Gupta, Rangan; Karmakar, Sayar; Pierdzioch, Christian

Book

Safe havens, machine learning, and the sources of geopolitical risk : a forecasting analysis using over a century of data

University of Pretoria
Department of Economics Working Paper Series

Safe Havens, Machine Learning, and the Sources of Geopolitical Risk: A Forecasting Analysis Using Over a Century of Data

Rangan Gupta
University of Pretoria
Sayar Karmakar
University of Florida
Christian Pierdzioch
Helmut Schmidt University
Working Paper: 2022-01
January 2022

Department of Economics
University of Pretoria
0002, Pretoria
South Africa
Tel: +27 12 420 2413

Safe Havens, Machine Learning, and the Sources of Geopolitical Risk: A Forecasting Analysis Using Over a Century of Data

January 2022

Rangan Gupta[*], Sayar Karmakar[†], Christian Pierdzioch[‡]

Abstract

We use monthly data covering a century-long sample period (1915-2021) to study whether geopolitical risk helps to forecast subsequent gold returns and gold volatility. We account not only for geopolitical threats and acts, but also for 39 country-specific sources of geopolitical risk. The response of subsequent returns and volatility is heterogeneous across countries and nonlinear. We find that accounting for geopolitical risk at the country level improves forecast accuracy especially when we use random forests to estimate our forecasting models. As an extension, we report empirical evidence on the predictive value of the country-level sources of geopolitical risk for two other candidate safe-haven assets, oil and silver, over the sample periods 1900–2021 and 1915–2021, respectively. Our results have important implications for the portfolio decisions of investors who seek a safe haven in times of heightened geopolitical tensions.

JEL Classifications: C22; D80; H56; Q02.

Keywords: Gold; Geopolitical Risk; Forecasting; Returns; Volatility; Random Forests

Conflicts of interest: The authors declare no conflict of interest.

[*]Department of Economics, University of Pretoria, Private Bag X20, Hatfield 0028, South Africa; Email address: rangan.gupta@up.ac.za

[†]Department of Statistics, University of Florida, 230 Newell Drive, Gainesville, FL, 32601, USA; Email address: sayarkarmakar@ufl.edu

[‡]Department of Economics, Helmut Schmidt University, Holstenhofweg 85, P.O.B. 700822, 22008 Hamburg, Germany; Email address: macroeconomics@hsu-hh.de.

1 Introduction

The role of gold as a "safe haven" in times of extreme jitters and disruptions in financial (stocks, bonds, and (crypto-)currencies) and commodity markets markets has been extensively studied in a large and significant literature and, thus, is a well-established research topic (see, for example, Baur and Lucey 2010, Baur and McDermott 2010, Reboredo 2013a, Agyei-Ampomah et al. 2014, Gürgün and Ünalmis 2014, Beckmann et al. 2015, Balcilar et al. 2020, Reboredo 2013b, Low et al., 2016, Tiwari et al. 2020). More recently, several systematic studies (such as Balcilar et al. 2016 and 2017, Bouoiyour et al. 2018, Beckmann et al. 2019, Boubaker et al. 2020, Huynh 2020, and Bouri et al. 2021) have been undertaken to better understand the role played by crises, general economic uncertainty, uncertainty due to the COVID-19 pandemic, investor sentiment, i.e., global shocks that adversely affect the markets for risky assets and commodities, and, in the process, act as drivers of gold prices in the context of its safe haven property. Not surprisingly, the market for gold is by now the world's largest metal market in terms of U.S. dollar, valued at 170 billion U.S. dollars per year at current spot prices, with a production of over 3,200 tons per annum (World Gold Council).

Building on the above-mentioned second line of research involving negative worldwide shocks, Baur and Smales (2020) find that gold serves as not only a hedge against geopolitical risk, but also its safe-haven property continues to hold under extreme geopolitical risk. A strong causal impact of geopolitical risk on gold returns has also been reported recently by Gozgor et al. (2019), Huang et al. (2021), and Li et al. (2021). Taken together, these findings hold important lessons given that central bankers, the financial press, and business investors have often cited geopolitical risk as a determinant of investment decisions (Caldara and Iacoviello, 2021). In this regard, when Gallup surveyed in 2017 more than 1,000 investors, 75 percent of respondents expressed concerns about the economic impact of the various military and diplomatic conflicts taking place around the world, and in this regard geopolitical risk ranked ahead of political and economic uncertainties. It is, therefore, not surprising that Carney (2016) includes geopolitical risk, along with economic and policy uncertainties, in the "uncertainty trinity", which could have significant adverse economic and financial effects. Moreover, in the April 2017 Economic Bulletin of the European Central Bank, and the October 2017 World Economic Outlook of the International Monetary Fund, geopolitical uncertainty is highlighted as a salient risk to the economic outlook. Besides geopolitical risk being the dominant form of uncertainty, researchers have shown that it contains leading information for not only real economic variables of advanced and emerging economies (Cheng and Chiu 2018, Clance et al. 2019, Caldara and Iacoviello 2021), but also for financial markets of these economies (Balcilar et al. 2018, Bouri et al. 2019, forthcoming, Gupta et al. 2021, Salisu et al. 2021a, 2022, Yang et al. 2021), and, in addition, for (co-)movements of commodity markets (Ding and Zhang

2021, Tiwari et al. 2021).

Against this backdrop, of in-sample evidence involving geopolitical risk and gold returns, our objective is to conduct an elaborate out-of-sample forecasting analysis of the role of such risk for the predictability of gold returns and volatility, since in-sample evidence does not necessarily translate into out-of-sample forecasting gains (Rapach and Zhou 2013). In this regard, Campbell (2008) points out that the ultimate test of any predictive model (in terms of the econometric methodology(ies) and the predictor(s) used) is in its out-of-sample performance. Besides the statistical significance of a forecasting exercise, real-time forecasting of gold returns is of much more value to investors, relative to full-sample predictions, in designing their optimal portfolios involving gold due its its ability to offer diversification and hedging benefits during periods of turmoil and heightened uncertainties in financial and commodity markets, and the economy in general, emanating from geopolitical risk.

As far as the forecasting experiment is concerned, we analyze the role of global geopolitical risk, and since throughout history the realization of adverse geopolitical events has often been the catalyst for increased fears about future adverse events, we also disaggregate the overall geopolitical risk into threats and realization of adverse geopolitical events, i.e., acts. Importantly, we also study the contribution of country-specific geopolitical risk involving 39 economies to the forecastability of gold returns and volatility. Given that many of the 39 economies considered here play an important role in the supply- and demand-side of the gold market, besides being vulnerable to consistent geopolitical risk at different points in time, the emphasis of our analysis to use country-specific data is likely to be more informative than overall geopolitical risk at the global level. Because analyzing so many predictors in a standard predictive regression framework comes at a cost of overparameterization, and, hence, poor out-of-sample performance, when we look at the role of the geopolitical risk of the 39 countries, we rely on two machine learning approaches. First, we use the least absolute shrinkage and selection operator (Lasso), proposed by Tibshirani (1996). The Lasso belongs to the spectrum of linear regression-analysis methods and performs both model selection and regularization in order to enhance the prediction accuracy and interpretability of the resulting forecasting model. Second, we switch to a nonlinear model and estimate random forests (Breiman, 2001), which, in turn, is a technique tailored to operate in settings featuring a large array of predictors, while simultaneously capturing the predictive value of any potential nonlinear links between the dependent variable and predictors, as well as any interaction effects between the predictors, as highlighted in the case of the relationship between gold market movements and geopolitical risk by Huang et al. (2021) and Li et al. (2021).

To the best of our knowledge, ours is the first paper to forecast real gold returns over the monthly period from February 1915 to September 2021. Such a sample period allows us to cover the longest possible high-frequency (monthly) data available for gold, and the associated predictive impact of various historical global and country-specific geopolitical risk, and in the process makes our analysis immune to any sample-selection bias (Hollstein et al. 2021). Our paper can be considered to add to the relatively large literature associated with the forecasting of gold price and/or returns based on a wide array of macroeconomic, financial, and behavioural predictors that rely on a large spectrum of linear and nonlinear univariate or multivariate models (see, for example, Pierdzioch et al. 2014a, 2014b, 2015a, 2015b, 2016a, 2020, Aye et al. 2015, Hassani et al. 2015, Sharma 2016, Bonato et al. 2018, Nguyen et al. 2019, and Dichtl 2020, with the last paper in particular providing a detailed review). Our paper goes beyond this earlier research in that we use the information content of country-level disaggregate geopolitical risk. It also should be noted that geopolitical risk has been shown to lead several of the predictors considered thus far in this literature. The only somewhat related paper is that of Gupta et al. (2017), wherein the authors use a quantile predictive regression approach to analyze whether terror attacks predict gold returns. They find that terror attacks have predictive power for the lower and particularly the upper quantiles of the conditional distribution of gold returns over the sample period from January 1986 to December 2009, which clearly underlines the importance to account for a potential nonlinear link between gold returns and uncertainty in our forecasting experiment.

While, in line with the safe haven property, the focus is on gold returns, Baur and Smales (2020) and Huang et al. (2021) also provide evidence of the impact of geopolitical risk on the volatility of gold returns. The second-moment impact is not surprising, since higher (lower) geopolitical risk serves as negative (positive) news, and results in higher (lower) trading activity, which in turn can translate into higher (lower) volatility in the gold market (Baur 2012). Realizing this, we also delve into the role of aggregate and disaggregate geopolitical risk in forecasting the conditional volatility of gold prices based on the same set of models used for gold returns. We obtain our metric of volatility by fitting to the gold returns data a standard generalized autoregressive conditional heteroskedasticity (GARCH) model and two recently-developed flexible variants of this model (Wu and Karmakar 2021a, 2021b), namely the time-varying parameter GARCH (TVPGARCH) model and the non-parametric GARCH (NPGARCH) model. We then determine the "optimal" model as the one that produces the lowest forecast errors for the univariate process of volatility, i.e., squared returns. In this regard, it should be noted that Gkillas et al. (2020) have used in recent research a quantile-regression heterogeneous autoregressive realized volatility (QR-HAR-RV) model to show that global geopolitical risk have predictive power for realized gold-returns volatility (estimated from intraday data) mainly at a longer forecast horizon (when one accounts for the potential asymmetry of the loss function a forecaster might use to evaluate forecasts), over the daily period from December 1997 to May 2017. Clearly, our paper can be considered to add to this recent strand of research, and the large gold volatility forecasting literature in general,[1] by predicting the future evolution of

[1]See Pierdzioch et al. 2016b, Salisu et al. 2020, and Luo et al. 2022 for detailed reviews in terms of predictors and

more than a century of gold returns and volatility, based on the information content of not only global but also country-specific geopolitical risk, using linear and nonlinear predictive regressions involving machine learning.

The remainder of the paper is organized as follows: Section 2 outlines our dataset, Section 3 presents the methodologies, Section 4 discusses the results, and Section 5 concludes.

2 Data

We retrieved the real gold price in U.S. dollars from Macrotrends[2], with the data starting from January 1915. With us working with log-returns, implies that the effective sample of our analyses covered February 1915. The sample period ends in September 2021, governed by data availability at the time of writing of this paper.

As far as our predictors are concerned, we rely on the work of Caldara and Iacoviello (2021), who create a new measure of adverse geopolitical events based on an analysis of newspaper articles covering geopolitical tensions. Their data starts in 1900. The geopolitical risk (GPR) index summarizes the results of an automated text search of the electronic archives of three newspapers (The New York Times, Chicago Tribune, and The Washington Post), with the authors calculating the index by counting the number of articles related to adverse geopolitical events in each newspaper for each month (as a share of the total number of news articles). The authors organize the search in eight categories: War Threats (Category 1), Peace Threats (Category 2), Military Buildups (Category 3), Nuclear Threats (Category 4), Terror Threats (Category 5), Beginning of War (Category 6), Escalation of War (Category 7), and Terror Acts (Category 8). Building on the results for these search categories, Caldara and Iacoviello (2021) also derive two sub-indexes, namely an index of geopolitical threats (GPT), which includes words belonging to categories 1 to 5 above, and an index of geopolitical acts (GPA), which covers words belonging to categories 6 to 8.[3]

Besides these global indexes, Caldara and Iacoviello (2021) construct country-specific indexes for 39 different countries,[4] For each of the 39 countries, the authors build the country-specific index (calculated as a monthly share of newspaper articles) by counting the monthly share of all newspaper articles that meet the criterion for inclusion in the GPR index and, in addition, either

alternative econometric frameworks used to forecast gold price volatility.

[2]Internet address: https://www.macrotrends.net/

[3]The data is available for download from the following internet page: https://www.matteoiacoviello.com//gpr.htm, where interested readers also can find further information on the construction of the index, graphs of the data, and links to the relevant literature.

[4]The countries are: North America: Canada, Mexico, US; South America: Argentina, Brazil, Chile, Colombia, Peru, Venezuela; Europe (North and East): Denmark, Finland, Norway, Russia, Sweden, Ukraine, the United Kingdom; Europe (South and West): Belgium, France, Germany, Italy, The Netherlands, Portugal, Spain, Switzerland; Middle East and Africa: Israel, Saudi Arabia, South Africa, Turkey; Asia and Oceania: Australia, China, Hong Kong, Japan, South Korea, The Philippines, Taiwan, Indonesia, India, Malaysia, Thailand. Data on the country-specific indexes can be downloaded from the following internet page: https://www.matteoiacoviello.com/gpr_country.htm

mention the country name or the names of its major cities.

3 Methodologies

3.1 Forecasting models

The purpose of a forecasting exercise is to use training data to estimate a general model

$$y_{t+h} = F(x_{1,t}, x_{2,t}, ..., x_{n,t}), \tag{1}$$

and then to use the estimates along with updated predictors to compute a forecast of the dependent variable:

$$\hat{y}_{t+h+1} = F(x_{1,t+1}, x_{2,t+1}, ..., x_{n,t+1}), \tag{2}$$

where y_t denotes the dependent variable, $x_{i,t}, i = 1, 2, ..., n$ denote the predictors, F is a function that links the dependent variable to the predictors, t denotes time, h denotes the forecast horizon, a hat over a variable denotes a forecast. The dependent variable in the context of our forecasting exercise is either returns or volatility, where we forecast average returns (volatility) when our forecast horizon is larger than one month, that is, when we set $h > 1$. Also, we construct the data such that the number of forecasts is the same for all forecast horizons.

In practice, a popular approach is to start with a small number of predictors and to use a simple linear approximation of the function, F. When only one predictor is considered, the resulting forecasting model is of the format

$$y_{t+h} = \beta_0 + \beta_1 x_{1,t} + e_{t+h}, \tag{3}$$

where e_t denotes a disturbance term and $\beta_i, i = 0, 1$ are the coefficients to be estimated. Estimation of this model can be done by the ordinary least-squared (OLS) technique. When the only predictor used to set up this model is given by lagged returns (volatility) then a simple autoregressive model obtains. We shall use such simple autoregressive model as one of our benchmark models.

The advantage of having a simple benchmark model is that we can compare its forecasting performance with the performance of slightly more complex models of the format:

$$y_{t+h} = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + e_{t+h}, \tag{4}$$

$$y_{t+h} = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \beta_3 x_{3,t} + e_{t+h}, \tag{5}$$

These somewhat more complex models retain the simple linear structure of the benchmark model, but add additional predictors on the right-hand side. In the context of our analysis, the additional predictors are either geopolitical risk (the case of one additional predictor) or geopolitical risk as decomposed into threats and acts (the case of two additional predictors). These two extended forecasting models can also be estimated by the OLS technique.

We can also use the various country-level sources of geopolitical risk to set up an even more advanced extended linear forecasting model, which can then be described by an equation of the following format:

$$y_{t+h} = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \beta_3 x_{3,t} + \beta_4 x_{4,t} + ... + \beta_n x_{n,t} + e_{t+h}. \tag{6}$$

While this forecasting model still can be estimated by the OLS techniques, things are likely to get unedifying when the number, n, of predictors in this equation gets large, which is the case when we study the various country-level sources of geopolitical risk. A model shrinkage and predictor selection technique can help at this stage of the empirical analysis. The technique that we shall use to this end is the least absolute shrinkage and selection operator (Lasso) estimator (Tibshirani, 1996). The Lasso technique chooses the coefficients, $\beta_i, i = 1, 2, ...n$ to be estimated so as to minimize the following expression (for a detailed discussion of the Lasso, see Hastie et al. 2009):

$$\sum_{t=1}^{N} \left(y_{t+h} - \beta_0 - \sum_{i=1}^{n} \beta_i x_{i,t} \right)^2 + \lambda \left(|\beta_0| + \sum_{i=1}^{n} |\beta_i| \right), \tag{7}$$

where N denotes the number of observations used for estimation of the model. The Lasso estimator, thus, adds to the standard quadratic loss function of the OLS technique a penalty term. The shrinkage parameter, λ, defines the weight attached to this penalty terms, and the penalty term itself is given by the sum of the absolute values of the coefficients. In other words, the Lasso estimator uses the L1 norm of the coefficient vectors to shrink the dimension of the estimated forecasting model. Depending on the magnitude of the shrinkage parameter, λ, the Lasso estimator shrinks the magnitude of the coefficients, or even sets to zero some of the coefficients. In the latter case, the Lasso estimator can be interpreted as a predictor-selection technique.

The Lasso estimator, as all other forecasting models that we have discussed so far, retains the assumption that the link between the dependent variable and its predictors is of a linear form, that is, the assumption is that the function, F, is linear. Moreover, the Lasso estimator retains the assumption that the various predictors enter the forecasting model in a additive and separable format. These two assumptions are problematic in a situation when the links between the dependent variable and its various predictors may be nonlinear, and when accounting for potential interaction effects between the predictors could help to improve forecasting performance. As we shall demonstrate when we describe our empirical results in Section 4, nonlinearities are widespread in our data. Moreover, the country-specific sources of geopolitical risk are likely to interact because political agents do not act in isolation and the risks originating in one country are likely to infect allies and adversaries.

In order to capture potential nonlinearities in the data as well as interaction effects among the predictors, we use random forests (Breiman 2001). A random forest belongs to the class of ensemble machine-learning technique because it additively combines a large number of individual regression trees, T. Hence, the basic idea is to approximate the function F, by means of an ensemble of m regression trees as follows:

$$F(x_{1,t}, x_{2,t}, ..., x_{n,t}) = \sum_i T_i, \quad i = 1, 2, ...m \tag{8}$$

As every tree in a natural forest, an individual regression tree consists of a root and several crotches and branches, which subdivide the space of the predictors, $\mathbf{x} = (x_1, x_2, ...)$, into l non-overlapping regions, R_l. These regions, which can be interpreted as the terminal leaves of a regression tree, are formed by applying a search-and-split algorithm in a recursive top-down fashion (for a textbook exposition, see Hastie et al. 2009). In order to describe this search-and-split algorithm, we start at the root of a regression tree. Our aim is to define a crotch in an optimal way such that we subdivide the space of predictors into a left region (that is, a branch), R_1, and a right region, R_2. To this end, we iterate over all predictors and use (in the simplest case) every realization of a predictor as a candidate splitting point. For every combination of a predictor and a splitting point, $\{s, p\}$, we then define the left and right branches, $R_1(s, p) = \{x_s | x_s \leq p\}$ and $R_2(s, p) = \{x_s | x_s > p\}$. In order to find the optimal combination of a predictor and a splitting point, we minimize standard squared-error loss function:

$$\min_{s,p} \left\{ \min_{\bar{y}_1} \sum_{x_s \in R_1(s,p)} (y_i - \bar{y}_1)^2 + \min_{\bar{y}_2} \sum_{x_s \in R_2(s,p)} (y_i - \bar{y}_2)^2 \right\}, \tag{9}$$

where i identifies those realizations of the dependent variable that belong to a half-plane, $\bar{y}_k = \text{mean}\{y_i \ | x_s \in R_k(s, p)\}, k = 1, 2$ denotes the region-specific mean of the dependent variable, and where where we have dropped the time index and the index for the forecast horizon to keep the notation as simple as possible. Hence, the outer minimization runs over all combinations of $\{s, p\}$, and for every single one of those combinations the inner minimization optimally selects the branch-specific means of the dependent variable so as to minimize the branch-specific squared error loss. The result of this inner and outer minimization yields an optimal top-level optimal splitting predictor, an optimal top-level splitting point (and, thus, two branches), and the two branch-specific means of the dependent variable (returns or volatility).

We already can use this rudimentary regression tree to forecast the dependent variable. To this end, we simply update the predictors, decide on whether the updated realization of the optimal top-level predictor belongs to the left or the right branch, and use the corresponding brach-specific mean of the dependent variable as our forecast. We can attempt to produce a better forecast, however, upon growing a larger regression tree. To this end, we apply the search-and-split algorithm to both the left and the right top-level branches, which gives us two second-level optimal splitting predictors and optimal splitting points, and four second-level branch-specific means of the dependent variable. Applying the search-and-split algorithm multiple times in an by now obvious way, we grow an increasingly complex regression tree. We stop this growth process when a regression tree has a preset maximum number of terminal crotches or every terminal branch has a minimum number of observations.

We now can use the complex regression tree that we have built in this way to send the predictors down the tree from its top level to the various leaves along the optimal partitioning points and branches. Equipped with this information, we then compute the optimal means of the dependent variable for the terminal regions and, hence, model the link between the dependent variable and the various predictors as follows:[5]

$$T\left(\mathbf{x}_i, \{R_l\}_1^L\right) = \sum_{l=1}^{L} \bar{y}_l \mathbf{1}(\mathbf{x}_i \in R_l), \tag{10}$$

where L denotes the number of regions and $\mathbf{1}$ denotes the indicator function. Upon updating the predictors, sending them down the tree, and using the optimal means of the dependent variable, we then can use this equation to compute a forecast of the dependent variable.

An obvious problem of this approach to computing forecasts of the dependent variable is that the complex hierarchical structure of a regression tree gives results in an overfitting and data-sensitivity problem, which most likely deteriorates the forecasting performance of a regression tree. It is at this stage of the analysis that the concept of a random forest enters the stage. A random forest resolves the overfitting and data-sensitivity problem by combining a large number of individual regression trees to an ensemble of trees. This ensemble is formed by applying a third-step approach. In the first step, a large number of bootstrap samples is obtained by resampling from the data. In the second step, a regression tree is fitted to every single one of the bootstrapped samples. Importantly, the regression tree that is being fitted is a so-called random regression tree. A random regression tree uses for every splitting step only a random subset of the predictors. Injecting randomness into the splitting process in this way mitigates the effect of influential predictors on tree building. In the third step, we combine the large number of the resulting bootstrapped random trees and, in this way, decorrelate the predictions of the dependent variable as obtained from the individual random regression trees. Moreover, averaging the predictions computed by means of the

[5] It should be noted that Equation (10) shows that random forests ensure that the forecasts of volatility are always non-negative, unlike volatility forecasts derived from models estimated by means of the ordinary least squares (OLS) technique. In our forecasting exercise, however, this is not a serious issue.

individual random regression trees stabilizes predictions.

We use the R language and environment for statistical computing (R Core Team 2021) to set up our forecasting experiment. We use the R add-on package "glmnet" (Friedman et al. 2010) to implement the Lasso estimator, where we use 10-fold cross-validation to identify the optimal shrinkage parameter that minimizes the mean cross-validated error. We use the R add-on package "grf" (Tibshirani et al. 2021) to estimate random forests. In our forecasting experiment, a random forest is built from 1,000 random regression trees (bootstrapping is done by sampling with replacement). We use cross-validation to select the tree parameters (number of randomly sampled predictors selected for tree building, minimum number of data at a terminal tree node, and maximum imbalance of a split at a node of a tree). We refer a reader for further technical details to the extensive documentations of these packages.

3.2 Conditional volatility models

Because one purpose of our forecasting exercise is to forecast the conditional volatility of real log-returns of gold, we need to specify models that render it possible to estimate conditional volatility. To this end, we estimate three alternative models, and then identify the "optimal" model as the one that yields the lowest mean-squared error (MSE), $\sum(y_i^2 - \hat{\sigma}^2)^2$.

We start off with the standard GARCH(1,1) model:

$$y_t \sim N(0, \sigma_t^2) \quad \text{with} \quad \sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \tag{11}$$

We use the "fgarch" add-on package in R to obtain our maximum-likelihood parameter estimates of α_0, α_1 and β_1, so as eventually to obtain $\hat{\sigma}^2$.

We then estimate a time-varying parameter GARCH (TVPGARCH) as follows:

$$y_t \sim N(0, \sigma_t^2) \quad \text{with} \quad \sigma_t^2 = \alpha_0(t/n) + \alpha_1(t/n) y_{t-1}^2 + \beta_1(t/n) \sigma_{t-1}^2, \tag{12}$$

In order to estimate the time-varying parameter functions, $\alpha_0(\cdot), \alpha_1(\cdot)$, and $\beta_1(\cdot)$, we use the kernel-based method described in detail in Karmakar et al. (2021). For a suitable choice of kernel K and bandwidth parameter, $b_n \in [0,1]$, we estimate $\theta = (\alpha_0, \alpha_1, \beta)$ using

$$\hat{\theta}_{b_n}(t) = \text{argmin}_{\theta \in \Theta} \sum_{i=1}^{n} K((t - i/n)/b_n) \ell(y_i, X_i, \theta), \qquad t \in [0,1], \tag{13}$$

where $\ell(\cdot)$ is the corresponding negative log-likelihood or quasi log-likelihood for estimating the GARCH parameters, X_i denotes the vector of covariates, which in the context of our empirical analysis will be y_{i-1} for an univariate GARCH model. In particular, for our estimation problem, ℓ is of the following format:

$$\ell(y_i, X_i, \theta') = -\frac{1}{2}\log(\sigma_i^2) + y_i^2/\sigma_i^2 \quad \text{with} \quad \sigma_i^2 = \alpha_0 + \alpha_1 y_{i-1}^2 + \beta_1 \sigma_{i-1}^2, \tag{14}$$

and we choose an Epanechnikov Kernel to specify K. Finally, with the estimated function, $\alpha_0(\cdot), \alpha_1(\cdot)$, and $\beta_1(\cdot)$ in hand, we compute $\hat{\sigma}^2$.

As our third model, we estimate a non-parametric GARCH (NPGARCH) model. This is a relatively new model-free approach of fitting a GARCH model from the perspective of superior prediction performance than standard GARCH or GARCH-type models. Politis (2015) proposes a model-free approach that relaxes the normality assumption of traditional GARCH models, and rather tries to recover the error process from the reconstructed residuals. Building on that approach, Chen and Politis (2019) propose a method named GE-NoVas curated to ARCH models, which, in turn, paved the path for the parsimonious GE-NoVas approach described by Wu and Karmakar (2021a). Adapted to the class of GARCH models, their approach led to the development of the parsimonious GARCH-NoVas (GA-NoVas in short) method by Wu and Karmakar (2021b), which shows superior predictive performance among its class of model-free approaches, as well as significantly beating the traditional GARCH-based method. Formally the NPGARCH can be described as follows:

$$y_t \sim N(0, \sigma_t^2) \text{ with } \sigma_t^2 = \left(1 - \frac{\alpha_1}{1-\beta_1}\right) s_{t-1}^2 + \alpha_1 y_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \tag{15}$$

where s_{t-1}^2 is the estimated variance using the data up to time period $t-1$. In a way, one can think of this as a special time-varying method because the intercept term is changing, but it is clearly different from the one we use in the TVPGARCH framework.

The MSE obtained under the GARCH, TVPGARCH, and NPGARCH model for real log-returns of gold is equal to 3194.4030, 3420.2440, and 3184.9460, i.e., it is lowest for the NPGARCH model, and so we use the fitted variance computed by means of this model in the forecasting exercise we shall describe in Section 4 below when we forecast the volatility of real gold returns.[6]

4 Empirical results

Table 1 summarizes results for the root-mean-squared forecast error (RMSFE) statistic based on recursive-window estimates of the various models, while Table 2 reports the corresponding results for a rolling-estimation window. Panel A depicts the results for returns and Panel B depicts the results for volatility. In order to be able to compare models, we document the RMSFE statistic as the ratio of the RMSFE statistic of a benchmark model and the RMSFE statistic computed for a rival

[6]Complete estimation details of the parameters of the three models are available upon request from the authors.

model. A ratio exceeding unity, thus, indicates that the rival model yield more accurate forecasts in terms of the RMSFE statistic. As a sensitivity check, we report the RMSFE ratio for two training periods in the case of a recursive window, and two rolling-estimation window.

– Please include Tables 1 and 2 about here. –

Turning first to the results for returns and a recursive-estimation window, we find that extending a simple autoregressive (AR) benchmark model to include geopolitical risk, or the corresponding threats and acts, we find that, for both training periods and for all four forecast horizons that we study, the rival model outperforms the benchmark model. Moreover, the advantage of the rival model over the benchmark model is increasing in the length of the forecast horizon. Similarly, the Lasso estimator and random forests, both estimated using the country-level shares of geopolitical risk, produce better forecasts than the autoregressive benchmark model, and again this relatively good performance strengthens when the forecast horizon gets longer.[7] Importantly, we observe that random forests produce the largest gains in terms of forecasting performance. Furthermore, when we include geopolitical risk as an additional predictor in the benchmark model then we observe a moderate improvement of forecasting performance when we consider threats and acts separately to set up our rival model, or when we study the Lasso estimator to model the link between returns and country-level shares of geopolitical risk. Random forests, in turn, produce relatively large forecasting gains for the intermediate and longer forecast horizons.

Turning next to the results for returns and the rolling-estimation windows in Table 2, the RMSFE ratios corroborate the relatively good forecasting performance of random forests. The RMSFE ratios for the other models, in turn, often fall short of unity, indicating that adding geopolitical risk to the vector of predictors does not improve their forecasting performance relative to a simple autoregressive benchmark. As in the case of a recursive-estimation window, the forecasitng performance of the rival models increases in the forecast horizon.

The results for volatility are similar for the recursive and the rolling-estimation window. Geopolitical risk and its components improve forecasting performance in the majority of model configurations and forecasting gains increase in the forecast horizon. Differentiating between threats and acts contributes to forecasting performance when the simple autoregressive model is the benchmark model. When the benchmark model also includes geopolitical risk as a predictor, however, accounting for threats and acts mainly helps to improve forecasting performance only when we study a rolling-estimation window, especially at the two long forecast horizons. Random forests clearly yield the largest and most robust forecasting gains in terms of the RMSFE statistic, irrespective of whether we study a recursive-estimation window or a rolling-estimation window.

[7] Because the country-specific GPR indexes are expressed as a monthly share of newspaper articles, when we study the Lasso and random forests, we use for our model comparisons the GPR index expressed as a share for the benchmark model (though the share, of course, is perfectly correlated with the GPR index).

– Please include Tables 3 and 4 about here. –

Tables 3 (recursive-estimation window) and 4 (rolling-estimation window) report the results (p-values computed using robust standard errors) of the regression-based Clark and West (2007) test for an equal (adjusted) mean-squared-prediction error (MSPE). The alternative hypothesis is that the rival model has a smaller MSPE than the benchmark model. The test, thus, is one-sided.

We find that, for returns, several tests yield significant results. Moreover, the test results are stronger for a recursive-estimation window than for a rolling-estimation window. For a a recursive-estimation window, all test results are significant, and for both training windows, when the simple autoregressive model is the benchmark model. Yet, for the recursive-estimation windows, half of the test results are insignificant when we differentiate between threats and acts when the benchmark model already includes geopolitical risk. Similarly, half of the test results for the Lasso estimator are insignificant.

For a rolling-estimation window, considering geopolitical risk as a predictor of returns is ineffective in terms of forecasting performance when we use a simple autoregressive benchmark as our benchmark model. Differentiating between threats and acts, in turn, mainly helps to compute better forecasts relative to the autoregressive benchmark and the autoregressive-cum-geopolitical-risk benchmark model for the two longer forecast horizons. The test results for the Lasso estimator and random forests are significant for all forecast horizons and for both rolling-estimation window.

As for volatility, we obtain significant test results when we use a recursive-estimation window. The only exceptions are the test results for the model that uses threats and acts as separate predictors when the benchmark model includes geopolitical risk. For a rolling-estimation window, in turn, we observe a tendency for the test results to strengthen as the forecast horizon increases, especially when we study the longer rolling-estimation window. We obtain relatively strong evidence that geopolitical risk, as differentiated across country sources, matters for forecasting volatility when we study the Lasso estimator and random forests, where random forests appear to produce results that are somewhat more robust than the Lasso results across the two rolling-estimation windows.

– Please include Tables 5 and 6 about here. –

The results for the Diebold and Mariano (1995) test, as modified by Harvey et al. (1997), that we report in Tables 5 (recursive-estimation window) and 6 (rolling-estimation window) corroborate the good forecasting performance of random forests. While the test results for the other model configurations, including the Lasso estimator, are only occasionally significant, we find that random forests compute better forecasts relative to the autoregressive benchmark and the autoregressive-cum-geopolitical-risk except for the short forecast horizon.

We next turn our attention to volatility. Evidence that geopolitical risk, and its decomposition into threats and acts, contributes to the predictability of volatility is strong when we study a recursive-estimation window, but only when the simple autoregressive model is the benchmark model. When the benchmark model already includes geopolitical risk as a predictor then decomposing risk into threats and acts does not further increase forecasting performance. When we turn to a rolling-estimation window, in turn, the results of the modified Diebold and Mariano (1995) test are significant neither for geopolitical risk nor its two components, threats and acts. Similarly, the test results for the Lasso estimator are insignificant for both the recursive-estimation window and the rolling-estimation window. The forecasts that we compute by means of random forests, however, yield significant test results in all cases when we study a recursive-estimation window, and in six out of eight cases when we study a rolling-estimation window.

– Please include Table 7 about here. –

The relatively good and robust performance of random forests in comparison to the Lasso estimator leads us to hypothesize that it is not only accounting for the country-sources of geopolitical risk that leverages forecasting performance, but rather that accounting also for potentially nonlinear links between returns and volatility and the country-specific sources of geopolitical risk, as well as potential interactions between the latter, that matters for forecasting performance. The results summarized in Table 7 support this hypothesis. Random forests outperform the Lasso estimator in terms of the RMSFE statistic for both returns and volatility, where the gains from using random forests are stronger for a rolling-estimation window than for a recursive-estimation window, and the relatively forecast performance increases in the forecast horizon. The results of the Clark and West (2007) test corroborate the results of for the RMSFE statistic. Similarly, the results of the modified Diebold-Mariano test yield strong evidence of the superior performance of random forests relative to the Lasso estimator.

– Please include Figures 1 and 2 about here. –

Figures 1 (for returns) and 2 (for volatility) illustrate that returns and volatility are linked to geopolitical risk (i) in a heterogeneous way across countries (and, thus, accounting for country-specific sources of geopolitical risk matters), and, (ii) these links can be strongly nonlinear. The two figures plot partial dependence functions that visualize how returns and volatility respond to a variation in the share of geopolitical risk that can be attributed to the US and China, two major powers on the international political scene.[8] The partial dependence functions for the US show that returns increase at a relatively low value of geopolitical risk that originates in the US except at the three-months forecast horizon. At the one month and six months forecast horizon, returns stay at this higher level when geopolitical risk increases further. For the long forecast horizon, we observe that

[8]The estimation of the partial dependence functions use the full sample of data. We computed the partial dependence functions by means of the R add-on package randomForestSRC (Ishwaran and Kogalur 2021). Sampling is with replacement, the minimum node size is five and one third of the predictors are used for splitting.

returns decrease again to a lower level for intermediate to high values of geopolitical risk. For all four forecast horizons, returns are more or less insensitive to high values of US geopolitical risk. As for China, we find a clear pattern that returns first tend to drop in the region of very low geopolitical risk, but they then start increasing as geopolitical risk increases, while the partial dependence functions become more or less flat for high values of geopolitical risk. The partial dependence functions further show that volatility is lower for higher than for lower values of US geopolitical risk for the intermediate and long forecast horizons. For the short forecast horizon, we find a partial dependence function that exhibits a U-shaped pattern at low und intermediate values of US geopolitical risk. The partial dependence functions for Chinese geopolitical risk, in sharp contrast, witness that volatility clearly is increasing when geopolitical risk increases from a low to an intermediate value, and then stays at this higher level when geopolitical risk increases further.

In order to get an idea of the potential economic benefits of accounting for geopolitical risk and its country-level sources, we study the directional accuracy of returns forecasts, motivated by the earlier works of Parisi et al. (2008), and Malliaris and Malliaris (2015) associated with the gold market. The directional accuracy of forecasts is an important element of the profitability of, for example, a simple trading rule that switches back and forth between an investment in gold and some riskless asset. In order to measure directional accuracy, we define a dummy variable that takes on the value one when actual returns are positive, and zero otherwise. We then compute for the benchmark and the rival model the area under the receiver operating characteristic (AUROC) curve. A receiver operating characteristic curve measures the directional accuracy of forecasts by confronting the rate of true signals (sensitivity) with the rate of false signals (one minus specificity) for alternative values of a decision criterion that defines when a forecast signals positive returns.[9] Finally, we compute the ratio of the AUROC statistic of the rival model and the AUROC statistic of the benchmark model. Hence, a ratio larger than unity shows that the rival model has a greater directional accuracy for returns than the benchmark model.

– Please include Table 8 about here. –

Table 8 summarizes the results for the AUROC statistic. Three results emerge. First, accounting for geopolitical risk improves the directional accuracy of returns forecasts, especially when we study a recursive-estimation window, but we also observe many AUROC ratios larger than unity for a rolling-estimation window. Second, random forests perform not only well, but also better than the Lasso estimator. Third, the contribution of geopolitical risk to directional forecast performance in general tends to become stronger when the forecast horizon increases.

Another question is whether the contribution of geopolitical risk and its country-specific sources to forecasting performance is mainly a phenomenon that can be observed during a view historical

[9]For a more detailed description of receiver operating characteristic curves and the (AUROC statistic as well as a recent application of these techniques in economics, see Pierdzioch and Gupta (2020).

episodes, or whether forecasting gains are more evenly spread across the century-long sample period that we study in our empirical research. In order to answer this question, we split our forecasting periods into ten sub-periods, compute for every sub-period the RMSFE statistics for the benchmark and rival models, and then count the number of sub-periods during which the resulting RMSFE ratio exceeds unity. The result of this count can take on a value between zero and ten, where a number of or close to ten shows that the rival model outperforms the benchmark model in several sup-periods and, thus, its superior forecast performance is spreads across the whole sample period.[10]

– Please include Tables 9 and 10 about here. –

Tables 9 (recursive-estimation window) and 10 (rolling-estimation window) summarize the results. The results show that, while geopolitical risk as well as threats and acts also contribute to forecasting performance in several sub-periods, the benefits from using random forests, when studied across returns and volatility, window types, and forecast horizons, tend to be more evenly spread across the sub-periods than for the other forecasting models.

As an extension, we report at the end of the paper (Appendix) results for two other commodities namely, silver and oil, that have also recently been shown to act as hedges, if not safe-havens, against geopolitical risk (see for example, Bouoiyour et al. 2019, Baur and Smales 2020, Li et al. 2020, Qin et al. 2020, Smales 2021).[11] For oil returns, we observe that random forests, in terms of the RMSFE statistic, tend to perform comparatively well, where the forecasting gains are concentrated at the long forecast horizon. For oil volatility, we also observe that random forests perform substantially better than the competing models, particularly when we study a rolling-estimation window and a long forecast horizon. For silver returns, we find that random forests perform comparatively well for both window types, where relative forecasting performance strengthens in the length of the forecast horizon. For silver volatility, in contrast, random forests do not lead to an improvement in forecasting performance in terms of the RMSFE statistic. For all model configuration, the RMSFE ratios in general are close to unity, indicating a small, if any, contribution of geopolitical risk and its country-specific components to the predictability of silver volatility. Taken together, the results for oil and silver show that, in terms of the RMSFE statistic, the gains in forecasting performance from accounting for geopolitical risk in general and its country-specific sources in particular

[10] It should be kept in mind that the count statistic does not indicate the magnitude of the forecasting gains that a forecaster can reap upon switching from the benchmark to the rival models.

[11] The oil data starts in January 1900 and ends in September 2021. The sample period for silver is the same as that for gold. The data source for the nominal West Texas Intermediate (WTI) oil price in U.S. dollars and the U.S. consumer price Index used to deflate the nominal price to get real values are from Global Financial Data (`https://globalfinancialdata.com/`), and the datasource for real silver price is Macrotrends. As with gold, we work with the real log returns of these two commodities, and the optimal conditional volatility models (i.e., based on the minimum MSE among the GARCH, TVPGARCH and NPGARCH models) for oil is the NPGARCH model (with MSE equal to 19965.9300 relative to 32658.5000 (GARCH) and 26905.0200 (TVPGARCH)), while that for silver is the standard GARCH model (with an MSE of 38626.0900 relative to 39249.1100 (TVPGARCH) and 41390.3100 (NPGARCH)).

tend to be smaller than those we observe for gold.[12]

5 Conclusion

Using country-specific sources of geopolitical risk is useful for forecasting gold returns and gold volatility, especially when random forests, which account for potentially complex nonlinear data links and predictor interactions in a flexible and purely data-driven way, are used for model estimation and forecasting. Random forests often perform better than various linear alternative models, including forecasting models estimated using the Lasso estimator, a popular model shrinkage and predictor selection technique. Random forests also render it possible via the instrument of partial dependence functions to track more closely than is possible in the case of other methods how exactly gold returns and gold volatility respond to different levels of geopolitical risk. This information is particularly useful for forecasters and investors in the pricing of related derivatives as well as for devising hedging strategies involving gold investments as a safe haven in times of heightened geopolitical stress. Moreover, given that gold returns and volatility tend to lead economic activity (Piffer and Podstawski 2017, Çepni et al. 2021, Salisu et al. 2021c), the real effect of geopolitical events could end up being more persistent due to the existence of the indirect channel involving the gold market-geopolitical risk nexus. In light of this, accurate forecasts of the first- and second-moment movements of gold returns due to geopolitical acts and threats, would help policymakers to design the size of their policy response to prevent deep recessions.

In future research, it is interesting to extend our research along several avenues. First, one could study the contribution of geopolitical risk at the country-level to forecast accuracy in case of other commodities. Second, the implications of country-level geopolitical risk for other asset prices could also be investigated. In this regard, it would be interesting to study, for example, whether relative country-specific risk helps to improve the accuracy of forecasts of exchange-rate changes. Yet another avenue for future research would be to analyze the safe-haven property of gold (oil and silver) using other machine-learning techniques besides random forests used in this paper.

[12] It should be noted that, while ours is the first paper to forecast historical real-returns and volatility of silver based on geopolitical risk, there exists a few papers that have forecasted the returns and volatility of the oil market, but based on data spanning the last three decades or so (see, for example, Liu et al. 2019, Plakandaras et al. 2019, Asai et al. 2020, Mei et al. 2020, and Salisu et al. 2021b).

References

Agyei-Ampomah, S., Gounopoulos, D. and Mazouz, K. (2014). Does gold offer a better protection against sovereign debt crisis than other metals? Journal of Banking and Finance, 40, 507–521.

Asai, M. Gupta, R., and McAleer, M. (2020). Forecasting volatility and co-volatility of crude oil and gold futures: Effects of leverage, jumps, spillovers, and geopolitical risks. International Journal of Forecasting, 36(3), 933–948.

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

Balcilar, M., Bonato, M., Demirer, R., and Gupta, R. (2017). The effect of investor sentiment on gold market return dynamics: Evidence from a nonparametric causality-in-quantiles approach? Resources Policy, 51, 77–84.

Balcilar, M., Bonato, M., Demirer, R., and Gupta, R. (2018). Geopolitical risks and stock market dynamics of the BRICS, Economic Systems, 42(2), 295–306.

Balcilar, M., Demirer, R., Gupta, R., and Wohar, M.E. (2020). The effect of global and regional stock market shocks on safe haven assets. Structural Change and Economic Dynamics, 54, 297–308.

Balcilar, M., Gupta, R., and Pierdzioch, C. (2016). Does uncertainty move the gold price? New evidence from a nonparametric causality-in-quantiles test. Resources Policy, 49(C), 74–80.

Baur, D.G. (2012). Asymmetric volatility in the gold market. Journal of Alternative Investments, 14, 26–38.

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

Baur, D.G., and Smales, L.A. (2020). Hedging geopolitical risk with precious metals. Journal of Banking and Finance, 117, 105823.

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

Beckmann, J., Berger, T., and Czudaj, R. (2019). Gold Price dynamics and the role of uncertainty. Quantitative Finance, 19(4), 663–681.

Bonato, M., Demirer, R., Gupta, R., and Pierdzioch, C. (2018). Gold futures returns and realized moments: A forecasting experiment using a quantile-boosting approach. Resources Policy, 57, 196–212.

Boubaker, H., Cunado, J., Gil-Alana, L.A., and Gupta, R. (2020). Global crises and gold as a safe haven: Evidence from over seven and a half centuries of data. Physica A: Statistical Mechanics and its Applications, 540, 123093.

Bouoiyour, J., Selmi, R., Hammoudeh, S., and Wohar, M.E. (2019). What are the categories of geopolitical risks that could drive oil prices higher? Acts or threats? Energy Economics, 84, 104523.

Bouoiyour, J., Selmi, R., and Wohar, M.E. (2018). Measuring the response of gold prices to uncertainty: An analysis beyond the mean. Economic Modelling, 75(C), 105–116.

Bouri, E., Çepni, O., Gabauer, D., and Gupta, R. (2021). Return connectedness across asset classes around the COVID-19 outbreak? International Review of Financial Analysis, 73, 101646.

Bouri, E., Çepni, O., Gupta, R., and Jalkh, N. (Forthcoming). Geopolitical risks and stock market volatility in the G7 countries: A century of evidence from a time-varying nonparametric panel data model. Handbook for the Economics of Terrorism, Edited by Atin Basuchoudhary and Gunther G. Schulze.

Bouri, E., Demirer, R., Gupta, R., and Marfatia, H.A. (2019). Geopolitical risks and movements in Islamic bond and equity markets: A note. Defence and Peace Economics, 30(3), 367–379.

Breiman, L. (2001). Random forests. Machine Learning, 45, 5–32.

Caldara, D., and Iacoviello, M. (2021). Measuring geopolitical risk. Working Paper, Board of Governors of the Federal Reserve.

Campbell, J.Y. (2008). Viewpoint: estimating the equity premium. Canadian Journal of Economics, 41, 1–21.

Carney, M. (2016). Uncertainty, the economy and policy. Speech at the Bank of England, London, 30 June 2016. Available for download from: https://www.bis.org/review/r160704c.pdf.

Çepni, O., Dul, W., Gupta, R., and Wohar, M.E. (2021). The dynamics of U.S. REITs returns to uncertainty shocks: A proxy SVAR approach. Research in International Business and Finance, 58(C), 101433.

Chen, J., and Politis, D.N. (2019). Optimal multi-step-ahead prediction of arch/garch models and novas transformation. Econometrics, 7, 34.

Cheng, C.H.J., and Chiu, C-W(J). (2018). How important are global geopolitical risks to emerging countries? International Economics, 156, 305–325.

Clance, M.W., Gupta, R., and Wohar, M.E. (2019). Geopolitical risks and recessions in a panel of advanced economies: Evidence from over a century of data. Applied Economics Letters, 26(16), 1317–1321.

Clark, T.D., and West, K.D. (2007). Approximately normal tests for equal predictive accuracy in nested Benchmark/rival model. Journal of Econometrics, 138, 29–311.

Diebold, F.X., and Mariano, R.S. (1995). Comparing predictive accuracy. Journal of Business and Economic Statistics, 13, 253–263.

Ding, Z., and Zhang, X (2021). The impact of geopolitical risk on systemic risk spillover in commodity market: An EMD-based network topology approach. Complexity, 2021, 2226944.

Dichtl, H. (2020). Forecasting excess returns of the gold market: Can we learn from stock market predictions?, Journal of Commodity Markets, 19, 100106.

Friedman, J. Hastie, T., and Tibshirani, R. (2010). Regularization paths for Generalized Linear Models via coordinate descent. Journal of Statistical Software, 33(1), 1–22.

Gkillas, K., Gupta, R., and Pierdzioch, C. (2020). Forecasting realized gold volatility: Is there a role of geopolitical risks?, Finance Research Letters, 35(C), 101280.

Gozgor, G., Lau, C.K.M., Sheng, X., and Yarovaya, L. (2019). The role of uncertainty measures on the returns of gold. Economics Letters, 185, 108680.

Gupta, R., Majumdar, A., Nel, J., and Subramaniam, S. (2021). Geopolitical risks and the high-frequency movements of the US term structure of interest rates. Annals of Financial Economics, 16(3), 2150012.

Gupta, R., Majumdar, A., Pierdzioch, C., and Wohar, M.E. (2017). Do terror attacks predict gold returns? Evidence from a quantile-predictive-regression approach. Quarterly Review of Economics and Finance, 65(C), 276–284.

Gürgün, G. and Ünalmis, I. (2014) Is gold a safe haven against equity market investment in emerging and developing countries? Finance Research Letters, 11, 341–348.

Harvey, D., Leybourne, S., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. International Journal of Forecasting, 13(2), 281–291.

Hassani, H., Silva, E.S., Gupta, R., and Segnon, M.K. (2015). Forecasting the price of gold. Applied Economics, 47(39), 4141–4152.

Hastie, T., Tibshirani, R., Friedman, J. (2009) The elements of statistical learning: Data mining, inference, and prediction, 2nd ed.: Springer: New York, NY, USA.

Hollstein, F., Prokopczuk, J., Tharann, B., and Wese Simen, C. (2021). Predictability in commodity markets: Evidence from more than a century. Journal of Commodity Markets, 24(C), 100171.

Huang, J., Li, Y., Suleman, M.T, and Zhang, H. (2021). Effects of geopolitical risks on gold market return dynamics: Evidence from a nonparametric causality-in-quantiles approach. Defence and Peace Economics. DOI: https://doi.org/10.1080/10242694.2021.2007333.

Huynh, T.L.D. (2020). The effect of uncertainty on the precious metals market: New insights from Transfer Entropy and Neural Network VAR. Resources Policy, 66, 101623.

Ishwaran H. and Kogalur U.B. (2021). Fast unified random forests for survival, regression, and classification (RF-SRC), R package version 2.12.1.

Karmakar, S., Richter, S., and Wu, W.B., 2021. Simultaneous inference for time-varying models. Journal of Econometrics. DOI: https://doi.org/10.1016/j.jeconom.2021.03.002.

Li, B., Chang, C-P., Chu, Y., and Sui, B. (2020). Oil prices and geopolitical risks: What implications are offered via multi-domain investigations? Energy and Environment, 31(3), 492–516.

Li, Y., Huang, J., and Chen, J. (2021). Dynamic spillovers of geopolitical risks and gold prices: New evidence from 18 emerging economies. Resources Policy, 70(C), 101938.

Liu, J., Ma, F., Tang, Y., and Zhang, Y. (2019). Geopolitical risk and oil volatility: A new insight. Energy Economics, 84, 104548.

Low, R.K.Y., Yao, Y., and Faff, R. (2016). Diamonds vs. precious metals: What shines brightest in your investment portfolio?, International Review of Financial Analysis, 43, 1–14.

Luo, J., Demirer, R., Gupta, R., and Ji, Q. (2022). Forecasting oil and gold volatilities with sentiment indicators under structural breaks. Energy Economics, 105, 105751.

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

Mei, D., Ma, F., Liao, Y., and Wang, L. (2020). Geopolitical risk uncertainty and oil future volatility: Evidence from MIDAS models. Energy Economics, 86, 104624.

Nguyen, D.B.B., Prokopczuk, M., and Wese Simen, C. (2019). The risk premium of gold. Journal of International Money Finance, 94, 140–159.

Parisi, A., Parisi, F., and Díaz, D. (2008). Forecasting gold price changes: rolling and recursive neural network models. Journal of Multinational Financial Management, 18, 477–487.

Plakandaras, V., Gupta, R., and Wong, W-K. (2019). Point and density forecasts of oil returns: The role of geopolitical risks. Resources Policy, 62(C), 580–587.

Pierdzioch, C., and Gupta, R. (2020). Uncertainty and forecasts of U.S. recessions. Studies in Nonlinear Dynamics and Econometrics, 24(4), 20180083.

Pierdzioch, C., and Risse, M. (2020). Forecasting precious metal returns with multivariate random forests. Empirical Economics, 58(3), 1167–1184.

Pierdzioch, C., Risse, M. and Rohloff, S. (2014a). 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. and Rohloff, S. (2014b). The international business cycle and gold-price fluctuations. Quarterly Review of Economics and Finance, 54(2), 292–305.

Pierdzioch, C., Risse, M. and Rohloff, S. (2015a). A real-time quantile-regression approach to forecasting gold returns under asymmetric loss. Resources Policy, 45(C), 299–306.

Pierdzioch, C., Risse, M. and Rohloff, S. (2015b). Forecasting gold-price fluctuations: A real-time boosting approach. Applied Economics Letters, 22(1), 46–50.

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

Pierdzioch, C., Risse, M., and Rohloff, S. (2016b). A boosting approach to forecasting the volatility of gold-price fluctuations under flexible loss. Resources Policy, 47, 95–107.

Piffer, M., and Podstawski, M. (2017). Identifying uncertainty shocks using the price of gold. Economic Journal, 128(616), 3266–3284.

Politis, D.N. (2015). The model-free prediction principle, in: Model-free prediction and regression. Springer, 13–30.

Qin, Y., Hong, K., Chen, J., and Zhang, Z. (2020). Asymmetric effects of geopolitical risks on energy returns and volatility under different market conditions. Energy Economics, 90, 104851.

Rapach, D.E., and Zhou, G. (2013). Forecasting stock returns. In: Elliott, G., Timmermann, A. (Eds.), Handbook of Economic Forecasting, vol. 2A. Elsevier, Amsterdam, 328–383.

Reboredo, J.C. (2013a). Is gold a safe haven or a hedge for the US dollar? Implications for risk management. Journal of Banking and Finance, 37(8), 2665–2676.

Reboredo, J.C. (2013b). Is gold a hedge or safe haven against oil price movements? Resources Policy 38(2), 130–137.

Salisu, A.A., Cunado, J., and Gupta, R. (2022). Geopolitical risks and historical exchange rate volatility of the BRICS. International Review of Economics and Finance, 77, 179–190.

Salisu, A.A., Gupta, Bouri, E. and Ji, Q. (2020). The role of global economic conditions in forecasting gold market volatility: Evidence from a GARCH-MIDAS approach. Research in International Business and Finance, 54, 101308.

Salisu, A.A., Gupta, R., Karmakar, S., and Das, S. (2021c). Forecasting output growth of advanced economies over eight centuries: The role of gold market volatility as a proxy of global uncertainty. Resources Policy, 75(C), 102527.

Salisu, A.A., Lukman, L., Tchankam, J.P. (2021a). Historical geopolitical risk and the behaviour of stock returns in advanced economies. European Journal of Finance. DOI: https://doi.org/10.1080/1351847X.2021.1968467.

Salisu, A.A., Pierdzioch, C., and Gupta, R. (2021b). Geopolitical risk and forecastability of tail risk in the oil market: Evidence from over a century of monthly data. Energy, 235, 121333.

Sharma, S.S. (2016). Can consumer price index predict gold price returns?, Economic Modelling, 55, 269–278.

Smales, L.A. (2021). Geopolitical risk and volatility spillovers in oil and stock markets. Quarterly Review of Economics and Finance Volume, 80, 358–366

Tibshirani, R. (1996) Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society, Series B, 58, 267–288.

Tibshirani, J., Athey, S. E. Sverdrup, and Wager, S. (2021). grf: Generalized Random Forests. R package version 2.0.2. https://CRAN.R-project.org/package=grf.

Tiwari, A.K., Aye, G.C., Gupta, R., and Gkillas, K. (2020). Gold-oil dependence dynamics and the role of geopolitical risks: Evidence from a Markov-switching time-varying copula model. Energy Economics, 88, 104748.

Tiwari, A.K., Boachie, M.K., Suleman, M.T., and Gupta, R. (2021). Structure dependence between oil and agricultural commodities returns: The role of geopolitical risks. Energy, 219, 119584.

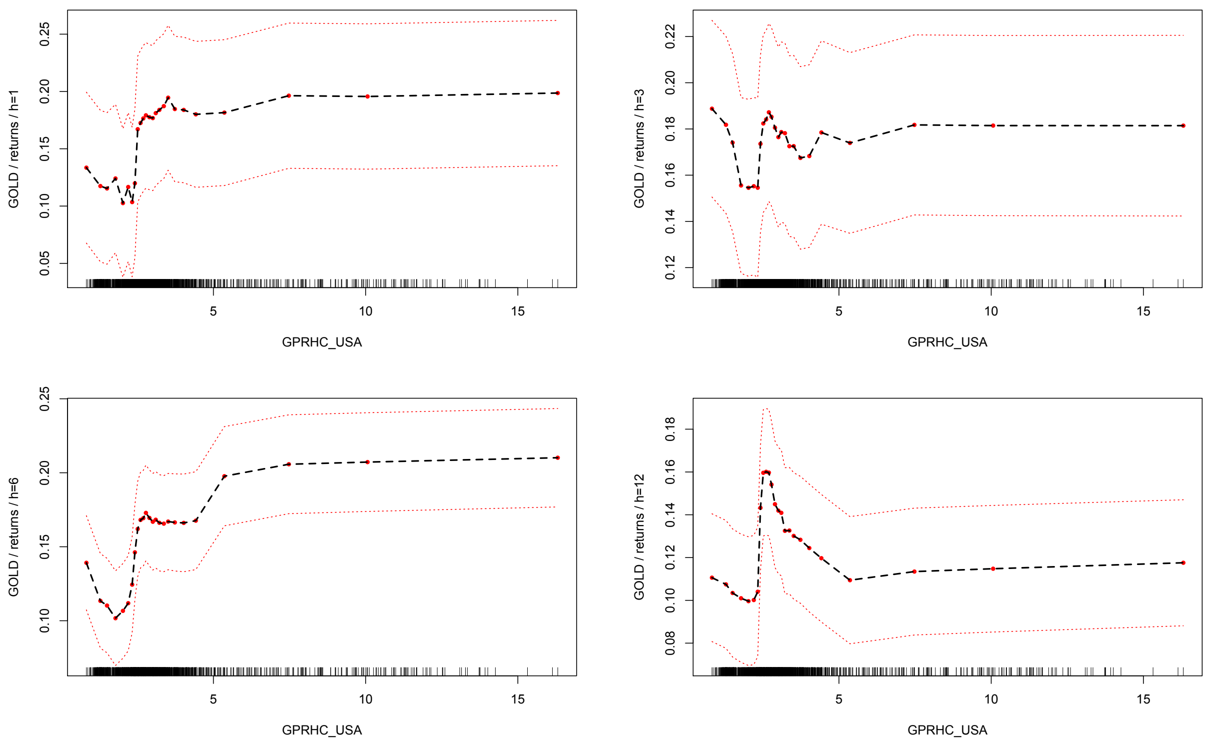
Wu, K., and Karmakar, S. (2021a). Model-free time-aggregated predictions for econometric datasets. Forecasting, 3(4), 920–933.

Wu, K., and Karmakar, S. (2021b). A new model-free prediction method: GA-NoVaS. DOI: https://arxiv.org/abs/2112.08601.

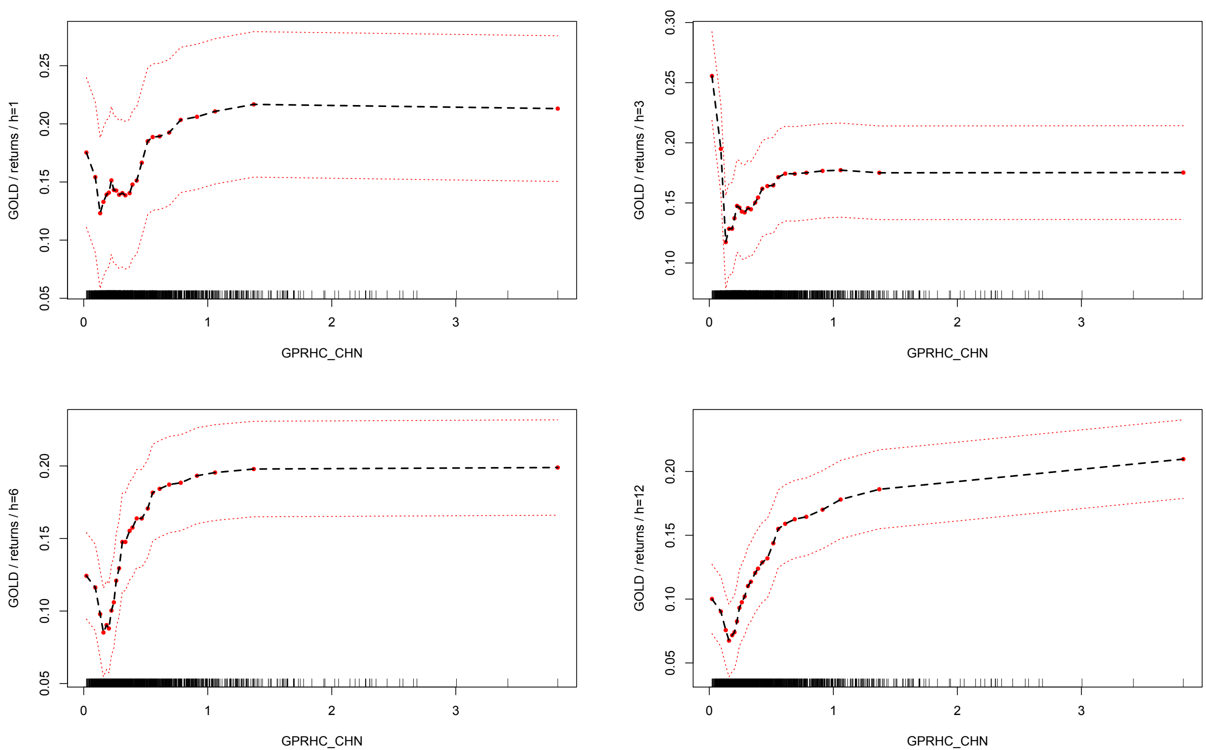
Yang, M., Zhang, Q., Yi, A., and Peng, P. (2021). Geopolitical risk and stock market volatility in emerging economies: Evidence from GARCH-MIDAS model. Discrete Dynamics in Nature and Society, 2021, 1159358.

Figure 1: Examples of Partial Dependence Functions for Returns

Panel A: U.S.



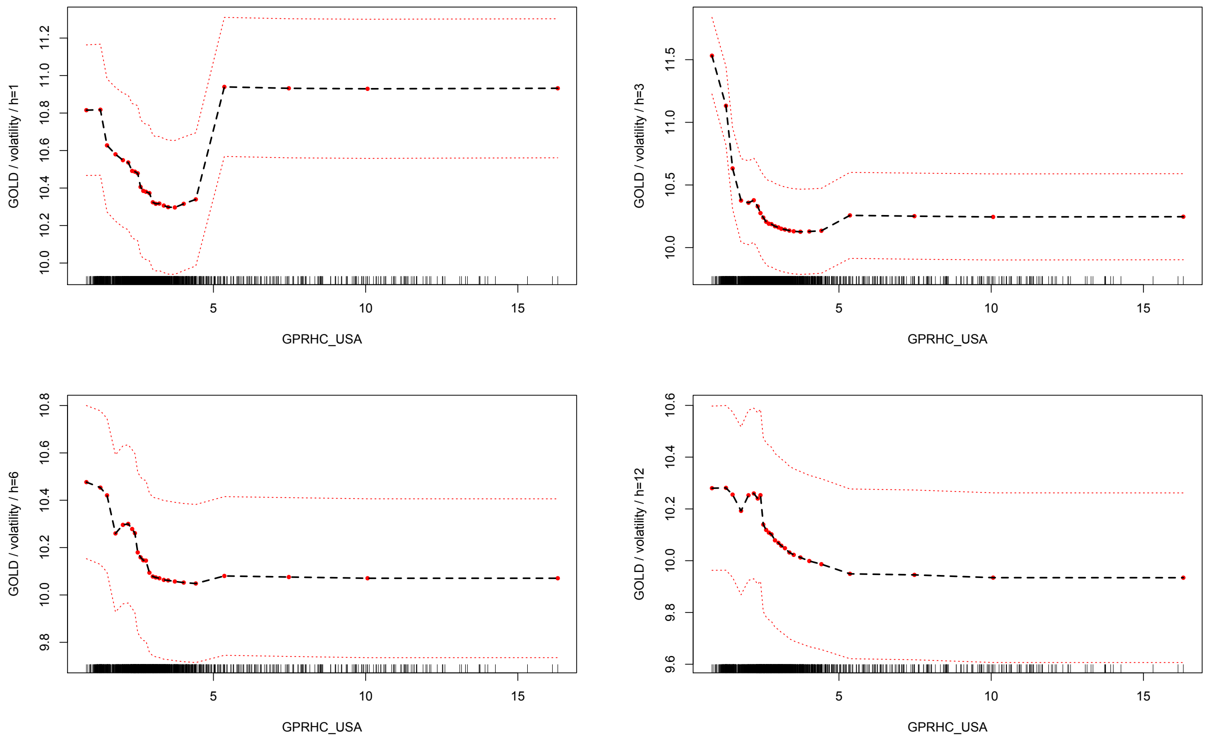
Panel B: China



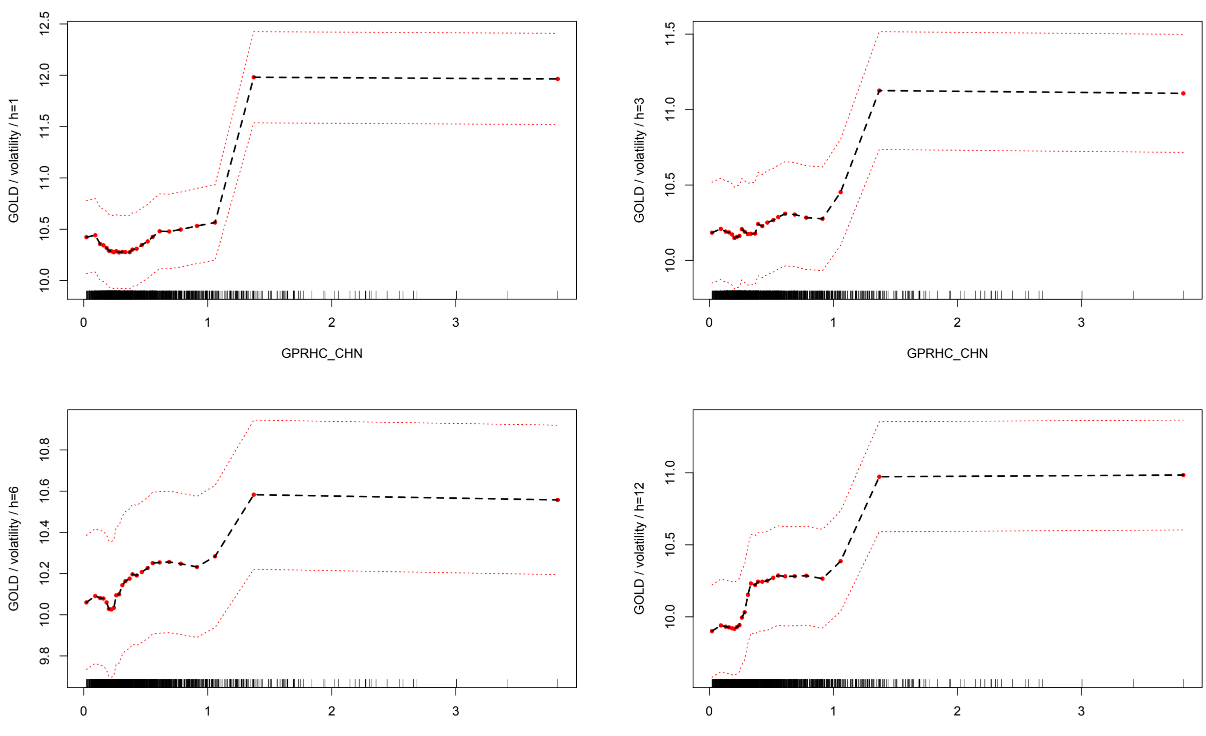
Note: The partial dependence functions are estimated based on the full sample of data. Red points/black dashed lines: partial values. Dashed red lines: smoothed error band (plus/minus two standard errors).

Figure 2: Examples of Partial Dependence Functions for Volatility

Panel A: U.S.



Panel B: China



Note: The partial dependence functions are estimated based on the full sample of data. Red points/black dashed lines: partial values. Dashed red lines: smoothed error band (plus/minus two standard errors).

Table 1: Results for RMSFE Ratios for Gold (Recursive)

Panel A: Returns

Benchmark / rival model	W	h=1	h=3	h=6	h=12
AR/ AR+ GPR	60	1.0035	1.0075	1.0114	1.0178
AR/ AR+ GPT and GPA	60	1.0042	1.0086	1.0127	1.0220
AR/ AR+ GPC (Lasso)	60	1.0022	1.0078	1.0198	1.0386
AR/ AR+ GPC (RF)	60	1.0047	1.0310	1.0500	1.1318
AR+ GPR / AR+ GPT and GPA	60	1.0007	1.0011	1.0013	1.0041
AR+ share / AR+ GPC (Lasso)	60	–	1.0002	1.0083	1.0204
AR+ share / AR+ GPC (RF)	60	1.0012	1.0233	1.0381	1.1120
AR/ AR+ GPR	120	1.0031	1.0061	1.0092	1.0131
AR/ AR+ GPT and GPA	120	1.0038	1.0070	1.0103	1.0173
AR/ AR+ GPC (Lasso)	120	1.0023	1.0066	1.0180	1.0332
AR/ AR+ GPC (RF)	120	1.0047	1.0297	1.0474	1.1247
AR+ GPR / AR+ GPT and GPA	120	1.0007	1.0009	1.0011	1.0041
AR+ share / AR+ GPC (Lasso)	120	–	1.0005	1.0087	1.0198
AR+ share / AR+ GPC (RF)	120	1.0016	1.0235	1.0378	1.1102

Panel B: Volatility

Benchmark / rival model	W	h=1	h=3	h=6	h=12
AR/ AR+ GPR	60	1.0047	1.0115	1.0165	1.0217
AR/ AR+ GPT and GPA	60	1.0045	1.0110	1.0151	1.0195
AR/ AR+ GPC (Lasso)	60	1.0187	1.0528	1.0962	1.1414
AR/ AR+ GPC (RF)	60	1.1239	1.2665	1.4237	1.6933
AR+ GPR / AR+ GPT and GPA	60	–	–	–	–
AR+ share / AR+ GPC (Lasso)	60	1.0139	1.0408	1.0785	1.1171
AR+ share / AR+ GPC (RF)	60	1.1186	1.2522	1.4006	1.6573
AR/ AR+ GPR	120	1.0047	1.0115	1.0164	1.0216
AR/ AR+ GPT and GPA	120	1.0045	1.0110	1.0151	1.0195
AR/ AR+ GPC (Lasso)	120	1.0187	1.0528	1.0962	1.1413
AR/ AR+ GPC (RF)	120	1.1239	1.2665	1.4236	1.6932
AR+ GPR / AR+ GPT and GPA	120	–	–	–	–
AR+ share / AR+ GPC (Lasso)	120	1.0139	1.0408	1.0785	1.1172
AR+ share / AR+ GPC (RF)	120	1.1186	1.2522	1.4007	1.6574

Note: A root-mean-squared-forecast-error (RMSFE) ratio exceeding unity indicates that the rival model produces a smaller RMSFE than the benchmark model. W = Training window. h = Forecast horizon. Only RMSFE ratios larger than unity are reported.

Table 2: Results for RMSFE Ratios for Gold (Rolling)

Panel A: Returns

Benchmark / rival model	W	h=1	h=3	h=6	h=12
AR/ AR+ GPR	60	–	–	–	–
AR/ AR+ GPT and GPA	60	–	–	–	–
AR/ AR+ GPC (Lasso)	60	–	–	–	1.1517
AR/ AR+ GPC (RF)	60	–	1.0409	1.0952	1.2182
AR+ GPR / AR+ GPT and GPA	60	–	–	1.0024	1.0093
AR+ share / AR+ GPC (Lasso)	60	–	1.0177	–	1.1650
AR+ share / AR+ GPC (RF)	60	1.0077	1.0735	1.1232	1.2323
AR/ AR+ GPR	120	–	–	–	–
AR/ AR+ GPT and GPA	120	–	–	–	1.0038
AR/ AR+ GPC (Lasso)	120	–	–	–	–
AR/ AR+ GPC (RF)	120	–	1.0272	1.0559	1.1396
AR+ GPR / AR+ GPT and GPA	120	–	–	1.0109	1.0167
AR+ share / AR+ GPC (Lasso)	120	–	–	–	1.0017
AR+ share / AR+ GPC (RF)	120	1.0034	1.0442	1.0698	1.1543

Panel B: Volatility

Benchmark / rival model	W	h=1	h=3	h=6	h=12
AR/ AR+ GPR	60	1.1246	1.1118	1.0798	1.0981
AR/ AR+ GPT and GPA	60	1.0733	–	1.1259	1.1434
AR/ AR+ GPC (Lasso)	60	1.3770	1.3326	1.3165	–
AR/ AR+ GPC (RF)	60	1.5142	1.4557	1.7732	1.8995
AR+ GPR / AR+ GPT and GPA	60	–	–	1.0427	1.0412
AR+ share / AR+ GPC (Lasso)	60	1.2245	1.1986	1.2193	–
AR+ share / AR+ GPC (RF)	60	1.3465	1.3093	1.6422	1.7298
AR/ AR+ GPR	120	–	–	–	1.0042
AR/ AR+ GPT and GPA	120	1.0036	1.0471	1.0599	1.0628
AR/ AR+ GPC (Lasso)	120	1.0969	1.0177	–	1.0782
AR/ AR+ GPC (RF)	120	1.1103	1.1279	1.2133	1.3264
AR+ GPR / AR+ GPT and GPA	120	1.0499	1.0684	1.0666	1.0584
AR+ share / AR+ GPC (Lasso)	120	1.1475	1.0385	–	1.0737
AR+ share / AR+ GPC (RF)	120	1.1616	1.1509	1.2209	1.3209

Note: A root-mean-squared-forecast-error (RMSFE) ratio exceeding unity indicates that the rival model produces a smaller RMSFE than the benchmark model. W = Rolling window. h = Forecast horizon. Only RMSFE ratios larger than unity are reported.

Table 3: Results of the Clark-West Test for Gold (Recursive)

Panel A: Returns

Benchmark / rival model	W	h=1	h=3	h=6	h=12
AR/ AR+ GPR	60	0.0030	0.0009	0.0014	0.0000
AR/ AR+ GPT and GPA	60	0.0011	0.0004	0.0004	0.0000
AR/ AR+ GPC (Lasso)	60	0.0378	0.0341	0.0004	0.0000
AR/ AR+ GPC (RF)	60	0.0011	0.0001	0.0001	0.0000
AR+ GPR / AR+ GPT and GPA	60	0.0498	–	–	0.0249
AR+ share / AR+ GPC (Lasso)	60	–	–	0.0035	0.0000
AR+ share / AR+ GPC (RF)	60	0.0040	0.0003	0.0010	0.0000
AR/ AR+ GPR	120	0.0078	0.0042	0.0055	0.0001
AR/ AR+ GPT and GPA	120	0.0029	0.0027	0.0025	0.0000
AR/ AR+ GPC (Lasso)	120	0.0408	0.0455	0.0006	0.0000
AR/ AR+ GPC (RF)	120	0.0013	0.0001	0.0003	0.0000
AR+ GPR / AR+ GPT and GPA	120	0.0480	–	–	0.0271
AR+ share / AR+ GPC (Lasso)	120	–	–	0.0035	0.0000
AR+ share / AR+ GPC (RF)	120	0.0037	0.0003	0.0012	0.0000

Panel B: Volatility

Benchmark / rival model	W	h=1	h=3	h=6	h=12
AR/ AR+ GPR	60	0.0000	0.0000	0.0000	0.0000
AR/ AR+ GPT and GPA	60	0.0001	0.0000	0.0000	0.0000
AR/ AR+ GPC (Lasso)	60	0.0145	0.0010	0.0000	0.0000
AR/ AR+ GPC (RF)	60	0.0001	0.0000	00.000	00.000
AR+ GPR / AR+ GPT and GPA	60	–	–	–	–
AR+ share / AR+ GPC (Lasso)	60	0.0203	0.0019	0.0000	0.0000
AR+ share / AR+ GPC (RF)	60	0.0002	0.0000	0.0000	0.0000
AR/ AR+ GPR	120	0.0000	0.0000	0.0000	0.0000
AR/ AR+ GPT and GPA	120	0.0001	0.0000	0.0000	0.0000
AR/ AR+ GPC (Lasso)	120	0.0144	0.0010	0.0000	0.0000
AR/ AR+ GPC (RF)	120	0.0001	0.0000	0.0000	0.0000
AR+ GPR / AR+ GPT and GPA	120	–	–	–	–
AR+ share / AR+ GPC (Lasso)	120	0.0203	0.0018	0.0000	0.0000
AR+ share / AR+ GPC (RF)	120	0.0002	0.0000	0.0000	0.0000

Note: Results (p-values) of the Clark-West test for an equal (adjusted) mean-squared-prediction error (MSPE) for alternative forecast horizons. The alternative hypothesis is that the rival model has a smaller MSPE than the benchmark model. Results are based on robust standard errors. W = Training window. h = Forecast horizon. Only p-values smaller than 10% are reported.

Table 4: Results of the Clark-West Test for Gold (Rolling)

Panel A: Returns

Benchmark / rival model	W	h=1	h=3	h=6	h=12
AR/ AR+ GPR	60	–	–	–	0.0515
AR/ AR+ GPT and GPA	60	–	–	0.0374	0.0021
AR/ AR+ GPC (Lasso)	60	–	0.0007	0.0012	0.0000
AR/ AR+ GPC (RF)	60	0.0023	0.0001	0.0000	0.0000
AR+ GPR / AR+ GPT and GPA	60	–	–	0.0435	0.0333
AR+ share / AR+ GPC (Lasso)	60	0.0205	0.0010	0.0004	0.0000
AR+ share / AR+ GPC (RF)	60	0.0029	0.0036	0.0000	0.0000
AR/ AR+ GPR	120	–	–	–	–
AR/ AR+ GPT and GPA	120	–	–	0.0453	0.0045
AR/ AR+ GPC (Lasso)	120	–	0.0019	0.0008	0.0000
AR/ AR+ GPC (RF)	120	0.0015	0.0000	0.0000	0.0000
AR+ GPR / AR+ GPT and GPA	120	–	–	0.0382	0.0172
AR+ share / AR+ GPC (Lasso)	120	0.0361	0.0021	0.0026	0.0000
AR+ share / AR+ GPC (RF)	120	0.0008	0.0003	0.0000	0.0000

Panel B: Volatility

Benchmark / rival model	W	h=1	h=3	h=6	h=12
AR/ AR+ GPR	60	–	–	–	0.0917
AR/ AR+ GPT and GPA	60	–	–	–	0.0710
AR/ AR+ GPC (Lasso)	60	–	0.0804	0.0603	–
AR/ AR+ GPC (RF)	60	–	0.0724	0.0755	0.0520
AR+ GPR / AR+ GPT and GPA	60	–	–	–	0.0297
AR+ share / AR+ GPC (Lasso)	60	0.0275	0.0694	0.0429	–
AR+ share / AR+ GPC (RF)	60	0.0405	0.0584	0.0623	0.0593
AR/ AR+ GPR	120	–	–	0.0519	0.0008
AR/ AR+ GPT and GPA	120	–	–	0.0509	0.0049
AR/ AR+ GPC (Lasso)	120	–	–	0.0258	0.0090
AR/ AR+ GPC (RF)	120	–	0.0583	0.0396	0.0088
AR+ GPR / AR+ GPT and GPA	120	–	–	–	0.0580
AR+ share / AR+ GPC (Lasso)	120	–	–	0.0584	0.0194
AR+ share / AR+ GPC (RF)	120	–	0.0830	0.0577	0.0169

Note: Results (p-values) of the Clark-West test for an equal (adjusted) mean-squared-prediction error (MSPE) for alternative forecast horizons. The alternative hypothesis is that the rival model has a smaller MSPE than the benchmark model. Results are based on robust standard errors. W = Rolling window. h = Forecast horizon. Only p-values smaller than 10% are reported.

Table 5: Results of the Modified Diebod-Mariano Test for Gold (Recursive)

Panel A: Returns

Benchmark / rival model	W	h=1	h=3	h=6	h=12
AR/ AR+ GPR	60	0.0452	0.0749	–	–
AR/ AR+ GPT and GPA	60	0.0353	0.0745	–	0.0808
AR/ AR+ GPC (Lasso)	60	–	–	–	–
AR/ AR+ GPC (RF)	60	–	0.0689	0.0704	0.0168
AR+ GPR / AR+ GPT and GPA	60	–	–	–	–
AR+ share / AR+ GPC (Lasso)	60	–	–	–	–
AR+ share / AR+ GPC (RF)	60	–	0.0957	0.0925	0.0178
AR/ AR+ GPR	120	0.0667	–	–	–
AR/ AR+ GPT and GPA	120	0.0500	–	–	–
AR/ AR+ GPC (Lasso)	120	–	–	–	–
AR/ AR+ GPC (RF)	120	–	0.0789	0.0830	0.0228
AR+ GPR / AR+ GPT and GPA	120	–	–	–	–
AR+ share / AR+ GPC (Lasso)	120	–	–	–	–
AR+ share / AR+ GPC (RF)	120	–	0.0956	0.0965	0.0203

Panel B: Volatility

Benchmark / rival model	W	h=1	h=3	h=6	h=12
AR/ AR+ GPR	60	0.0000	0.0000	0.0000	0.000
AR/ AR+ GPT and GPA	60	0.0000	0.0000	0.0004	0.0017
AR/ AR+ GPC (Lasso)	60	–	–	0.0701	0.0722
AR/ AR+ GPC (RF)	60	0.0001	0.0000	0.0000	0.0006
AR+ GPR / AR+ GPT and GPA	60	–	–	–	–
AR+ share / AR+ GPC (Lasso)	60	–	–	–	–
AR+ share / AR+ GPC (RF)	60	0.0002	0.0000	0.0000	0.0008
AR/ AR+ GPR	120	0.0000	0.0000	0.0000	0.0000
AR/ AR+ GPT and GPA	120	0.0000	0.0000	0.0004	0.0017
AR/ AR+ GPC (Lasso)	120	–	–	0.0702	0.0722
AR/ AR+ GPC (RF)	120	0.0001	0.0000	0.0000	0.0005
AR+ GPR / AR+ GPT and GPA	120	–	–	–	–
AR+ share / AR+ GPC (Lasso)	120	–	–	–	–
AR+ share / AR+ GPC (RF)	120	0.0002	0.0000	0.0000	0.0008

Note: Results (p-values) of the modified Diebold-Mariano test for equal accuracy of forecasts for alternative forecast horizons. The alternative hypothesis is that the rival model performs better than the benchmark model. W = Training window. h = Forecast horizon. Only p-values smaller than 10% are reported.

Table 6: Results of the Modified Diebod-Mariano Test for Gold (Rolling)

Panel A: Returns

Benchmark / rival model	W	h=1	h=3	h=6	h=12
AR/ AR+ GPR	60	–	–	–	–
AR/ AR+ GPT and GPA	60	–	–	–	–
AR/ AR+ GPC (Lasso)	60	–	–	–	0.0149
AR/ AR+ GPC (RF)	60	–	0.0455	0.0079	0.0043
AR+ GPR / AR+ GPT and GPA	60	–	–	–	–
AR+ share / AR+ GPC (Lasso)	60	–	–	–	0.0112
AR+ share / AR+ GPC (RF)	60	–	0.0339	0.0012	0.0037
AR/ AR+ GPR	120	–	–	–	–
AR/ AR+ GPT and GPA	120	–	–	–	–
AR/ AR+ GPC (Lasso)	120	–	–	–	–
AR/ AR+ GPC (RF)	120	–	0.0372	0.0156	7e-04
AR+ GPR / AR+ GPT and GPA	120	–	–	–	–
AR+ share / AR+ GPC (Lasso)	120	–	–	–	–
AR+ share / AR+ GPC (RF)	120	–	0.0198	0.0020	0.0004

Panel B: Volatility

Benchmark / rival model	W	h=1	h=3	h=6	h=12
AR/ AR+ GPR	60	–	–	–	–
AR/ AR+ GPT and GPA	60	–	–	–	–
AR/ AR+ GPC (Lasso)	60	–	–	–	–
AR/ AR+ GPC (RF)	60	–	0.0793	0.0777	0.0527
AR+ GPR / AR+ GPT and GPA	60	–	–	–	0.0811
AR+ share / AR+ GPC (Lasso)	60	0.0827	–	–	–
AR+ share / AR+ GPC (RF)	60	0.0428	0.0615	0.0624	0.0583
AR/ AR+ GPR	120	–	–	–	–
AR/ AR+ GPT and GPA	120	–	–	–	–
AR/ AR+ GPC (Lasso)	120	–	–	–	–
AR/ AR+ GPC (RF)	120	–	–	0.0638	0.0342
AR+ GPR / AR+ GPT and GPA	120	–	–	–	–
AR+ share / AR+ GPC (Lasso)	120	–	–	–	–
AR+ share / AR+ GPC (RF)	120	–	–	0.0774	0.0411

Note: Results (p-values) of the modified Diebold-Mariano test for equal accuracy of forecasts for alternative forecast horizons. The alternative hypothesis is that the rival model performs better than the benchmark model. W = Rolling window. h = Forecast horizon. Only p-values smaller than 10% are reported.

Table 7: Comparing the Lasso and Random Forests for Gold

Panel A: Returns

Test (window type)	W	h=1	h=3	h=6	h=12
RMSFE ratio (recursive)	60	1.0025	1.0231	1.0296	1.0898
RMSFE ratio (rolling)	60	1.0337	1.0548	1.1262	1.0577
Clark-West test (recursive)	60	0.0088	0.0026	0.0000	0.0000
Clark-West test (rolling)	60	0.0110	0.0000	0.0001	0.0000
Modified Diebod-Mariano test (recursive)	60	–	–	0.0621	0.0005
Modified Diebod-Mariano test (rolling)	60	0.0627	0.0034	0.0022	0.0049
RMSFE ratio (recursive)	120	1.0024	1.0230	1.0288	1.0886
RMSFE ratio (rolling)	120	1.0477	1.0914	1.1982	1.1523
Clark-West test (recursive)	120	0.0096	0.0030	0.0000	0.0000
Clark-West test (rolling)	120	0.0289	0.0237	0.0771	0.0139
Modified Diebod-Mariano test (recursive)	120	–	–	0.0685	0.0006
Modified Diebod-Mariano test (rolling)	120	0.0339	0.0295	0.0988	0.0528

Panel B: Volatility

Test (window type)	W	h=1	h=3	h=6	h=12
RMSFE ratio (recursive)	60	1.1033	1.2030	1.2987	1.4836
RMSFE ratio (rolling)	60	1.0997	1.0924	1.3469	1.9092
Clark-West test (recursive)	60	0.0000	0.0006	0.0000	0.0000
Clark-West test (rolling)	60	0.0019	0.0125	0.0332	0.0656
Modified Diebod-Mariano test (recursive)	60	0.0001	0.0015	0.0000	0.0002
Modified Diebod-Mariano test (rolling)	60	0.0011	–	0.0416	0.0738
RMSFE ratio (recursive)	120	1.1033	1.2030	1.2987	1.4836
RMSFE ratio (rolling)	120	1.0123	1.1083	1.2976	1.2302
Clark-West test (recursive)	120	0.0000	0.0006	0.0000	0.0000
Clark-West test (rolling)	120	0.0002	0.0004	0.0074	0.0001
Modified Diebod-Mariano test (recursive)	120	0.0001	0.0015	0.0000	0.0002
Modified Diebod-Mariano test (rolling)	120	–	0.0058	0.0135	0.0090

Note: Benchmark model: AR+ GPC (Lasso). Rival model: AR+ GPC (RF). A root-mean-squared-forecast-error (RMSFE) ratio exceeding unity indicates that the rival model produces a smaller RMSFE than the benchmark model. Only RMSFE ratios larger than unity are reported. Results (p-values, robust standard errors) of the Clark-West test for an equal (adjusted) mean-squared-prediction error (MSPE) for alternative forecast horizons. The alternative hypothesis is that the rival model has a smaller MSPE than the benchmark model. Results (p-values) of the modified Diebold-Mariano test for equal accuracy of forecasts for alternative forecast horizons. The alternative hypothesis is that the rival model performs better than the benchmark model. W = Training window (recursive) / rolling window (rolling). h = Forecast horizon. Only p-values smaller than 10% are reported.

Table 8: Results for AUROC Ratios for Gold Returns

Panel A: Recursive

Benchmark / rival model	W	h=1	h=3	h=6	h=12
AR / AR + GPR	60	1.0657	1.0936	1.0519	1.1056
AR / AR + GPT and GPA	60	1.0892	1.1317	1.0978	1.1456
AR / AR + GPC (Lasso)	60	1.0663	1.1062	1.1606	1.1873
AR / AR + GPC (RF)	60	1.0866	1.1465	1.1752	1.2395
AR + GPR / AR + GPT and GPA	60	1.0220	1.0348	1.0436	1.0362
AR + share / AR + GPC (Lasso)	60	1.0005	1.0115	1.1034	1.0738
AR + share / AR + GPC (RF)	60	1.0196	1.0484	1.1172	1.1211
AR / AR + GPR	120	1.0536	1.0604	1.0425	1.0922
AR / AR + GPT and GPA	120	1.0778	1.0959	1.0901	1.1333
AR / AR + GPC (Lasso)	120	1.0637	1.0802	1.1566	1.1691
AR / AR + GPC (RF)	120	1.0841	1.1115	1.1619	1.2129
AR + GPR / AR + GPT and GPA	120	1.0230	1.0334	1.0456	1.0376
AR + share / AR + GPC (Lasso)	120	1.0096	1.0187	1.1094	1.0704
AR + share / AR + GPC (RF)	120	1.0290	1.0481	1.1145	1.1105

Panel B: Rolling

Benchmark / rival model	W	h=1	h=3	h=6	h=12
AR / AR + GPR	60	1.0187	1.0657	1.0578	1.0603
AR / AR + GPT and GPA	60	1.0189	1.0596	1.0444	1.0592
AR / AR + GPC (Lasso)	60	–	1.0570	1.1024	1.1723
AR / AR + GPC (RF)	60	1.0599	1.1026	1.1567	1.1816
AR + GPR / AR + GPT and GPA	60	1.0002	–	–	–
AR + share / AR + GPC (Lasso)	60	–	–	1.0422	1.1056
AR + share / AR + GPC (RF)	60	1.0405	1.0346	1.0936	1.1144
AR / AR + GPR	120	1.0653	1.0631	1.0699	1.0797
AR / AR + GPT and GPA	120	1.0174	1.0530	1.0687	1.0865
AR / AR + GPC (Lasso)	120	–	1.0494	1.1287	1.1724
AR / AR + GPC (RF)	120	1.0695	1.1064	1.1797	1.2098
AR + GPR / AR + GPT and GPA	120	–	–	–	1.0063
AR + share / AR + GPC (Lasso)	120	–	–	1.0550	1.0859
AR + share / AR + GPC (RF)	120	1.0040	1.0407	1.1026	1.1206

Note: An AUROC ratio exceeding unity indicates that the rival model produces better forecasts in terms of directional accuracy than the benchmark model. W = Training (rolling) window. h = Forecast horizon. Only AUROC ratios larger than unity are reported.

Table 9: Sub-Periods During Which the Rival Models for Gold Outperform the Benchmark Models (Recursive)

Panel A: Returns

Benchmark / rival model	W	h=1	h=3	h=6	h=12
AR / AR + GPR	60	5	4	4	4
AR / AR + GPT and GPA	60	7	5	4	5
AR / AR + GPC (Lasso)	60	4	5	4	4
AR / AR + GPC (RF)	60	5	6	6	8
AR + GPR / AR + GPT and GPA	60	5	5	6	4
AR + share / AR + GPC (Lasso)	60	5	4	4	5
AR + share / AR + GPC (RF)	60	4	7	6	9
AR / AR + GPR	120	5	4	4	4
AR / AR + GPT and GPA	120	5	4	4	5
AR / AR + GPC (Lasso)	120	5	4	4	4
AR / AR + GPC (RF)	120	4	6	6	5
AR + GPR / AR + GPT and GPA	120	4	4	5	6
AR + share / AR + GPC (Lasso)	120	7	5	6	6
AR + share / AR + GPC (RF)	120	5	7	6	9

Panel B: Volatility

Benchmark / rival model	W	h=1	h=3	h=6	h=12
AR / AR + GPR	60	6	7	7	7
AR / AR + GPT and GPA	60	5	5	6	6
AR / AR + GPC (Lasso)	60	5	5	5	7
AR / AR + GPC (RF)	60	7	9	9	9
AR + GPR / AR + GPT and GPA	60	3	3	4	3
AR + share / AR + GPC (Lasso)	60	3	4	5	7
AR + share / AR + GPC (RF)	60	8	9	9	9
AR / AR + GPR	120	7	7	7	7
AR / AR + GPT and GPA	120	6	6	6	6
AR / AR + GPC (Lasso)	120	4	5	4	5
AR / AR + GPC (RF)	120	8	9	9	9
AR + GPR / AR + GPT and GPA	120	4	4	5	5
AR + share / AR + GPC (Lasso)	120	3	4	5	5
AR + share / AR + GPC (RF)	120	8	9	9	9

Note: The number of sub-periods during which the rival model outperforms the benchmark model in terms of the RMSFE statistic. The total number of sup-periods is 10. W = Training window. h = Forecast horizon.

Table 10: Sub-Periods During Which the Rival Models for Gold Outperform the Benchmark Models (Rolling)

Panel A: Returns

Benchmark / rival model	W	h=1	h=3	h=6	h=12
AR / AR + GPR	60	4	4	6	5
AR / AR + GPT and GPA	60	3	3	3	5
AR / AR + GPC (Lasso)	60	4	5	6	9
AR / AR + GPC (RF)	60	6	9	8	9
AR + GPR / AR + GPT and GPA	60	2	3	4	3
AR + share / AR + GPC (Lasso)	60	4	7	5	10
AR + share / AR + GPC (RF)	60	6	9	10	10
AR / AR + GPR	120	0	4	3	2
AR / AR + GPT and GPA	120	1	3	5	4
AR / AR + GPC (Lasso)	120	4	4	4	5
AR / AR + GPC (RF)	120	5	8	9	9
AR + GPR / AR + GPT and GPA	120	4	4	5	4
AR + share / AR + GPC (Lasso)	120	6	4	4	6
AR + share / AR + GPC (RF)	120	7	8	9	10

Panel B: Volatility

Benchmark / rival model	W	h=1	h=3	h=6	h=12
AR / AR + GPR	60	4	6	5	7
AR / AR + GPT and GPA	60	1	3	4	6
AR / AR + GPC (Lasso)	60	5	4	7	7
AR / AR + GPC (RF)	60	3	8	9	10
AR + GPR / AR + GPT and GPA	60	1	4	3	7
AR + share / AR + GPC (Lasso)	60	6	5	6	6
AR + share / AR + GPC (RF)	60	6	9	10	10
AR / AR + GPR	120	4	4	5	7
AR / AR + GPT and GPA	120	4	5	6	6
AR / AR + GPC (Lasso)	120	4	5	6	7
AR / AR + GPC (RF)	120	6	9	9	9
AR + GPR / AR + GPT and GPA	120	6	7	7	6
AR + share / AR + GPC (Lasso)	120	4	6	6	6
AR + share / AR + GPC (RF)	120	7	10	10	10

Note: The number of sub-periods during which the rival model outperforms the benchmark model in terms of the RMSFE statistic. The total number of sup-periods is 10. W = Rolling window. h = Forecast horizon.

Appendix

Table A1: Results for RMSFE Ratios for Oil (Recursive)

Panel A: Returns

Benchmark/rival model	Window	h=1	h=3	h=6	h=12
AR/ AR+ GPR	60	–	–	–	–
AR/ AR+ GPT and GPA	60	–	–	–	–
AR/ AR+ GPC (Lasso)	60	–	–	–	1.0068
AR/ AR+ GPC (RF)	60	–	–	1.0173	1.0561
AR+ GPR / AR+ GPT and GPA	60	–	–	–	1.0010
AR+ share / AR+ GPC (Lasso)	60	–	–	1.0004	1.0138
AR+ share / AR+ GPC (RF)	60	–	–	1.0235	1.0634
AR/ AR+ GPR	120	–	–	–	–
AR/ AR+ GPT and GPA	120	–	–	–	–
AR/ AR+ GPC (Lasso)	120	–	–	–	1.0080
AR/ AR+ GPC (RF)	120	–	–	1.0192	1.0589
AR+ GPR / AR+ GPT and GPA	120	–	–	–	1.0011
AR+ share / AR+ GPC (Lasso)	120	–	–	–	1.0120
AR+ share / AR+ GPC (RF)	120	–	–	1.0235	1.0630

Panel B: Volatility

Benchmark/rival model	Window	h=1	h=3	h=6	h=12
AR/ AR+ GPR	60	–	1.0002	1.0018	1.0056
AR/ AR+ GPT and GPA	60	–	–	1.0009	1.0043
AR/ AR+ GPC (Lasso)	60	–	–	1.0236	1.0572
AR/ AR+ GPC (RF)	60	–	–	1.0543	1.1844
AR+ GPR / AR+ GPT and GPA	60	–	–	–	–
AR+ share / AR+ GPC (Lasso)	60	–	–	1.0217	1.0513
AR+ share / AR+ GPC (RF)	60	–	–	1.0524	1.1778
AR/ AR+ GPR	120	–	1.0002	1.0018	1.0055
AR/ AR+ GPT and GPA	120	–	–	1.0009	1.0042
AR/ AR+ GPC (Lasso)	120	–	–	1.0235	1.0570
AR/ AR+ GPC (RF)	120	–	–	1.0540	1.1839
AR+ GPR / AR+ GPT and GPA	120	–	–	–	–
AR+ share / AR+ GPC (Lasso)	120	–	–	1.0216	1.0513
AR+ share / AR+ GPC (RF)	120	–	–	1.0521	1.1774

Note: A root-mean-squared-forecast-error (RMSFE) ratio exceeding unity indicates that the rival model produces a smaller RMSFE than the benchmark model. W = Training window.

Table A2: Results for RMSFE Ratios for Silver (Recursive)

Panel A: Returns

Benchmark/rival model	Window	h=1	h=3	h=6	h=12
AR/ AR+ GPR	60	–	–	–	–
AR/ AR+ GPT and GPA	60	–	–	–	–
AR/ AR+ GPC (Lasso)	60	1.0014	–	–	1.0249
AR/ AR+ GPC (RF)	60	1.0012	1.0043	1.0328	1.0870
AR+ GPR / AR+ GPT and GPA	60	–	–	–	1.0010
AR+ share / AR+ GPC (Lasso)	60	1.0026	–	–	1.0286
AR+ share / AR+ GPC (RF)	60	1.0024	1.0054	1.0348	1.0909
AR/ AR+ GPR	120	–	–	1.0000	1.0006
AR/ AR+ GPT and GPA	120	–	–	–	–
AR/ AR+ GPC (Lasso)	120	1.0015	–	–	1.0269
AR/ AR+ GPC (RF)	120	1.0024	1.0062	1.0342	1.0857
AR+ GPR / AR+ GPT and GPA	120	–	–	–	–
AR+ share / AR+ GPC (Lasso)	120	1.0018	–	–	1.0263
AR+ share / AR+ GPC (RF)	120	1.0026	1.0066	1.0343	1.0851

Panel B: Volatility

Benchmark/rival model	Window	h=1	h=3	h=6	h=12
AR/ AR+ GPR	60	1.0006	1.0033	1.0043	1.0037
AR/ AR+ GPT and GPA	60	–	1.0034	1.0052	1.0032
AR/ AR+ GPC (Lasso)	60	–	–	–	1.0100
AR/ AR+ GPC (RF)	60	–	–	–	–
AR+ GPR / AR+ GPT and GPA	60	–	1.0001	1.0009	–
AR+ share / AR+ GPC (Lasso)	60	–	–	–	1.0063
AR+ share / AR+ GPC (RF)	60	–	–	–	–
AR/ AR+ GPR	120	–	1.000	1.0001	1.0008
AR/ AR+ GPT and GPA	120	–	–	1.0001	1.0004
AR/ AR+ GPC (Lasso)	120	–	–	–	1.0067
AR/ AR+ GPC (RF)	120	–	–	–	–
AR+ GPR / AR+ GPT and GPA	120	–	–	1.0000	–
AR+ share / AR+ GPC (Lasso)	120	–	–	–	1.0059
AR+ share / AR+ GPC (RF)	120	–	–	–	–

Note: A root-mean-squared-forecast-error (RMSFE) ratio exceeding unity indicates that the rival model produces a smaller RMSFE than the benchmark model. W = Training window.

Table A3: Results for RMSFE Ratios for Oil (Rolling)

Panel A: Returns

Benchmark/rival model	Window	h=1	h=3	h=6	h=12
AR/ AR+ GPR	60	–	–	–	–
AR/ AR+ GPT and GPA	60	–	–	–	–
AR/ AR+ GPC (Lasso)	60	–	–	–	1.0439
AR/ AR+ GPC (RF)	60	–	1.0183	1.0262	1.1092
AR+ GPR / AR+ GPT and GPA	60	–	–	–	1.0234
AR+ share / AR+ GPC (Lasso)	60	–	–	1.0311	1.0885
AR+ share / AR+ GPC (RF)	60	1.0179	1.0583	1.0640	1.1565
AR/ AR+ GPR	120	–	–	–	1.0114
AR/ AR+ GPT and GPA	120	–	–	–	1.0005
AR/ AR+ GPC (Lasso)	120	–	–	–	1.0461
AR/ AR+ GPC (RF)	120	–	–	1.0237	1.1010
AR+ GPR / AR+ GPT and GPA	120	–	–	–	–
AR+ share / AR+ GPC (Lasso)	120	–	–	–	1.0343
AR+ share / AR+ GPC (RF)	120	–	1.0109	1.0309	1.0886

Panel B: Volatility

Benchmark/rival model	Window	h=1	h=3	h=6	h=12
AR/ AR+ GPR	60	–	1.0031	1.0049	1.0550
AR/ AR+ GPT and GPA	60	–	1.0102	1.0097	1.1002
AR/ AR+ GPC (Lasso)	60	1.3185	1.0604	–	1.2255
AR/ AR+ GPC (RF)	60	1.3417	1.2906	1.3300	1.5559
AR+ GPR / AR+ GPT and GPA	60	–	1.0070	1.0047	1.0428
AR+ share / AR+ GPC (Lasso)	60	1.3246	1.0571	–	1.1616
AR+ share / AR+ GPC (RF)	60	1.3479	1.2866	1.3235	1.4748
AR/ AR+ GPR	120	–	–	–	1.0195
AR/ AR+ GPT and GPA	120	–	–	–	1.0448
AR/ AR+ GPC (Lasso)	120	–	1.0130	–	1.0869
AR/ AR+ GPC (RF)	120	–	1.0570	1.1538	1.3589
AR+ GPR / AR+ GPT and GPA	120	–	1.0065	–	1.0248
AR+ share / AR+ GPC (Lasso)	120	–	1.0209	–	1.0662
AR+ share / AR+ GPC (RF)	120	–	1.0653	1.1635	1.3330

Note: A root-mean-squared-forecast-error (RMSFE) ratio exceeding unity indicates that the rival model produces a smaller RMSFE than the benchmark model. W = Rolling window.

Table A4: Results for RMSFE Ratios for Silver (Rolling)

Benchmark/rival model	Window	h=1	h=3	h=6	h=12
Panel A: Returns					
AR/ AR+ GPR	60	–	–	–	–
AR/ AR+ GPT and GPA	60	–	–	–	–
AR/ AR+ GPC (Lasso)	60	–	–	1.0043	1.0626
AR/ AR+ GPC (RF)	60	–	1.0318	1.0654	1.1806
AR+ GPR / AR+ GPT and GPA	60	–	–	–	–
AR+ share / AR+ GPC (Lasso)	60	1.0127	–	1.0469	1.0903
AR+ share / AR+ GPC (RF)	60	1.0093	1.0748	1.1106	1.2114
AR/ AR+ GPR	120	–	–	–	–
AR/ AR+ GPT and GPA	120	–	–	–	–
AR/ AR+ GPC (Lasso)	120	–	–	–	1.0259
AR/ AR+ GPC (RF)	120	–	1.0192	1.0482	1.1190
AR+ GPR / AR+ GPT and GPA	120	–	–	–	1.0012
AR+ share / AR+ GPC (Lasso)	120	1.0044	–	–	1.0345
AR+ share / AR+ GPC (RF)	120	–	1.0346	1.0622	1.1284
Panel B: Volatility					
AR/ AR+ GPR	60	–	–	–	–
AR/ AR+ GPT and GPA	60	–	–	–	–
AR/ AR+ GPC (Lasso)	60	–	1.0513	1.0468	–
AR/ AR+ GPC (RF)	60	–	–	–	1.0197
AR+ GPR / AR+ GPT and GPA	60	–	–	–	–
AR+ share / AR+ GPC (Lasso)	60	–	1.0722	1.0545	–
AR+ share / AR+ GPC (RF)	60	–	–	–	1.0224
AR/ AR+ GPR	120	–	–	–	–
AR/ AR+ GPT and GPA	120	–	–	1.0080	1.0043
AR/ AR+ GPC (Lasso)	120	–	1.0427	1.0216	1.0036
AR/ AR+ GPC (RF)	120	–	–	–	–
AR+ GPR / AR+ GPT and GPA	120	–	1.0053	1.0164	1.0175
AR+ share / AR+ GPC (Lasso)	120	–	1.0495	1.0301	1.0168
AR+ share / AR+ GPC (RF)	120	–	–	–	–

Note: A root-mean-squared-forecast-error (RMSFE) ratio exceeding unity indicates that the rival model produces a smaller RMSFE than the benchmark model. W = Rolling window.