Forecasting Precious Metal Returns
With Multivariate Random Forests

May 2017

Abstract

We use multivariate random forests to compute out-of-sample forecasts of a vector of returns of four precious metal prices (gold, silver, platinum, and palladium). We compare the multivariate forecasts with univariate out-of-sample forecasts implied by random forests independently fitted to every single returns series. Using univariate and multivariate forecast evaluation criteria, we show that multivariate forecasts are more accurate than univariate forecasts.

Keywords: Precious metals; Forecasting; Random forests

* Christian Pierdzioch, Department of Economics, Helmut Schmidt University, Holstenhofweg 85, P.O.B. 700822, 22008 Hamburg, Germany

Marian Risse, Department of Economics, Helmut Schmidt University, Holstenhofweg 85, P.O.B. 700822, 22008 Hamburg, Germany

* Corresponding author. E-mail: c.pierdzioch@hsu-hh.de. Phone: +49-40-6541-2879.

Electronic copy available at: https://ssrn.com/abstract=3160014

1 Introduction

Empirical research on precious metals has mushroomed in recent years. Researchers have studied the hedging and safe-haven properties of precious metals (Baur and McDermott 2010, Baur and Lucey 2010, Beckmann, Berger, and Czudaj 2015) and various aspects of the links between precious metal prices and important economic variables like exchange rates, interest rates, and the oil price (Beckmann and Czudaj 2013, Pukthuanthong and Roll 2011, Reboredo 2013a, b, Pierdzioch, Rohloff, and Risse 2016a, b, Balcilar, Gupta, and Pierdzioch 2017, to name just a few). Several other researchers have analyzed the long-run equilibirum links between different precious-metal prices (see, for example, Baur and Tran 2014, Pierdzioch, Rohloff, and Risse 2015, Kucher and McCoskey 2017) and still other researchers have examined whether it is possible to forecast movements of precious-metal prices (Gupta, Majumdar, Pierdzioch, and Wohar 2017, Pierdzioch, Rohloff, and Risse 2014, 2016c), a question that is particularly relevant for market participants, traders, and other practitioners interested in developments in precious-metals markets.

We reexamine the predictability of the returns of the prices of four key precious metals (gold, silver, platinum, and palladium) in a multivariate modeling framework. Using a multivariate modeling framework is warranted giving recent evidence of potentially time-varying linkages between precious-metal markets. For example, Batten, Ciner, and Lucey (2015) study the spillover effects between the returns and the volatilities of precious-metal prices. They find that gold and silver share a relatively close relationship in returns and especially in volatility. They also report evidence of time-varying spillover effects, which can be interpreted as evidence of time-varying market integration. Hammoudeh, Yuan, McAleer, and Thompson (2010) estimate various multivariate GARCH models to study the interdependencies between gold, silver, platinum, and palladium in terms of their conditional volatilities and correlations. They control for the impact of exchange-rate movements, monetary policy, and geopolitical events and examine (inter-)dependencies to news and past volatility. Their findings imply that the conditional correlations between the metal prices show substantial swings over time, where the responsiveness to own shocks, the persistence of volatility, and short-run cross-shock effects vary across metals. Cross-

Electronic copy available at: https://ssrn.com/abstract=3160014

shock effects become stronger in the long run. Sari, Hammoudeh, and Soytas (2010), in turn, study the short-run and long-run linkages between the prices of gold, silver, platinum, and palladium, on the one hand, and the oil price and the dollar/euro exchange rate, on the other hand. While there turns out to be no strong evidence of a long-run link, they find significant but temporary short-run movements of the precious metal prices to innovations in the respective other metal prices and the exchange rate. The response to innovations of the oil price is positive but insignificant in some cases , and there is also evidence that metal prices may overreact to innovations.

We show that, using a common set of predictors of metal returns that are standard in the literature (Pierdzioch, Rohloff, and Risse 2016b), the out-of-sample forecasts of a vector of metal returns computed by means of a multivariate modeling framework are more accurate than the forecasts of a corresponding univariate framework estimated separately on the returns of the four precious metals. The specific multivariate modeling framework that we use are multivariate random forests (on multivariate trees, see Segal 1992, De'ath 2002; on multivariate random forests, see Segal and Xiao 2011). Random forests consist of an ensemble of individual regression trees. Regression trees use recursive binary splits to subdivide the space of predictors of metal returns into non-overlapping regions, where the region-specific forecast of metal returns is set to some region-specific terminal-node constant (on regression trees, see Breiman et al. 1983). Random forests then build forecasts of metal returns from these terminal-node constants by growing and averaging across several regression trees, where bootstrapping decorrelates the forecasts from individual trees (on bootstrap aggregation, see Breiman 1996) and further stabilization of forecasts is achieved by randomly selecting subsets of predictors for identifying tree nodes (on random forests, see Breiman 2001).

In a recent study, Malliaris and Malliaris (2015) use regression trees to model gold returns. They use regression trees to study variables that help to classify positive versus negative gold returns. Using daily data for the sample period 2004−2014, they study how the financial crisis and the accompanying economic downturn of 2007−2009 have affected classification results and importance of predictors. We go beyond their research in that (i) we study random forests rather than individual regression trees, (ii) we model four precious metals simultaneously, (iii) we model returns rather than only their upticks and downticks, and, (iv) we focus on out-of-sample forecasting rather than in-sample prediction.

We show that multivariate random forests allow more accurate out-of-sample forecasts of precious-metal returns to be computed than random forests fitted to every metal returns series separately. For growing multivariate random forests, we use a common set of standard predictor variables. Multivariate random forests then translate the predictive information embedded in this common set of standard predictor variables into different commodity-specific forecasts, which explain why multivariate forecasting outperforms univariate forecasting even though results reported in recent research indicate that commodities cannot be considered as a single asset class (Erb and Harvey 2006, Batten, Ciner, and Lucey. 2010) and appear to have different safe-haven and hedging properties (Lucey and Sile 2015, Agyei-Ampomah, Gounopoulos, and Mazou 2014). In order to account for the possibility that the degree of integration of precious-metals markets is time-varying (Batten, Ciner, and Lucey 2015), we use a rolling-estimation window to demonstrate the superior performance of multivariate out-of-sample forecasts.

In Section 2, we describe the methods that we use in our research. In Section 3, we describe our data. In Section 4, we describe our results. In Section 5, we conclude.

2 Methods

2.1 Random Forests

A regression tree consists of a root, interior nodes, and terminal nodes. The depth of a tree measures the distance from the root to the terminal nodes. The nodes are the result of binary recursive splits that partition the space of predictors into non-overlapping regions. Splits are computed by optimizing a split function. The split function measures the homogeneity of data assigned to a region, R (for a useful textbook exposition of regression trees and random forests, see Hastie, Tibshirani, and Friedman 2009).

In the univariate case, and using a notation similar to that used by Segal and Xiao (2011), the split function for a region defined by a parent node can be expressed as $\phi(P,R) = SS(R) - SS(R_-) -$

$SS(R_+)$, where P denotes the predictor used to form a split, R_- and R_+ denote the left and right regions resulting from a split, and $SS(R)$ is a region-impurity measure. Assuming $L2$ loss, the region-impurity measure is defined as $SS(R) = \sum_{i \in R} (y_i - \mu(R))^2$, where y_t denotes the returns of the metal price being studied, t denotes time, and $\mu(R)$ denotes the mean of metal returns in region R. In a similar vein, we define $SS(R_-) = \sum_{i \in R_-} (y_i - \mu(R))^2$ and $SS(R_+) = \sum_{i \in R_+} (y_i - \mu(R_+))^2$. Given a predictor, P, considered as a candidate to form a split, metal returns allocated to the left region, R_-, are those returns for which we observe $x_{t,P} \leq c$, and metal returns allocated to the right region, R_+, are those returns for which $x_{t,P} > c$, where $x_{t,P}$, with $P = 1, 2, ..., n$, denotes the period-t realizations of the predictors (assumed to be the same for all metal returns) and c denotes the split point of the predictor, P, considered for forming a split. Maximization of $\phi(P, R)$ requires searching over all predictors and all possible split points.

Given the links between precious metals, it may be possible to reap predictive gains by modeling metal returns in a unified multivariate framework rather than separately. In the multivariate case, we observe a vector, $\mathbf{y}_t$, of metal returns in every period, t. The region-impurity measure is then computed as $SS(R) = \sum_{i \in R} (\mathbf{y}_i - \mu(R))' \mathbf{V}^{-1} (\mathbf{y}_i - \mu(R))$, where $\mu(R)$ now denotes a vector of means and $\mathbf{V}$ denotes the variance-covariance matrix of metal returns (for the case of a region-specific covariance matrix, see Segal 1992). The split function is the same as in the univariate case.

Once the predictor variables used for splitting and the corresponding split points have been identified in this way for the nodes of a given hierarchical level of a regression tree, we can continue by applying the same computations to partition the next hierarchical level of the tree. This recursive process of binary splits stops once the number of observations allocated to a terminal node reaches some minimum. The recursive hierarchical ordering of splits, however, gives rise to the problem that a (univariate or multivariate) regression tree is a high-variance predictor. Random forests overcome this problem by forming an ensemble of regression trees. In a first step, the data are bootstrapped a certain number of times. In a second step, random regression trees are estimated on the bootstrapped data, where the randomness stems from the fact that only a random subset of the predictors is used for forming splits. In a third step, a forecast of metal returns is computed by averaging the forecasts implied by the estimated random regression trees.

2.2 Forecast Evaluation

We evaluate forecasts in different ways. First, we use two univariate measures of forecast accuracy to statistically evaluate the forecasts: the root mean-squared error ($RMSE$) and the mean absolute error (MAE). For the sake of comparing multivariate with univariate forecasts, we compute for every metal the relative measures $RMSE_m/RMSE_u$ and MAE_m/MAE_u, where the index m denotes the component of the vector of multivariate forecasts for the metal being studied and u denotes the corresponding univariate forecast. Second, we use two multivariate measures of forecast accuracy. To this end, we consider the trace of the mean squared error matrix (TMSE; see also Zeng and Swanson 1997). Like Sinclair, Stekler, and Müller-Dröge (2016), we compute $TMSE = \text{trace}(\mathbf{F}'\mathbf{F})$, where $\mathbf{F}$ denotes the matrix of forecast errors. For the multivariate case, this $T \times 4$ matrix consists of the T out-of-sample forecasts of the 4 metal returns. In the univariate case, the columns of this matrix consist of the time series of 4 univariate out-of-sample forecasts of metal returns. Again, we consider the relative measure $TMSE_m/TMSE_u$. Moreover, in order to account for nonzero correlation of metal returns, we consider the Mahalanobis distance defined as $d = (\bar{\hat{\mathbf{y}}}_j - \bar{\mathbf{y}})'\mathbf{V}^{-1}(\bar{\hat{\mathbf{y}}}_j - \bar{\mathbf{y}})$, where $\bar{\hat{\mathbf{y}}}_j$ denotes the vector of means of the multivariate ($j = m$) or the univariate ($j = u$) forecasts, $\bar{\mathbf{y}}$ denotes the vector of means of metal returns, and $\mathbf{V}^{-1}$ denotes the inverse of the variance-covariance matrix of metal returns estimated on metal returns for the period of time for which out-of-sample forecasts are computed. For recent applications of the Mahalanobis distance, see also Sinclair, Stekler, and Carnow (2015) and Sinclair, Stekler, and Müller-Dröge (2016). Again, we consider the relative measure d_m/d_u. Finally, we compare the forecasts implied by multiariate and univariate random forests with random-walk forecasts, and we use the Diebold and Mariano (1995) test to evaluate relative forecast performance.

3 Data

Like Pierdzioch, Roholoff, and Risse (2016b) and Gupta, Majumdar, Pierdzioch, and Wohar (2017), we study weekly precious-metal returns (last trading day of a week). Continuously compounded returns (gold, silver, palladium, and platinum) are computed by taking the first difference of the natural logs of the metal prices. The sample period is 1999/01/15−2017/04/28.

Figure 1 plots the metal prices (scaled to start at 100) and their returns and Table 1 summarizes the sample estimates of the correlations of the four metal returns.[1]

– Please include Figure 1 and Table 1 about here. –

Table 1 shows the sample correlation matrix of metal returns. The correlation of gold and silver returns assumes a value of approximately 0.7 and is higher than the correlations of the other metal returns. Palladium and platinum are also highly correlated (correlation coefficient of 0.60). The correlation of platinum with gold, in turn, assumes a value of 0.55, while the correlation between gold and palladium returns is only roughly 0.38. The correlation matrix indicates that a multivariate approach to forecasting metal returns is likely to improve forecast performance.

Our list of predictor variables comprises the predictors studied by Pierdzioch, Roholoff, and Risse (2016b). Similar predictors have been studied by Reboredo (2013a). We study the following predictors: weekly returns of the S&P-500 index, weekly returns of the GSCI commodity-price index, the Chicago Board Options Exchange Volatility Index (VIX), the term spread (computed as the 10-year treasury constant maturity rate minus the 3-months treasury constant maturity rate), and the corporate bond (CB) spread (computed as Moody's Baa rate minus Moody's Aaa rate), weekly returns of the exchange rate of the dollar vis-á-vis the yen, the euro, the British pound, the Canadian dollar, the Australian dollar, and an aggregate trade-weighted exchange rate index (Broad Trade Weighted Exchange Index of the Fed, TWEXB, available as Wednesday data). In addition, we include metal returns in our list of predictors to capture the predictive value of any temporal (cross-) dependence of returns. In order to account for potential lagged effects of the predictors on metal returns, we include up to 6, 9, and 12 lags of the predictors in the list of predictors.

4 Results

Because results reported by Batten, Ciner, and Lucey (2015) show that the integration of precious metals is likely to undergo changes over time, we compute forecasts by estimating random forests

[1] The data are from Datastream (metal prices) and the FRED board of the Fed of St. Louis (predictors).

on a rolling-estimation window. We consider a window size of 52 weeks (that is, approximately one year of data), and we forecast one-week-ahead metal returns. For constructing regression trees, we allow for a minimum of 5 data points per node. Our random forests consist of 50, 75, and 100 regression trees. We shall also present results of robustness checks that are based on alternative choices for the window length, the forecast horizon, and the number of regression trees.

We set the number of predictors randomly selected for forming splits equal to the square root of the number of the total number of predictors (Hastie, Tibshirani, and Friedman 2009, page 589). We use the variance-covariance matrix estimated on the data for a rolling-estimation window to construct the region-impurity measure in the multivariate case. We use the R programming environment for statistical computing (R Core Team 2017) for our empirical study, and the add-on package "MultivariateRandomForest" (Rahman 2017) for estimating (multivariate and univariate) random forests.

– Please include Tables 2 and 3 about here. –

Tables 2 and 3 summarize our main results. The main message to take home from the tables is that the relative measures of forecast performance assume values smaller than unity, indicating that the univariate forecasts perform worse than the multivariate forecasts. In the case of the *RMSE* there is only one exception for Palladium when the number of lags of predictors is 12 and the random forests consist of an ensemble of 100 regression trees. Both relative univariate measures of forecast performance tend to assume smaller values when we consider the parsimonious models that use only 6 lags of the predictors and 50 regression trees to build random forests rather than the more complex models that feature more lags and more regression trees. On balance, the benefits of multivariate forecasting appear larger for gold, silver, and platinum than for palladium. The results for the $TMSE$ measure of forecast performance confirm that model parsimony tends to strengthen the better performance of multivariate relative to univariate forecasts. The relative Mahalanobis distance, in turn, corroborates that small random forests produce comparatively better multivariate forecasts, but now the performance of multivariate relative to univariate forecasts is best if the list of predictors comprises up to 9 lags of the predictors.

– Please include Figure 2 about here. –

Figure 2 plots the cumulated squared multivariate (black lines) and univariate (gray lines) forecast errors for every metal. The message conveyed by the figure is that, across all for metals, the multivariate forecasts started to dominate the univariate forecasts after 2005, when the prices of gold and silver and, to a lesser extent, that of platinum started to accelerate. Hence, another finding is that multivariate forecasts tend to benefit to a larger extent from the timely identification of rapid changes in market conditions.

– Please include Table 4 about here. –

The results of some robustness checks, summarized in Table 4, confirm this finding. The timely adaptation of a forecasting model to rapid changes in market conditions becomes more difficult when a forecaster opts for a longer rolling-estimation window. We conjecture that the performance of the multivariate relative to the univariate forecasts should be worse than in Table 3 when we choose a longer rolling-estimation window. The results for a rolling-estimation window of length 100 weeks given in Table 2 confirm this conjecture. The relative $TMSE$ and d statistics still show that the multivariate forecasts dominate the univariate forecasts, but as compared to the results given in Table 3, both statistics now assume larger numerical values.

As two additional robustness checks, we extend the forecast horizon from one-week-ahead returns to four-week-ahead returns and we delete all lagged (that is, $t-1$, $t-2$, and so on) predictors from our forecasting models. Results confirm the superior accuracy of the multivariate over the univariate forecasts.

– Please include Table 5 about here. –

Table 5 summarizes p-values of the Diebold and Mariano (1995) test for equal out-of-sample predictive ability. The average squared (absolute) loss of univariate out-of-sample forecasts exceeds in all but one cases studied in the table the average squared (absolute) loss of multivariate out-of-sample forecasts (results not shown in the table). The p-values indicate that the null hypothesis of equal predictive value can be rejected in the majority of cases. On balance, the evidence against the null hypothesis is stronger for gold, silver, and platinum than for palladium, in line with the results summarized in Table 2.

– Please include Figure 3 about here. –

In order to put our results into perspective, we compare in Figure 3 the forecasts implied by multivariate/unviarate random forests with forecasts implied by a random-walk model, where we vary the number of regression trees that form a random forest step by step from 1 to 100. Two results emerge. First, both multivariate and univariate random forests outperform the random-walk forecasts for all precious metals (when the random forests feature more than a single regression tree). Second, the forecasts implied by multivariate random forests outperform the random-walk forecasts to a stronger extent than the forecasts implied by univariate random forests for almost all numbers of regression trees (gold: 87%, silver: 94%, palladium: 90%, platinum: 89%).

5 Concluding Remarks

We have used multivariate random forests to jointly forecast the returns of four precious metals (gold, silver, palladium, and platinum). Using weekly data and applying a rolling-estimation approach, the main result of our empirical analysis is that during the sample period from January 1999 to April 2017 multivariate out-of-sample forecasts tend to be more accurate than univariate out-of-sample forecasts extracted from random forests fitted to every returns series separately.

In future research, our empirical analysis can be extended in several directions. First, it is interesting to adapt multivariate random forests to model price movements of other metals and commodities, perhaps at other data frequencies. Second, it is interesting to use multivariate random forecasts to analyze the hedging and safe-haven properties of precious metals in general and gold in particular. Our results do not inform about the hedging and safe-haven properties of precious metals because we have studied the usefulness of multivariate random forests as a tool to forecast a vector of precious-metal returns rather than as a tool to capture the linkages between precious metals and other asset prices.

Another direction for future research is to compare the out-of-sample forecast performance of multivariate/univariate random forests with the performance of other recent techniques for forecasting metal returns like dynamic model averaging (Aye, Gupta, Hammoudeh, and Kim 2015) and boosting (Pierdzioch, Risse, and Rohloff 2016c). In this regard, a natural extension of our research is to compare the performance of these and other popular techniques with the performance of multivariate/univariate random forests in terms of their implications for trading strategies and excess portfolio returns.

References

Agyei-Ampomah, S., Gounopoulos, D., & Mazouz, K. (2014). Does gold offer a better protection against losses in sovereign debt bonds than other metals?, *Journal of Banking and Finance*, 50, 507−521.

Aye, G. C., Gupta, R., Hammoudeh, S., & Kim, W. J. (2015). Forecasting the price of gold using dynamic model averaging, *International Review of Financial Analysis*, 41, 257−266

Balcilar, M., Gupta, R., & Pierdzioch, C. (2017). On exchange-rate movements and gold-price fluctuations: Evidence for gold-producing countries from a nonparametric causality-in-quantiles test, *International Economics and Economic Policy* (forthcoming).

Batten, J. A., Ciner, C., & Lucey, B. M. (2010). The macroeconomic determinants of volatility in precious metals markets, *Resources Policy*, 35, 65−71.

Batten, J. A., Ciner, C., & Lucey, B. M. (2015). Which precious metals spill over on which, when and why? Some evidence, *Applied Economics Letters*, 22, 466−473.

Baur, D. G., & McDermott, K. (2010). Is gold a safe haven? International evidence, *Journal of Banking and Finance*, 34, 1886−1898.

Baur, D. G., & Lucey, B. M. (2010). Is gold a hedge or a safe haven? An analysis of stocks, bonds and gold, *Financial Review*, 45, 217−229.

Baur, D. G., & Tran, D. (2014). The long-run relationship of gold and silver and the influence of bubbles and financial crises, *Empirical Economics*, 47, 1525−1541.

Beckmann, J., & Czudaj, R. (2013). Oil and gold price dynamics in a multivariate cointegration framework, *International Economics and Economic Policy*, 10, 453−468.

Beckmann, J., Berger, T., & Czudaj, R. (2015). Does gold act as a hedge or a safe haven for stocks? A smooth transition approach, *Economic Modelling*, 48, 16−24.

Breiman, L. (1996). Bagging predictors, *Machine Learning*, 24, 123−140.

Breiman, L. (1996). Random forests, *Machine Learning*, 45, 5−32.

Breiman, L., Friedman, J. H., Oshen, R., & Stone, C. (1983). *Classification and regression trees*. New York: Chapman & Hall.

De'ath, G. (2002). Multivariate regression trees: A new technique for modeling species-environment relationships, *Ecology*, 83, 1105−1117.

Diebold, F. X., Mariano, R. S. (1995). Comparing Predictive Accuracy, *Journal of Business and Economic Statistics*, 13, 253−263.

Erb, C. B., & Harvey, C. R. (2006). The strategic and tactical value of commodity futures, *Financial Analysts Journal*, 62, 69−97.

Gupta, R., Majumdar, A., Pierdzioch, C., & Wohar, M. E. (2017). Do terror attacks predict gold returns? Evidence from a quantile-predictive-regression approach, *Quarterly Review of Economics and Finance* (forthcoming).

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning. Data mining, inference, and prediction*. New York: Springer Science and Business Media, LLC.

Hammoudeh, S. M., Yuan, Y., McAleer, M., & Thompson, M. A. (2010). Precious metals-exchange rate volatility transmissions and hedging strategies, *International Review of Economics and Finance*, 19, 633−647.

Kucher, O., & McCoskey, S. (2017). The long-run relationship between precious metal prices and the business cycle, *Quarterly Review of Economics and Finance* (forthcoming).

Lucey, B. M., & Sile, L. (2015). What precious metals act as safe havens, and when? Some US evidence, *Applied Economics Letters*, 22, 35−45.

Malliaris, A. G., & Malliaris, M. (2015). What drives gold returns? A decision tree analysis, *Finance Research Letters*, 13, 45−53.

Pierdzioch, C., Risse, M., & Rohloff, S. (2014). On the efficiency of the gold market: Results of a real-time forecasting approach, *International Review of Financial Analysis*, 32, 95−108.

Pierdzioch, C., Risse, M., & Rohloff, S. (2015). Cointegration of the prices of gold and silver: RALS-based evidence, *Finance Research Letters*, 15, 133−137

Pierdzioch, C., Risse, M., & Rohloff, S. (2016a). Fluctuations of the real exchange rate, real interest rates, and the dynamics of the price of gold in a small open economy, *Empirical Economics*, 51, 1481−1499.

Pierdzioch, C., Risse, M., & Rohloff, S. (2016b). Are precious metals a hedge against exchange-rate movements? An empirical exploration using Bayesian additive regression trees, *North American Journal of Economics and Finance*, 38, 27−38.

Pierdzioch, C., Risse, M., & Rohloff, S. (2016c). A quantile-boosting approach to forecasting gold returns, *North American Journal of Economics and Finance*, 35, 38−55.

Pukthuanthong, K., & Roll, R. (2011). Gold and the dollar (and the euro, pound, and yen), *Journal of Banking and Finance*, 35, 2070−2083.

Rahman, R. (2017). MultivariateRandomForest: Models multivariate cases using random forests, `https://CRAN.R-project.org/package=MultivariateRandomForest`. R package version 1.1.5

R Core Team (2017). *R: A language and environment for statistical computing*, Vienna, Austria: R Foundation for Statistical Computing. URL http://www.R-project.org/. R version 3.3.3.

Reboredo, J. C. (2013a), Is gold a safe haven or a hedge for the U.S. dollar? Implications for risk management, *Journal of Banking and Finance*, 37, 2665−2676.

Reboredo, J. C. (2013b). Is gold a hedge or safe haven against oil price movements?, *Resources Policy*, 38, 130−137.

Sari, R., , Hammoudeh, S., & Soytas, U. (2010). Dynamics of oil price, precious metal prices, and exchange rate, *Energy economics*, 32, 351−362.

Segal, M. R. (1992). Tree-structured methods for longitudinal data, *Journal of the American Statistical Association*, 87: 407−418.

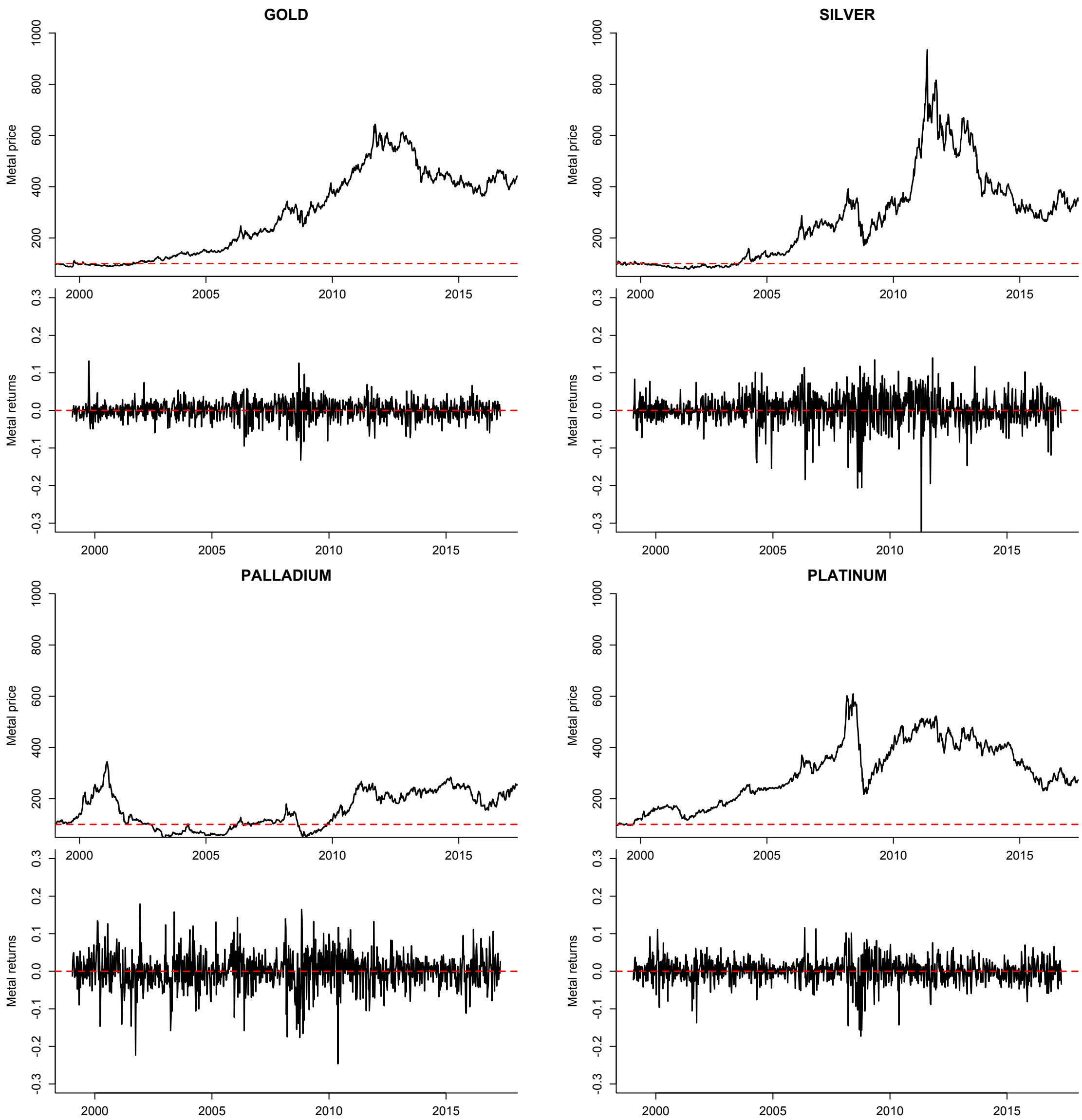
Segal, M., & Xiao, Y. (2011). Multivariate random forests, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1, 80−87.

Sinclair, T. M., Stekler, H. O. & Carnow, W. (2015). Evaluating a vector of the Fed's forecasts, *International Journal of Forecasting*, 31, 157−164.

Sinclair, T. M., Stekler, H. O. & Müller-Dröge, C. (2016). Evaluating forecasts of a vector of variables: A German forecasting competition, *Journal of Forecasting*, 35, 493−503.

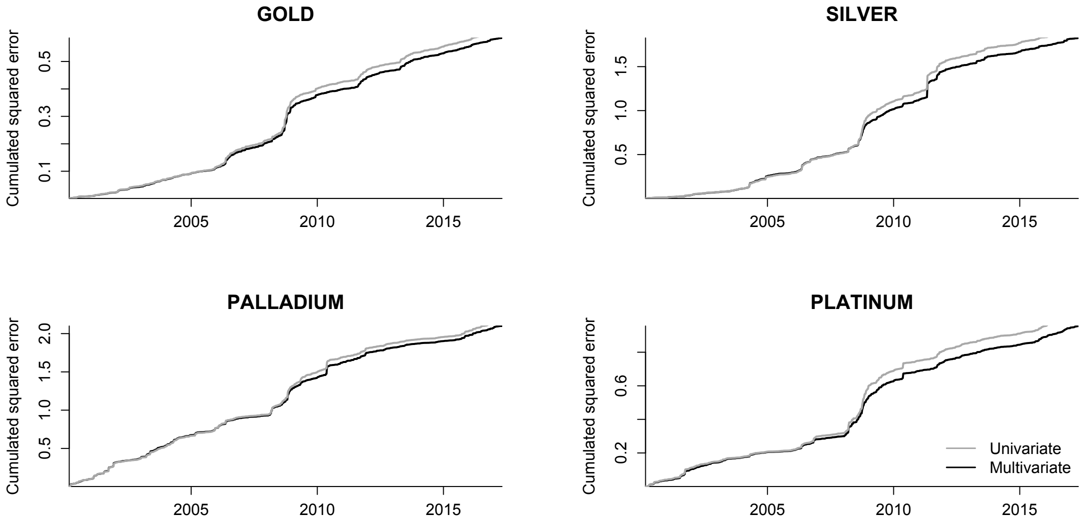
Zeng, T. & Swanson, N. R. (1997). Predictive evaluation of econometric forecasting models in commodity futures markets, *Studies in Nonlinear Dynamics and Econometrics*, 2, 159−177.

Figure 1: Precious Metals Prices and Returns



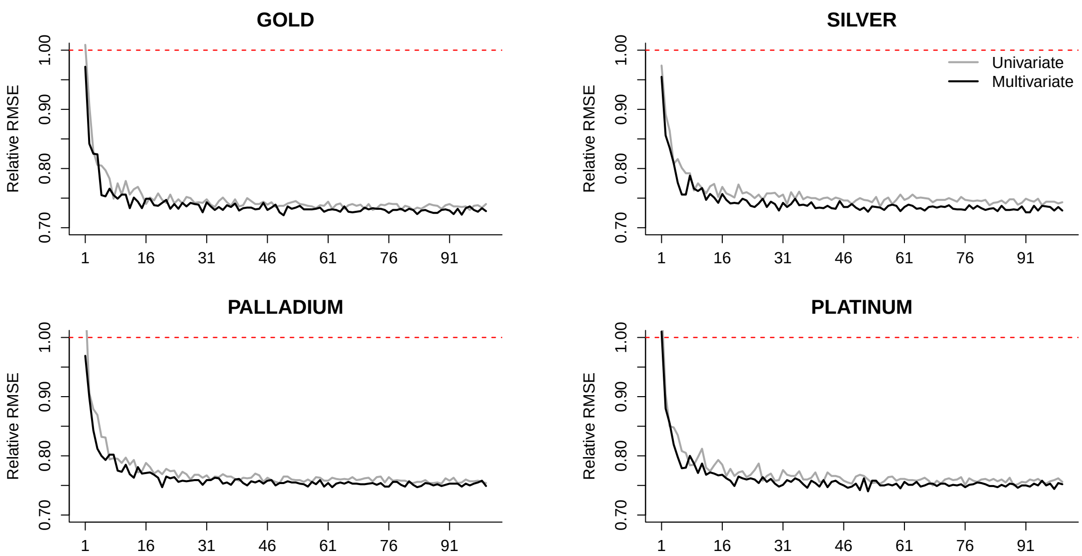
Note: Precious metal prices are scaled to start at 100. The dashed horizontal lines are the 100 (prices) and 0 (returns) lines.

Figure 2: Cumulated Squared Forecast Errors



Note: The cumulated squared forecast errors are based on random forests that consist of 50 regression trees. The random forests use 6 lags of the predictors. Results are for one-week-ahead forecasts and are based on a rolling-estimation window of length 52 weeks.

Figure 3: Comparison With Random-Walk Forecasts



Note: The horizontal line shows the number of regression trees that form a random forest. The random forests use 6 lags of the predictors. Results are for one-week-ahead forecasts and are based on a rolling-estimation window of length 52 weeks.

Table 1: Sample correlation matrix of metal returns

Correlation	**Gold**	**Silver**	**Palladium**	**Platinum**
Gold	1.0000	0.6985	0.3777	0.5490
Silver	0.6985	1.0000	0.4903	0.5744
Palladium	0.3777	0.4903	1.0000	0.6030
Platinum	0.5490	0.5744	0.6030	1.0000

Table 2: Univariate Measures of Forecast Accuracy

Panel A: Relative *RMSE*

Metal	**Gold**			**Silver**			**Palladium**			**Platinum**		
Trees	50	75	100	50	75	100	50	75	100	50	75	100
Lags 6	0.9787	0.9893	0.9838	0.9739	0.9728	0.9811	0.9855	0.9905	0.9924	0.9658	0.9838	0.9950
Lags 9	0.9868	0.9929	0.9880	0.9723	0.9839	0.9771	0.9858	0.9747	0.9849	0.9892	0.9890	0.9812
Lags 12	0.9804	0.9883	0.9877	0.9789	0.9907	0.9782	0.9971	1.0024	0.9898	0.9799	0.9881	0.9889

Panel B: Relative *MAE*

Metal	**Gold**			**Silver**			**Palladium**			**Platinum**		
Trees	50	75	100	50	75	100	50	75	100	50	75	100
Lags 6	0.9809	0.9870	0.9878	0.9783	0.9723	0.9791	0.9852	0.9978	0.9903	0.9765	0.9908	0.9851
Lags 9	0.9821	0.9962	0.9888	0.9826	0.9742	0.9805	0.9969	0.9749	0.9870	0.9833	0.9891	0.9802
Lags 12	0.9843	0.9881	0.9940	0.9747	0.9961	0.9829	0.9924	0.9987	0.9907	0.9765	0.9884	0.9956

Note: $RMSE$ denotes the root mean-squared error. MAE denotes the mean absolute error. Both measures of forecast accuracy are expressed in relative terms as $RMSE_m/RMSE_u$ and MAE_m/MAE_u, where the index m denotes the component of the vector of multivariate forecasts for the metal being studied and u denotes the corresponding univariate forecast. Results are for one-week-ahead forecasts and are based on a rolling-estimation window of length 52 weeks. "Lags" denote the number of lags of the predictors used to estimate random forests.

Table 3: Multivariate Measures of Forecast Accuracy

Panel A: Relative $TMSE$

Trees	**50**	**75**	**100**
Lags 6	0.9552	0.9665	0.9764
Lags 9	0.9643	0.9648	0.9642
Lags 12	0.9722	0.9887	0.9712

Panel B: Relative d

Trees	**50**	**75**	**100**
Lags 6	0.3769	0.3079	0.6327
Lags 9	0.2311	0.2938	0.3548
Lags 12	0.3225	0.3455	0.4560

Note: $TMSE$ denotes the trace of the mean square error matrix. d denotes the Mahalanobis distance. Both measures of forecast accuracy are expressed in relative terms as $TMSE_m/TMSE_u$ and d_m/d_u, where m denotes the multivariate forecasts and u denotes the univariate forecasts. Results are for one-week-ahead forecasts and are based on a rolling-estimation window of length 52 weeks. "Lags" denote the number of lags of the predictors used to estimate random forests.

Table 4: Robustness Checks

Statistic	**Relative $TMSE$**	**Relative d**
Rolling window (100 weeks)	0.9947	0.9060
Four-weeks-ahead forecasts	0.9801	0.6035
No lagged predictors	0.9721	0.6256

Note: $TMSE$ denotes the trace of the mean square error matrix. d denotes the Mahalanobis distance. Both measures of forecast accuracy are expressed in relative terms as $TMSE_m/TMSE_u$ and d_m/d_u, where m denotes the multivariate forecasts and u denotes the univariate forecasts. All robustness checks use random forecasts that consist of 50 regression trees.

Table 5: Diebold-Mariano Test (p-Values)

Panel A: Quadratic loss

Metal	**Gold**			**Silver**			**Palladium**			**Platinum**		
Trees	50	75	100	50	75	100	50	75	100	50	75	100
Lags 6	0.0110	0.0761	0.0158	0.0112	0.0045	0.0008	0.0468	0.1010	0.1011	0.0014	0.0424	0.2019
Lags 9	0.0962	0.2087	0.0313	0.0059	0.0365	0.0002	0.0625	0.0002	0.0039	0.1020	0.0794	0.0064
Lags 12	0.0049	0.0913	0.0466	0.0118	0.1466	0.0004	0.3626	0.6474	0.0563	0.0153	0.0559	0.0401

Panel B: Absolute loss

Metal	**Gold**			**Silver**			**Palladium**			**Platinum**		
Trees	50	75	100	50	75	100	50	75	100	50	75	100
Lags 6	0.0176	0.0447	0.0387	0.0175	0.0005	0.0023	0.0632	0.3930	0.0817	0.0088	0.1388	0.0157
Lags 9	0.0191	0.3158	0.0397	0.0489	0.0011	0.0062	0.3619	0.0007	0.0275	0.0253	0.0798	0.0015
Lags 12	0.0383	0.0555	0.1921	0.0033	0.3306	0.0158	0.1997	0.4301	0.0978	0.0056	0.0676	0.2648

Note: p-values are computed for a one-sided test. The alternative hypothesis is that the multivariate forecast is more accurate than individual forecast. Results are for one-week-ahead forecasts and are based on a rolling-estimation window of length 52 weeks. “Lags” denote the number of lags of the predictors used to estimate random forests.