APPLICATION OF TIME SERIES ANALYSIS METHODS TO GOLD PRICE FORECASTING

YULIIA YORDANOVA

ABSTRACT: *This paper applies time series methods to forecast gold prices using weekly LBMA spot prices (USD/troy oz), January 1980-April 2026 (2411 observations), through descriptive diagnostics, stationarity/autocorrelation tests (ADF, ACF/PACF, Ljung-Box), ARIMA(1,1,1)-GARCH(1,1) modelling with Student-t innovations, and 10,000 Monte Carlo simulations. The series is I(1), with strong conditional heteroscedasticity (ARCH χ^2 = 191.92, $p < 0.001$) and leptokurtic returns (kurtosis 6.34). ARIMA(1,1,1) is selected (AIC = −11,080.99), while GARCH(1,1) (ν = 6.20) captures volatility clustering and fat tails. Monte Carlo 95% CIs at a 4-week horizon are [4,346–5,243] USD (ARIMA) and [4,106–5,551] USD (GARCH). An out-of-sample check on seven weekly observations (19 April – 31 May 2026) gives MAE = 167.32 USD (3.7%) for ARIMA and 186.73 USD (4.1%) for GARCH; a second check on 16 weeks (June–September 2026) shows that neither model outperforms a naive forecast in point accuracy, while the GARCH 95% intervals covered all realised prices. Results confirm the hypothesis; the forecast evaluation is preliminary and specific to the periods studied.*

KEYWORDS: *gold price forecasting, time series analysis, ARIMA, GARCH(1,1), Monte Carlo simulation, volatility, safe haven asset, conditional heteroscedasticity*.

DOI: https://doi.org/10.46687/DIUK8136

ПРИЛОЖЕНИЕ НА МЕТОДИТЕ ЗА АНАЛИЗ НА ДИНАМИЧНИ РЕДОВЕ ЗА ПРОГНОЗИРАНЕ НА ЦЕНАТА НА ЗЛАТОТО[*]

ЮЛИЯ ЙОРДАНОВА

РЕЗЮМЕ: *Настоящата статия прилага методи за анализ на времеви редове за прогнозиране цената на златото чрез седмични спот цени от LBMA (USD/трой унция), януари 1980-април 2026 (2411 наблюдения), посредством описателна диагностика, тестове за стационарност и автокорелация (ADF, ACF/PACF, Ljung-Box), моделиране ARIMA(1,1,1)-GARCH(1,1) с остатъци с разпределение на Student-t и 10 000 симулации по Монте Карло. Редът е I(1), със силна условна хетероскедастичност (ARCH χ^2 = 191.92, $p < 0.001$) и разпределение на доходностите (ексцес 6.34). ARIMA(1,1,1) е избраният модел (AIC = −11 080.99), а GARCH(1,1) (ν = 6.20) улавя клъстеризацията на волатилността и тежките опашки. 95% доверителни интервали по Монте Карло за 4-седмичен хоризонт са [4 346–5 243] USD (ARIMA) и [4 106–5 551] USD (GARCH). Проверката извън извадката върху седем седмични наблюдения (19 април – 31 май 2026 г.) дава MAE = 167.32 USD (3.7%) за ARIMA и 186.73 USD (4.1%) за GARCH, а втората проверка върху 16 седмици (юни–септември 2026 г.) показва, че нито един модел не превъзхожда наивната прогноза по точност, докато 95% интервалите на GARCH обхващат всички реални цени. Резултатите потвърждават хипотезата; оценката на прогнозите е предварителна и се отнася за разгледаните периоди.*

КЛЮЧОВИ ДУМИ: *прогнозиране на цена на злато, анализ на динамични редове, ARIMA, GARCH(1,1), симулация на Монте Карло, волатилност, защитен актив, условна хетероскедастичност*

[*] This work was partially supported by the Konstantin Preslavsky University of Shumen under project No. RD-08-124/06.02.2026.

1 Introduction

Forecasting the prices of financial assets is one of the central tasks of applied econometrics. Gold occupies a unique position among tradable assets: it simultaneously serves as a commodity, an investment instrument, and an international reserve asset. This triple function makes it exceptionally sensitive to macroeconomic shocks, geopolitical disruptions, and shifts in monetary policy. The price dynamics of the metal provide a vivid illustration. From approximately 270 USD per troy ounce at the dawn of the twenty-first century, the price rose above 5,000 USD by 2026, traversing numerous sharp swings driven by the global financial crisis (2007-2009), the European sovereign debt crisis (2010-2012), the COVID-19 pandemic (2020), and geopolitical tensions after 2022 [1]. Each of these episodes reinforces the role of gold as a safe haven asset - one whose correlation with broad equity indices is historically weak or negative during periods of stress.

From the perspective of statistical modelling, gold price series exhibit three characteristic features that complicate forecasting. First, the level series is non-stationary, following an integrated process of order one, I(1). Second, the variance of log-returns is not constant but time-varying and clustered - a property known as conditional heteroscedasticity. Third, the return distribution has fat tails: extreme price movements occur far more often than a Gaussian model would predict. These stylised facts are well established in the empirical finance literature [1, 3, 4, 7].

The methodological toolkit for addressing these features has matured over the past four decades. The ARIMA class of models, proposed by Box and Jenkins [2], provides a rigorous framework for modelling the conditional mean of an integrated time series. The ARCH model introduced by Engle [4] and its generalisation GARCH(1,1) developed by Bollerslev [3] capture conditional variance dynamics and have become the industry standard for volatility modelling. Monte Carlo simulation, rooted in the foundational work of Metropolis and Ulam [9], transforms point-in-time model estimates into probability distributions over future paths, enabling risk quantification beyond simple confidence intervals.

Despite the richness of this methodology, comprehensive empirical studies applying all three components jointly to gold prices have not been encountered in the reviewed Bulgarian academic literature on gold price modelling. The present paper addresses this gap by developing, estimating, and validating an integrated ARIMA-GARCH-Monte Carlo framework on 46 years of weekly gold price data. The working hypothesis is that weekly gold prices follow a non-stationary process with conditional heteroscedasticity and fat-tailed log-return distributions, so that adequate modelling requires a two-component structure - ARIMA for the conditional mean and GARCH for the conditional variance - while reliable probabilistic forecasts can only be obtained via Monte Carlo simulation.

The remainder of the paper is organized as follows. Section 2 describes the data and outlines the four-stage methodology. Section 3 presents the descriptive statistical analysis and stationarity tests. Section 4 reports the ARIMA and GARCH estimation results. Section 5 presents the Monte Carlo forecasting experiment and its results. Section 6 validates the forecasts and discusses forecast accuracy. Section 7 concludes with the main findings, scientific contributions, and directions for future research.

2 Data and Methodology

2.1 Data

The study uses weekly LBMA spot prices of gold in USD per troy ounce, downloaded from Investing.com [5]. The LBMA Gold Price, fixed twice daily in London, is the global

benchmark for physical gold transactions. Weekly frequency was chosen to capture short-term dynamics and volatility structure while reducing intra-day noise. The full sample spans January 1980 to 12 April 2026, comprising N = 2,411 observations. The last recorded value is P_0 = 4,770.32 USD/oz. All computations were performed in the R statistical environment using the forecast, rugarch, and tseries packages.

The working variable for modelling is the series of log-returns, defined as

(1) $$r_t = \ln(P_t) - \ln(P_{t-1}),$$

where P_t denotes the price at time t. Log-returns are additive across periods, exhibit a more symmetric distribution than simple returns, and are approximately stationary even when the price series is I(1) [7, 8]. These properties make them the standard transformation in financial time series analysis.

2.2 Descriptive analysis and normality testing

The first methodological stage comprises descriptive statistics (mean, standard deviation, skewness, kurtosis), a time-series plot, a histogram, a Q-Q plot against a normal reference, and the Jarque-Bera test. The null hypothesis of the Jarque-Bera test is that the distribution is normal; rejection indicates departure from normality, typically through fat tails or skewness.

2.3 Stationarity and autocorrelation analysis

The Augmented Dickey-Fuller (ADF) test is applied to both the price level and the log-return series. The autocorrelation function (ACF) and the partial autocorrelation function (PACF) are examined at lags 1 through 20. The Ljung-Box statistic

(2) $$Q(m) = N(N+2)\Sigma^{k=1}{}^{m}\ \hat{\rho}^{k2}/(N-k)$$

is computed for lags m = 10 and m = 20, both for the residuals and for their squares. Significant autocorrelation in squared residuals is the standard diagnostic for ARCH effects.

2.4 ARIMA modelling

The ARIMA(p, d, q) class, developed by Box and Jenkins [2], models the conditional mean of a d-times-differenced series through

(3) $$\varphi(B)(1-B)^{d}Y_t = \theta(B)\varepsilon_t,$$

where $\varepsilon_t \sim WN(0, \sigma^2)$. The integration order d is determined by the ADF test. Competing specifications are compared by the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). Model adequacy is assessed by the Ljung–Box test on residuals.

2.5 GARCH(1,1) modelling

The GARCH(1,1) equation is

(4) $$\sigma^2_t = \omega + \alpha_1\varepsilon^2_{t-1} + \beta_1\sigma^2_{t-1}.$$

Covariance stationarity requires $\alpha_1 + \beta_1 < 1$; the unconditional variance is

(5) $$\sigma^2 = \omega/(1-\alpha_1-\beta_1).$$

Given the established fat tails, innovations follow a Student-t distribution with ν degrees of freedom. All parameters are estimated by maximum likelihood.

2.6 Monte Carlo simulation

The Monte Carlo method [9] generates K = 10,000 independent simulated price paths for each forecast horizon h = 4 and h = 8 weeks. In the ARIMA-based simulation, paths of the log-price $y_t = \ln P_t$ are generated recursively from the estimated ARIMA(1,1,1) model, $\Delta y_{t+h} = \hat{\varphi}_1\Delta y_{t+h-1} + \varepsilon_{t+h} + \hat{\theta}_1\varepsilon_{t+h-1}$, with $\varepsilon_{t+h} \sim N(0, \hat{\sigma}^2)$, starting from the last observed log-return and residual. In the GARCH-based simulation, $r_{t+h} = \mu + a_1(r_{t+h-1} - \mu) + \varepsilon_{t+h}$, $\varepsilon_{t+h} = \sigma_{t+h}z_{t+h}$, and the conditional variance is updated recursively by (4). The innovations are $z_{t+h} = t_{t+h}\cdot\sqrt{((\nu - 2)/\nu)}$, where t_{t+h} follows a Student-t distribution with the estimated ν, so that $Var(z_{t+h}) = 1$. Simulated

prices are obtained as $P_{t+h} = P_t \cdot \exp(r_{t+1} + \ldots + r_{t+h})$. The 95% and 80% confidence intervals are computed as empirical percentiles of the 10,000 simulated terminal prices.

3 Statistical Analysis of the Price Series

3.1 Descriptive statistics

Table 1 reports the main descriptive statistics for the weekly log-return series r_t over the full sample 1980–2026.

Table 1

Descriptive statistics for weekly gold log-returns, 1980–2026

Statistic	Value
Number of observations (N)	2,411
Mean (μ)	0.0008 per week
Standard deviation (σ)	2.43%
Minimum	−13.24%
Maximum	13.82%
Skewness	−0.04
Kurtosis	6.34
Jarque–Bera (p-value)	p < 0.001

The mean weekly log-return is close to zero ($\mu \approx 0.0008$), corresponding to approximately 4.2% per annum. The kurtosis of 6.34 substantially exceeds the theoretical value of 3 for the normal distribution, confirming leptokurtosis. The Jarque-Bera test rejects normality at any conventional level ($p < 0.001$). The time-series plot reveals a pronounced stochastic upward trend with increasing volatility and clear volatility clustering across three major episodes: 1979–1980, 2007–2012, and 2019–2026.

3.2 Stationarity tests

Table 2 summarises the ADF test results.

Table 2

Augmented Dickey–Fuller test results

Series	ADF statistic	p-value	Conclusion
Gold price, level	4.19	0.99	Non-stationary, I(1)
Log-returns (d = 1)	−13.95	0.01	Stationary, I(0)

The price series is I(1), requiring d = 1 in the ARIMA specification. After first differencing, the log-return series is stationary ($p = 0.01$).

3.3 ARCH effects and autocorrelation structure

The Ljung-Box statistic on squared log-returns is strongly significant ($\chi^2 = 656$; $p < 0.001$), and the ARCH test with 12 lags yields $\chi^2 = 191.92$ ($p < 0.001$). These results confirm conditional heteroscedasticity and motivate GARCH modelling. The ACF and PACF show significant autocorrelation at lags 1 and 2, pointing towards low-order AR and MA components.

4 Econometric Models

4.1 ARIMA model identification and estimation

Following the Box-Jenkins procedure [2], we compared several candidate ARIMA specifications on the log-transformed price series. Table 3 reports the key diagnostic statistics.

Table 3

Comparison of ARIMA specifications

Model	AIC	BIC	Ljung-Box p (residuals)	Ljung-Box p (residuals^2)	Assessment
ARIMA(1,1,0)	−11,081.73	−11,070.16	0.7163	< 0.001	Good
ARIMA(1,1,1)	−11,080.99	−11,063.63	0.8767	< 0.001	Selected
ARIMA(1,2,1)	−11,074.05	−11,056.69	0.6694	< 0.001	Adequate

ARIMA(1,1,0) has the lowest AIC and BIC. However, the AIC difference with ARIMA(1,1,1) is only 0.74, so the two specifications are statistically close; ARIMA(1,1,1) was selected because its residuals show the weakest autocorrelation (Ljung–Box p = 0.8767 vs 0.7163). The estimated parameters are: AR coefficient $\varphi_1 = 0.7953$ (SE = 0.152; $p < 0.001$) and MA coefficient $\theta_1 = -0.8098$ (SE = 0.147; $p < 0.001$). The two coefficients nearly cancel, so the model behaves close to a random walk. The Ljung–Box test on residuals gives p = 0.8767, indicating no residual autocorrelation. The Ljung–Box test on squared residuals yields $p < 0.001$, confirming unmodelled ARCH effects.

Remark 1. *The AIC values of ARIMA and GARCH are not directly comparable, since ARIMA is estimated on $\ln(P_t)$ while GARCH is estimated on r_t. The two models serve complementary purposes: ARIMA describes the conditional mean, GARCH describes the conditional variance.*

4.2 GARCH(1,1) estimation

Table 4 reports the maximum likelihood estimates of the AR(1)-GARCH(1,1) model with Student-t innovations.

Table 4

AR(1)-GARCH(1,1) with Student-t innovations - estimated parameters

Parameter	Estimate	Std. Error	p-value
μ (mean)	0.000436	0.000373	0.243
a_1 (AR(1))	−0.0169	0.0208	0.414
ω (constant)	0.000008	0.000004	0.063
α_1 (ARCH effect)	0.0879	0.0140	< 0.001
β_1 (GARCH effect)	0.9037	0.0143	< 0.001
ν (degrees of freedom)	6.20	0.80	< 0.001

The parameters α_1, β_1 and ν are significant at the 0.1% level, ω is significant at the 10% level ($p = 0.063$), and the mean parameters μ and a_1 are not significant. The volatility persistence $\alpha_1 + \beta_1 = 0.9916$ is very close to unity. The unconditional long-run weekly standard deviation is $\sqrt{(0.000008/0.0084)} \approx 3.09\%$; it exceeds the sample value (2.43%) because, with $\alpha_1 + \beta_1$ close to one, the estimate $\omega/(1 - \alpha_1 - \beta_1)$ is highly sensitive to the parameter estimates. The estimated degrees of freedom $\nu = 6.20$ formally confirm leptokurtosis. The Ljung–Box test on standardised residuals gives p = 0.8932.

Proposition 1. *For a GARCH(1,1) process with $0 < \alpha_1 + \beta_1 < 1$, the h-step-ahead forecast of the conditional variance satisfies $\sigma^2_{t+h}|_t - \sigma^2 = (\alpha_1 + \beta_1)^{h-1}(\sigma^2_{t+1}|_t - \sigma^2)$, where $\sigma^2 = \omega/(1 - \alpha_1 - \beta_1)$. Hence the forecast converges geometrically to the unconditional variance at rate $(\alpha_1 + \beta_1)^{h-1} = 0.9916^{h-1}$. The statement concerns the forecast path, not the realised conditional variances.*

5 Monte Carlo Forecasting

Using the estimated ARIMA(1,1,1) and GARCH(1,1)–Student-t models, we generated K = 10,000 price trajectories for forecast horizons h = 4 and h = 8 weeks from the base date 12 April 2026 (P_0 = 4,770.32 USD/oz). Table 5 presents the summary statistics of the simulated terminal price distributions.

Table 5

Monte Carlo forecast results - ARIMA(1,1,1) vs GARCH(1,1), K = 10,000 simulations (P_0 = 4,770.32 USD)

Statistic	ARIMA 4 wk.	ARIMA 8 wk.	GARCH 4 wk.	GARCH 8 wk.
Mean forecast (USD)	4,775.75	4,777.57	4,795.06	4,815.47
Median (USD)	4,772.14	4,767.74	4,783.18	4,791.57
95% lower bound (USD)	4,345.73	4,184.47	4,105.74	3,848.67
95% upper bound (USD)	5,242.99	5,425.44	5,550.84	5,941.58
Width of 95% CI (USD)	897.26	1,240.97	1,445.10	2,092.91
P(price > P_0)	50.3%	49.7%	51.7%	51.8%

The GARCH-based 95% interval at the 4-week horizon (width: 1,445 USD) is about 1.6 times as wide as the ARIMA-based interval (width: 897 USD). At the 8-week horizon the GARCH interval (width: 2,093 USD) is about 1.7 times as wide as the ARIMA interval (width: 1,241 USD). The probability of the price exceeding P_0 is close to 50% under both models, reflecting the approximately random-walk behaviour of the series.

6 Forecast Validation and Accuracy Analysis

Forecast accuracy was evaluated through out-of-sample validation against seven realised weekly gold prices for the period 19 April - 31 May 2026 (Table 6). These observations were withheld from the training sample and serve as independent test data. Model parameters were estimated once on data up to 12 April 2026 and kept fixed; all seven forecasts (horizons h = 1, …, 7 weeks) were generated from this single origin, without re-estimation. Table 7 presents the error metrics.

Table 6

Out-of-sample validation - point forecast vs realised price (19 Apr - 31 May 2026)

Date	Realised price	ARIMA forecast	ARIMA 95% interval	GARCH forecast	GARCH 95% interval
19 Apr 2026	4,724.40	4,770.74	[4,550.33; 4,997.41]	4,776.17	[4,415.88; 5,154.99]
26 Apr 2026	4,644.50	4,771.78	[4,463.58; 5,095.91]	4,783.70	[4,282.61; 5,325.62]
03 May 2026	4,730.70	4,773.61	[4,401.35; 5,170.20]	4,786.82	[4,167.31; 5,452.27]

10 May 2026	4,561.90	4,773.58	[4,351.13; 5,236.25]	4,791.72	[4,090.15; 5,570.37]
17 May 2026	4,523.20	4,775.37	[4,306.42; 5,283.28]	4,800.08	[4,021.11; 5,678.38]
24 May 2026	4,593.00	4,774.41	[4,256.11; 5,328.33]	4,803.44	[3,966.11; 5,786.86]
31 May 2026	4,466.59	4,776.02	[4,222.80; 5,383.17]	4,809.49	[3,910.91; 5,860.27]

All seven realised prices fell within the 95% intervals of both models. This is consistent with adequate interval forecasts, but seven observations are too few to test the nominal 95% coverage: even with a true coverage of 80%, the probability of seven hits would be $0.8^7 \approx 0.21$. However, a consistent pattern of over-prediction is evident: all forecast errors are negative - both models predicted prices above the subsequently realised values in every week. This systematic upward bias reflects a genuine downward trend in gold prices over the validation period (from 4,770 USD to 4,467 USD), which neither model anticipated. ARIMA(1,1,1) is estimated without drift, so its point forecast stays close to the last observed price; neither model was able to capture the directional correction that materialised during this specific seven-week window.

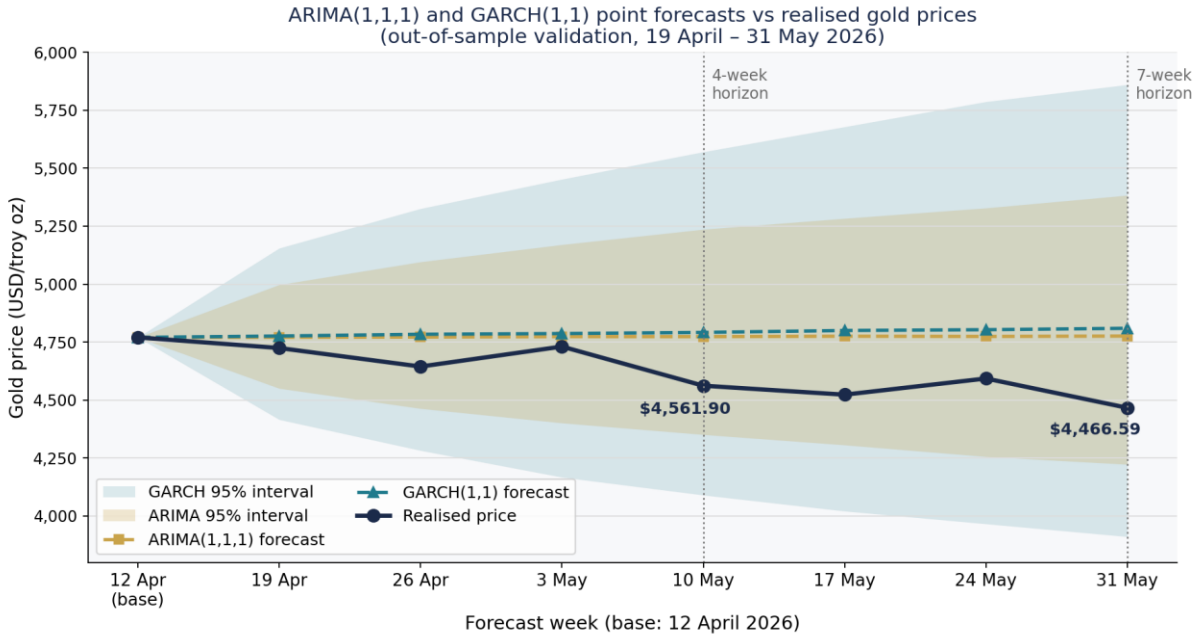


Figure 1

Figure 1. ARIMA(1,1,1) and GARCH(1,1) point forecasts vs realised gold prices, 19 April - 31 May 2026. Shaded bands represent 95% Monte Carlo confidence intervals. Dotted vertical lines mark the 4-week and 7-week forecast horizons.

Figure 1 illustrates the divergence between the model forecasts and the realised price trajectory. The ARIMA and GARCH point forecast lines remain close to each other and approximately flat, anchored near the base price of 4,770 USD/oz, while the realised price drifts

steadily downward. The width difference between the two confidence bands is clearly visible: the gold-shaded ARIMA 95% band is substantially narrower than the teal-shaded GARCH band at every horizon. Notably, even the ARIMA confidence interval contains all realised observations, but it provides a less conservative and therefore less informative risk assessment than the GARCH interval. At the 4-week horizon the realised price of 4,561.90 USD lies 212 USD below the ARIMA point forecast yet comfortably within the confidence interval; by week 7, the realised price of 4,466.59 USD is 309 USD below the ARIMA forecast and 343 USD below the GARCH forecast. The figure therefore provides visual confirmation of the key finding: the point forecasts are systematically upward-biased during a corrective episode, while the interval forecasts correctly capture the range of plausible outcomes.

Table 7

Forecast error analysis by week - ARIMA(1,1,1) vs GARCH(1,1)

Week	Date	Realised	ARIMA error	ARIMA %	GARCH error	GARCH %
1	19 Apr	4,724.40	−46.34	−1.0%	−51.77	−1.1%
2	26 Apr	4,644.50	−127.28	−2.7%	−139.20	−3.0%
3	3 May	4,730.70	−42.91	−0.9%	−56.12	−1.2%
4	10 May	4,561.90	−211.68	−4.6%	−229.82	−5.0%
5	17 May	4,523.20	−252.17	−5.6%	−276.88	−6.1%
6	24 May	4,593.00	−181.41	−3.9%	−210.44	−4.6%
7	31 May	4,466.59	−309.43	−6.9%	−342.90	−7.7%
MAE	—	—	167.32 USD	3.7%	186.73 USD	4.1%

The MAE of the ARIMA point forecast is 167.32 USD (3.7%), compared with 186.73 USD (4.1%) for GARCH; the RMSE is 191.70 and 212.65 USD respectively. ARIMA outperforms GARCH as a point predictor over this period because the GARCH specification incorporates a slightly positive estimated mean log-return ($\mu = 0.000436$), leading to systematically higher point forecasts. Both models exhibit monotonically increasing absolute errors across weeks 1, 2, 4, 5, 7, with temporary reductions at weeks 3 and 6 where the realised price briefly recovered. This error pattern is consistent with the known property of ARIMA-GARCH point forecasts: they converge to the unconditional mean at longer horizons and cannot track directional corrections. The validation results thus reinforce the complementary interpretation: ARIMA provides adequate point accuracy for short horizons (weeks 1-3, errors below 3%), while the GARCH confidence bands provide the more conservative and informative risk measure across all seven weeks.

6.1 Second validation, June–September 2026

To extend the test period, both models were re-estimated on data up to 31 May 2026 and compared with 16 new weekly prices (7 June – 20 September 2026), again from a single forecast origin. A naive forecast (price unchanged from 31 May 2026) serves as a benchmark. The ARIMA intervals are analytical, and the GARCH intervals are obtained from 10,000 Monte Carlo paths with standardised Student-t innovations. Table 8 shows that neither model outperforms the naive benchmark in point accuracy, which is consistent with the near-random-walk behaviour of gold prices. The GARCH 95% interval covered all 16 realised prices, whereas

the ARIMA interval covered 14 of 16 (87.5%); coverage of the 80% intervals was below nominal for both models.

Table 8

Out-of-sample accuracy, 7 June – 20 September 2026 (16 weeks)

Model	MAE, USD	RMSE, USD	MAPE, %	80% coverage	95% coverage
ARIMA(1,1,1)	214.72	261.55	5.14	50.0% (8/16)	87.5% (14/16)
AR(1)-GARCH(1,1), Student-t	235.91	277.84	5.64	68.8% (11/16)	100% (16/16)
Naive forecast	208.98	255.73	5.00	-	-

7 Conclusion

This study has developed and empirically validated an integrated ARIMA-GARCH-Monte Carlo framework for short-term forecasting of weekly gold prices over the period 1980–2026. The working hypothesis is confirmed in full across all four components: the gold price series is I(1) (ADF p = 0.99); conditional heteroscedasticity is strongly confirmed (ARCH $\chi^2 = 191.92$, $p < 0.001$); fat tails are established (kurtosis 6.34; $\nu = 6.20$); and the GARCH intervals are about 1.6 times as wide as ARIMA intervals at 4 weeks.

ARIMA(1,1,1) is selected as the conditional mean specification (AIC = −11,080.99; Ljung–Box p = 0.8767), statistically close to ARIMA(1,1,0). GARCH(1,1) with Student-t innovations captures volatility clustering ($\alpha_1 + \beta_1 = 0.9916$) and leptokurtosis ($\nu = 6.20$). Out-of-sample validation against seven weekly observations shows that all realised prices fall within the 95% intervals of both models, while point forecast errors range from 0.9% to 6.9% for ARIMA and from 1.1% to 7.7% for GARCH (MAE = 167.32 USD, 3.7%, for ARIMA). A second validation on 16 weeks (June–September 2026) shows that neither model outperforms a naive forecast in point accuracy, whereas the GARCH 95% interval covered all realised prices. These results are preliminary and specific to the periods studied.

The scientific contributions of this paper are threefold. First, an integrated ARIMA–GARCH–Monte Carlo framework for gold price analysis is developed and empirically validated in a configuration not encountered in the reviewed Bulgarian academic literature on gold price modelling. Second, a quantitative comparison reveals that GARCH intervals are about 1.6 times as wide as ARIMA intervals at 4 weeks and 1.7 times at 8 weeks. Third, an empirical out-of-sample validation, visualised in Figure 1, provides reproducible evidence of both the forecasting capability and the limitations of the two-component model during directional price corrections.

Directions for future research include: asymmetric GARCH specifications (GJR-GARCH, EGARCH) to capture leverage effects; multivariate frameworks (VAR, DCC-GARCH) for joint modelling with related assets; exogenous macroeconomic regressors (real interest rates, geopolitical risk indices, central bank purchase flows); rolling-window re-estimation over a longer out-of-sample period; and a systematic comparison with machine learning methods (LSTM networks, gradient boosting).

REFERENCES:

[1] Bodi, Z., Kane, A. and Marcus, A. (2014) Investments. 5th edn. Moscow: Williams. [in Russian]

[2] Box, G.E.P. and Jenkins, G.M. (1974) Time Series Analysis: Forecasting and Control. Trans. from English. Moscow: Mir. [in Russian]

[3] Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31(3), pp. 307–327.

[4] Engle, R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica, 50(4), pp. 987–1007.

[5] Investing.com (2026) Gold Spot Price – Historical Data [Online dataset]. London Bullion Market Association via Investing.com. Available at: https://www.investing.com/commodities/gold [Accessed: April 2026].

[6] Yordanova, P. (2015) Time Series Analysis [Unpublished teaching materials]. Faculty of Mathematics and Informatics, Shumen University "Bishop Konstantin Preslavski", Shumen, Bulgaria.

[7] Brockwell, P.J. and Davis, R.A. (2002) Introduction to Time Series and Forecasting. 2nd edn. New York: Springer.

[8] Shiryaev, A.N. (1998) Foundations of Stochastic Financial Mathematics. Vol. 1. Moscow: FAZIS. [in Russian]

[9] Metropolis, N. and Ulam, S. (1949) The Monte Carlo method. Journal of the American Statistical Association, 44(247), pp. 335–341.

[10] Gmurman, V.E. (2003) Probability Theory and Mathematical Statistics. 10th edn. Moscow: Vysshaya Shkola. [in Russian]

[11] Lukasevich, I.Ya. (2010) Financial Management. 2nd edn. Moscow: Eksmo. [in Russian]

[12] MacKinnon, J.G. (1996) Numerical distribution functions for unit root and cointegration tests. Journal of Applied Econometrics, 11(6), pp. 601–618.

Yuliia Yordanova
Master's student in Statistics and Econometrics
Shumen University "Episkop Konstantin Preslavski"
Faculty of Mathematics and Informatics
E-mail: yuliiayordanova2025@gmail.com