Proceedings of CONF-MPCS 2026 Symposium: Theoretic Physics and Plasma Physics
DOI: 10.54254/2753-8818/2026.DL33463

Analysis of Volume-Price Relationship of Gold: A Study Based on MLR and Robust Tests

Kepu Ai

School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, China
kepuai_2005@stu.xjtu.edu.cn

Abstract. The relationship between gold price and its trading volume has long been a key part of financial research. This study investigates the relationship between trading volume and price movements in China's gold futures market, using daily data from the primary contract on the Shanghai Futures Exchange, spanning from 3 January 2020 to 25 February 2026. A multiple linear regression (MLR) framework with White's robust standard errors is constructed to correct for heteroskedasticity. Broader macroeconomic influences are controlled for using the US Dollar Index (DXY). The results, on the one hand, show a strong, positive relationship between volume and price volatility, indicating that greater trading volume is associated with more volatile gold prices. On the other hand, the DXY was shown to have opposite effects, while its increment is associated with reduction gold's volatility. These findings highlight the important role of trading volume in predicting prices, and are useful for investors who want to optimize their risk portfolios and for policymakers who want to stabilize market regulations.

Keywords: gold market, price-volume relationship, volatility, multiple linear regression, US Dollar Index

1. Introduction

The relationship between trading volume and yield is still a basic way to understand market momentum. This relationship between price and volume has long been a key part of financial research because it can be used to predict trends in the stock market [1]. It is important to understand this relationship, as it can be helpful to understand how the market works, see how easily assets can be sold, and set up a basic system for managing risk [2]. A recent study has shown this to be true, especially when it comes to gold. Whether predicting future price movements [3] or measuring the effects of economic uncertainty [4] , volatility remains a key determinant in the gold pricing mechanism. This has become even more unpredictable since the global impact of the Coronavirus pandemic, which has caused big, irregular changes in international gold markets.

When examining the statistical behavior of gold, prior studies offer varying perspectives. In order to analyze the volume-price relationship in the Shanghai spot market, He and Xu [5] employed quantile regression. While providing highly granular details should be encouraged, care must be taken to avoid the risk of overfitting that can result from excessive parameter tuning. Consequently, such strategies frequently prove ineffective during out-of-sample testing. Guo et al. [6] adopted an

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2026.DL33463

alternative approach by employing ARMA (2,2)-GARCH (1,1) and VAR (2) models to examine the relationship between volume and open interest with regard to gold futures volatility. Their research showed that the current trading volume is a good way to explain changes in price. But these older models have one big problem: they focus too much on only three factors, namely volume, price, and open interest. It is interesting to note that the U.S. Dollar Index (DXY) is often not included in calculations of the overall economy. Gold is basically a dollar asset, so this is very important. Research shows that the value of gold is affected by global economic policies and the U.S. dollar [7].

This study examines how often gold futures contracts are traded on the Shanghai Futures Exchange, aiming to investigate the complex relationships among trading volume, the DXY, and the volatility of the gold price. This article also presents a method for using multiple linear regression (MLR) together with robust tests to assess the reliability of a study's results. In particular, the standard errors are estimated using White's standard error [8] to deal with the violation of homoscedasticity, which is a common issue in financial data. Thus, this practical design addresses the uncertainty about the role of trading volume as a leading indicator of price changes, providing insights that are particularly useful for people looking to improve their investments and for regulatory bodies monitoring stability.

2. Materials and methods

2.1. Data acquisition

To collect the amount of traded gold each day, the Shanghai Futures Exchange's gold futures contract (Main Continuous Contract) was selected as a representative. This feature reflects the concentrated volume of domestic trading activity and thus accurately reflects broader market trends in China, including both volume and price.

The daily trading information for the gold futures contract used in this study was obtained from Sina Finance. The data was processed in bulk using the AkShare [9] interface to make sure the data was cleaned in the same way. The U.S. Dollar Index (DXY) is an important measure of the U.S. economy. It was obtained from the Stooq database (https://stooq.com/), a well-known source for international research.

2.2. Data preprocessing

The dataset includes 1,476 pieces of information, covering every day that the Shanghai Futures Exchange was open between 3 January 2020 and 25 February 2026. The U.S. and Chinese markets have different holiday schedules. Because of this, there were a few days of U.S. Dollar Index data missing. To fix this, the missing values were filled in using the data from the previous day. The next step is to clean the data. The next step is to delete the duplicate records so that each row had its own timestamp. Also, any price quotes that were too high or too low were removed using the standard 3-sigma rule. The daily log return $R_t=\ln(P_t/P_{t-1})$, where P_t is the closing price for the current period is then calculated, together with the absolute value $|R_t|$ as a volatility indicator. Using absolute daily returns as a robust, model-free proxy for realized volatility is a well-established practice in modern financial econometrics, as it effectively captures the stylized facts and sudden variation jumps inherent in commodity futures markets [10]. Natural logarithm of trading volume V_t as $\ln(V_t)$ and the U.S. Dollar Index DXY_t as $\ln(DXY_t)$ are employed to reduce potential heteroscedasticity and transform the data into a linear relationship.

DOI: 10.54254/2753-8818/2026.DL33463

2.3. MLR

In MLR, a model is established to characterize a linear relationship between an outcome and a few features. Let $|R_t|$ be the response variable $y=\{y_i\}_{i=1}^n$ with n observations, the term $\ln(V_t)$ is included as an explanatory variable $x_1=\{x_{i1}\}_{i=1}^n$ and $\ln(DXY_t)$ is inputted as a control $x_2=\{x_{i2}\}_{i=1}^n$. The random errors $\varepsilon=\{\varepsilon_i\}_{i=1}^n$ are assumed to be independent and identically distributed, with $E(\varepsilon_i)=0$, $Var(\varepsilon_i)=\sigma^2$. The MLR model is formulated as follows:

$$Y=\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}=X\beta+\varepsilon=\begin{bmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} \end{bmatrix}\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}+\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}, \tag{1}$$

where β is the regression coefficient to be estimated with β_0 being the intercept. Let $\hat{\beta}=\left[\hat{\beta}_0,\ \hat{\beta}_1,\ \hat{\beta}_2\right]^T$ be the ordinary least-squares (OLS) estimate of the parameter β, the fitted value then becomes:

$$\hat{y}_k=\hat{\beta}_0+\hat{\beta}_1 x_{1k}+\hat{\beta}_2 x_{2k} \tag{2}$$

Using the OLS, the estimated coefficient should minimize the sum of squared errors:

$$\hat{\beta}=\underset{\beta}{\operatorname{argmin}}\frac{1}{n}\sum_{i=1}^n(y_i-\hat{y}_i)^2=\underset{\beta}{\operatorname{argmin}}\frac{1}{n}(Y-X\beta)^\top(Y-X\beta), \tag{3}$$

leading to the estimation of the regression coefficient $\hat{\beta}=(X^\top X)^{-1}X^\top Y$.

2.4. Heteroskedasticity-consistent estimator type 1 (HC1) correction

In financial time-series regression models, the residuals often exhibit heteroskedasticity, that is, they vary with the levels of the explanatory variables, which means that the assumption of homoskedasticity underlying the traditional OLS method no longer holds. In particular, although the coefficient estimator remains consistent, its covariance matrix will be biased, thereby leading to spurious significance in the empirical conclusions.

To address this issue, this paper adopts the heteroscedastic robust covariance matrix estimation from the HC1 correction. While White's original estimator laid the theoretical foundation, recent empirical studies on high-frequency financial time series increasingly emphasize the necessity of finite-sample corrections like HC1. Modern literature confirms that such adjustments are indispensable to strictly prevent the over-rejection of the null hypothesis caused by severe volatility clustering and heavy-tailed distributions during global market turbulence [11].

In the presence of heteroscedasticity, $\widehat{\beta}$ is represented by the true covariance matrix:

$$Var(\hat{\beta})=(X^TX)^{-1}(X^T\Omega X)(X^TX)^{-1},\ \ X^T\Omega X=\sum_{i=1}^n\sigma_i^2x_ix_i^T, \tag{4}$$

DOI: 10.54254/2753-8818/2026.DL33463

where $\Omega=E(\varepsilon\varepsilon^T)$ is a heteroscedastic matrix. The true variance of error is approximated by its unbiased estimate from the residuals ($e_i^2=\left(y_i-\hat{y}_i\right)^2$). Substituting this into the above equation yields White's heteroscedastic robust estimator (HC0):

$$\widehat{Var}_{HC0}\left(\hat{\beta}\right)=(X^TX)^{-1}\left(\sum_{i=1}^{n}e_i^2x_ix_i^T\right)(X^TX)^{-1}. \tag{5}$$

In finite sample, to eliminates this downward bias caused, the modified version of HC1 introduces a scaling factor:

$$\widehat{\Omega}_{HC1}=\frac{n}{n-k}\widehat{\Omega}_{HC0}\ , \tag{6}$$

where n is the number of observations, and k is the number of parameters in the regression equation (including the intercept term). Variance estimate for HC1 after substitution is then

$$\widehat{Var}_{HC1}(\hat{\beta})=\frac{n}{n-k}(X^TX)^{-1}\left(\sum_{i=1}^{n}e_i^2x_ix_i^T\right)(X^TX)^{-1}. \tag{7}$$

2.5. Hypothesis test

To verify the relationship between volume and price in the gold futures market, this paper sets up two-tailed hypothesis tests for the volume coefficient β_1 and the US Dollar Index coefficient β_2 respectively:

$$H_0:\beta_j=0\ \text{v.s.}\ H_1:\beta_j\neq0,\quad j=1,2, \tag{8}$$

based on the HC1 heteroscedastic robust standard error derived, the following z-statistic is constructed:

$$z_j=\frac{\hat{\beta}_j}{\sqrt{\widehat{Var}_{HC1}(\hat{\beta})_{jj}}}\sim N(0,1),\quad j=1,2, \tag{9}$$

and the p -values are calculated $p_j=2\times[1-\Phi(|z_j|)]$. To perform robust inference on the parameters, this paper sets the significance levels α at 1%, 5% and 10%, respectively. In the test results, statistical significance is denoted by *** ($P_j<0.01$), ** ($P_j<0.05$) and * ($P_j<0.10$).

Besides, this paper uses F-test to assess the overall explanatory power of the multiple linear regression model. The null hypothesis and alternative hypothesis regarding the overall significance are given as follows:

$$H_1:\beta_1=\beta_2=0\ \text{v.s.}\ H_1:\text{At least one}\ \beta_j\neq0,\ j\in\{1,2\}. \tag{10}$$

The total variation in an observation series can be decomposed into an explainable component and a residual component, namely:

$$SST=SSE+SSR, \tag{11}$$

DOI: 10.54254/2753-8818/2026.DL33463

where total sum of squares ($SST=\sum_{i=1}^{n}(y_i-\bar{y})^2$) measures the total level of variation in the dependent variable $|R_t|$; Sum of Squares Explained ($SSE=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$) is a measure of the portion of volatility explained jointly by trading volume (V_t) and the US Dollar Index (DXY_t); Sum of squared residuals ($SSR=\sum_{i=1}^{n}e_i^2$) is a measure of the random noise that the model cannot explain. Then, the F-statistic is constructed as:

$$F=\frac{SSE/k}{SSR/(n-k-1)}\sim F(k,n-k-1) \tag{12}$$

In this context, $k=2$ represents the number of constraint variables to be tested, and $n=1476$ represents the effective sample size. At a given significance level α (set at 0.01), if the calculated F-statistic satisfies $F>F_\alpha(k,n-k-1)$,then the null hypothesis H_0 is rejected.

2.6. Multicollinearity diagnosis

To ensure that the marginal effects of trading volume and the US Dollar Index on fluctuations in the price of gold are statistically independent within the model, this paper employs the Variance Inflation Factor (VIF) for diagnostic purposes. For the j- th explanatory variable x_j in the model, its VIF is defined as:

$$VIF_j=\frac{1}{1-R_j^2}, \tag{13}$$

where R_j^2 is the coefficient of determination from the auxiliary regression of x_j on all remaining independent variables, calculated as:

$$R_j^2=1-\frac{\sum_{i=1}^{n}(x_{ij}-\hat{x}_{ij})^2}{\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2}. \tag{14}$$

Drawing on the classic empirical guidelines from econometrics [12]: $VIF_j<5$ indicates that no severe multicollinearity and the parameter estimates are robust; $VIF_j>10$ is generally regarded as a warning sign of severe multicollinearity, and consideration should be given to dropping variables or performing variable transformations.

3. Results

3.1. Descriptive statistics

Descriptive statistics are presented for each variable is provide in Table 1. The mean absolute value of gold returns is 0.00087, with a relatively small standard deviation. However, the maximum value reaches 0.13, indicating the presence of extreme volatility within the sample period and exhibiting the typical 'fat-tailed' characteristics of financial time series. The distributions of the independent variables $\ln(V_t)$ and $\ln(DXY_t)$ are relatively normal, with no significant outliers, thus meeting the basic requirements for subsequent linear regression.

DOI: 10.54254/2753-8818/2026.DL33463

Table 1. Descriptive statistics for gold

Variable	Sample size	Mean	Standard deviation	Minimum value	Maximum value
$\lvert R_t \rvert$	1476	0.00087	0.01059	0.00000	0.13158
$\ln(V_t)$	1476	11.9007	0.54164	10.0885	14.1268
$\ln(DXY_t)$	1476	4.12311	0.04747	4.02747	4.23623

As shown in Figure 1, the price of gold followed an overall upward trend with fluctuations between 2020 and 2025. It is important to note that times when prices change a lot, like the liquidity shock in early 2020 and the big rise since 2024, often see a lot more trading. This shows that the market often follows a pattern where price and volume are connected.

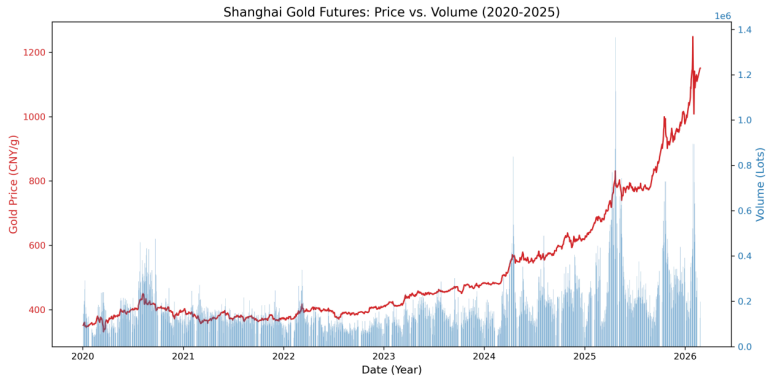


Figure 1. Temporal trend plot of the price of gold

3.2. Regression results and robustness tests

Table 2 shows the results of the regression estimates that have been adjusted using the HC1 robust standard error. The F-statistic is as high as 38.69, which is significant at the 1% level. This strongly rejects the idea that the coefficients of all explanatory variables are zero at the same time. This shows that the regression model built in this paper is statistically valid and can explain things well. The regression results show that the regression coefficient for the trading volume variable $\ln(V_t)$ is 0.0059, and it passed the z-test at the 1% significance level ($z=8.324$, $p<0.001$). This positive coefficient indicates that there is a significant positive correlation between increased trading activity in gold futures and heightened price volatility. The regression coefficient for the US Dollar Index variable $\ln(DXY_t)$ is -0.0112, and it passes the significance test at the 1% level ($z=-3.689$, $p<0.001$). This negative coefficient clearly demonstrates the strong transmission mechanism of the US dollar as the currency in which gold is priced within the international macroeconomic context. The VIF values for the variables $\ln(V_t)$ and $\ln(DXY_t)$ are both 1.0008, which is well below thew empirical threshold of 5.

DOI: 10.54254/2753-8818/2026.DL33463

Table 2. Results for MLR Regression

Variable symbol	Regression coefficient	HC1 error	Z-statistic (HC1)	p-value	VIF
β_0	− 0.0170	0.015	-1.165	0.244	
$\ln(V_t)$	0.0059***	0.001	8.234	<0.001	1.000816
$\ln(DXY_t)$	− 0.0112***	0.003	-3.689	<0.001	1.000816
n	1476				
F-statistic	38.69***			<0.001	

3.3. Residual diagnoses

To further verify whether the assumptions underlying the linear regression model hold, this paper plots a scatter plot of the model's residual distribution (as shown in Figure 2). The horizontal axis of the plot represents the model's fitted values for gold volatility (Fitted Value), whilst the vertical axis represents the standardised residuals.

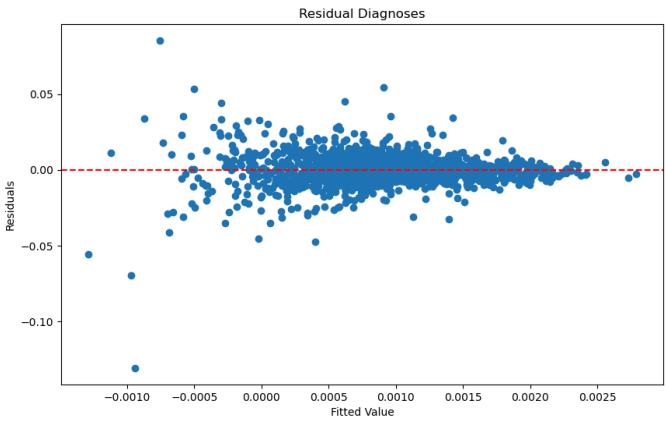


Figure 2. Plot of residuals from gold data

First of all, in terms of the characteristics of the random distribution, the residual data points are generally scattered on either side of the zero line and do not exhibit a clear non-linear trend. Besides, regarding the test for heteroscedasticity, observation of the residual plot reveals that the amplitude of residual fluctuations is not entirely constant; in particular, there are a small number of outliers in the lower-left quadrant, exhibiting a certain degree of clustering in volatility. This phenomenon is extremely common in high-frequency financial time series, confirming that there is indeed significant heteroscedasticity in gold market returns. Under these circumstances, if one were to insist on using homoscedastic standard errors under the traditional assumptions, this would inevitably lead

DOI: 10.54254/2753-8818/2026.DL33463

to the failure of the parameter significance tests. However, this is precisely the core motivation for introducing HC1 in this paper.

4. Discussion

This study employs an adjusted linear regression model with robust error estimates to investigate the relationship between trading volume and the US Dollar Index regarding the gold price volatility, based on daily data collected from the Shanghai Futures Exchange between 2020 and 2025.

The main findings show that trading volume is a leading indicator of market fluctuations. This is likely due to the high trading activity, which represents a growing market divergence and a rapid influx of information, accumulated as trading momentum and translating into wide price fluctuations at the macro level. A similar mechanism was also reported in the microstructure literature, where the trading volume serves as a primary conduit for resolving heterogeneous investor beliefs [13].

The result also implies a countercyclical regulatory effect of the US dollar cycle on gold price volatility. This is also supported by previous evidence and relevant studies, which show that gold's safe-haven properties are not static but are dynamically influenced by global dollar liquidity and geopolitical risk premiums [14]. In particular, during periods of dollar strength, speculative capital withdrawals generally suppress gold's financial attributes, thereby reducing market volatility. In contrast, periods of dollar weakness often trigger a rebound in gold's safe-haven appeal, leading to higher price volatility. The association with the US dollar index also confirms that, despite a degree of domestic independence, the Chinese gold market remains closely tied to the global pricing system.

Collectively, although the volume-price model in this study explains approximately 16% of the variation of volatility in the gold market, the strong statistical significance under the robust error estimator still highlights the key roles of trading volume and the US Dollar Index. These are actually key representations from two primary sources: internal trading activity and external macroeconomic constraints. Future studies are recommended to include more indices in these fields and to employ other flexible statistical tools, such as time-series models, to gain deeper insights.

Based on the aforementioned findings, the rule for risk management strategies is thus suggested to be "volume precedes price". Since trading volume provides a rather reliable representation of market sentiment, in cases of rising volume and stable prices, a sudden, extreme price shift can probably follow. Reducing market exposure and balanced gold portfolios with US dollar assets to hedge against currency-driven volatility is then suggested.

In terms of regulatory frameworks, considering the strong association between trading volume and gold volatility, early-warning systems are encouraged to incorporate volume deviations as a key feature. Additionally, to address a potential validity issue driven by foreign exchange shocks, information disclosure is suggested. In situations of shocks driven by domestic speculation, regulatory measures such as increased margin requirements could be more appropriate. Finally, to enhance the pricing independence of Shanghai Gold and increase China's influence in global gold pricing, more CNY-denominated gold derivatives can be considered to reduce the reliance on the US dollar cycle.

5. Conclusion

This paper investigates the volume-price relationship in China's gold futures market from January 2020 to February 2026, using an MLR model with HC1 robust standard errors. The results confirm that both trading volume and the US Dollar Index are significant positive leading indicators of price

DOI: 10.54254/2753-8818/2026.DL33463

volatility. The findings suggest that risk management and market regulation should transition from a reactive focus on price movements to a monitoring of abnormal volume and external dollar cycles.

References

[1] Blume, L., Easley, D., and O'Hara, M. (1994) Market statistics and technical analysis: The role of volume. The Journal of Finance, 49, 153-181.
[2] Amihud, Y. and Mendelson, H. (1986) Asset pricing and the bid-ask spread. Journal of Financial Economics, 14, 223-249.
[3] Liang, C., Wei, Y., Ma, F., and Li, X. (2022) Forecasting the volatility of China's gold futures: New evidence from model reduction methods. Chinese Journal of Management Science, 30, 30-41.
[4] Yu, J., Yu, C., and Wei, J. (2024) Global political and economic uncertainty and gold price volatility: An empirical study based on the GARCH-MIDAS model. The Journal of Quantitative Economics, 15, 132-152.
[5] He, X. and Xu, Y. (2012) Analysis of price-volume relationship in the gold market: An empirical study based on quantile regression model. Journal of Guangdong University of Finance & Economics, 3, 25-31.
[6] Guo, S., Yuan, T., and He, Z. (2014) Analysis of the volume-price relationship of China's gold futures. Times Finance, 8X, 5.
[7] Bouri, E., Lucey, B., Saeed, T., and Vo, X.V. (2021) The realized volatility of commodity futures: Interconnectedness and determinants. International Review of Economics & Finance, 73, 139-151.
[8] White, H. (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica, 48, 817-838.
[9] King, A. and Zhang, Y. (2022) AKShare [Computer software]. GitHub. (URL removed to match Maganioti style — no URLs)
[10] Fan, J.H. and Zhang, T. (2020) The untold story of commodity futures in China. Journal of Futures Markets, 40, 671-706.
[11] Bouras, C., Christou, C., Gupta, R., and Suleman, T. (2019) Geopolitical risks, returns, and volatility in emerging stock markets: Evidence from a panel GARCH model. Emerging Markets Finance and Trade, 55, 1841-1856.
[12] Hair, J.F., Black, W.C., Babin, B.J., and Anderson, R.E. (2010) Multivariate data analysis (7th ed.). Pearson.
[13] Giglio, S., Maggiori, M., Stroebel, J., and Utkus, S. (2021) Five facts about beliefs and portfolios. American Economic Review, 111, 1481-1522.
[14] Mensi, W., Aslan, A., Vo, X.V., and Kang, S.H. (2023) Time-frequency spillovers and connectedness between precious metals, oil futures and financial markets: Hedge and safe haven implications. International Review of Economics & Finance, 83, 219-232.