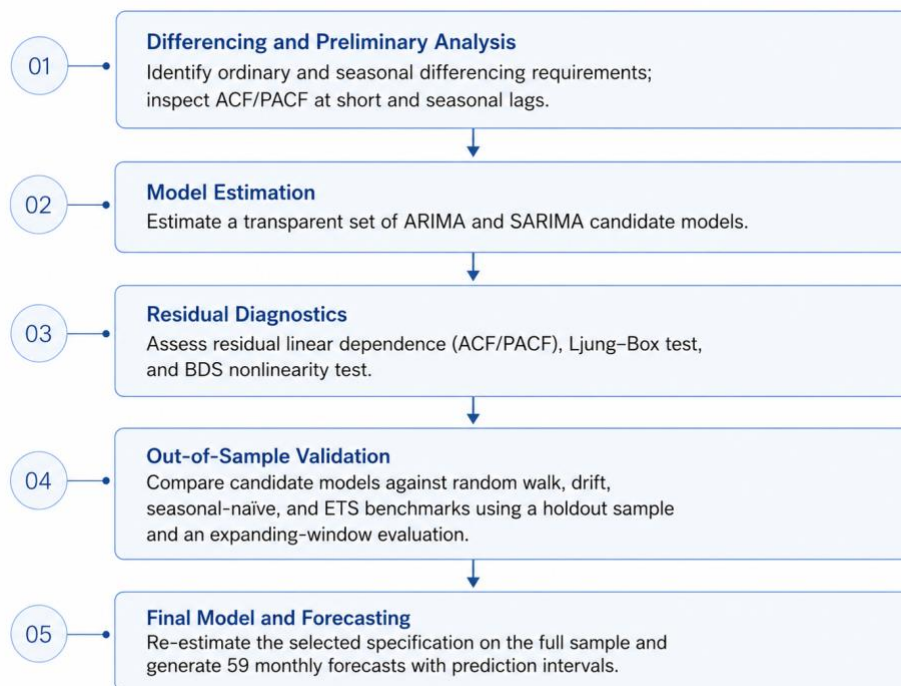
$$Q(m) = n(n+2) \sum_{k=1}^m \frac{\hat{\rho}_k^2}{n-k} \quad (7)$$

Residual heteroskedasticity is examined with an ARCH LM diagnostic (Engle, 1982), and remaining nonlinear or distributional dependence is assessed using the BDS independence test (Broock et al., 1996). BDS is used only as a residual independence diagnostic, not as a stationarity test.

Model Construction and Validation

The empirical procedure has five linked stages: (i) identify ordinary, seasonal differencing requirements, inspect the ACF/PACF at short, seasonal lags; (ii) estimate a transparent set of ARIMA, SARIMA candidates; (iii) assess residual linear dependence, other diagnostics; (iv) compare candidates against random-walk, drift, seasonal-naïve, and ETS benchmarks on a genuine holdout, an expanding-window exercise; (v) re-estimate the selected specification on the full sample, generate 59 monthly forecasts with prediction intervals. This validation design follows the general principle that forecast models must be judged on unseen observations, against explicit benchmarks (Bergmeir et al., 2018; Tashman, 2000). Figure 1 summarizes the Box–Jenkins workflow applied to Iraq’s monthly M2 series, integrating differencing and seasonality assessment, candidate-model estimation, residual diagnostics, out-of-sample validation, and final forecasting. This sequential procedure ensures that model selection reflects both statistical adequacy and predictive performance.

Figure 1: Box–Jenkins Workflow for Monthly Iraq M2



Source: Authors' design for the empirical forecasting procedure.

3. Empirical Application and Results

This section applies the Box–Jenkins framework to the monthly Iraq M2 dataset. All statistics, diagnostics, validation results, and forecasts reported below are calculated from the 265-observation series covering 2004M01–2026M01.

Descriptive Statistics

Across the 265 observations, the mean M2 is 85,372.32 billion dinars, the median is 88,855.35, the standard deviation is 51,557.49, the minimum is 7,445, and the maximum is 181,092. The Jarque–Bera statistic is 11,677 ($p = 0.0029$), so normality of the level distribution is rejected at conventional levels. Normality of the raw level series is not a prerequisite for ARIMA estimation; the more relevant diagnostics concern integration, serial dependence, and final-model residual behaviour.

Time-Series Persistence, Stationarity and Seasonality

The level-series correlogram is used as an initial diagnostic of persistence rather than as a stand-alone stationarity test. The null hypothesis for the Ljung–Box statistic is that autocorrelations through the tested lag are jointly zero. Persistent, slowly decaying autocorrelations in the M2 level series are consistent with non-stationarity and motivate formal unit-root testing.

Because the series is monthly, annual dependence is examined explicitly. In levels, $ACF(12) = 0.867$. After first differencing, $ACF(12)$ falls to 0.199 but remains statistically relevant. A joint regression test of monthly indicators in $\Delta M2$ gives $F = 3.336$ ($p = 0.00026$). OCSB selects $D=0$, indicating that ordinary differencing is sufficient for integration while seasonal AR/MA terms should still be evaluated.

Table 1: Autocorrelation and Partial Autocorrelation of Iraq's M2 Level Series

Lag	ACF	PACF	Ljung–Box Q	p-value
1	0.9901	0.9901	262.76	$p < 0.001$
2	0.9802	-0.0079	521.26	$p < 0.001$
3	0.9698	-0.0336	775.24	$p < 0.001$
6	0.9374	-0.0175	1509.43	$p < 0.001$
12	0.8668	0.0035	2845.51	$p < 0.001$
18	0.7866	-0.0224	3993.33	$p < 0.001$
24	0.6981	-0.0256	4936.04	$p < 0.001$

Source: Authors' calculations from Central Bank of Iraq (2026) monthly M2 data using statsmodels 0.14.6 and pmdarima 2.1.1.

Table 1 shows very high and slowly declining autocorrelation in the level series: $ACF(1) = 0.990$ and $ACF(12) = 0.867$. The Ljung–Box statistics reject joint zero autocorrelation at the displayed lags, which is consistent with a highly persistent non-stationary level process.

First differencing leaves 264 usable observations from 2004M02 through 2026M01. Table 2 reports the correlogram of the differenced series.

Table 2: First-Difference Correlogram for Iraq's M2

Lag	ACF	PAC	Ljung–Box Q	p-value
1	0.1966	0.1966	10.32	0.0013
2	0.1292	0.0942	14.80	$p < 0.001$
3	0.0235	-0.0189	14.94	0.0019
6	0.2270	0.1836	34.79	$p < 0.001$
12	0.1985	0.1379	55.18	$p < 0.001$
18	0.1599	0.0767	68.09	$p < 0.001$
24	0.0595	0.0254	76.34	$p < 0.001$

Source: Authors' calculations from Central Bank of Iraq (2026) monthly M2 data (265 observations, 2004M01–2026M01).

Short-lag autocorrelation is much smaller after differencing, although significant dependence remains, including at lag 12. This motivates estimation of ARIMA/SARIMA terms; stationarity does not require all autocorrelations to equal zero.

Unit-Root and Stationarity Diagnostics

Three ADF deterministic specifications are reported: no deterministic terms, an intercept, and an intercept plus trend. The PP $Z(\alpha)$ statistic is calculated under an intercept-and-trend regression with Newey–West bandwidth 5. KPSS with an intercept is reported as a complementary test whose null is stationarity.

Table 3: Unit-Root and Stationarity Diagnostics at the M2 Level

Test / specification	Statistic	p-value	5% critical value / null	Decision
ADF: none	0.8450	0.8926	-1.9421	Do not reject unit root
ADF: intercept	-0.7546	0.8321	-2.8731	Do not reject unit root
ADF: intercept + trend	-2.6107	0.2750	-3.4281	Do not reject unit root
PP Z(α): intercept + trend	-4.0261	0.8845	≈ -21.31	Do not reject unit root
KPSS: intercept	2.3437	$p < 0.01$	H0 = stationarity	Reject stationarity

Source: Authors' calculations from Central Bank of Iraq (2026) monthly M2 data; ADF lag length selected by AIC; PP Z(α) bandwidth = 5.

At level. all ADF variants fail to reject a unit root: $p = 0.893$ without deterministic terms. $p = 0.832$ with an intercept. and $p = 0.275$ with intercept. trend. The PP Z(α) statistic is -4.026 ($p = 0.885$), while KPSS rejects stationarity (2.344; $p < 0.01$). Taken together, the level series is non-stationary.

Table 4: Unit-Root and Stationarity Diagnostics for the First Difference of M2

Test/ specification	Statistic	p-value	5% critical value/ null	Decision
ADF: none	-2.0782	0.0361	-1.9421	Reject unit root
ADF: intercept	-2.7187	0.0709	-2.8731	Do not reject at 5%; marginal
ADF: intercept + trend	-2.6799	0.2445	-3.4281	Do not reject
PP Z(α): intercept + trend	-226.8633	$p < 0.01$	≈ -21.31	Reject unit root
KPSS: intercept	0.1531	> 0.10	H0 = stationarity	Do not reject stationarity

Source: Authors' calculations from Central Bank of Iraq (2026) monthly M2 data.

After first differencing, PP Z(α) = -226.863 ($p < 0.01$), decisively rejecting the unit-root null, and KPSS = 0.153 ($p > 0.10$), failing to reject stationarity. The ADF result is more sensitive to deterministic terms: ADF = -2.078 ($p = 0.036$) without deterministic terms -2.719 ($p = 0.071$) with an intercept, and -2.680 ($p = 0.245$) with intercept trend. The combined evidence supports ordinary integration order $d = 1$ while retaining the specification-sensitive ADF evidence explicitly.

Seasonal Diagnostics for Monthly M2

The first-differenced ACF is 0.199 at lag 12, and the joint month-dummy test is significant ($F = 3.336$, $p = 0.00026$). However, OCSB selects $D = 0$, so no seasonal difference is imposed. Seasonal AR/MA terms with period 12 are instead compared directly with non-seasonal ARIMA alternatives. This distinguishes seasonal correlation from a seasonal unit root.

Residual Independence and Nonlinearity Diagnostics

For the selected SARIMA specification, the BDS test uses embedding dimensions 2–6 and an epsilon of 1.5 residual standard deviations. BDS rejects the i.i.d.. null at every reported dimension (all $p < 0.003$). indicating remaining nonlinear or distributional dependence even though Ljung–Box tests find no significant linear residual autocorrelation at lags 12.24. The

BDS evidence is so treated as a model limitation and motivation for structural-break or nonlinear extensions.

Table 5: Residual Diagnostics for the Selected SARIMA Model

Diagnostic	Statistic	p-value	Interpretation
Ljung–Box Q(12)	18.499	0.1014	No significant linear residual autocorrelation
Ljung–Box Q(24)	32.752	0.1094	No significant linear residual autocorrelation
ARCH LM(12)	19.854	0.0699	No ARCH rejection at 5%
Jarque–Bera	18.487	$p < 0.001$	Residual normality rejected
BDS m=2	3.030	0.00245	Reject i.i.d.
BDS m=3	3.889	0.00010	Reject i.i.d.
BDS m=4	3.732	0.00019	Reject i.i.d.
BDS m=5	3.605	0.00031	Reject i.i.d.
BDS m=6	3.865	0.00011	Reject i.i.d.

Source: Authors' residual diagnostics for SARIMA(0,1,1)(0,0,1)₁₂, excluding the first 24 initialization residuals; BDS distance = 1.5 residual SDs.

The selected linear model is adequate with respect to residual linear autocorrelation, but the residuals are not fully i.i.d. according to BDS. Stationarity is established with unit-root and stationarity tests, whereas model adequacy is assessed with residual diagnostics; these are distinct inferential tasks.

Model Identification and Candidate Specification

A univariate AR lag-order screen is used only to guide the range of candidate autoregressive orders. Because it cannot identify the moving-average order q or seasonal orders P . Q . the final ARIMA/SARIMA specification is selected from directly estimated non-seasonal. seasonal candidates.

Table 6: Role of the AR-Lag Screen in Candidate Specification

Diagnostic screen	Result	Role in model identification
Univariate AR lag-order screen	Minimum AIC at lag 7	Used only to guide the range of candidate AR orders
Identification limitation	Does not identify MA order q or seasonal orders	Final ARIMA/SARIMA order selected from directly estimated candidates

Source: Authors' model-identification procedure.

Candidate-Model Estimation and Selection

Candidate specifications include ARIMA(0,1,0), ARIMA(1,1,0), ARIMA(2,1,0), ARIMA(0,1,1), ARIMA(0,1,2), ARIMA(1,1,1), ARIMA(2,1,1), ARIMA(3,1,2) and parsimonious SARIMA alternatives with period 12. Drift and no-drift versions are checked where relevant. Because the historical drift performs poorly during the 2023–2026 holdout, the selected specification excludes drift. R-squared is not used as a model-selection criterion.

Table 7: Candidate ARIMA/SARIMA Comparison and Model Selection

Model	AIC	BIC	LB p(24)	Fixed-holdout RMSE	MAPE %	Decision
ARIMA(1,1,0)	4641.33	4648.48	0.0014	4,487.5	1.995	Good accuracy; residual autocorrelation remains
ARIMA(2,1,0)	4635.66	4646.39	0.0139	5,211.9	2.494	Residual autocorrelation remains
ARIMA(0,1,1)	4648.34	4655.50	$p < 0.001$	5,798.9	2.622	Residual autocorrelation remains
ARIMA(0,1,2)	4640.01	4650.74	0.0198	4,474.6	1.991	Good accuracy; residual autocorrelation remains
ARIMA(1,1,1)	4616.39	4627.12	0.0673	35,080.4	16.767	Low AIC but unstable holdout
ARIMA(3,1,2)	4616.00	4637.46	0.2961	34,299.7	16.368	Adequate residual diagnostics but poor holdout performance
SARIMA(0,1,1)(0,0,1) ₁₂	4634.26	4644.99	0.1094	4,279.3	2.201	Selected: best expanding-window accuracy with residual adequacy
SARIMA(1,1,1)(1,0,0) ₁₂	4613.21	4627.51	0.2435	35,864.2	17.124	Lowest AIC class but unstable holdout
ARIMA(0,1,0) + drift	4627.61	4634.76	$p < 0.001$	12,512.8	5.328	Historical drift performs poorly

Source: Authors' maximum-likelihood estimates and 37-month fixed-holdout evaluation.

The preferred specification is SARIMA(0,1,1)(0,0,1)₁₂ without drift. Its non-seasonal MA coefficient is 0.2064 (SE = 0.0532, $p < 0.001$), and its seasonal MA(12) coefficient is 0.2279 (SE = 0.0638, $p < 0.001$). Both are statistically significant, and the model is invertible (minimum MA-root modulus = 1.131). Some higher-order candidates have slightly lower in-sample AIC/BIC but indicate poor holdout stability. The selected model combines the best expanding-window forecast accuracy with acceptable Ljung–Box residual diagnostics.

In lag-operator form, the fitted model is:

$$(1 - L)M2_t = (1 + 0.2064L)(1 + 0.2279L^{12})\varepsilon_t \quad (8)$$

Expanding the multiplicative moving-average operators gives:

$$\Delta M2_t = \varepsilon_t + 0.2064\varepsilon_{t-1} + 0.2279\varepsilon_{t-12} + 0.0470\varepsilon_{t-13} \quad (9)$$

The multiplicative cross-term is $0.2064 \times 0.2279 = 0.04703856$, reported as 0.0470. No constant or drift is included. The innovation variance is $\sigma^2 = 2,633,186$. Multi-step forecasts are generated dynamically by the fitted state-space recursion and returned directly to M2 levels.

Out-of-Sample Forecast Validation

The primary holdout uses 2004M01–2022M12 for estimation and 2023M01–2026M01 (37 months) for evaluation. A second expanding-window exercise re-estimates each model each month and produces one-step-ahead forecasts over the same validation period. This reduces dependence on a single forecast origin and more closely approximates operational monthly updating.

Table 8: Out-of-Sample Forecast Evaluation and Benchmark Comparison

Model / benchmark	RMSE	MAE	MAPE %	Theil U	Expanding-window RMSE
Selected SARIMA(0,1,1)(0,0,1) ₁₂	4,279.3	3,819.6	2.201	0.01229	2,043.4
ARIMA(1,1,0)	4,487.5	3,520.8	1.995	0.01299	2,082.1
ARIMA(0,1,2)	4,474.6	3,512.6	1.991	0.01295	2,124.7
Random walk	6,815.7	5,753.8	3.264	0.01991	2,121.5
Random walk with drift	12,512.8	9,132.2	5.328	0.03515	2,218.5
Seasonal naïve	24,647.0	23,230.1	13.332	0.07582	—
ETS additive trend-seasonal	45,036.1	36,623.0	21.241	0.11669	—
Holt linear trend	46,565.7	38,312.6	22.213	0.12011	—

Source: Authors' 37-month fixed-holdout and expanding-window validation calculations.

In the fixed holdout, the selected SARIMA model records RMSE=4,279.3, MAE=3,819.6, MAPE = 2.201%, and Theil U = 0.01229. The random-walk benchmark gives RMSE = 6,815.7. MAPE = 3.264%. while a random walk with historical drift performs materially worse (RMSE = 12,512.8). In the expanding-window evaluation, the selected SARIMA again has the lowest RMSE among the principal candidates at 2,043.4. compared with 2,121.5 for the random walk. 2,082.1 for ARIMA(1,1,0). Model selection therefore balances information criteria, statistically meaningful parameters, residual adequacy, parsimony, and out-of-sample accuracy.

Forecasts for February 2026–December 2030

The forecasting horizon comprises 59 monthly forecasts from February 2026 through December 2030. SARIMA(0,1,1)(0,0,1)₁₂ is re-estimated on all 265 observations and used for dynamic multi-step forecasting. Because the model contains no drift, the central forecast does not mechanically increase by a fixed amount each month. Seasonal innovation effects fade and the point forecast converges towards approximately 169,348 billion dinars, while prediction intervals widen with the horizon.

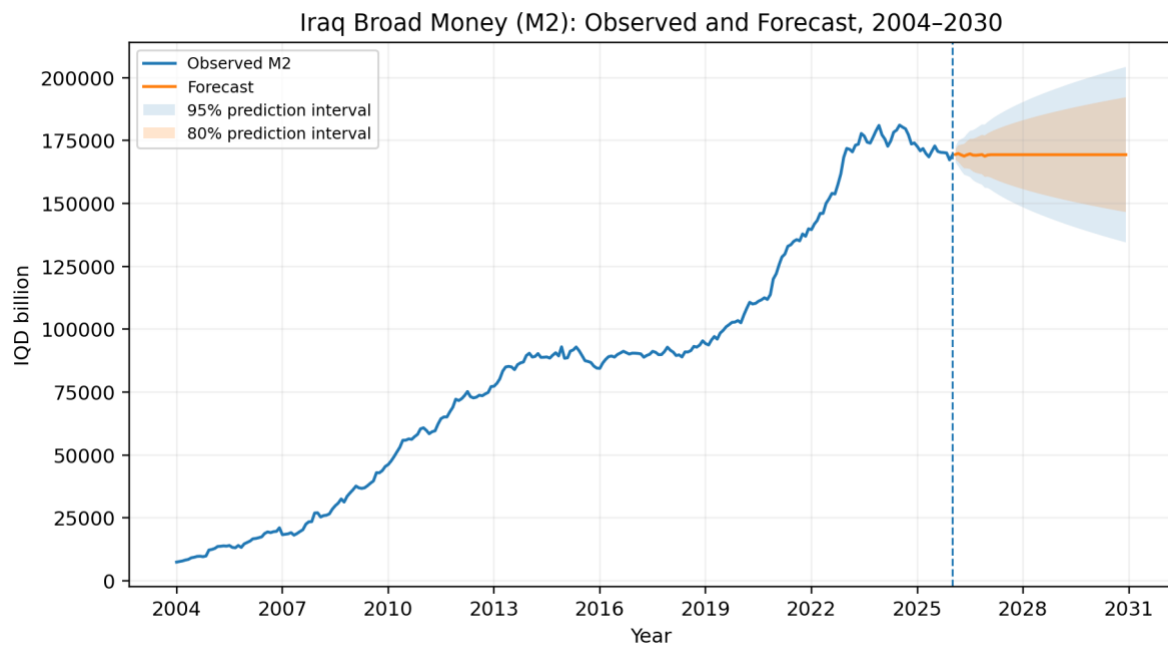
Table 9: Selected Forecast Horizons for Iraq's Broad Money (M2), February 2026–December 2030

Horizon	Point forecast	80% interval	95% interval	Unit
2026M02	169,455.2	167,375.6–171,534.8	166,274.8–172,635.7	IQD bn
2026M12	168,762.4	160,561.0–176,963.8	156,219.4–181,305.4	IQD bn
2027M12	169,348.4	156,031.4–182,665.4	148,981.8–189,715.0	IQD bn

Horizon	Point forecast	80% interval	95% interval	Unit
2028M12	169,348.4	152,283.5–186,413.3	143,249.9–195,446.9	IQD bn
2029M12	169,348.4	149,221.8–189,474.9	138,567.5–200,129.3	IQD bn
2030M12	169,348.4	146,568.0–192,128.8	134,508.8–204,188.0	IQD bn

Source: Authors' calculations

Figure 2: Observed and Forecast Iraq M2 (2004–2030) with 80% and 95% Prediction Intervals



Source: Authors' calculations from Central Bank of Iraq (2026) monthly M2 data using SARIMA(0,1,1)(0,0,1)₁₂.

Table 10: Annual Summary of M2 Forecasts

Year	Forecast months	Annual average	End-year forecast	End-year 95% interval
2026	11 (Feb–Dec)	169,263.1	168,762.4	156,219.4–181,305.4
2027	12	169,338.1	169,348.4	148,981.8–189,715.0
2028	12	169,348.4	169,348.4	143,249.9–195,446.9
2029	12	169,348.4	169,348.4	138,567.5–200,129.3
2030	12	169,348.4	169,348.4	134,508.8–204,188.0

Note: Central forecasts are in billion Iraqi dinars, with only ~0.07% change through 2030; long-term uncertainty dominates. The forecast begins at 169,455.2 billion dinars in 2026M02, is 168,762.4 in 2026M12 and converges to about 169,348.4 from late 2027 onward. By 2030M12, the 80% prediction interval is 146,568.0–192,128.8, the 95% interval is 134,508.8–204,188.0 billion dinars. The widening intervals are central to the policy interpretation.

Conclusion

Four findings emerge. First, Iraq's M2 level series is non-stationary. while combined PP. KPSS evidence supports one ordinary difference ($d = 1$); the ADF result after differencing remains specification-sensitive. Second, monthly dependence remains after ordinary differencing, but OCSB indicates that a seasonal unit root is unnecessary ($D = 0$). Third. SARIMA(0,1,1)(0,0,1)₁₂ without drift is preferred because it combines significant parameters. acceptable Ljung–Box residual diagnostics. parsimony. and the best expanding-window forecast accuracy among the principal candidates. Fourth, the 59-month central forecast is broadly stable near 169.3 trillion dinars, while the 95% interval widens markedly through 2030. BDS rejection of i.i.d. residuals reinforce the interpretation of the model as a conditional linear baseline rather than a complete causal description of Iraq's monetary dynamics.

Monetary Policy Implications

The M2 forecast is best used as a baseline for liquidity monitoring rather than as a monetary target. Material deviations above the forecast range would warrant examination of fiscal injections, public-sector payment cycles, reserve, deposit creation, bank credit and foreign-exchange operations; persistent downside deviations could indicate liquidity withdrawal or weaker balance-sheet expansion. This interpretation is consistent with the emphasis on fiscal-monetary coordination, liquidity conditions and monetary-policy transmission in Iraq (International Monetary Fund, 2024), while the univariate model itself cannot identify causality.

Prediction intervals are essential for policy application. The December 2030 central forecast is 169,348.4 billion dinars, but the 95% range is approximately 134,508.8–204,188 billion dinars. The point forecast must therefore not be treated as a deterministic policy target. As new monthly observations arrive, the model must be re-estimated and benchmark accuracy ought to be monitored.

Limitations and Future Research

The analysis is univariate and cannot identify the causal drivers of M2. Residual Ljung–Box diagnostics are satisfactory at lags 12. 24 and the ARCH LM test is not significant at 5% ($p = 0.070$) but residual normality is rejected (Jarque–Bera $p < 0.001$). BDS rejects i.i.d. residuals across embedding dimensions 2–6. These results indicate remaining nonlinear or distributional structure and possible regime effects. The 59-month horizon additionally compounds parameter and model uncertainty. Future work ought to consider structural-break models, multivariate monetary determinants, nonlinear methods and forecast combinations while retaining simple benchmarks as operational references (Bai and Perron, 1998; Casini and Perron, 2019; Lim and Zohren, 2021; Wang et al., 2023).

For operational forecasting, SARIMA(0,1,1)(0,0,1)₁₂ ought to be re-estimated whenever new monthly M2 observations become available, with accuracy rechecked against a random walk. other simple benchmarks. The absence of drift is an empirical result for this sample, not a permanent monetary-law assumption; major fiscal, exchange-rate, oil-market or banking-sector changes need to trigger renewed model identification.

CRedit Authorship Contribution Statement

Ashoor, E. J.: Conceptualization, Methodology, Formal analysis, Writing – original draft. Hussein, B. O.: Validation, Investigation, Writing – review & editing. Akawee, O. M.: Data curation, Software, Visualization, Validation, Writing – review & editing. All authors reviewed and approved the final manuscript.

Acknowledgments / Funding

The authors report no external funding for this study and no additional acknowledgments.

Conflict of Interest Statement

The authors declare no commercial or financial relationships that could be construed as a potential conflict of interest.

Data Availability Statement

The monthly broad money supply (M2) series used in this study was obtained from the Central Bank of Iraq (CBI), Statistics and Research Department, Monthly Statistical Bulletin (النشرة الإحصائية الشهرية), official source page: <https://cbi.iq/page/122>. At the time of access, that page linked the official downloadable Excel workbook at <https://cbi.iq/static/uploads/up/file-178962848411389.xlsx>. The estimation dataset was restricted to 2004M01–2026M01; observations published after 2026M01 were not used. The analysed working dataset supporting the reported results is available from the corresponding author upon reasonable request.

Ethical Approval Statement

This study uses aggregate secondary macroeconomic time-series data and does not involve human participants, personal data, or animal subjects. Therefore, ethical approval was not required.

References

- Abdullah, A. W., & Daoud, S. S. (2024). Measuring and analysing the response of the money supply to the targeted monetary policy in Iraq. *Khazayin of Economic and Administrative Sciences*, 1(1), 25–40. <https://doi.org/10.69938/Keas.2401013>
- Akawee, O. M., Yas, H. R. A., & Kolaib, A. G. (2024). The impact of monetary policy on gross domestic product in Iraq: New evidence using the autoregressive distributed lag model. *International Journal of Economics and Finance Studies*, 16(3), 16–33. <https://doi.org/10.34109/ijefs.202416302>
- Ali, S. Q., & Ghafor, N. M. (2024). Analyzing and measuring the effect of the velocity of money on the money demand in Iraq during the period (1990–2022). *Journal of University of Human Development*, 10(1), 22–30. <https://doi.org/10.21928/juhd.v10n1y2024.pp22-30>
- Bai, J., & Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1), 47–78. <https://doi.org/10.2307/2998540>
- Barnett, W. A., & Chauvet, M. (2011). How better monetary statistics could have signaled the financial crisis. *Journal of Econometrics*, 161(1), 6–23. <https://doi.org/10.1016/j.jeconom.2010.09.002>
- Bergmeir, C., Hyndman, R. J., & Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, 120, 70–83. <https://doi.org/10.1016/j.csda.2017.11.003>
- Broock, W. A., Scheinkman, J. A., Dechert, W. D., & LeBaron, B. (1996). A test for independence based on the correlation dimension. *Econometric Reviews*, 15(3), 197–235. <https://doi.org/10.1080/07474939608800353>
- Central Bank of Iraq. (2026). Monthly Statistical Bulletin [official Excel workbook; broad money (M2) monthly series]. Statistics and Research Department. <https://cbi.iq/page/122>

- Casini, A., & Perron, P. (2019). Structural breaks in time series. *Oxford Research Encyclopedia of Economics and Finance*. <https://doi.org/10.1093/acrefore/9780190625979.013.179>
- Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366a), 427–431. <https://doi.org/10.1080/01621459.1979.10482531>
- Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), 253–263. <https://doi.org/10.1080/07350015.1995.10524599>
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4), 987–1007. <https://doi.org/10.2307/1912773>
- Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477), 359–378. <https://doi.org/10.1198/016214506000001437>
- Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2), 357–384. <https://doi.org/10.2307/1912559>
- Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13(2), 281–291. [https://doi.org/10.1016/S0169-2070\(96\)00719-4](https://doi.org/10.1016/S0169-2070(96)00719-4)
- Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, 27(3), 1–22. <https://doi.org/10.18637/jss.v027.i03>
- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688. <https://doi.org/10.1016/j.ijforecast.2006.03.001>
- International Monetary Fund. (2024). Iraq: 2024 Article IV consultation - Press release; staff report; and statement by the Executive Director for Iraq (IMF Country Report No. 2024/128). <https://doi.org/10.5089/9798400275357.002>
- Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series carry a unit root? *Journal of Econometrics*, 54(1–3), 159–178. [https://doi.org/10.1016/0304-4076\(92\)90104-Y](https://doi.org/10.1016/0304-4076(92)90104-Y)
- Lim, B., & Zohren, S. (2021). Time-series forecasting with deep learning: A survey. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379(2194), 20200209. <https://doi.org/10.1098/rsta.2020.0209>
- Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2), 297–303. <https://doi.org/10.1093/biomet/65.2.297>
- Lucas, R. E., Jr., & Nicolini, J. P. (2015). On the stability of money demand. *Journal of Monetary Economics*, 73, 48–65. <https://doi.org/10.1016/j.jmoneco.2015.03.005>
- Makridakis, S., & Hibon, M. (2000). The M3-competition: Results, conclusions and implications. *International Journal of Forecasting*, 16(4), 451–476. [https://doi.org/10.1016/S0169-2070\(00\)00057-1](https://doi.org/10.1016/S0169-2070(00)00057-1)
- Makridakis, S., & Winkler, R. L. (1983). Averages of forecasts: Some empirical results. *Management Science*, 29(9), 987–996. <https://doi.org/10.1287/mnsc.29.9.987>

- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods: Concerns and ways forward. *PLOS ONE*, 13(3), e0194889. <https://doi.org/10.1371/journal.pone.0194889>
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 Competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, 36(1), 54–74. <https://doi.org/10.1016/j.ijforecast.2019.04.014>
- Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3), 703–708. <https://doi.org/10.2307/1913610>
- Osborn, D. R., Chui, A. P. L., Smith, J. P., & Birchenhall, C. R. (1988). Seasonality and the order of integration for consumption. *Oxford Bulletin of Economics and Statistics*, 50(4), 361–377. <https://doi.org/10.1111/j.1468-0084.1988.mp50004002.x>
- Perron, P. (1989). The Great Crash, the oil price shock, and the unit root hypothesis. *Econometrica*, 57(6), 1361–1401. <https://doi.org/10.2307/1913712>
- Phillips, P. C. B., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75(2), 335–346. <https://doi.org/10.1093/biomet/75.2.335>
- Seabold, S., & Perktold, J. (2010). statsmodels: Econometric and statistical modeling with Python. Proceedings of the 9th Python in Science Conference, 92–96. <https://doi.org/10.25080/Majora-92bf1922-011>. Soft version used: 0.14.6; documentation: <https://www.statsmodels.org/stable/>
- Smith, T. G., et al. (2017–). pmdarima: ARIMA estimators for Python. Software version used: 2.1.1; documentation: <https://alkaline-ml.com/pmdarima/>
- Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: An analysis and review. *International Journal of Forecasting*, 16(4), 437–450. [https://doi.org/10.1016/S0169-2070\(00\)00065-0](https://doi.org/10.1016/S0169-2070(00)00065-0)
- Wang, X., Hyndman, R. J., Li, F., & Kang, Y. (2023). Forecast combinations: An over 50-year review. *International Journal of Forecasting*, 39(4), 1518–1547. <https://doi.org/10.1016/j.ijforecast.2022.11.005>

Appendix A. Complete Monthly Forecasts

Table S1: Complete Monthly M2 Forecasts, 2026M02–2030M12

Month	Forecast	80% lower	80% upper	95% lower	95% upper
2026M02	169,455.2	167,375.6	171,534.8	166,274.8	172,635.7
2026M03	169,795.9	166,537.3	173,054.5	164,812.3	174,779.5
2026M04	169,250.6	165,138.1	173,363.0	162,961.1	175,540.0
2026M05	168,782.8	163,965.6	173,600.1	161,415.4	176,150.3
2026M06	169,295.8	163,864.4	174,727.2	160,989.2	177,602.4
2026M07	169,649.2	163,666.4	175,632.0	160,499.2	178,799.1
2026M08	169,163.6	162,676.1	175,651.2	159,241.8	179,085.4
2026M09	169,104.8	162,149.1	176,060.5	158,467.0	179,742.6
2026M10	169,230.5	161,836.2	176,624.8	157,921.9	180,539.1
2026M11	169,403.2	161,594.9	177,211.5	157,461.4	181,344.9
2026M12	168,762.4	160,561.0	176,963.8	156,219.4	181,305.4
2027M01	169,224.4	160,647.9	177,801.0	156,107.7	182,341.1
2027M02	169,348.4	160,268.0	178,428.8	155,461.1	183,235.6
2027M03	169,348.4	159,759.7	178,937.0	154,683.8	184,013.0
2027M04	169,348.4	159,277.1	179,419.7	153,945.7	184,751.1
2027M05	169,348.4	158,816.5	179,880.2	153,241.3	185,455.5
2027M06	169,348.4	158,375.3	180,321.5	152,566.5	186,130.3
2027M07	169,348.4	157,951.1	180,745.7	151,917.7	186,779.0
2027M08	169,348.4	157,542.2	181,154.6	151,292.3	187,404.5
2027M09	169,348.4	157,146.9	181,549.9	150,687.8	188,008.9
2027M10	169,348.4	156,764.1	181,932.7	150,102.3	188,594.4
2027M11	169,348.4	156,392.5	182,304.2	149,534.1	189,162.6
2027M12	169,348.4	156,031.4	182,665.4	148,981.8	189,715.0
2028M01	169,348.4	155,679.7	183,017.0	148,444.0	190,252.8
2028M02	169,348.4	155,336.9	183,359.8	147,919.7	190,777.1
2028M03	169,348.4	155,002.3	183,694.5	147,408.0	191,288.8
2028M04	169,348.4	154,675.3	184,021.4	146,907.9	191,788.9
2028M05	169,348.4	154,355.5	184,341.3	146,418.7	192,278.1
2028M06	169,348.4	154,042.3	184,654.5	145,939.7	192,757.0

Month	Forecast	80% lower	80% upper	95% lower	95% upper
2028M07	169,348.4	153,735.4	184,961.4	145,470.4	193,226.4
2028M08	169,348.4	153,434.4	185,262.4	145,010.1	193,686.7
2028M09	169,348.4	153,139.0	185,557.7	144,558.3	194,138.5
2028M10	169,348.4	152,848.9	185,847.9	144,114.6	194,582.1
2028M11	169,348.4	152,563.8	186,132.9	143,678.6	195,018.2
2028M12	169,348.4	152,283.5	186,413.3	143,249.9	195,446.9
2029M01	169,348.4	152,007.7	186,689.1	142,828.1	195,868.7
2029M02	169,348.4	151,736.2	186,960.5	142,412.9	196,283.9
2029M03	169,348.4	151,468.9	187,227.9	142,004.0	196,692.8
2029M04	169,348.4	151,205.4	187,491.3	141,601.2	197,095.6
2029M05	169,348.4	150,945.8	187,751.0	141,204.1	197,492.7
2029M06	169,348.4	150,689.8	188,007.0	140,812.5	197,884.3
2029M07	169,348.	150,437.2	188,259.6	140,426.2	198,270.6
2029M08	169,348.4	150,188.0	188,508.8	140,045.0	198,651.7
2029M09	169,348.4	149,941.9	188,754.9	139,668.8	199,028.0
2029M10	169,348.4	149,699.0	188,997.8	139,297.2	199,399.6
2029M11	169,348.4	149,459.0	189,237.8	138,930.1	199,766.6
2029M12	169,348.4	149,221.8	189,474.9	138,567.5	200,129.3
2030M01	169,348.4	148,987.5	189,709.3	138,209.1	200,487.7
2030M02	169,348.4	148,755.8	189,941.0	137,854.7	200,842.1
2030M03	169,348.4	148,526.6	190,170.1	137,504.3	201,192.5
2030M04	169,348.4	148,300.0	190,396.8	137,157.7	201,539.1
2030M05	169,348.4	148,075.8	190,621.0	136,814.8	201,882.0
2030M06	169,348.4	147,853.9	190,842.8	136,475.5	202,221.3
2030M07	169,348.4	147,634.3	191,062.4	136,139.6	202,557.2
2030M08	169,348.4	147,416.9	191,279.9	135,807.1	202,889.7
2030M09	169,348.4	147,201.6	191,495.1	135,477.9	203,218.9
2030M10	169,348.4	146,988.4	191,708.3	135,151.8	203,545.0
2030M11	169,348.4	146,777.2	191,919.5	134,828.8	203,868.0
2030M12	169,348.4	146,568.0	192,128.8	134,508.8	204,188.0

Note: Values are billion Iraqi dinars. Forecasts are dynamic multi-step projections from SARIMA(0,1,1)(0,0,1)₁₂ without drift, re-estimated on all 265 observations through 2026M01.