

Analyzing factors influencing global precious metal markets: A feature selection study

Stephanie Stephanie^{1,*}, Dhanuskodi Rengasamy², Jobrun Nandong³, Andrew Brennan⁴, and Filbert H. Juwono⁵

¹Dept. of Economics and Finance, Curtin University Malaysia, Miri 98009, Sarawak, Malaysia

²Dept. of Acct., Finance, and Economics, Curtin University Malaysia, Miri 98009, Sarawak, Malaysia

³Dept. of Chemical Engineering, Curtin University Malaysia, Miri 98009, Sarawak, Malaysia

⁴Dept. of Acct., Finance, and Economics, Curtin University, Perth 6102, Western Australia, Australia

⁵Computer Science Program, University of Southampton Malaysia, Iskandar Puteri 79100, Johor, Malaysia

Abstract. Precious metals are valuable commodities providing superior protection against risky financial exposure. Identifying factors influencing the market is crucial for anticipating changes. Forecast applications utilize stochastic models capable of learning from historical data to project future values. The dataset is a vital component for prediction tools since all estimations begin with constructing the appropriate information. Detecting the association between input and output is essential to filter data, as including unrelated variables could destabilize the response. Feature selection considers removing uncorrelated attributes before incorporating them as inputs to the predictor. This study employs three regression-based algorithms to examine 58 precious assets from gold, silver, platinum, and palladium markets against several variables cited in the literature. Relationships were detected using regressive feature selection methods, known as least absolute shrinkage and selection operator (LASSO), ridge, and elastic net (EN). Results demonstrate that the proposed algorithms achieved satisfactory performance on 42 assets, justified through a reliable fit and acceptable error. The remaining 16 assets exhibited large deviations with considerably poor regression quality, indicating considerable nonlinearity. Attributes were selected with a detailed emphasis on those exerting the most substantial impact on a particular metal. Based on computational analysis, most investments are susceptible to macroeconomic factors. Some assets may present hedging capabilities towards key features, including stock index, exchange rates, and bond yield. An assessment of common variables among each metal revealed that real GDP growth and interest rates are vital indicators for the precious metal market. Overall, the simulation outcomes show no consistent commonalities amongst attributes within the same asset class in a country. Feature selection from this research offers necessary information regarding time-series dynamics, serving as a basis to project trends. The filtered dataset is expected to enhance the reliability of nonlinear predictive algorithms by removing inaccurate correlations to lower computational load. Furthermore, the outcome provides information regarding correlations affecting global precious metal investments over five-year period. These discussions are necessary for investors considering such commodities as potential portfolio diversifiers.

*Corresponding author: stephanie123@postgrad.curtin.edu.my

1 Introduction

Gold, silver, platinum, and palladium are precious metals with high investment value. The physical properties, such as corrosion resistance and conductivity, are ideal for a wide range of industrial applications. Their appearance and rarity create massive demand for numerous manufacturing products, and high market liquidity entices investors internationally [1, 2]. Precious metals are safe havens during economic downturns, making them preferable protective assets. Haven behavior appears briefly, requiring a critical observation for maximizing investment returns using these commodities [3]. Projecting future trends is vital in financial analysis, delivering valuable information regarding market behavior. Forecasting prices based on foreseeable trends provides the foundation for developing diversification strategies [4].

Achieving accurate estimates is relevant to generate positive yields, which begins with the construction of a dataset that has a profound influence on determining the accuracy. Financial time series contain large volumes of data, requiring complex repeated iterations to compute the changes. Preparing input data is a fundamental aspect of predictive modeling that conceivably impacts its performance. When uncorrelated variables are included in an algorithm, it may falsely map relationships, leading to inaccurate predictions. Consequently, input datasets should be filtered to extract relevant features upon conducting forecast analysis. Least absolute shrinkage and selection operator (LASSO) and ridge are two commonly employed regression models providing effective feature selection among statistical data explored in the literature. Several applications, including energy management, soft sensors, and assessing financial credit risk, have demonstrated the practicability of these models for improving the precision of respective projections [5–7]. Unfortunately, they suffer from performance limitations attributed to particular data types, restricting their applications to varying nonlinearities. LASSO and ridge can be integrated using an elastic net (EN) algorithm with a tuning parameter to optimize the estimates [8]. The broad applicability of EN presents suitable characteristics for observing numerous assets susceptible to external influences [9].

This article presents a detailed comparison between the three feature selection techniques (LASSO, ridge, and EN) to filter the factors affecting precious metal market prices. A set of 58 assets from global exchanges have been selected for analysis over five years between July 2016 - June 2021. Best-performing configurations for each output were examined to provide essential information regarding these markets. Several studies have attempted to predict precious metal prices, while none have focused on an analytical feature selection for data filtering. Therefore, this study contributes to investigating variables influencing precious metal assets, which hold high regard for investors worldwide.

2 Market Research and Data Preparation

Constructing a dataset for feature selection involves a comprehensive analysis of the market. Commodities related to precious metals are traded globally through financial markets where futures, equities, and exchange-traded funds (ETFs) can be purchased [10]. Within these markets, we appoint 58 assets from 12 countries for the current investigation based on the

criteria achievement in 2021, which includes holding the highest market capitalization or production. The assets are traded on the country's primary commodity market or major national stock exchanges. The daily historical data collected between July 2016 to September 2021 were obtained from trading websites ¹. The financial data selected includes open, high, low, close, and volume to predict the adjusted close price as the output. Describing the input features requires a review of the literature to identify variables influencing market prices.

Correlations with national stock indices, exchange rates, and oil prices (Brent, WTI, and OPEC) have been reviewed within the literature, representing possible relationships [2, 11]. The financial markets possess strong connections to the country's economy and often observe shocks in response to an economic downturn [12]. Hence, the influence of macro and microeconomic variables such as supply-demand, real gross domestic product (GDP), inflation, interest rates, and unemployment are essential factors to consider in feature selection. The industry generates considerable demand for precious metals, consequently impacting their value significantly. Gold has extensive applications in jewelry, and silver shows high demand in the electronics industry. Platinum and palladium are widely used in manufacturing automobile catalytic converters [13]. Famous precious metals product manufacturers such as Chow Tai Fook jewelry group (gold), Foxconn technology (silver), Faurecia (platinum), and BASF (palladium) are included in the dataset to observe their correlation with respective metals along with supply and demand. Each asset is evaluated according to the country's leading stock index, total reserves (foreign exchange (FX) and gold), real GDP growth rate, inflation, interest rate, unemployment rate, 10-year bond yield, and the exchange rate with US Dollar except for USA exchange rates against the Chinese yuan. Equity assets, namely NEM, AMSJ, SBSW, and PTM, are mining corporations that produce multiple metals, indicating that additional factors potentially influence their values. Gold prices are additionally affected by the national gold reserves, which are included in the current analysis. The final dataset associates 17 inputs for gold and 16 for white metals (silver, platinum, and palladium).

The asset prices are converted to US Dollars, and the precious metal commodity futures are measured in US Dollars per troy ounce. Historical data for inputs and outputs utilized within this study are available over varying frequencies (daily, monthly, quarterly, or annually). Moreover, international markets have different operating days resulting in missing values that statistical regression models cannot process. In most cases, missing information is discarded, and the function is defined to begin from the next available value. This analysis provides a rigid prediction model with the possibility of producing a significant error when applied to high-frequency data. Overcoming such difficulties requires imputing missing values using statistical tools such as quadratic extrapolation. Modified akima piecewise cubic hermite interpolation (MAKIMA) is a tool utilizing third-order polynomials to compute data points. MAKIMA controls overshoots in estimations by balancing the undulations and rigidity, thus providing conservative estimates [14]. Existing literature has used this method to treat numerous applications with nonlinear missing data, confirming that MAKIMA is an effective imputation strategy [15–17]. Several input variables within the defined dataset fluctuate at distinct scales. For instance, values of GDP usually oscillate in millions of dollars, whereas changes in interest rates are relatively low. High estimation errors could arise from a wide disparity in magnitude, which might weaken the correlations developed within regression. Before conducting regression analysis, all inputs are regularised within a defined range to prevent the loss of correlations from small-scale data. In addition to

¹<https://www.finance.yahoo.com>, <https://www.investing.com>

preserving low correlation values, this technique can avoid unstable behavior triggered by significant variances. Min-max is a standard regularization method to scale data within the range [-1,1], thus improving regression efficiency. This study performs MAKIMA interpolation and min-max scaling for feature selection dataset preparation using MATLAB R2020a software.

3 Feature Reduction Methods

Regression is a widespread statistical approach to approximate the numerical values describing any input's influence over the output. The calculated values can be used to describe a function that varies with time while minimizing the forecast error. Ordinary least squares (OLS) regression computes an error function (LOLS) on an estimation of output (y), described through n inputs over the entire dataset as shown in *equation 1* [18].

$$L_{OLS} = \sum_{i=1}^n (y_i - \beta_{x_i})^2 \quad (1)$$

Where β is the set of correlations representing the input's impact on the output, variable selection encompasses the removal of specific inputs (x_i) based on low values of β .

3.1 LASSO

LASSO is a popular feature selection method that modifies the loss function with an additional L1 penalty, represented by the second term on right-hand side of *equation 2* [19].

$$L_{LASSO} = \sum_{i=1}^n (y_i - \beta_{x_i})^2 + \lambda \sum_{j=1}^m |\beta_j| \quad (2)$$

Here, λ is the shrinkage coefficient, defined to reduce the loss function using correlations. The coefficient reduction provides the selection basis by removing the features with β approaching zero. Large values can lead to bias overestimation resulting in an insufficient model fit. Data representing high variance can produce errors in LASSO since it utilizes simple linear regression modeling. The presence of relationships between inputs (multicollinearity) might further diminish regressive capacities and generate substantial estimation inaccuracies [20].

3.2 Ridge

Ridge is a regression model incorporating an L2 penalty term added to the total loss function. This term includes the squared sum of coefficients described in *equation 3*.

$$L_{Ridge} = \sum_{i=1}^n (y_i - \beta_{x_i})^2 + \lambda \sum_{j=1}^m \beta_j^2 \quad (3)$$

The squared coefficient values lower the penalty on overall loss functions, avoiding overestimation of bias. Moreover, ridge analyzes each coefficient separately, which enables the utilization of data exhibiting multicollinearity. Although it amplifies the regression performance (quality of fit), this method might underestimate the penalty, ultimately reducing

selection efficiency. With small values of λ , the majority of the coefficients do not achieve a 0 value, leading to a larger dataset [21]. This limitation leads to the underperformance of ridge regression in analyzing numerous features.

3.3 Elastic net

Models described in sections 3.1 and 3.2 suffer constraints that hinder their use across extensive data ranges. The financial dataset contains several variables depicting varying levels of nonlinearities, which can be challenging to analyze using LASSO or ridge. The EN algorithm can address these shortcomings by employing a trade-off between LASSO and ridge estimations. The ratios of both regression estimates can be adjusted using the tuning parameter (α) as described in equation 4 [22].

$$L_{EN} = \sum_{i=1}^n (y_i - \beta_{x_i})^2 + \alpha\lambda \sum_{j=1}^m |\beta_j| + (1 - \alpha)\lambda \sum_{j=1}^m \beta_j^2 \tag{4}$$

The tuning parameter can be effectively adjusted ($0 < \alpha < 1$) according to the data type, making EN applicable in various applications. Several studies in the literature have successfully implemented this algorithm for high-accuracy variable selection [23, 24]. The current dataset can be evaluated using several EN configurations to determine optimum α values. The tuning parameter value can provide necessary information regarding the type of predictive model which could be employed to forecast the output data. Performances of regression models can be tracked through the mean squared error (MSE) described in equation 5 [24].

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \beta_{x_i})^2 \tag{5}$$

The regression coefficient (R^2) is another accuracy measurement, describing the quality of fit over the data. The values close to 1 represent an accurate model and are often included as a performance measure along with errors [25]. Several works of literature utilize lasso, ridge, and elastic net to identify factors influencing widespread malaria, glioma grading, and detecting egg quality [23, 26, 27]. Existing studies indicate best-performing model varies based on its application, suggesting the existence of multicollinearity in data. When high multicollinearity is present, ridge may outperform LASSO and EN. However, LASSO shows a superior capability to select features, while EN improves estimations over datasets with unknown variance [21].

4 Results and Discussions

This research conducts feature selection analysis for the selected 58 precious metal assets based on the detailed review by performing repeated simulations in a one-factor-at-a-time (OFAT) approach. The EN algorithm is tested at varying configurations of α under a small step size (0.1), where α values of 0 and 1 indicate ridge and LASSO, respectively. Configuration results from ridge, LASSO, and EN for each asset are described in table 1, 3, 5, and 7. The selected features based on the best-performing algorithm are enlisted in table 2, 4, 6, and 8.

Table 1. Feature selection performance result (Gold)

Asset name	Ridge ($\alpha = 0$)		Elastic net ($\alpha = 0.1 - 0.9$)			LASSO ($\alpha = 1$)	
	MSE	R ²	α	MSE	R ²	MSE	R ²
Gold Futures							
[USA] Gold Futures (GC=F)	0.00801	0.96053	0.5	0.00096	0.97462	0.00095	0.97032
[CHINA] Gold Futures (SHAUc1)	0.01080	0.96462	0.5	0.00383	0.97659	0.00382	0.97481
[INDIA] Gold Futures (MGPLc1)	0.00730	0.96359	0.1	0.00230	0.99159	0.00229	0.99147
Gold ETF							
[USA] SPDR Gold Shares (GLD) - [Commodities]	0.00740	0.96805	0.5	0.00096	0.98586	0.00095	0.98232
[USA] iShares Gold Trust (IAU) - [Commodities]	-	(Negative)	-	-	(Negative)	-	(Negative)
[INDIA] Nippon India ETF Hang Seng BeES (HNGS) – [Equity]	-	(Negative)	-	-	(Negative)	-	(Negative)
[CHINA] Huanan Gold ETF Fund (518880) - [Commodities]	0.00998	0.96689	0.3	0.00268	0.97489	0.00268	0.97265
Gold Equity							
[USA] Newmont Corporation (NEM)	0.00930	0.95907	0.7	0.00387	0.98574	0.00375	0.98571
[USA] Barrick Gold (GOLD)	0.01455	0.88419	0.8	0.00398	0.83494	0.00403	0.83537
[RUSSIA] Polyus PAO (PLZL)	0.01178	0.91025	0.6	0.00512	0.97166	0.00516	0.97167
[USA] AngloGold Ashanti Ltd. (AU)	-	(<0.7)	-	-	(<0.7)	-	(<0.7)
[USA] Gold Fields Limited (GFI)	0.00936	0.84264	0.8	0.00447	0.94012	0.00453	0.94023
[USA] Kinross Gold Corporation (KGC)	0.00995	0.94183	0.1	0.00444	0.97902	0.00445	0.97891
[AUSTRALIA] Newcrest Mining Limited (NCM)	0.04100	0.74424	0.1	0.02849	0.83756	0.02846	0.83136
[CHINA] Zijin Mining Group Co Ltd Class A (601899)	-	(Negative)	-	-	(Negative)	-	(Negative)
[CANADA] Endeavour Mining Corp (EDV)	0.02923	0.72088	-	-	(<0.7)	-	(<0.7)
[AUSTRALIA] Northern Star Resources Ltd (NST)	0.03025	0.86464	-	-	(<0.7)	-	(<0.7)

Table 2. Features selected for gold

Best α	Features selected*	Asset name
Gold Futures		
0.5	1,9,12,14,16,17,18,19,20	[USA] Gold Futures (GC=F)
0.5	5,9,12,14,16,18,19	[CHINA] Gold Futures (SHAUc1)
0.1	1,2,5,6,8,9,10,11,12,13,14,15,16,17,18,19,20	[INDIA] Gold Futures (MGPLc1)
Gold ETF		
0.5	1,2,6,8,9,12,14,15,16,17,18,19,20	[USA] SPDR Gold Shares (GLD) - [Commodities]
-	-	[USA] iShares Gold Trust (IAU) - [Commodities]
-	-	[INDIA] Nippon India ETF Hang Seng BeES (HNGS) – [Equity]
0.3	2,5,6,8,9,10,12,14,16,17,18,19	[CHINA] Huanan Gold ETF Fund (518880) - [Commodities]
Gold Equity		
0.7	1,2,3,4,6,7,8,10,12,15,16,17,18,19,20	[USA] Newmont Corporation (NEM)
0.8	1,2,6,8,9,10,11,12,13,14,15,16,17,18,19,20	[USA] Barrick Gold (GOLD)
1	1,2,5,6,8,9,10,11,12,13,14,15,16,17,18,19,20	[RUSSIA] Polyus PAO (PLZL)
-	-	[USA] AngloGold Ashanti Ltd. (AU)
1	2,6,8,9,10,11,12,14,15,16,17,18,19,20	[USA] Gold Fields Limited (GFI)
0.1	1,2,6,8,9,10,11,12,13,14,15,16,17,18,19,20	[USA] Kinross Gold Corporation (KGC)
0.1	1,2,6,8,9,10,11,12,14,15,16,17,18,19,20	[AUSTRALIA] Newcrest Mining Limited (NCM)
-	-	[CHINA] Zijin Mining Group Co Ltd Class A (601899)
0	1,2,5,6,8,9,10,11,12,13,14,15,16,17,18,19,20	[CANADA] Endeavour Mining Corp (EDV)
0	1,2,6,8,9,10,11,12,13,14,15,16,17,18,19,20	[AUSTRALIA] Northern Star Resources Ltd (NST)

*1: Supply, 2: Demand, 3: Supply (Silver), 4: Demand (Silver), 5: Gold Reserve, 6: Market, 7: Market (Silver), 8: Dow Jones Precious metal Index, 9: Dow Jones Precious metal commodity Index, 10: Stock Index, 11: OPEC, 12: WTI, 13: Brent, 14: Total Reserves (FX and gold), 15: Exchange Rate, 16: Real GDP Growth, 17: Inflation, 18: Interest rate, 19: Unemployment rate, 20: 10y Bond yield

Table 3. Feature selection performance result (Silver)

Asset name	Ridge ($\alpha = 0$)		Elastic net ($\alpha = 0.1 - 0.9$)		LASSO ($\alpha = 1$)	
	MSE	R ²	MSE	R ²	MSE	R ²
Silver Futures						
[USA] Silver Futures (SI=F)	0.01472	0.79911	0.1	0.00521	0.96645	0.00498
[CHINA] Silver Futures (SAGc1)	0.01839	0.78906	0.2	0.00765	0.92952	0.00757
[INDIA] Silver Futures (MSVc1)	0.02113	0.82861	0.6	0.00645	0.96998	0.00663
Silver ETF						
[USA] iShares Silver Trust (SLV) - [Commodities]	0.01631	0.80959	0.1	0.00605	0.96028	0.00603
[USA] Global X Silver Miners ETF (SIL) - [Equity]	0.02408	0.88818	0.1	0.00809	0.93868	0.00782
[USA] Aberdeen Std. Physical Silver Shares ETF (SIVR) - [Commodities]	0.01619	0.80556	0.1	0.00616	0.96693	0.00592
[USA] ETFMG Prime Junior Silver Miners ETF (SILJ) - [Equity]	0.02797	0.83820	0.5	0.01191	0.90887	0.01198
[USA] iShares MSCI Global Silver and Metals Miners ETF (SLVP) - [Equity]	0.01666	0.90640	0.1	0.00646	0.96903	0.00659
[AUSTRALIA] ETFs Physical Silver ETC (ETPMAG) - [Commodities]	-	(<0.7)	0.6	0.00692	0.95740	0.00731
Silver Equity						
[UK] Fresnillo plc (FRES)	0.05136	0.73518	0.2	0.03657	0.76017	0.03649
[POLAND] KGHM Polska Miedz S.A. (KGH)	0.02334	0.80544	0.1	0.01150	0.91959	0.01149
[UK] Glencore plc. (GLEN)	-	(<0.7)	0.7	0.01113	0.93808	0.01115
[INDIA] Hindustan Zinc Limited (HZNC)	0.03391	0.62682	0.9	0.02314	0.82532	0.02296
[UK] Polymetal International plc (POLYP)	0.01451	0.82623	0.1	0.00664	0.97786	0.00667
[CANADA] Pan American Silver Corp. (PAAS)	0.01916	0.64149	0.9	0.01041	0.92608	0.01047
[USA] Southern Copper Corporation (SCCO)	-	(Negative)	-	-	(<0.7)	-
[MEXICO] Industrias Penoles, S.A.B. de C.V. (PEOLES)	0.03880	0.81099	0.9	0.02451	0.89262	0.02332
[AUSTRALIA] South32 Limited (S32)	0.03092	0.19968	0.9	0.01786	0.90578	0.01803
[JAPAN] Sumitomo Corp. (8053)	-	(<0.7)	0.7	0.01966	0.84518	0.01944
[USA] Compania de Minas Buenaventura S.A.A. (BVN)	0.03943	0.73206	0.3	-	(<0.7)	0.02023
[USA] Hecla Mining Company (HL)	-	(<0.7)	0.8	0.01541	0.80821	0.01535
[USA] First Majestic Silver Corp. (AG)	-	(Negative)	-	-	(Negative)	-
[USA] BHP Group (BHP)	0.01370	0.66606	0.2	0.00714	0.96004	0.00718
[UK] Hochschild Mining plc (HOCM)	-	(<0.7)	-	-	(<0.7)	-
[SWEDEN] Boliden AB (publ) (BOL)	-	(<0.7)	0.4	0.01397	0.79578	0.01399
[USA] Teck Resources Limited (TECK)	0.02393	0.75719	0.9	0.02018	0.86925	0.02024

Table 4. Features selected for silver

Best α	Features selected*	Asset name
Silver Futures		
0.1	1,2,6,8,9,10,12,13,14,15,16,17,18,19,20	[USA] Silver Futures (SI=F)
0.2	1,2,8,9,10,12,14,15,16,17,18,20	[CHINA] Silver Futures (SAGc1)
0.6	1,6,9,10,12,14,15,16,17,18,19,20	[INDIA] Silver Futures (MSVc1)
Silver ETF		
0.1	1,2,6,8,9,10,12,13,14,16,17,18,19,20	[USA] iShares Silver Trust (SLV) - [Commodities]
0.1	1,2,6,8,9,10,12,13,14,16,17,18,19,20	[USA] Global X Silver Miners ETF (SIL) - [Equity]
0.1	1,2,6,8,9,10,12,14,15,16,17,18,19,20	[USA] Aberdeen Std. Physical Silver Shares ETF (SIVR) - [Commodities]
1	1,2,6,8,9,10,12,13,14,15,16,17,18,19,20	[USA] ETFMG Prime Junior Silver Miners ETF (SILJ) - [Equity]
0.1	1,2,6,8,9,10,11,12,13,14,15,16,17,18,19,20	[USA] iShares MSCI Global Silver and Metals Miners ETF (SLVP) - [Equity]
1	1,2,6,8,9,12,14,15,16,17,18,19,20	[AUSTRALIA] ETFs Physical Silver ETC (ETPMAG) - [Commodities]
Silver Equity		
0.2	1,2,6,8,11,14,15,16,17,18,19	[UK] Fresnillo plc (FRES)
0.1	1,2,6,9,10,11,12,14,15,16,17,18,19,20	[POLAND] KGHM Polska Miedz S.A. (KGH)
0.7	1,2,6,8,9,10,12,13,14,15,16,17,18,19,20	[UK] Glencore plc. (GLEN)
1	1,2,6,8,9,10,11,12,13,14,15,16,17,18,19,20	[INDIA] Hindustan Zinc Limited (HZNC)
0.1	1,2,6,8,9,10,12,13,14,15,16,17,18,19	[UK] Polymetal International plc (POLYP)
0.9	1,2,6,8,9,10,13,14,15,16,17,18,19,20	[CANADA] Pan American Silver Corp. (PAAS)
-	-	[USA] Southern Copper Corporation (SCCO)
0.9	2,6,9,10,11,12,13,14,15,16,17,18,19,20	[MEXICO] Industrias Penoles, S.A.B. de C.V. (PEOLES)
0.9	1,2,6,8,10,11,12,13,14,15,16,17,18,19,20	[AUSTRALIA] South32 Limited (S32)
0.7	1,2,6,8,10,11,13,14,16,17,18,19,20	[JAPAN] Sumitomo Corp. (8053)
1	1,2,6,8,9,10,11,14,15,16,17,18,19,20	[USA] Compania de Minas Buenaventura S.A.A. (BVN)
0.8	1,2,6,8,9,10,11,12,13,14,15,16,17,18,19,20	[USA] Hecla Mining Company (HL)
-	-	[USA] First Majestic Silver Corp. (AG)
0.2	1,2,6,8,9,10,11,12,14,15,16,17,18,19,20	[USA] BHP Group (BHP)
-	-	[UK] Hochschild Mining plc (HOCM)
0.4	1,2,6,8,9,10,11,12,14,15,16,17,18,19,20	[SWEDEN] Boliden AB (publ) (BOL)
1	1,2,6,8,9,10,12,13,15,16,17,18,20	[USA] Teck Resources Limited (TECK)

*1: Supply, 2: Demand, 6: Market, 8: Dow Jones Precious metal Index, 9: Dow Jones Precious metal commodity Index, 10: Stock Index, 11: OPEC, 12: WTI, 13: Brent, 14: Total Reserves (FX and gold), 15: Exchange Rate, 16: Real GDP Growth, 17: Inflation, 18: Interest rate, 19: Unemployment rate, 20: 10y Bond yield

Table 5. Feature selection performance result (Platinum)

Asset name	Ridge ($\alpha = 0$)		Elastic net ($\alpha = 0.1 - 0.9$)			LASSO ($\alpha = 1$)	
	MSE	R^2	α	MSE	R^2	MSE	R^2
Platinum Futures							
[USA] Platinum Futures (PL=F)	-	(Negative)	-	-	(<0.7)	-	(<0.7)
[JAPAN] Platinum Futures (JPLc1)	-	(<0.7)	-	-	(<0.7)	-	(<0.7)
Platinum ETF							
[USA] Aberdeen Std. Physical Platinum Shares ETF (PPLT) – [Commodities]	-	(<0.7)	-	-	(<0.7)	-	(<0.7)
[SOUTH AFRICA] NewGold Platinum ETF (NGPLTJ) – [Commodities]	0.02428	0.73263	0.1	0.01706	0.83589	0.01701	0.83554
Platinum Equity							
[SOUTH AFRICA] Anglo American Platinum Limited (AMSJ)	-	(<0.7)	-	-	(<0.7)	-	(<0.7)
[SOUTH AFRICA] Impala Platinum Holdings Limited (IMPJ)	-	(<0.7)	-	-	(<0.7)	-	(<0.7)
[USA] Sibanye Stillwater Limited (SBSW)	-	(Negative)	-	-	(Negative)	-	(Negative)
[USA] Vale S.A. (VALE)	-	(<0.7)	0.7	0.00805	0.93624	0.00805	0.93029
[CANADA] Platinum Group Metals (PTM)	-	(Negative)	-	-	(Negative)	-	(Negative)

Table 6. Features selected for platinum

Best α	Features selected*	Asset name
Platinum Futures		
-	-	[USA] Platinum Futures (PL=F)
-	-	[JAPAN] Platinum Futures (JPLc1)
Platinum ETF		
-	-	[USA] Aberdeen Std. Physical Platinum Shares ETF (PPLT) – [Commodities]
0.1	1,2,6,8,9,10,11,12,13,14,15,16,17,18,19,20	[SOUTH AFRICA] NewGold Platinum ETF (NGPLTJ) – [Commodities]
Platinum Equity		
-	-	[SOUTH AFRICA] Anglo American Platinum Limited (AMSJ)
-	-	[SOUTH AFRICA] Impala Platinum Holdings Limited (IMPJ)
-	-	[USA] Sibanye Stillwater Limited (SBSW)
0.7	1,2,6,8,9,10,11,12,13,15,16,17,18,19,20	[USA] Vale S.A. (VALE)
-	-	[CANADA] Platinum Group Metals (PTM)

*1: Supply, 2: Demand, 3: Supply (Palladium), 4: Demand (Palladium), 6: Market, 7: Market (Palladium), 8: Dow Jones Precious metal Index, 9: Dow Jones Precious metal commodity Index, 10: Stock Index, 11: OPEC, 12: WTI, 13: Brent, 14: Total Reserves (FX and gold), 15: Exchange Rate, 16: Real GDP Growth, 17: Inflation, 18: Interest rate, 19: Unemployment rate, 20: 10y Bond yield

Table 7. Feature selection performance result (Palladium)

Asset name	Ridge ($\alpha = 0$)		Elastic net ($\alpha = 0.1 - 0.9$)			LASSO ($\alpha = 1$)	
	MSE	R^2	α	MSE	R^2	MSE	R^2
Palladium Futures							
[USA] Palladium Futures (PA=F)	0.02510	0.89900	0.1	0.01262	0.93697	0.01250	0.87932
[JAPAN] Palladium Futures (JPAc1)	0.02725	0.90638	0.1	0.01624	0.91921	0.01454	0.90529
Palladium ETF							
[USA] Aberdeen Std. Physical Palladium Shares ETF (PALL) – [Commodities]	0.02804	0.89668	0.1	0.01453	0.92062	0.01468	0.86716
[SOUTH AFRICA] NewGold Palladium ETF (NGPLDJ) – [Commodities]	0.02705	0.90857	0.6	0.01616	0.94287	0.01550	0.91838
Palladium Equity							
[CANADA] Ivanhoe Mines Ltd. (IVN)	-	(<0.7)	-	-	(<0.7)	-	(<0.7)
[CAD] New Age Metals (NAM)	-	(Negative)	-	-	(<0.7)	-	(<0.7)

Table 8. Features selected for palladium

Best α	Features selected*	Asset name
Palladium Futures		
0.1	1,2,6,8,9,10,11,12,14,15,16,17,18,19	[USA] Palladium Futures (PA=F)
0.1	2,6,9,10,14,15,16,17,18,19,20	[JAPAN] Palladium Futures (JPAc1)
Palladium ETF		
0.1	1,2,6,8,9,10,11,12,14,15,16,17,18,19	[USA] Aberdeen Std. Physical Palladium Shares ETF (PALL) – [Commodities]
0.6	1,2,6,8,9,10,11,12,14,16,17,18,19,20	[SOUTH AFRICA] NewGold Palladium ETF (NGPLDJ) – [Commodities]
Palladium Equity		
-	-	[CANADA] Ivanhoe Mines Ltd. (IVN)
-	-	[CAD] New Age Metals (NAM)

*1: Supply, 2: Demand, 6: Market, 8: Dow Jones Precious metal Index, 9: Dow Jones Precious metal commodity Index, 10: Stock Index, 11: OPEC, 12: WTI, 13: Brent, 14: Total Reserves (FX and gold), 15: Exchange Rate, 16: Real GDP Growth, 17: Inflation, 18: Interest rate, 19: Unemployment rate, 20: 10y Bond yield

Results from feature selection reveal that only two assets (EDV and NST) depicted the best performance through ridge regression ($\alpha = 0$). LASSO and EN produced a relatively inadequate fit over the two assets ($R^2 < 0.7$), indicating severe multicollinearity within the data. LASSO outperformed other models for seven assets (PLZL, GFI, SILJ, ETPMAG, HZNC, BVN, and TECK), implying that these datasets contain low variance and provide effective feature selection. Most of the best-performing LASSO assets (six out of seven) produced close approximation to EN with high α values (0.6 - 0.9), suggesting the presence of ridge lowers the model fit. BVN produced a poor fit using EN algorithm, whereas LASSO and ridge provided acceptable approximations. This observation signifies possibility of less volatility and multicollinearity, which could be interpreted on either end. The EN exceeded the performance of LASSO, and ridge for 33 out of 58 assets, proving the effectiveness of tuning parameters to adjust model complexity.

The three algorithms could not achieve desired accuracy within 16 assets due to inferior regression fit. Among these low-performing simulations, ten outputs demonstrated a low regression coefficient ($R^2 < 0.7$), which might depict the poor regression capabilities of selected models under nonlinear data. Furthermore, the remaining six targets reflected negative R^2 values, denoting that no detectable correlations might be present. These observations could be justified through the haven aspects of precious metals, where there is a lower probability that external influence may affect the market. On the contrary, these results also signal the presence of significant nonlinearities in input-output correlations, which could not be captured through the algorithms employed in this study. Confirming these speculations would require a deeper regressive analysis using hybrid models combined with more complex algorithms such as kernel function and recursive elimination [28–30]. Comparing the three feature selection methods provides crucial information regarding data variance, complexity, and multicollinearity, to construct an appropriate dataset.

Investigating the overall performance of the conducted feature selection shows evidence that EN algorithm can improve the applicability of LASSO and ridge for a wide range of data. Although few features were eliminated from certain assets, a substantial number of input variables are available in the filtered dataset. Numerous features would require an advanced predictive model with high memory component. Based on the dataset from our feature selection analysis, neural networks can be applied in future research to project the value of precious metal assets. If the desired accuracy is not obtained during prediction, subsequent data filtering should be attempted through a more aggressive optimization approach. Aside from the tuning parameter, the shrinkage coefficient can be included as a supplementary performance variable. In the current research, we apply the values of λ provided within MATLAB defaults (10^{-1} to 10^{-4}) to remove uncorrelated variables. However, incorporating a slight additional bias could improve the feature selection performance, resulting in the removal of low-impact variables. Careful tuning of shrinkage is necessary to ensure the stability and accuracy of the regression model. A threshold for this parameter must be defined, above which the regression performance declines.

Effect of multiple performance parameters (λ and α) should not be evaluated through OFAT due to the possible combined influence of both variables on the algorithm performance. Multi-objective optimization, such as response surface methodology, can overcome the drawbacks of OFAT through a statistical design of experiments [30]. Employing a higher degree of optimization tends to improve the accuracy of statistical models, although the added complexity may require higher computational ability. In order to create successful real-world applications, data analysts must determine the acceptable trade-off between complexity and precision.

The US gold futures and GLD (ETF) are not correlated with SP 500, one of the country's major market indices. Other assets, including ETPMAG (Australia) and FRES (UK) exhibit no relationship to their respective national stock indices as well. Gold futures and two silver ETFs (SLV and SIL) from the US are unrelated to the Chinese Yuan exchange rate fluctuations. Additionally, the USD volatility did not present any impact on China gold futures, 518880 (China), 8053 (Japan), and NGPLDJ (South Africa). Furthermore, the 10-year bond yield demonstrated no discernable effect over China gold futures, USA palladium futures, 518880 (China), and PALL (USA). All assets except Japanese palladium futures reveal a linkage to at least one oil price (OPEC, Brent, or WTI). The absence of input-output connections indicates that the assets discussed may be an effective hedge against crucial financial and economic variables. This correlation analysis provides beneficial details for investors seeking to diversify their portfolios across precious metals. From the presented results, it is possible to identify factors exerting significant influence on each asset.

List of common variables for each metal is presented in table 9, consisting of important information to evaluate the selected assets. These essential features represent the majority of correlated inputs for an overall market assessment. All four metals appear to have connections with economic factors such as real GDP growth rate and interest rate, implying that a country's economic health could determine precious metal market behavior during the observed duration. WTI crude oil price and unemployment rates may significantly impact gold price movements within the selected period. The silver market observes considerable effects from inflation in conjunction with the common economic variables for all metals (GDP and interest rate). Palladium markets are found to be correlated with inflation and several other features such as demand, commercial application (BASF), Dow Jones indices (Commodity and equity), stock index, and total reserves and unemployment rate. Two platinum assets (NGPLTJ and VALE) show reliable results in table 8, and the selected features involved almost the entire input dataset (15 out of 17 features). These outcomes are justified only through two estimates (for NGPLTJ and VALE), while others were unreliable ($R^2 < 0.7$). Findings indicate potential nonlinearities which cannot be approximated using selected regression models. Nonlinear estimators capable of solving information with higher complexity may verify these implications. A more advanced model could be employed to generate market forecasts, serving as prospects for further research in this area. This comprehensive statistical analysis of the market correlations among global precious metal markets is presented in this study.

Table 9. Correlated features of precious metals

Precious metal	Common features
Gold	WTI, Real GDP growth, Interest rate, Unemployment rate
Silver	Real GDP growth, Inflation, Interest rate
Platinum	Supply, Demand, Market (Faurecia), Dow Jones Precious metal Index, Dow Jones Precious metal commodity Index, Stock Index, OPEC, WTI, Brent, Exchange Rate, Real GDP Growth, Inflation, Interest rate, Unemployment rate, 10y Bond yield
Palladium	Demand, Market (BASF), Dow Jones Precious metal commodity Index, Stock Index, Total Reserves (FX and gold), Real GDP Growth, Inflation, Interest rate, Unemployment rate

5 Conclusion

This report demonstrated a feature selection analysis for 58 global precious metal assets. Three regression-based algorithms (LASSO, ridge, and EN) were employed to identify the correlation of selected input factors from existing literature. Daily trends from July 2016 to June 2021 were investigated, representing short-to-medium-term relationships. Each asset’s highest-performing algorithm was examined to determine valuable information regarding data complexity by removing uncorrelated variables. EN outperformed LASSO and ridge for 33 out of 58 assets, proving its efficiency for a wide range of data. Seven assets are observed to achieve higher accuracy utilizing LASSO, implicating data with low variance and relatively linear relationships. Ridge estimations were utilized in two assets with modest fit ($R^2 < 0.9$), suggesting that a high degree of nonlinearity might exist.

Verifying these results requires deploying nonlinear predictive models adopting the filtered dataset developed in this study. Minimizing the computational load is possible by discarding low-impact variables to reduce the number of inputs in a dataset. The process of elimination incorporates additional bias to increase the shrinkage parameter. When overestimating bias occurs, the network may show a sign of being underfitted, causing lower accuracy. Consequently, the coefficient of shrinkage (λ) could be employed as an additional performance parameter. However, implementing multiple parameters leads to a non-convex optimization issue, where multi-objective techniques provide a viable solution. Applying complex configurations adds further computational load, resulting in an extension of testing time. Findings indicate that optimum modeling results in a trade-off between complexity and accuracy. Solving this subject requires interpreting two vital questions: “What is the achievable complexity?” and “What is the desired accuracy?”. Investigating the answers to these questions requires examining feature selection’s impact through performance optimization of the adopted techniques. These extensive evaluations describe the prospects of this study, which are expected to reveal further insight into the dynamic behavior of precious metal markets.

Analyzing financial markets using regression models requires rigorous testing and optimization. Utilizing advanced approaches assists in improving the accuracy at the expense of extending the computational complexity and the time required. EN algorithm successfully achieved higher feature selection performance across a wide dataset range. The observation agrees with previous research stating EN’s superiority in identifying correlations with economic factors [24]. The absence of variable selection studies attempting to predict precious metal prices could prevent reliable estimation from being generated over extended periods. These methods could improve the quality of forecasts generated within these markets through their dataset.

Simulation results were further investigated to outline the relationships between the precious metal market and its features. The undetected correlation with national stock indices, exchange rates, bond yield, and crude oil implies potential hedging qualities of various assets. From an overall perspective, the haven behavior of precious metal investments shows no foreseeable pattern. Hence, the stochastic modeling procedures serve a practical purpose in identifying risk diversifiers. Common factors from each metal were identified through its assets, offering valuable details for preliminary investment evaluation. These variables are considered crucial market indicators, providing input data information to estimate price movements. Correlations could only be detected among two platinum assets, with the majority of features (15 out of 16) included in the final dataset. Based on these observations, further assessment of platinum market was recommended due to the presence of large nonlinearities. Relevant factors for individual assets vary substantially depending on the location an instrument is traded and the analysis duration. Therefore, it is necessary to validate connections among input-output data throughout any particular timeframe before employing prediction tools. This study presents several findings relating to precious metal investments through significant spillovers from several attributes. These implications represent the contributions of this study, offering critical insights for international investors.

References

- [1] Y. Seo, S. Morimoto, *Resources* **6**, 61 (2017)
- [2] J. Popp, J. Oláh, M.F. Fekete, Z. Lakner, D. Máté, *Energies* **11**, 2392 (2018)
- [3] A. Lahiani, S. Mefteh-Wali, D.G. Vasbieva, *Resources Policy* **74**, 102340 (2021)
- [4] V. Ta, C. Liu, D.A. Tadesse, *Applied Sciences***10**, 437 (2020)
- [5] F. Zhang, K. Sun, X. Wu, *Neurocomputing* **361**, 110 (2019)
- [6] A. Rosato, M. Panella, A. Andreotti, O.A. Mohammed, R. Araneo, *Applied Energy* **291**, 116852 (2021)
- [7] L. Cui, L. Bai, Y. Wang, X. Jin, E.R. Hancock, *Pattern recognition* **114**, 107835 (2021)
- [8] F. Amini, G. Hu, *Expert Systems with Applications* **166**, 114072 (2021)
- [9] Stephanie, D. Rengsamy, F.H. Juwono, J. Nandong, A. Brennan, L. Gopal, *Optimizing NARX-RNN Performance to Predict Precious Metal Futures market*, in Proceedings of the 2022 International Conference on Green Energy, Computing and Sustainable Technology, GECOST, 26-28 October 2022, Miri Sarawak, Malaysia (2022)
- [10] E. Balarie, *Commodities for Every Portfolio: How You Can Profit from the Long-Term Commodity Boom* (John Wiley & Sons, Hoboken, New Jersey, 2007) 61
- [11] M. Eryiğit, *Economic research-Ekonomiska istraživanja* **30**, 449 (2017)
- [12] W. Mensi, J.A. Hernandez, S.M. Yoon, X.V. Vo, S.H. Kang, *International Review of Financial Analysis* **74**, 101672 (2021)
- [13] S. Hammoudeh, F. Malik, M. McAleer, *The Quarterly Review of Economics and Finance* **51**, 435 (2011)
- [14] J. Panda, S. Meher, *Displays* **74**, 102196 (2022)
- [15] E.L. Dan, M. Dînsoreanu, R.C. Mureşan, *Accuracy of six interpolation methods applied on pupil diameter data*, in Proceedings of the 2020 IEEE international conference on automation, quality and testing, robotics, AQTR, 21-23 May 2020, Cluj-Napoca, Romania (2020)
- [16] M. Weber, T. Wilhelm, V. Schmidt, *Image Analysis & Stereology* **40**, 85 (2021)
- [17] A. Jaffar, N.M. Thamrin, M.S.A.M. Ali, M.F. Misnan, A.I.M. Yassin, N.M. Zan, *Bulletin of Electrical Engineering and Informatics* **11**, 2368 (2022)

- [18] F.L. Huang, The Journal of Experimental Education **86**, 265 (2018)
- [19] F.H. Juwono, W.K. Wong, H.T. Pek, S. Sivakumar, D.D. Acula, Biomedical Signal Processing and Control **77**, 103785 (2022)
- [20] Y. Takashi, M. Ueki, M. Yamada, G. Tamiya, I.N. Motoike, D. Saigusa, M. Sakurai, F. Nagami, S. Ogishima, S. Koshiba, and others, Translational psychiatry **10**, 157 (2020)
- [21] W. Wang, J. Liang, R. Liu, Y. Song, M. Zhang, Mathematics **10**, 2985 (2022)
- [22] S. Hammoudeh, P.A. Santos, A. Al-Hassan, The North American Journal of Economics and Finance **25**, 318 (2013)
- [23] J.M.K. Aheto, H.O. Duah, P. Agbadi, E.K. Nakua, Preventive Medicine Reports **23**, 101475 (2021)
- [24] S. Altelbany, Journal of Applied Economics and Business Studies **5**, 131 (2021)
- [25] W. Liu, Q. Li, PloS one **12**, e0171122 (2017)
- [26] M.N. Çiftsüren, S. Akkol, Archives Animal Breeding **61**, 279 (2018)
- [27] Y. Wu, B. Liu, W. Wu, Y. Lin, C. Yang, M. Wang, Journal of Ambient Intelligence and Humanized Computing **9**, 1671 (2018)
- [28] Z. Liu, J. Qu, M.J. Zuo, H.B. Hong, The International Journal of Advanced Manufacturing Technology **67**, 1217 (2013)
- [29] H. Jeon, S. Oh, Applied Sciences **10**, 3211 (2020)
- [30] P. Fathiraja, S. Gopalrajan, M. Karunanithi, M. Nagarajan, M.C. Obaiah, S. Durairaj, N, Neethirajan, Polymer Bulletin **79**, 6211 (2022)