Erasmus University Rotterdam
Erasmus School of Economics
Bachelor Thesis BSc^2 Econometrics and Economics

Optimal portfolio weights for a gold-and-stock portfolio with random forests for macroeconomic features

Author: Annegien Smilde (493320)
Date: July 2nd 2023
Supervisor: Dr. A. Tetereva
Second assessor: Dr. E.P. O'Neill

Abstract

This research examines whether random forests can predict optimal portfolio weights as functions of macroeconomic variables for a gold-and-stock portfolio. Three portfolios are created, one where the optimal weights for the gold index are modelled independently and a second portfolio for the stock index. The third portfolio models the assets together in terms of utility, where this is modelled as the portfolio variance. The weights are uniquely reconstructed from the forecasted utility to form the third utility. Based on relative variable importance and Accumulated Local Effect (ALE) plots, it becomes clear that the index level of narrow money (M1) and the annual growth in housing prices are most important in determining the stock and gold weights. Narrow money and broad money (M3) are the most important features of the utility model. The portfolios are compared to three benchmark portfolios, a Markowitz portfolio, a minimum variance portfolio, and a portfolio based on the optimization process without the use of random forests. The stock and gold index models offer slightly higher Sharpe ratios than the two simple benchmark portfolios, but they do not outperform these significantly. The portfolio based on the reconstructed weights from utility performs similarly to the benchmark portfolios.

Contents

1 Introduction

It has been an ongoing practice to find optimal investment strategies over time. One of many ways this can be done is by continuously rebalancing portfolio weights. A well-balanced portfolio includes assets that deliver high returns while preventing significant losses. This could be a portfolio with a stock and a bond in its simplest form, yet the relationship between these assets is far from simple. The asset categories are complementary and substitutes (Boucher & Tokpavi, 2019), where the sign of the correlation depends heavily on market conditions. In times of economic growth or volatility shocks, the correlation is negative, while positive with inflation shocks (Ilmanen, 2003). An extreme example of this correlation is the flight to safety in times of economic downturn with a significant decline (rise) in stock (bond) prices. Similar patterns are observed for gold, often considered a safe-haven asset (Ang, 2014). This implies that the short- and long-run relationships between gold and stock indices can vary in different market states (Chirwa & Odhiambo, 2020), such that optimal portfolio weights for these assets also differ over time. Furthermore, Pellegrino (2021) describes how financial returns are linked to macroeconomic fundamentals. Ultimately, the portfolio weights with a maximum Sharpe ratio can be described as a function of macroeconomic variables.

Model estimation and statistical inference for this problem are challenging for various reasons. First, macroeconomic variables are endogenous and highly correlated with one another. It is trivial that without the proper adjustments, biases and multicollinearity occur. While well-studied adjustments such as panel IV estimation can be a solution, IV estimators are complicated to construct, and biases remain hard to avoid. Second, an increasingly large number of factors can be used as explanatory characteristics, causing high dimensionality. Third, due to the high correlation between economic variables, the interaction effects of the explanatory characteristics should be adequately modelled, but the relations usually have non-linearities (Pellegrino, 2021). These difficulties could result in spurious overfitting and model misspecification when conventional cross-sectional regression is used.

A natural solution could be implementing Markov-switching models, which capture the joint distribution of asset returns in various market states. They allow for different relationships between the financial returns in bear and bull markets, are more diversified, and provide a better risk-return relationship compared to traditional mean-variance portfolios (Oliveira & Valls Pereira, 2018). Furthermore, Ang and Bekaert (2004) show how a regime-switching model can be used to ensure dynamic portfolio allocation. They find that switching between different financial assets can add substantial value. However, Markov-switching models have mixed out-of-sample forecasting performances (Guidolin, 2011). And most importantly, regime-switching models only offer solutions as a discrete response variable. Yet, this research paper's optimal portfolio weights are a continuous response variable.

Regression trees can offer an alternative to Markov-switching models and are beneficial for various reasons. First and most importantly, they allow optimal portfolio weights to be continuous response variables. Furthermore, it is a widely-applied machine learning technique with an intuitive economic interpretation, using recursive binary splitting on given criteria (Loh, 2011). In their research, Samitas and Armenatzoglou (2014) conclude that the regression tree model outperforms the Markov regime-switching model. Regression trees are a suitable solution for the problem since they can model the complex and non-linear relationships between financial returns and macroeconomic variables (Carrizosa, Molero-Río & Morales, 2021). The general problem is that regression trees have a high chance of overfitting the data in the training sample and tend to provide an unstable estimation, reducing forecast accuracy. Random forests offer improvements upon these problems while having the same benefits as regression trees do (Breiman, 2001). This paper takes inspiration from Bryzgalova, Pelger and Zhu (2021), who implement the idea of regression trees to build cross-sections of stock returns. Similarly, this paper implements the idea of regression trees to model optimal portfolio weights. More specifically, this research aims to use random forests of macroeconomic variables to find how the save-haven asset gold reacts to various market conditions in relation to a stock index, leading to the following research question:

> *To what extent can random forests predict the optimal portfolio weights as a function of macroeconomic variables for a portfolio with a gold index and a stock index?*

The relevance of this paper is twofold. First, it contributes to the existing academic literature by extending the knowledge of the relationships between financial asset returns, optimal portfolio weights, and macroeconomic factors. Furthermore, it assesses random forests' applicability, interpretability, and predictive performance for these relationships. While there is literature on the effects of market conditions (e.g., business cycles) on asset returns and stock volatility, no existing literature currently takes the regression tree approach to predict optimal portfolio weights in relation to macroeconomic factors. Second, this paper has social relevance. By defining the portfolio weights as a function of market conditions and assessing the predictive performance of regression trees with respect to these portfolios, trading strategies can be developed or improved for practitioners worldwide to keep portfolio allocations optimal.

This research employs two datasets. The first set contains the monthly prices for the S&P GSCI Gold Index and the S&P 500 Stock Index from January 1980 to December 2022, yielding 43 years of observations. Monthly returns are calculated for both indices and are added to the dataset. The second set contains 42 macroeconomic variables from December 1989 to November 2022, yielding 32 years of observations. The first ten years of the portfolio data are used as an estimation sample for expected returns and the sample covariance matrix. The next twenty years of observations from both datasets are used as the training sample for the random forests. The last thirteen years of both datasets form the test sample that assesses the performance of the portfolios created from the random forests.

The monthly returns for the gold and stock indices are used to calculate optimal portfolio weights at every time t in the sample. Optimal weights are defined as weights that maximize the Sharpe ratio net of transaction costs. Following Kazak, Li, Nolte and Nolte (2022), a utility function is constructed, which directly follows from the maximization of the Sharpe ratio. Assuming that the investor is infinitely risk averse translates the optimization problem into the minimization of the sample covariance matrix at every time t conditional on information known at time t-1. Then the expected utility obtained from holding a portfolio with weights at time t is equivalent to the portfolio variance. After obtaining optimal portfolio weights, two random forests with macroeconomic features are trained on the optimal portfolio weights in the training sample – one for the gold weights and one for the stock weights. In addition, a random forest with macroeconomic features is trained on the expected utility obtained from the optimization problem.

These random forests create three portfolios for the stock-and-gold index over the test sample. The macro features' variable importance is measured for each random forest, and their Accumulated Local Effects (ALE) are plotted and assessed. This allows for interpreting the various macro features' effects on the portfolio weights. Moreover, the out-of-sample performances of the three portfolios are compared to three benchmark portfolios. Two are relatively simple portfolios, a Markowitz (1952) portfolio – an equally weighted portfolio – and a minimum variance portfolio estimated once over the training sample without using random forests and kept constant over the test sample. The third portfolio is created from the optimal weights created without using a random forest. The differences in Sharpe ratios for the portfolios are statistically compared using the robust Sharpe ratio test from Ledoit and Wolf (2008).

Generally, the level of narrow money available in the economy and the annual growth in housing prices is most important in determining the optimal portfolio weights for the independently modelled stock and gold weights. For utility, narrow and broad money are the most important macro features of the model. Narrow money has a step-wise positive effect on the gold weights and a step-wise negative effect on the stock weights. A steep decline in gold weights occurs when the annual growth of housing prices is larger than 5%. This is combined with a steep increase in the stock weights when the annual growth of housing prices is larger than 5%. The ALE plots also show that portfolio variance is minimal when the indices for narrow (M1) and broad money (M3) are around 35.

The out-of-sample Sharpe ratios of the random forests for the gold and stock weights are slightly larger than the simple benchmark portfolios. However, the difference is minimal, and the portfolios do not significantly outperform the benchmark portfolios according to the test of Ledoit and Wolf (2008). The portfolio based on the reconstructed weights from the forecasted utility performs similarly to the benchmark portfolios. Unfortunately, the improvement that random forests offer in other situations does not apply here, and the extra computational steps do not achieve better out-of-sample results.

The remainder of this paper discusses the relevant literature on portfolio optimization and regression trees in Section 2. Section 3 outlines the data, followed by the methodology in Section 4. Results are presented in Section 5, and Section 6 provides a discussion and conclusion of the research.

2 Literature Review

One of the main practices in financial markets is the optimal allocation of assets, where an investor wants to ensure more attractive returns while limiting her risks. Markowitz (1952) was the first to state that agents should minimize risks by diversifying their investments by creating portfolios. In his paper, Markowitz (1952) developed a mean-variance model that describes the risk-return trade-off problem, which forms the basis of Modern Portfolio Theory. Hence, portfolio allocation depends on agents' preferences regarding expected risks and returns (Oliveira & Valls Pereira, 2018). Optimal portfolios are said to be on the efficient frontier: the set of risk and return combinations that establish minimum risk allocations for a given average portfolio return (Markowitz, 1952). Furthermore, all portfolios on the efficient frontier have optimal Sharpe ratios (Sharpe, 1964). In other words, an investor aims to maximize the expected utility that her portfolio yields at time t (Sharpe, 2007). This research assumes this is always done out-of-sample. That is, the investor rebalances her portfolio at time t - 1 conditional on the information available then. Defining utility as a function for portfolio returns penalized by portfolio volatility is equivalent to optimizing the Sharpe ratio over the portfolio (Kazak et al., 2022).

Economic theory suggests that stock prices should reflect expectations about future corporate performance while corporate profits generally reflect the level of economic activities (Neifar et al., 2021). This implies that financial asset prices vary under different market conditions. Empirical applications show that the state of the economy can significantly impact the performance of different financial assets. For example, riskier assets generally perform poorly and are much more volatile during periods of low economic growth. In the past, stocks have shown to be more volatile in times of economic downturn. At the same time, the gold price has increased substantially with only the smallest increase in the likelihood of disaster (Ang, 2014). In line with these observations, the investor can minimize her risks by investing more in gold during economic downturn, i.e. the portfolio weight for gold should increase. Therefore, based on theories from Ang (2014) and empirical applications, there is reason to believe that optimal portfolio weights can be described as a function of macroeconomic variables.

2.1 The behaviour of gold and stock prices with macroeconomic variables

In their letter for the Federal Reserve Bank of Chicago, Barsky, Epstein, Lafont-Mueller and Yoo (2021) conclude that gold prices reflect protection against bad economic times and that pessimism about future economic activity affects gold prices. Moreover, Barsky et al. (2021) find that long-term real interest rates negatively affect gold prices, and gold can be considered an inflation hedge. Similar results follow from a recent article in Forbes, where Iuorio (2023) discusses how gold performs with recession, inflation, and stagnation. First, during the past recessions between 1973 and 2020, gold has outperformed the S&P500 six out of eight times by 37% on average. Second, in the period of hyperinflation between 1973 and 1979, gold gained a 35% annual return. During the 2021 inflation, gold started performing well, but its value dropped by 20% later. According to Iuorio (2023), this was due to the Federal Reserve's aggressive policy to raise interest rates. Nevertheless, gold benefits from an overall increase in the money supply (Iuorio, 2023). Finally, during stagflation in the mid to late 1970s, i.e. stagnant growth, high inflation, and high unemployment, the Federal Reserve stimulus and fiscal intervention caused investors to shy away from individual company stocks, where gold is often the alternative beneficiary. According to Iuorio (2023), changes in the gold price are not necessarily caused by market expectations and sentiment but more so by Federal Reserve policies. Gold prices react negatively to an increase in the (long-term) interest rates but positively to an increase in money (M1). This implies that a portfolio partly invested in gold is likely to react to changes in these macroeconomic factors, and thus describing the optimal weights as a function of macro factors is possible.

As Iuorio (2023) also discusses, individual company stocks tend to perform oppositely in the scenarios mentioned above. As equities are riskier, they also tend to attain higher returns, especially when times are good (Ang, 2014). These patterns are observed in the data set discussed in the next section. Choudry, Hassan and Shabi (2015) provide a somewhat opposing view, as they find that gold might not be a good safe haven asset during the financial crisis due to the strong interdependence between gold and stock returns and stock market volatility. Yet, Choudry et al. (2015) conclude that gold may be a hedge against stock returns and volatility in more stable financial conditions. Al-Ameer, Hammad, Ismali and Hamdan (2018) confirm the idea that there is a correlation between the gold and stock market that differs each period, in size and sign. Based on this behaviour of the gold and stock prices under various market conditions, there is reason to believe that a function of macro factors for the portfolio weights can give interesting insights into portfolio performance when altering the weight of these two asset classes. Oliveira and Valls Pereira (2018) discuss that optimal diversification choices using the mean-variance framework assume risk and return parameters of the assets to be known. Yet, the optimization is based on moment estimates with sampling errors that differ from actual risks and returns. Furthermore, mean-variance portfolios themselves do not consider the different states of the market. For example, the joint distribution of asset returns differs in bull and bear markets. Therefore, the fact that risks and returns depend on the states of the market and that the joint distribution of the returns is not independent and identically distributed (i.i.d.) should be addressed when optimizing portfolio weights (Oliveira & Valls Pereira, 2018).

Following this line of arguments, a logical possibility would be using regime-switching models, as Oliveira and Valls Pereira (2018) do. This captures the joint distribution of the portfolio assets for each regime and the probability of switching from one regime to another. However, as discussed in the previous section, Markov-switching regimes have significant disadvantages. The return of a discrete response variable makes it difficult to apply these models to this research, as portfolio weights are continuous variables. Moreover, they generally tend to be outperformed by regression trees (Samitas & Armenatzoglou, 2014).

2.2 Regression trees and random forests

Regression trees are commonly used in machine learning and are decision trees for a continuous dependent variable. Regression trees offer an alternative method to capture non-linear relations. By recursively partitioning the data with a binary choice at every node, complex regression models are simplified to a point where the data is tame enough to apply a simple regression model. With that model, the value of the dependent variable can be estimated. The dependent variable can be estimated as the sample mean of earlier predictions at a specific node. This is why regression trees can fit almost any traditional model (Loh, 2014) and can thus be applied to the problem at hand. The split decision in the tree is based on a threshold value. Common practice is to use the minimum sum of squared errors as the criteria. While regression trees can model complex interactions without assuming a model or distribution, they quickly tend to overfit data in the training set, reducing forecasting accuracy. Furthermore, regression trees can be unstable, as small data changes may lead to larger changes in the model (Rokach, 2016). A solution would be to prune the regression trees, but these issues can better be tackled by improving the regression tree through bagging or bootstrapping, e.g. with a random forest. A random forest consists of multiple trees, i.e. an ensemble of trees, in which their predictions are combined into one final prediction. In the algorithm developed by Breiman (2001), B different bootstrap samples are generated, where a separate tree is independently fitted for each sample. The variables used to grow the tree in each sample are randomly chosen and thus vary across the forest. With one regression tree, splits are done based on the best variable choice among the data. In the random forest of Breiman (2001), splits are done based on the best variable choice among a subset of predictors randomly chosen at that node. This is also called "attribute bagging". In doing so, the random forest can significantly reduce a single regression tree's variance. As the random forest method forces an ensemble of different trees, the variance reduction is larger compared to standard bagging.

Besides the computational benefits of trees and random forests, tree-based machine-learning models are preferred over alternative machine-learning methods for portfolio allocation problems. As Pinelis and Ruppert (2022) explain, tree-based models are non-metric because there are no inherent assumptions of distributions in the data. Furthermore, decision trees are scale-invariant, and the number of parameters that typically need to be optimized in a Random Forest is fewer than in many other machine-learning models. Moreover, it is well-known that financial data and especially expected returns are extremely noisy, reducing the out-of-sample performance of many models. This is often due to changing relationships between predictors and the target variable and potential shifts in the data distributions. To align this with the behaviour of gold and stock prices, as Iuorio (2023) explains, the rise in gold prices could be explained by the inflation that occurred in some recessions. In other recessions, gold prices quickly fell again due to the Federal Reserve's aggressive policy to raise interest rates. Sometimes, the increase in the money supply (M1) was the cause for the high returns on gold. According to Pinelis and Ruppert (2022), Random Forests can mitigate these complex and varying relationships between the splitting variables and the response variable. Suppose one tree is grown to capture the relationship between gold and inflation with the money supply. In that case, the tree may accurately predict optimal weights for the gold index in some market environments but not in all. Other variables, such as long-term interest rates or volatility, may correlate more with excess market returns in specific periods. Since a Random Forest grows many trees with different variables, changes in the data distributions or relationships may cause some trees to perform poorly. Still, the results of the entire forest should remain mostly unchanged. Thus the forest helps to reduce noisy data. Finally, random forests allow for easy interpretation, as it is relatively straightforward to calculate variable importance and partial dependence for each feature (Breiman, 2001).

2.3 Transaction costs

An essential factor that must be considered when rebalancing the portfolio over time to maintain optimal weights is that transaction costs are an essential part of the investment strategy. While continuous rebalancing to maximize the Sharpe ratio at all times seems very promising, a penalty needs to be introduced to account for the transaction cost of buying or selling (parts) of the portfolio. Transaction costs can significantly impact performance (Detzel, Novy-Marx & Velikov, 2023) and can be accounted for by only rebalancing the portfolio if the new optimal portfolio weights offer higher excess returns than the current portfolio weights plus a penalty for the transaction costs. Unfortunately, there is no real agreement in the literature on what a suitable quantification of transaction costs, i.e. the penalty, should be. This paper uses the rule of thumb as mentioned by De Nard, Ledoit and Wolf (2021) that the return loss (per month) due to portfolio turnover is twice the amount of portfolio turnover times the chosen transaction cost. Previous literature usually dealt with transaction costs of fifty basis points (bps) (DeMiguel, Garlappi & Uppal, 2009). Nowadays, transaction costs are closer to five bps (De Nard et al., 2021). This research assumes that transaction costs are five bps and thus ten bps of turnover, equivalent to 0.1% of portfolio turnover. Portfolio turnover can be calculated as the sum of the difference in weights between time t and t - 1. De Nard et al. (2021) also note that rebalancing only monthly significantly diminishes the problem of transaction costs. Since this research is restricted to data availability on macroeconomic variables, the portfolio weights can only be rebalanced monthly, but transaction costs should still be accounted for.

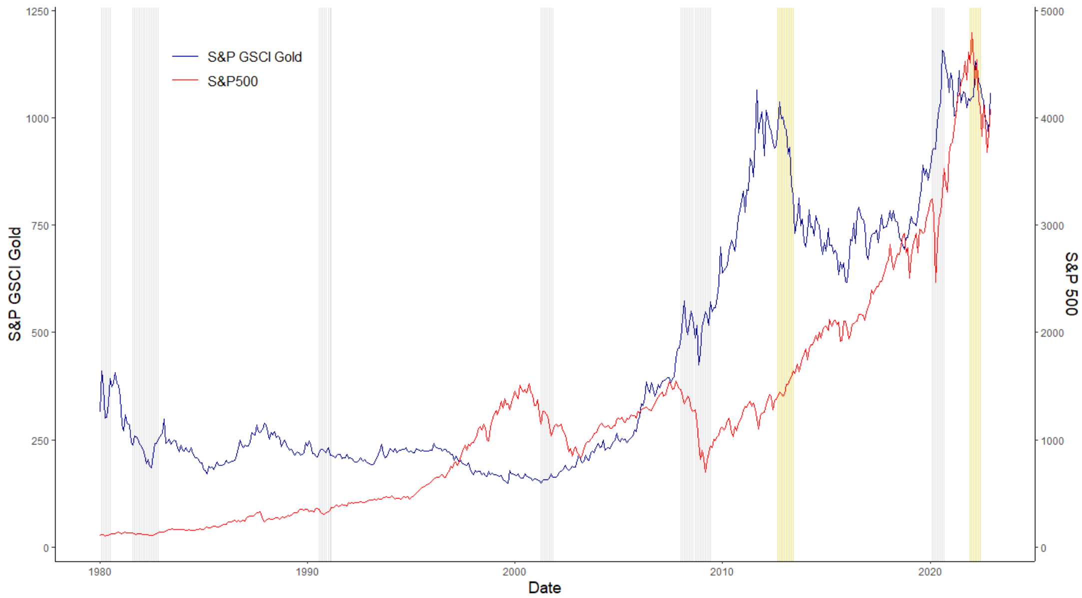
3 Data

This research consists of two datasets. The first set contains financial data on **the monthly prices for the S&P GSCI Gold index and the S&P 500 stock index**, which is retrieved from Refinitv Datastream (2023) over the period January 1980 to December 2022, yielding 43 years of monthly observations. The second dataset contains macroeconomic data on **42 different macro factors** obtained from Refinitiv Datastream (2023), OECD Database (2023), the Chicago Board Options Exchange (2023), and Federal Reserve Economic data from the Federal Reserve Bank of St. Louis (2023). The macroeconomic data is collected over the time period from December 1989 to November 2022. The macroeconomic data is applied to the portfolio weights out-of-sample with one lag, that is, weights determined for time t, are related to the macroeconomic data at time t - 1. Throughout this research, the portfolio data is split into an estimation sample of ten years – ranging from January 1980 to December 1989, yielding 119 observations; a training sample of 20 years – from January 1990 to December 2009, yielding 240 observations; and a test sample of 13 years – from January 2010 to December 2022, yielding 156 observations. The macroeconomic data is split two ways: a training sample of 20 years, yielding 240 observations, and a testing sample of 13 years, yielding 156 observations. The portfolio data has no missing values, whereas a few occur in the macroeconomic data. This problem is mitigated by setting the missing values to the average of the four neighbouring values.

3.1 Portfolio data

The S&P 500 Index summarizes the performances of the 500 largest individual company stocks in the United States and is chosen as a proxy for the performance of stocks. It is assumed that the index is affected by market conditions in a similar manner as stocks would. The S%P GSCI Gold Index is a renowned index chosen such that both indices considered in this research are from the same provider. This ensures pricing is done similarly for both indices, making them comparable. Figure 1 shows the price levels of both indexes over the entire sample, where interesting patterns in the gold and stock index prices can be seen. The grey bars in the graph indicate recessions based on National Bureau of Economic Research (NBER) data. The first

*Figure 1. Monthly Price Level for the S&P GSCI Gold Index (in **blue** on the left axis) and the S&P500 Index (in **red** on the right axis).*



Note. Grey strokes in the background indicate months in which a recession occurred, based on the NBER Recession Indicator (National Bureau of Economic Research, 2023). Yellow background strokes indicate interesting events not based on the NBER Recession Indicator.

recession noted was in 1980, after hyperinflation in the 1970s. Gold prices had risen during the 1970s, explaining the high GSCI Gold index price at the beginning of the 1980s. It was only after the announcement of the aggressive Federal Reserve policy to raise interest rates that gold prices fell sharply. Another recession followed quickly after, but this time the GSCI Gold index dropped further due to the persistent negative effect of long-term interest rates on the gold price (Iuorio, 2023). In 1991 a mild recession occurred, and gold prices did not spike again. The recessions of 1981 and 1991 are the two past recessions in which the GSCI Gold index did not outperform stocks. In all other past recessions, gold did. After the burst of the dotcom bubble in 2001, stock prices fell quickly while gold steadily increased and captured an annual return of approximately 6.9%. The largest reactions and interactions between the stock market, gold, and recessions have occurred in recent years.

During the Great Financial Crisis (GFC) from 2007 to 2009, a sharp increase in the gold price occurred right at the start of the recession in September 2007, even though the gold price also fell with the largest decline in the S&P 500 index towards the end of the financial crisis. This was due to the immense impact that the GFC had on the stock market, as any positive outlook on the market was gone at that time. To end the crisis in 2009, the FED cut the funds rate to almost 0%, causing the GSCI Gold index to shoot back up. Afterwards, the gold price kept rising, partly due to the low funds rate and market sentiment. The GFC caused a great recession in Europe, leading to the (near) default of Greece and other European countries. This kept investors afraid of market instability and the possibility of another crash soon, which fed the rise of the GSCI Gold index, as gold was truly seen as a safe-haven asset these days.

In 2013, marked by the first yellow stroke in Figure 1, gold prices finally dropped significantly, bursting the "gold bubble". Interestingly, this event does not coincide with a reaction from the stock market, as seen in Figure 1. A reason for the drop in the GSCI Gold index is the recovery of trust in the markets in both the United States and Europe, which would also explain why the stock index steadily increased during this time. Nevertheless, the Covid-19 pandemic and the subsequent market crash in 2020 showed investors' trust in the commodity gold. While the S&P 500 falls, the GSCI Gold index shoots up, reaching an all-time high, and only slightly diminishes in value after the financial crisis is more or less over. Finally, the S&P 500 experienced a drop over the summer of 2022 due to the high inflation during that time. This is marked by the second yellow stroke in Figure 1. Consequently, the GSCI Gold index rises but then also falls. This is due to the FED's policy to raise long-term interest rates, as Iuorio (2023) explains. To summarize, the patterns observed over the past 43 years show how gold and stock prices interact with one another in relation to macroeconomic features. This suggests a relation between the optimal gold and stock portfolio weights and macroeconomic variables. This relation should be modelled to see if it can give new insights or improve trading strategies.

The monthly asset returns for asset i at time t, $R_{i,t}$, are calculated as:

$$R_{i,t} = \frac{P_{t,i} - P_{t-1,i}}{P_{t-1,i}}, \tag{1}$$

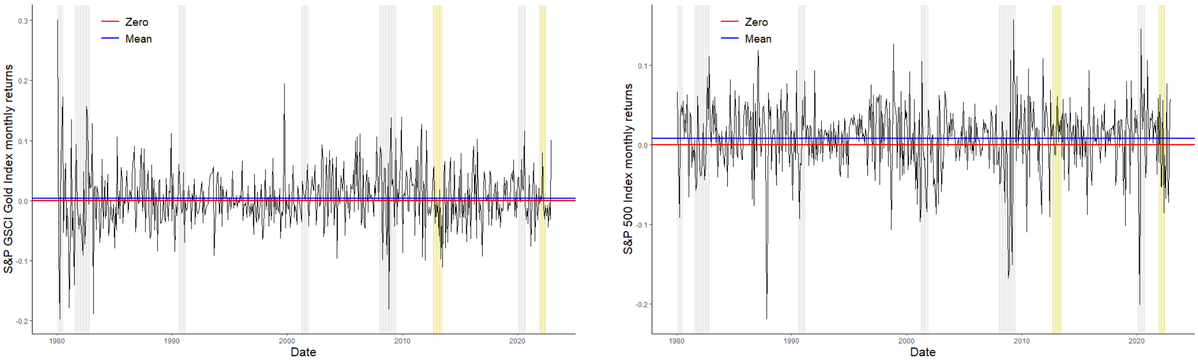
where $P_{t,i}$ denotes the price at time t. This results in a TxN matrix, where $N = 2$ denotes the number of assets included in the portfolio, and $T = 516$ is the total monthly observations. Figures 2a and 2b show the monthly returns for the two assets. Both assets display volatility clustering when a recession occurs (as marked by the grey strokes in the background). The largest negative returns for the gold index are recorded during the hyperinflation period in the 1980s and the GFC. Stock returns display large negative returns more often during recessions, especially during the GFC and the Covid-19 crises. From Figure 2, it is clear that the S&P 500 index has slightly higher returns on average (mean = 0.008) but also suffers from larger negative returns. In other words, the S&P 500 index is more volatile and thus has a higher risk premium. The S&P GSCI Gold index is less volatile but offers lower returns (mean = 0.003). The descriptive statistics of both assets can be found in Table B2 in Appendix B.

3.2 Macroeconomic data

There is a large variety of possible macro factors available that can influence asset returns. For example, Cutler et al. (1988) suggested the well-known variables Inflation, Volatility, Long term interest rates, Short term interest rates, Real Money, Industry Production, and Real Dividends. Empirical applications especially underline the importance of Inflation, Interest Rates, Real GDP, Consumption, and the Volatility Index (VIX) (Ang, 2014). Maio and Philip (2013) apply a factor representation of 107 macro factors to predict cross-sections in stock returns. Commonly used factors applied in their paper are different variables of Housing Units started, Orders for Durable Goods, Manufacturing, and Retail Trade. Maio and Philip (2013) also use Unemployment rates, Foreign Exchange rates, and the Consumer and Producer Price Indices. Finally, market sentiment indexes can impact asset returns, of which the most well-known example is the CBOE Volatility Index (VIX). Even small movements to the VIX can significantly impact

Figure 2. Monthly Returns for the Financial Assets included in the Portfolio.

(a) S&P GSCI Gold Index *(b) S&P 500 Index*



Note. Grey strokes in the background indicate months in which a recession occurred, based on the NBER Recession Indicator (National Bureau of Economic Research, 2023). Yellow background strokes indicate interesting events not based on the NBER Recession Indicator. The **red** line represents a 0% return. The **blue** line represents the mean of asset returns.

asset returns (Ang, 2014). Pooling the four databases used in this research, many macroeconomic variables are available after filtering for data availability on the sample. It is important to consider the maximum number of variables that can be applied in the model, as including many variables increases the risk of overfitting and causes poor out-of-sample performance. Random forests can mitigate this problem, whereas single regression trees cannot. This allows for a great selection of variables in the model. Nevertheless, correlation among all variables should be assessed before including variables in the model, as highly correlated variables are likely to decrease predictive accuracy. The result is a TxK matrix, where $K = 42$ denotes the variables in the data set and $T = 396$ denotes the total number of observations over time. For some variables, transformations are necessary to include the variable. An overview of the included macroeconomic features, their transformations, frequency, and their databases is presented in Table A1 in Appendix A. Furthermore, the descriptive statistics of the macro features are presented in Table B1 in Appendix B.

4 Methodology

This section starts by discussing portfolio optimization, then continues with the model for transaction costs, and finishes with the implementation of random forests. First, the notation that is used throughout the paper is introduced. At time t, optimal weights ω_t^* are found for a portfolio with a stock and gold index, that is $\omega_t^* = (\omega_{G,t}^* \quad \omega_{S,t}^*)'$ respectively. The portfolio should always be fully invested, such that the sum of weights equals 1, or $\omega_{S,t} = 1 - \omega_{G,t}$. Formally, this can be denoted as $\omega_t \in S$, where $S = \{\omega \in \mathbb{R}^N : \omega' \iota = 1\}$. Furthermore, the weights for an asset i are bounded by three times their leverage, that is $-3 \leq \omega_{i,t} \leq 3$. Important to note is that throughout this research, portfolio weights are optimized out-of-sample. For every time t, the optimal weights are estimated conditional on the information known at time t - 1. This requires time-varying estimates for the mean and covariance matrix of the assets. Similarly to Kazak et al. (2022), the following assumption is made.

Assumption 1 *First, consider the returns from the portfolio to be a random walk with a trend. Then, on a filtered probability space, define a 2-dimensional vector for the monthly return process* $r_{t\{t=1,2,...\}}$ *for the two assets. For every time interval* $[t\text{ - }1, t]$, r_t *is generated as:*

$$r_t \equiv \mu_t + \int_{t-1}^{t} \Theta_s dW_s, \tag{2}$$

where μ_t *is a bounded random variable,* Θ_s *is a continuous spot covolatility process, and* W_s *is a Brownian motion process. Then* $\Sigma_t = \int_{t-1}^{t} \Theta_s \Theta_s' ds$ *is the quadratic covariation of* r_t *on the time interval. For all* $\omega \in \mathbb{R}^N$ *and* t*, it is assumed that* $0 < \omega' \Sigma_t \omega < \infty$ *and that* Σ_t *is weakly stationary and ergodic.*

This assumption implies that $E_{t-1}[r_t] = \mu_t$ and $V_{t-1}[r_t] = E_{t-1}[\Sigma_t]$ by construction, where $E_{t-1}[\cdot] = E[\cdot|\mathcal{F}_{t-1}]$ and $\mathcal{F}_{t-1}$ denotes the information set at time t - 1. Finally, set $\Omega_t = E_{t-1}[\Sigma_t]$ for notational convenience. Monthly returns for an asset i obtained at time t, $R_{i,t}$ are multiplied with the chosen portfolio weights to construct portfolio returns $R_{p,t} = \omega_t' r_t$.

4.1 Portfolio optimization

At time t, an investor wants to rebalance her portfolio according to the information set $\mathcal{F}_{t-1}$ to keep her portfolio weights optimal. Based on the above assumption and notation, the von Neumann-Morgenstern theorem (von Neumann & Morgenstern, 1994) states that the investor should maximize the conditional expectation of her utility for holding the portfolio $\omega_t' r_t$. Solving for the portfolio weights implies:

$$\omega_t^* = \underset{\omega_t \in S}{\arg\max}\, E_{t-1}[\mathcal{U}(\omega_t' r_t)], \quad s.t.\ \omega_t'\iota = 1, \tag{3}$$

where $\mathcal{U}(\cdot)$ is the investor's utility function, and the constraint ensures that a full investment is made at time t. Similar to Kazak et al. (2022), this paper assumes a simple mean-variance conditional utility function:

$$E_{t-1}[U_t(\omega_t; \gamma)] = \omega_t' \mu_t - \frac{\gamma}{2} \omega_t' \Omega_t \omega_t, \tag{4}$$

where γ is the Arrow-Pratt risk-aversion coefficient. A choice must be made as it is almost impossible to maximize returns and minimize risks simultaneously. Modelling μ_t with little noise is very difficult (Bryzgalova et al., 2021; Jagannathan & Ma, 2003). Even more so, the Global Minimum Variance Portfolio (GMVP) tends to have a better out-of-sample Sharpe ratio than when expected returns are also included in the model (Jagannathan & Ma, 2003; Chakrabarti, 2021). Therefore, this research assumes $\gamma = \infty$, i.e. the investor is infinitely risk-averse. Then the utility function can essentially be written as:

$$U_t(\omega_t) = -\omega_t' \hat{\Omega}_t \omega_t, \tag{5}$$

where $\hat{\Omega}_t = E_{t-1}[\Sigma_t] = E[\Sigma_t|\mathcal{F}_{t-1}]$ is estimated as the sample covariance matrix at time t - 1. While much literature is spent on estimating conditional covariance matrices, this paper assumes a simple model to focus on implementing regression trees. At the start of the training sample, the estimation sample is used to obtain estimates for the sample mean $\hat{\mu}_t$ and covariance matrix $\hat{\Omega}_t$ to optimize the portfolio weights for the first data point in the training sample. Afterwards, a rolling window with ten years of observations estimates $\hat{\mu}_t$ and $\hat{\Omega}_t$.

4.2 Transaction costs

As discussed in Section 2, transaction costs play an important role in a model for rebalancing portfolio returns. For each portfolio re-balancement, the transaction costs (TC) can be calculated following De Nard et al. (2021); DeMiguel et al. (2009):

$$TC_t = 2c \cdot \sum_{i=1}^{N} |\omega_{j,t} - \omega_{j,t-1}| \tag{6}$$

where c denotes the basis points, which are assumed to be $c = 0.005$, $\omega_{i,t-1}$ denotes the weight of asset i before rebalancing at time t - 1, and $\omega_{i,t}$ denotes the weight after rebalancing. The portfolio variance is minimized for a target mean μ_0 larger than transaction costs to implement transaction costs in the portfolio optimisation algorithm. This ensures that the expected return gain from rebalancing at least covers transaction costs and is, thus, more likely to give a net positive return. The portfolio optimization problem can then be written as follows:

$$\begin{aligned} \min_{\omega_t} \quad & \omega_t' \hat{\Omega}_t \omega_t \\ \text{subject to} \quad & \omega_t' \mathbb{1} = 1 \\ & \omega_t' \hat{\mu}_t \geq TC_t, \end{aligned} \tag{7}$$

where $\mathbb{1}$ denotes a $2x1$ vector of ones, and TC_t are the transaction costs for rebalancement at t. Maximizing the utility function from Equation 5 is equivalent to minimizing this function multiplied by -1, and thus also equivalent to optimizing the Sharpe ratio and tracing out portfolios on the efficient frontier for a given target mean μ_0 (Bryzgalova et al., 2021).

4.3 Random forests

In the training sample, two random forests are trained separately on the sets of optimal portfolio weights obtained using Equation 7, such that there is one forest for the gold index and one for the stock index. The optimal portfolio weights for an asset i are defined as a function of the macro features in the random forest as follows:

$$\begin{aligned} \hat{\omega}_{i,t} &= \mathcal{F}_i(X_t) \\ \text{s.t.} \quad & \min_{\omega_t} \quad \omega_t' \hat{\Omega}_t \omega_t, \end{aligned} \tag{8}$$

where $\mathcal{F}_i$ denotes the random forest for asset i, and X_t denotes the macro features used in the random forest. Note that this does not necessarily satisfy the investment constraint of the model and, thus, cannot be directly implemented as a trading strategy. However, it is also interesting to see what the random forest models can predict for the portfolio weights and how these interact with the model's features. The implementation of modelling the two financial assets together is discussed later.

4.3.1 Tuning hyperparameters and model estimation

As explained in Section 2, applying random forests to financial or macroeconomic data significantly limits the risk of overfitting and instability that would occur when employing simple regression trees. Therefore, pruning is necessary for regression trees, while this is not the case for random forests. Even more so, Coulombe (2020) shows the out-of-sample performance of pruned forests is equivalent to fully grown forests if and only if the forest is constructed of sufficiently diversified trees. A well-diversified forest can be obtained by choosing the hyperparameters of the forest correctly. For the random forests in this research, three parameters are considered, i.e. the number of trees included in the model, B, the fraction of features used in each tree as a subset of the total set of features, `mtry`, and the minimum sample size at each of the terminal nodes, `nodesize`. Using previous literature from Coulombe (2021), `nodesize` should be set to 15 for monthly data. Then, `mtry` is tuned for each random forest employing five-fold cross-validation over the training sample with a random grid search for $g = \{1, 2, 3, ...10\}$. The optimal value of `mtry` is assessed using three characteristics, Root Mean Squared Error (RMSE), R-squared (R^2), and Mean Absolute Error (MAE). For all random forests, RMSE was used to select the final value of `mtry`. This five-fold cross-validation uses the `R` package `caret`. Finally, B is chosen by assessing the out-of-bag error over an increasing number of trees. The error does not significantly diminish for all forests after bagging 100 trees, as shown in Figure C1 in Appendix C. This is also a commonly used value for B in previous literature (Pinelis & Ruppert, 2022). Therefore, $B = 100$ is selected for all forests. After parameter tuning, the model with optimal parameters is estimated using the `R` package `randomForest`. For each tree, nodes are split on the best improvement in the Mean Squared Error (MSE) using a subset of the macro features of size $\frac{1}{\texttt{mtry}}$. Thus, at every step in the tree of random forest $\mathcal{F}_i$, the splitting rule aims to find the tree's optimal feature and splitting value to minimise the overall MSE. This can be formulated as follows:

$$\min_{c \in \mathbb{R},\, k \in Z} \left(\min_{\hat{\omega}_i} \frac{1}{T} \sum_{t|X_{k,t}<c} \varepsilon_{i,t}^2 + \min_{\hat{\omega}_i} \frac{1}{T} \sum_{t|X_{k,t}\geq c} \varepsilon_{i,t}^2 \right) \tag{9}$$

where $\varepsilon_{i,t}^2$ denotes the squared error term obtained from the model in Equation 8. Furthermore, k represents a macro feature from the set Z_t, the subset of X_t randomly chosen for this tree using `mtry`. c denotes the value on which the feature is split. This splitting procedure is recursively applied until the minimum `nodesize` of 15 is reached. Then, the estimates from each tree are averaged over all the trees in the forest to find the final estimates. The results of parameter tuning and model estimation are presented in the Results section.

The independent random forest models predict new values for the optimal portfolio weights in the test sample. Transaction costs are considered when calculating the optimal portfolio weights in the training set, such that the weights cannot differ significantly between time t -

1 and time t. This setup causes the random forest to predict the portfolio weights for time $t + 1$ considerably close to weights at time t, automatically mitigating the transaction costs problem. For asset class with predicted portfolio weight $\hat{\omega}^i_{t+1}$, the weight for the opposite asset j is calculated using the full investment constraint, $\hat{\omega}^j_{t+1} = 1 - \hat{\omega}^i_{t+1}$. This method gives two sets of predicted portfolio weights. For both sets, the portfolio returns are calculated over the test sample as $R^i_{p,t+1} = \hat{\omega}^i_{t+1} r^i_{t+1} + \hat{\omega}^j_{t+1} r^j_{t+1}$, where the superscript i denotes that the weights for the asset i are forecasted. The weights for the asset j are inferred. The performance of both portfolios is evaluated over the testing sample as described in Section 4.3.4.

4.3.2 Modelling the financial assets together

A third random forest is grown on the investor's utility captured from the optimal portfolio weights in the training sample. The expected utility as in Equation 5 is calculated for the training set from the optimal weights as defined by the optimization problem from Equation 7. This essentially leads to a random forest based on the following formulation:

$$E[U_t] = \min_{\omega} \omega_t' \hat{\Omega}_t \omega_t = \mathcal{F}_u(X_t) \tag{10}$$

where $\mathcal{F}_u$ represents the random forest for the utility model with macroeconomic features X_t. This results in a splitting rule that differs slightly from the splitting rule defined in Equation 9, in the sense that the squared error terms are now obtained from Equation 10. Before estimating the random forest, the hyperparameters of the forests are tuned using the approach described in Section 4.3.1 in this subsection. After model estimation, utility is predicted using the test sample of features and optimal portfolio weights. The portfolio weights at time $t + 1$ can then be uniquely reconstructed by solving the following system of equations for the weight ω_{t+1}:

$$\begin{aligned} \hat{\omega}^{*\prime}_{t+1} \Omega_{t+1} \hat{\omega}^*_{t+1} &= \hat{u}_{t+1} \\ \hat{\omega}^{*\prime}_{t+1} \mathbb{1} &= 1, \end{aligned} \tag{11}$$

where $\hat{u}_{t+1}$ denotes the predicted utility at time $t + 1$ using the random forest model. Furthermore, $\Omega_{t+1} = E_t[\Sigma_{t+1}] = E[\Sigma_{t+1}|\mathcal{F}_t]$, and $\mathbb{1}$ denotes a vector ones similar to Equation 7. Like the weights obtained from the independently modelled random forest, the weights are used to assess portfolio performance over the test sample. For notational convenience, the returns at time t obtained from this portfolio are denoted as $R^*_{p,t+1} = \hat{\omega}^*_{G,t+1} r_{G,t+1} + \hat{\omega}^*_{t+1} r_{S,t+1}$.

4.3.3 Interpreting random forests

The intuitive interpretations of random forests are instrumental in this research to assess the individual effects of the macroeconomic features on the portfolio weights with an estimate for variable importance. Variable importance uses the internal out-of-bag (OOB) observations created by the random forest, i.e. the observations not used in growing the trees due to sampling with replacement (Breiman, 2001). For every tree, a subsample is considered OOB, for which its prediction accuracy is calculated using MSE. Then the values of every k-th variable are randomly permuted in the OOB sample, and the model is re-estimated with this OOB data, where the MSE is obtained. Variable importance (VI) is now computed as the percentage increase in the MSE when the variable is included in the tree versus when the variable is permuted. Variables with a relatively large increase in MSE are considered more important. VI is averaged over all the trees and normalized by the standard deviation of the differences to obtain the final estimate (Breiman, 2001). The drawback of VI is that the estimates tend to be more biased when features are highly correlated, which is the case in this research. Randomly permuting the variables more than once can establish more stable results, but these improvements tend to be small. Therefore, the results from variable importance should be treated with caution. Estimates of variable importance are presented in the Results section.

This paper set out to find the optimal portfolio weights as functions of the macroeconomic variables to forecast the optimal portfolio weights. The marginal relation between the features and the response variables in each forest is assessed using Accumulated Local Effects (ALE) plots, a correlation-robust version of the more well-known Partial Dependence (PD) plots (Apley & Zhu, 2020). PD plots estimate the marginal distribution of variable k from the complete set of K macro features relative to all other features excluding k. An estimation of the partial dependence between variable k and the portfolio weights is obtained by averaging over the trees. While this leaves for intuitive interpretation, PD plots are based on the assumption that the features from the random forest are independent. However, dealing with macroeconomic variables implies a high correlation among the features. ALE plots do not suffer from this disadvantage, as ALE plots focus on the difference in the average prediction that results from the feature (Apley & Zhu, 2020). In other words, an effect of 0.1 in the ALE plot for a particular value of a feature implies that that value of the feature causes the response variable to be 0.1 larger than its average prediction. Since ALE plots are better suited for the dataset used in this research, ALE plots instead of PD plots are used. The ALE plots for each model's nine most important features are presented in the Results section. The other plots are shown in Appendix C.

4.3.4 Out-of-sample performance of the portfolios

To summarize, three portfolios are created for the test sample. Two portfolios are based on the independently modelled trees for the gold and the stock index, where each portfolio is created by forecasting the weights for one of the two financial assets and inferring the other weight from these forecasts. The third portfolio is created by modelling utility and uniquely reconstructing its weights. Furthermore, a benchmark portfolio is created from the weights obtained from the quadratic programming problem described by Equation 7. The performance of each portfolio p is estimated by tracking the portfolio returns over time and by tracking estimates for the expected return, i.e. $\hat{\mu}_p$, and the standard deviation (volatility), i.e. $\hat{\Sigma}_p$. Finally, each portfolio's Sharpe ratio is calculated as $\widehat{SP}_p = \hat{\mu}_p/\hat{\Sigma}_p$. The performance of the portfolios is compared with each other. It is also benchmarked against the portfolio based on the weights from Equation 7 and two relatively simple portfolios. The first is a Markowitz portfolio, i.e. an equally-weighted portfolio, and the second is a portfolio where the optimal weights are estimated once over the training sample and are kept constant throughout the test sample.

It is also of interest to statistically compare the performance of the portfolios. The Sharpe ratios of the six portfolios are compared using the robust prewhitened HAC test from Ledoit and Wolf (2008) for hypothesis testing with the Sharpe ratio. For each portfolio p and q, the following null hypothesis is tested:

$$H_0: \qquad \widehat{SR}_p - \widehat{SR}_q = 0, \tag{12}$$

such that rejecting the null hypothesis implies that one portfolio significantly outperforms the other. The comparison of the out-of-sample performance of the portfolios and the corresponding test statistics are presented in the Results section.

5 Results

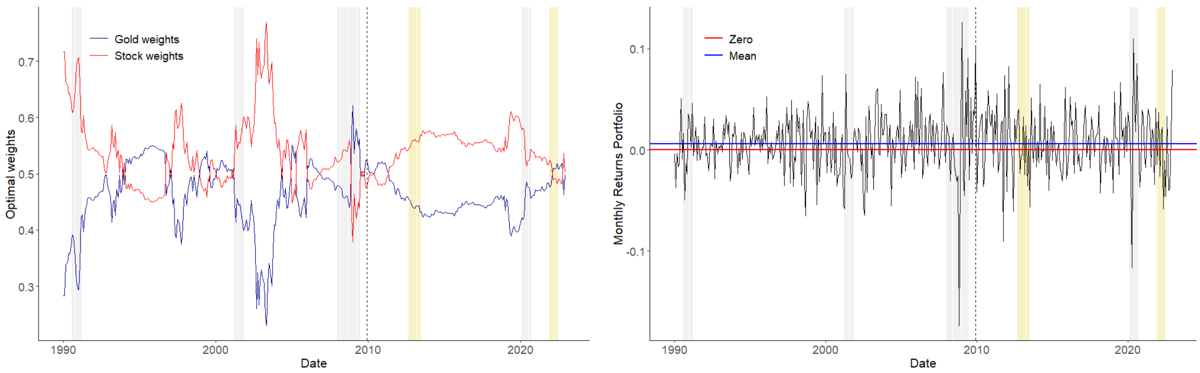
Equation 7 in the previous section describes the optimization process used in this research. Note that the objective of the optimization process was to minimize the variance to maximize the Sharpe ratio. Figure 3a shows the results from this optimization process for the portfolio weights. The weight for the stock index tends to have the upper hand in the allocation, whereas the weight for the gold index is only larger for a few occurrences. Often the gold weight quickly decreases after that. However, sharp increases in the gold weight do not seem to coincide with the recessions, as noted by the NBER Recession Indicator. Only in 2020, the increase in the gold weight coincides with the recession. Here, a flight-to-safety effect can be observed, as the weights for the gold index increase sharply during the crisis while the stock weights decrease. However, there seems to be a lagged effect during the 1990 recession and the Great Financial Crisis (GFC) from 2007 until 2009, but there was a decrease in the weight during the recession in 2001. These lagged effects are also visible in Figure 3b, where significant losses were still incurred during the crises in 2009 and 2020. Random forests can offer a solution where the idea is that changes in macroeconomic variables imply changes in portfolio weights. In this manner, the weights can be predicted more accurately, and the lagged effect becomes smaller. Therefore, this section continues by describing the parameter tuning and model estimation. Afterwards, the interpretation of the random forests is discussed. This section concludes with the performance of the portfolios created in this research.

5.1 Tuning hyperparameters and model estimation

The parameter `mtry` describes the fraction of random features used in creating a single tree in the random forest and is tuned to its optimal value using a grid search with five-fold cross-validation on the training sample. Table 1 shows the results for each model, where an `mtry` of 6 implies that $\frac{1}{6}$ of the features is randomly selected to grow the trees in the forest. Furthermore, the number of trees is set to 100, as is often done in previous literature (Pinelis & Ruppert, 2022). Figure C1 in Appendix C supports this choice, showing that the OOB errors stabilize when 100 trees or more are used. Following Coulombe (2021), the minimum node size is set to 12, as this paper deals with monthly returns. After parameter tuning, the random forests are re-estimated with the chosen parameters and the training sample. The last column of Table 1 represents the MSE obtained for the last tree (100^{th} tree). Note that the values for the MSE of the gold and stock weights are quite small. This seems promising, but caution must be taken here because the differences in weights between observations are already minimal, leading to smaller squared errors. Even more so for the utility model, as these contain even smaller values for the response

Figure 3. Optimal Weights and their Returns for the Training and Test Sample based solely on the Portfolio Optimization Problem from Equation 7

(a) Optimal Portfolio Weights over Time *(b) Returns for the Optimal Weights*



Note. Grey strokes in the background indicate months in which a recession occurred, based on the NBER Recession Indicator (National Bureau of Economic Research, 2023). Yellow background strokes indicate interesting events not based on the NBER Recession Indicator. In Figure 3a on the left, the gold weights are in **blue** and the stock weights in **red**). In Figure 3b on the right, the **red** line represents a 0% return. The **blue** line represents the mean of asset returns. The **black** dotted line represents the split for the training and test sample.

variable. The "pseudo" R^2 of the model is also noted. This is $R^2 = 1 - MSE/Var(y)$, where y denotes the response variable. The out-of-sample performance for the random forests is assessed using the Sharpe ratios for the test sample, which is discussed in Section 5.4.

Table 1. *Results for Hyperparameter Tuning of the Random Forests and Mean Squared Error (MSE) After Parameter Tuning*

Model	`mtry`	Nr. of trees	Min. node size	MSE	R^2
Weights for gold index	6	100	12	$6.876 * e^{-4}$	0.856
Weights for stock index	5	100	12	$6.578 * e^{-4}$	0.862
Utility model	4	100	12	$2.406 * e^{-9}$	0.965

Note. `mtry` represents the number of random features used in each tree as the fraction of the total number of features.

5.2 Variable importance

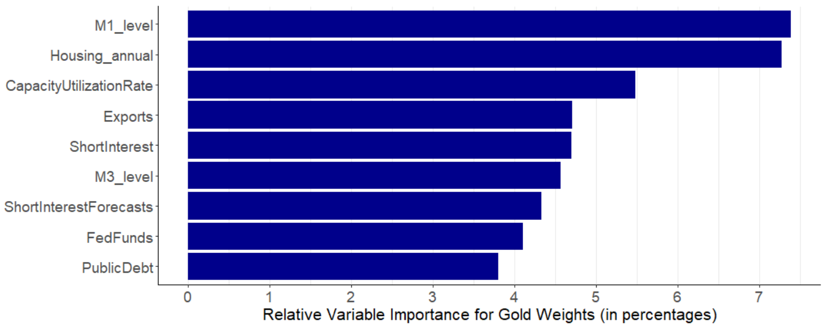
The estimator for variable importance (VI) assesses the increase in MSE when a feature is randomly permuted compared to when it is included in the forest, such that variables with a relatively large increase in MSE are considered more important. Since the features used in this research are often correlated, each variable is permuted three times, and the average increase in MSE over these three observations is taken. This should stable the estimators for VI a bit, as correlation can often cause unstable estimations of VI if the variable is permuted only once. However, these VI results should still be interpreted with caution. Figures 4a and 4b show the relative VI, i.e. the variable importance of a feature as a percentage of the total variable importance for all the features, for the nine most important features from the random forests models with gold and stock weights. The complete overviews of variable importance for both models are presented in Figures C2 and C3 in Appendix C. Furthermore, the increase in MSE is reported for both models in Tables D1 and D2 in Appendix D. For both stocks and gold, M1 or narrow money (without any transformations) – with a relative VI of 7.389% for gold and a relative VI of 6.681% for the stock weights – and the annual growth in housing prices – with a relative VI of 7.278% for gold and a relative VI of 5.382% for the stock weights – are the most dominant features in the data set. This partly confirms the statements of Iuorio (2023), as he discusses how the performance of the commodity gold is dependent on changes in money (M1).

Other important variables for both models are the capacity utilization rate, short-term interest rates, the quarterly forecast of these rates, the Fed Funds rate, and broad money (M3). For gold, exports and public debt are also considered important. Construction expenditures and personal consumption also seem relatively important for the stock weights. Interestingly, inflation does not seem important when assessing the portfolio weights, even though gold is often seen as an inflation hedge (Barsky et al., 2021). Similarly, long-term interest rates do not seem as important for this set-up's gold and stock weights, while short-term interest rates rank among the most important features. This also contradicts some of the remarks from Barsky et al. (2021). The unemployment rate and the VIX also do not seem important, ranking relatively low for both models. Moreover, GDP lists as the 12^{th} and 10^{th} most important variable for the stock and gold weights, respectively. Even though the literature describes the level of GDP to be quite influential (Ang, 2014). Finally, it is interesting that the NBER Recession is not considered important in any of the three models considered in this research, as it ranks last or second-to-last in every model.

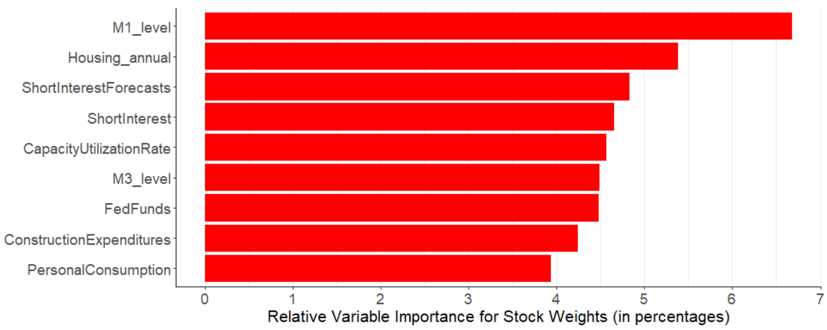
Relative variable importance for the nine most essential features from the random forest model with utility is shown in Figure 4c. The total figure is presented in Figure C4 in Appendix C, and the whole table can be found as Table D3 in Appendix D. Note that the utility, according to Equation 5, is the portfolio variance multiplied by minus one, but that here it is modelled as the portfolio variance itself. That is, a lower portfolio variance increases utility. This is done for interpretational convenience. Therefore, the relative VI shows what variables are important for reducing the variance of the portfolio. Figure 4c shows that narrow (M1) and broad money (M3) are especially important in determining the variance of a gold and stock portfolio with a relative VI of 7.117% for M1 and 6.867% for M3. The top five most important variables are public debt, Gross Domestic Product (GDP), and housing prices. Furthermore, sales of new family houses, government consumption and investment, unit labour costs and private fixed investments make up the top nine. Interestingly, GDP is considered more important in this model than the previous two models, with a relative VI of 6.602%/. Multiple variables differ from the most important ones for the gold and stock weights. For example, housing prices and sales of new family houses rank much lower, while they are in the top nine for the utility model. A possible explanation for this phenomenon is that the utility modelled in the random forest is already as small as possible as it is the decreased portfolio variance. This implies that some effects from the macro features are already controlled by choosing optimal weights, and thus effects cancel out. As the variables are all correlated, this might leave room for different features to become more critical.

Figure 4. Relative Variable Importance for the Macro Features from the Random Forests

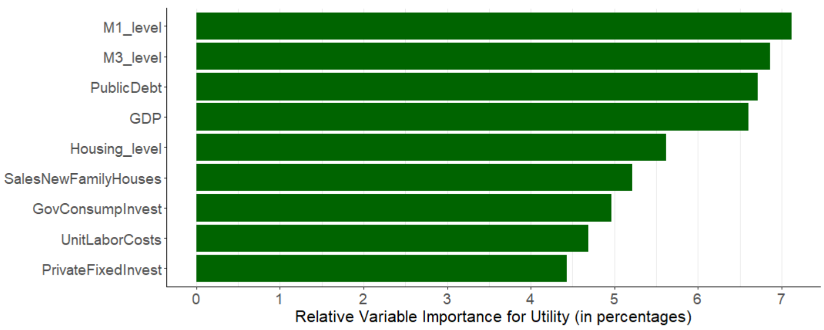
(a) Relative Variable Importance for the Gold Weights



(b) Variable Importance for the Stock Weights



(c) Variable Importance for Utility

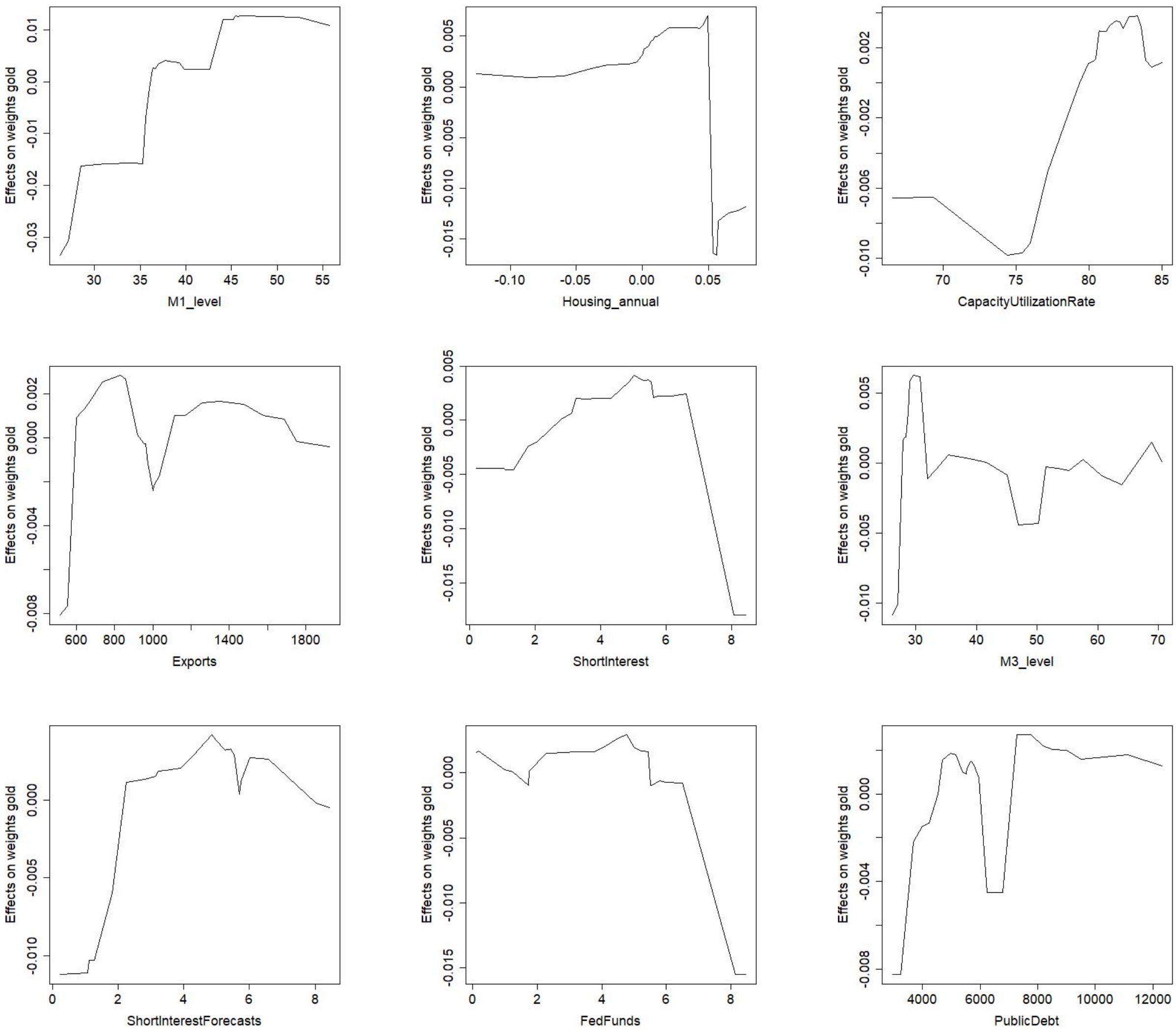


Note. The full figures with all features are presented in Appendix C.

5.3 Accumulated Local Effects

Using the relative importance discussed in the previous subsection, the nine most important features are selected for each model, and their accumulated local effects (ALE) are analyzed using ALE plots. The plots for the remaining 33 features in each of the models are presented in

Figure 5. The ALE Plots for the Nine Most Important Features from the Random Forests Model for the Gold Weights



Figures C5, C8, and C11 in Appendix C. As described in the Methodology section, accumulated local effects should be interpreted as the main effect of the feature at a specific value, compared to the average prediction of the response variable. Any time an effect of a feature on the response variable is described in this section, it is meant that this is the 'extra' effect of the feature on the average prediction, not the effect of the feature itself.

Figure 6. The ALE Plots for the Nine Most Important Features from the Random Forests Model for the Stock Weights

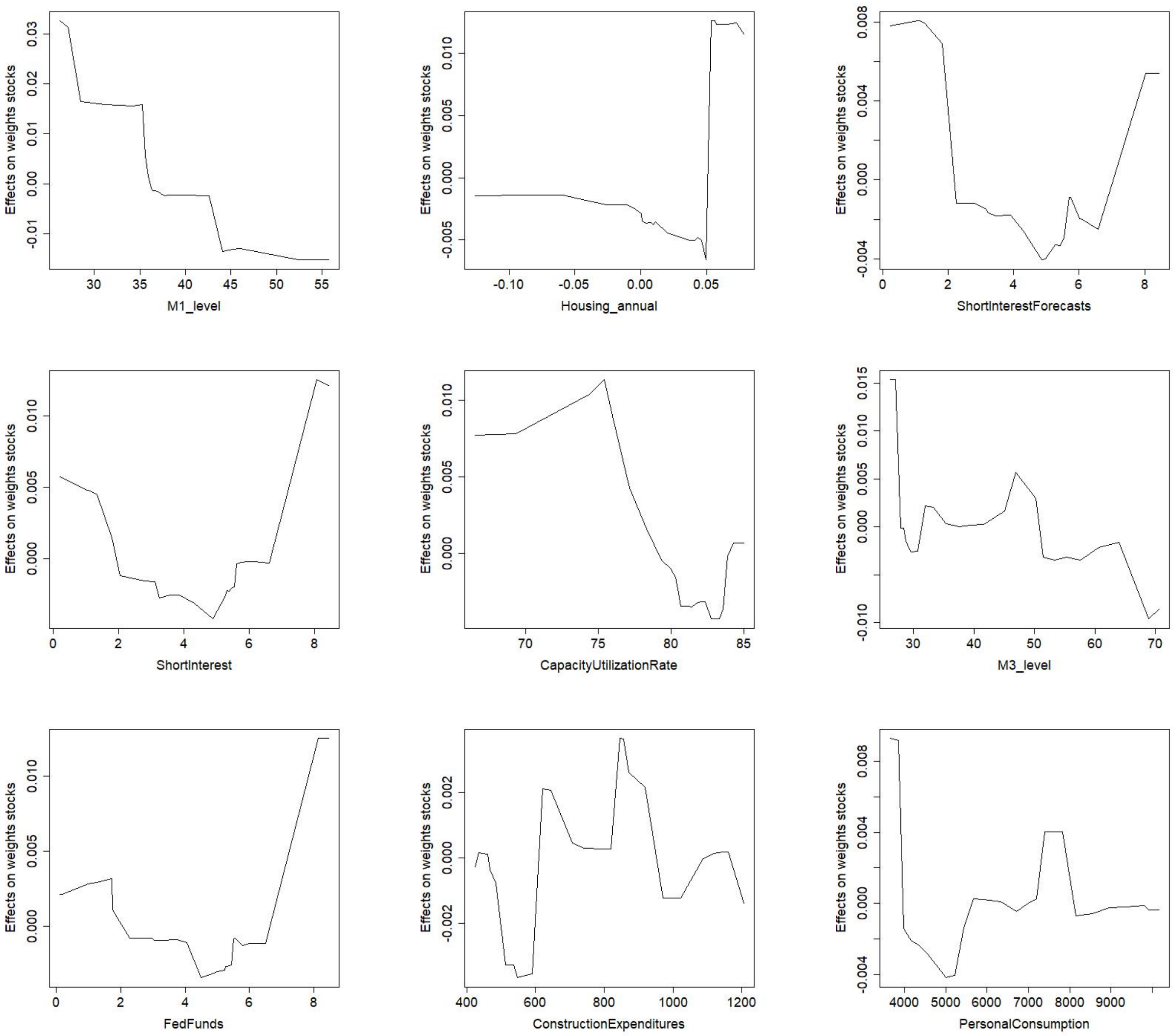
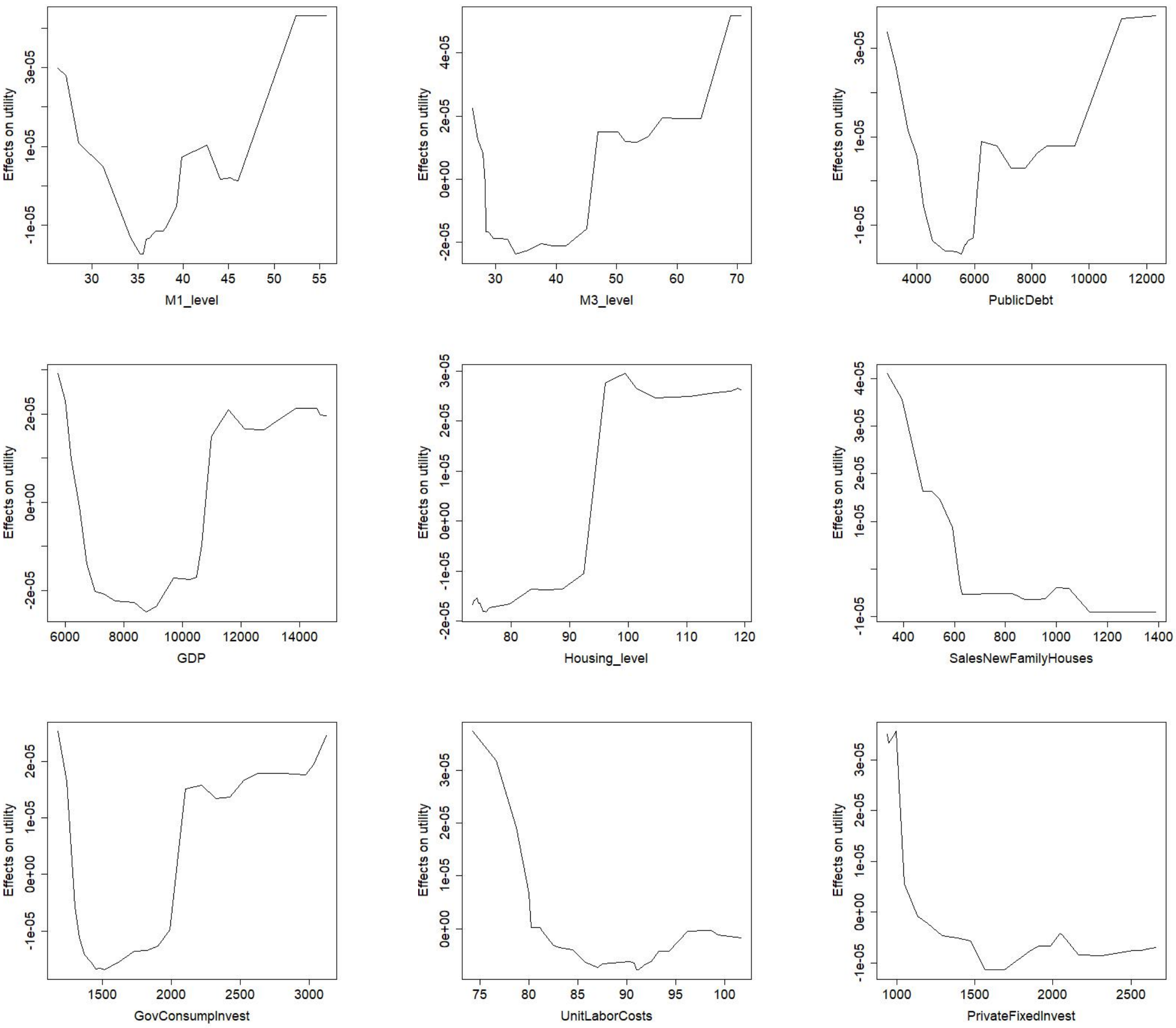


Figure 5 shows narrow money has a step-wise positive effect on the weights for the gold index. Annual growth in housing prices shows a significant drop in gold weights if prices increase by more than five per cent in one year. Increases in the capacity utilization rate seem to impact the weights for the gold index positively if the capacity utilization rate is sufficiently large. In contrast, the short-term interest rates and the fed funds rates have a significant adverse effect, especially when larger than seven per cent. Interestingly, the latter is not observed for short-term interest rate forecasts. One of the reasons could be that this variable is forward-looking and might already contain information that the short-term interest rates and fed funds rates do not have. Exports, broad money (M3) and public debt have a somewhat alternating effect on the optimal weights for the gold index, where they do not increase the average prediction much but can have a negative impact between specific ranges.

Figure 7. The ALE Plots for the Nine Most Important Features from the Random Forests Model for Utility



In general, for the effects observed in Figure 5, the inverse is observed for the stock weights in Figure 6. Where there is a step-wise positive effect of narrow money on the gold weights, this effect is step-wise negative for the stock weights. Similarly, a significant increase occurs for the stock weights when annual growth in housing prices is larger than five per cent, while a significant decrease occurs for the gold prices. The capacity utilization rate, short-term interest rates, fed funds rates, and short-term interest rate forecasts show similar inverse behaviour, where the short-term rates positively affect the stock weights, and the capacity utilization rate negatively affects the stock weights. M3 has an alternating effect over different values of M3 but seems to have an overall negative effect on the stock weights. This corresponds to the adverse effect observed for M1. Instead of public debt and exports, construction expenditures and personal consumption are among the most important variables for the stock weights. Personal consumption has an alternating but slightly more positive effect, whereas construction expenditures seem to alternate more around the average prediction.

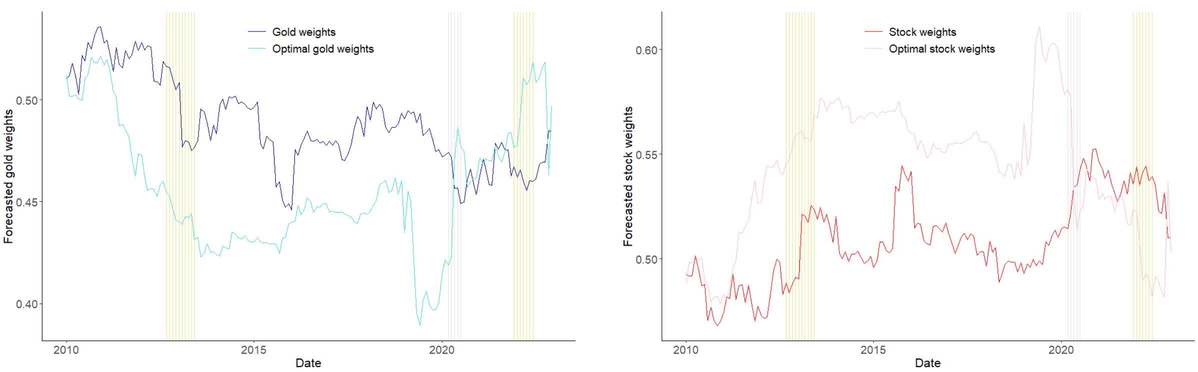
The ALE plots for the utility model in Figure 7 show that the portfolio variance is lowest when the narrow and broad money indices are around 35. A somewhat parabolic relation is generally observed for many of the essential variables. M1, M3, public debt, GDP and government consumption and investment all have an optimal range in which the features decrease the portfolio variance the most. Sales of new family houses, unit labour costs, and private fixed investments have a hyperbolic relation to the portfolio variance, where an increase in these features decreases portfolio variance. Housing prices seem to increase the portfolio variance, especially when the price index is around 95.

5.4 Out-of-sample performance of the portfolios

This research essentially creates three portfolios, one where the gold weights are modelled, and the stock weights are inferred, one that works vice versa, and one where utility is modelled, and both weights are reconstructed. Furthermore, the weights obtained from Equation 7 are also summarized in a portfolio to serve as a benchmark for comparison. The out-of-sample performances of these portfolios are assessed in multiple manners. First, the weights and portfolio returns are compared to the estimated weights and returns in Figures 3a and 3b. Figure 8 shows the forecasted weights from the random forest models. Remember that the weights presented in Figure 3a solve the quadratic programming problem from Equation 7 and that these weights are determined out-of-sample. As shown in Figure 8, there is quite a difference between the weights defined as optimal by the random forest models and the quadratic programming problem. This does not necessarily mean that one or the other is better. As mentioned, the optimal weights from Equation 7 seem to have some lagged effects, while the random forests could tackle this problem. In general, the random forests seem to predict a higher weight for the gold index and a lower weight for the stock index compared to Figure 3a. Interestingly, the random forest models do not necessarily predict a flight-to-safety for the gold index, as no increase was predicted for these weights during the Covid-19 crisis in 2020. Nevertheless,

Figure 8. Forecasted weights from the Random Forest Models for Gold and Stocks.

(a) Forecasted Gold Weights *(b) Forecasted Stock Weights*



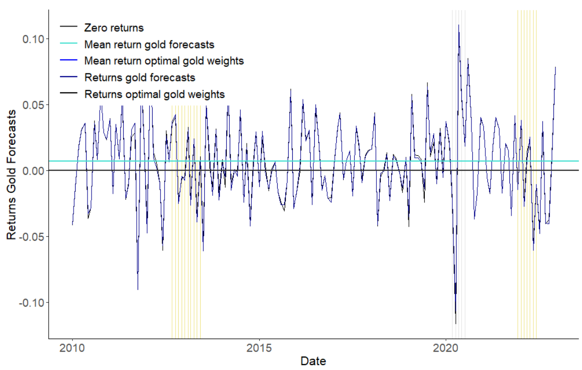
Note. Forecasted values for the gold weigths are in **blue** and its optimal values according to Figure 3a are in **turquoise**. Forecasted values for the stock weights are in **red** and its optimal values in **pink**.

both for the gold and stock weights, the burst of the gold bubble – as marked by the first yellow strokes in both figures – is predicted accurately by the random forest models, with a sharp drop in the gold weights and a simultaneous sharp increase in the stock weights. A similar occurrence can be observed during the second set of yellow strokes, where the gold price fell after the announcement of the aggressive policies of the FED. Since this research shows that the gold and stock weights primarily depend on the level of narrow money and broad money, this could be a direct reason for the quick adjustment that the random forests predict, and again confirms the ideas of Iuorio (2023) that the performance of the gold index is especially dependent on measures taken by the FED, which in this case is increasing the money in the economy. On the other hand, the NBER recession indicator ranks lowest or second-to-lowest in all of the models, which explains why the gold and stock weights, as predicted by the random forest, do not react much to the Covid-19 crisis. This is possibly why flight-to-safety is not observed for the random forest forecasts during this crisis.

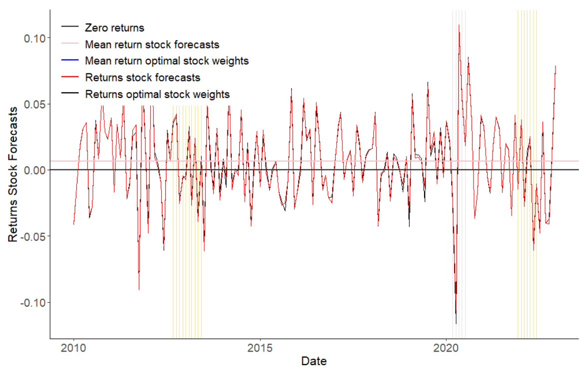
Figure 9 shows the realized returns for both portfolios with the forecasted weights from Figure 8. Here, it becomes clear that the forecasted weights do not significantly gain from the random forest predictions, compared to the realized returns obtained using only Equation 7. Furthermore, it seems to be the case that the portfolio is still exposed to a significant amount of market risk, as a significant loss occurs during the Covid-19 crisis. This implies that the random forest is not good at mitigating this market risk, given the features used in this research. An explanation could be that the relative VI is small for the NBER Recession indicator, such that the random forest misinterprets moments in which flight-to-safety should occur to minimize losses. Therefore, the random forests do not provide a significant benefit over the simple optimization problem from Equation 7. This is also confirmed by the Sharpe ratios presented in Table 2,

Figure 9. Returns Obtained from Forecasted weights for the Random Forest Models for Gold and Stocks.

(a) Returns for Forecasted Gold Weights



(b) Returns for Forecasted Stock Weights



Note. The **dark grey** line represents a 0% return. The **blue** line represents the mean of asset returns obtained from the optimal weights from Figure 3a. The **turquoise** and **pink** lines represent the mean return realized by the forecasted weights. The **dark blue** and **red** lines represent the realized returns from the forecasted weights. The **black** lines represent the returns as in Figure 3b.

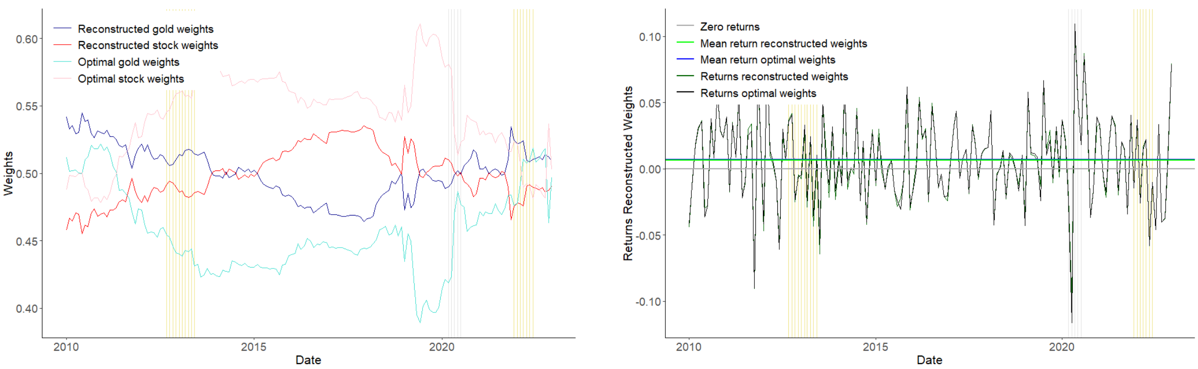
where the forecasted gold and stock weights from the random forests provide a Sharpe ratio of 0.201 for the gold weights and 0.201 for the stock weights.

Figure 10a shows the weights as reconstructed from the utility model following Equation 11. The patterns for the reconstructed weights and those found in Figure 3a are somewhat similar. Yet, the effect seems dimmed in that weight changes are less significant than those obtained from the quadratic programming problem. Again, the reconstructed weights adjust pretty well to the periods marked by the yellow strokes but do not seem to react as much to periods indicated as a recession by the NBER.

The consequence of the random forest not being able to recognize times of recession is the significant loss obtained during the beginning of 2020, as shown in Figure 10b. Also, the realized returns for the reconstructed weights are volatile in other periods, leading to the portfolio's lesser performance. This is also confirmed by Table 2, which shows that the portfolio with the reconstructed weights has a low Sharpe ratio of 0.196. This implies that it is slightly outperformed by the independently modelled weights and the optimization problem without the random forest.

Figure 10. Reconstructed Weights from the Utility Model and the Realized Returns from the Reconstructed Weights.

(a) Reconstructed Weights *(b) Realized Returns for the Reconstructed Weights*



Note. The **dark grey** line represents a 0% return. The **blue** line represents the mean of asset returns obtained from the optimal weights from Figure 3a. The **green** line represents the mean return realized by the reconstructed weights. The **dark green** lines represents the realized returns from the reconstructed weights, whereas the **black** lines represent the returns as in Figure 3b.

5.4.1 Comparison of the out-of-sample performances

Finally, comparing the performance of the portfolios against two benchmark portfolios is interesting. The first is the Markowitz portfolio, and the second is a minimum variance portfolio (MVP) estimated once over the training sample. Table 2 presents each portfolio's estimated mean return, standard deviation and Sharpe ratio. Overall, the mean return of the Markowitz and MVP portfolios is slightly smaller than those created in this research paper. However, the standard deviation for the benchmark portfolios is often slightly smaller than those from this research. Hence, no portfolio necessarily outperforms the benchmark portfolios, confirmed by the slight differences in Sharpe ratios. Unfortunately, the portfolio with the reconstructed weights does not outperform any other portfolios, while the independently modelled portfolios seem to have a slight gain compared to the benchmark portfolios.

Table 2. *Out-of-Sample Performance Comparison for the Three Portfolios and Simple Portfolios.*

Portfolio	Mean returns	St. Deviation	Sharpe ratio
Gold weights forecasted	0.007	0.034	0.201
Stock weights forecasted	0.007	0.034	0.201
Reconstructed weights	0.007	0.033	0.196
Weights from Equation 7	0.007	0.034	0.206
Markowitz Portfolio	0.006	0.033	0.197
Minimum Variance Portfolio	0.006	0.033	0.196

Note. The numbers are rounded to three decimal places.

Table 3 presents the test statistics from the robust test from Ledoit and Wolf (2008) for the difference in Sharpe ratios between a portfolio p on the rows and a portfolio q in the columns. In this manner, every test statistic represents the null hypothesis as presented in Equation 12. As shown in Table 2, there are only minor differences in Sharpe ratios. The test statistics from Table 3 confirm this, as none of the portfolio differences is significant. Based on these results, it can be concluded that the benefits of continuously rebalancing the portfolio and modelling the weights for the portfolios with random forests are minimal.

Table 3. *Test Statistics for the Difference in Sharpe Ratios with the Prewhitened HAC Test from Ledoit and Wolf (2008).*

Portfolios	Gold forecasted	Stock forecasted	Reconstructed weights	Weights from Equation 7
Gold forecasted	-	0.000 (0.001)	0.005 (0.005)	-0.005 (0.007)
Stock forecasted	0.000 (0.001)	-	0.004 (0.012)	-0.005 (0.007)
Reconstructed weights	-0.005 (0.005)	-0.004 (0.012)	-	-0.010 (0.009)
Weights from Equation 7	0.005 (0.007)	0.005 (0.007)	0.010 (0.009)	-
Markowitz Portfolio	-0.004 (0.004)	-0.004 (0.003)	0.001 (0.003)	-0.009 (0.009)
Minimum Variance Portfolio	-0.005 (0.005)	-0.005 (0.004)	0.000 (0.003)	-0.010 (0.010)

Note. The test statistic represents the difference in Sharpe ratios between portfolio p on the rows and portfolio q as the columns, such that the null hypothesis is as in Equation 12. Standard deviations are in brackets, where * marks significance at the 5% level. Numbers are rounded to three decimal places.

6 Discussion and Conclusion

This paper set out to find optimal portfolio weights for a gold-and-stock portfolio using random forests to answer the following research question:

> *To what extent can random forests predict the optimal portfolio weights as a function of macroeconomic variables for a portfolio with a gold index and a stock index?*

To answer this research question, two datasets with monthly observations are used. One with portfolio data for the gold and stock prices from January 1980 to December 2022, yielding 43 years of observations. The other dataset consists of the macroeconomic variables from December 1989 to November 2022, yielding 33 years of observations. Optimal weights are created out-of-sample, serving as the random forests' training sample. Three random forests are modelled, one for the weights of the gold index, one for the stock index, and a third for the utility obtained from holding the portfolio with optimal weights. Based on these random forests, predictions for the optimal portfolio weights are made for the test sample.

In general, narrow money (M1) and the annual growth in housing prices are the most important features for predicting the stock and gold weights in the random forests model. Other important features are short-term interest rates, short-term forecasted interest rates, the Fed funds rate, the capacity utilization rate, and broad money (M3). For utility, i.e. the variance of the portfolio, narrow and broad money are the most important. Other important variables are public debt, GDP, unit labour costs, and housing prices. Narrow money and the capacity utilization rate have an overall positive effect on the gold weights. In contrast, the annual growth in housing prices, short-term interest rates, and the Fed funds rate have an overall negative effect. For the stock weights, these relationships are inverted. In the utility model, the features seem to have a parabolic or hyperbolic relationship with the portfolio variance. Narrow and broad money, public debt, and GDP are parabolic, whereas unit labour costs have a hyperbolic relationship with the portfolio variance.

The independently forecasted weights for the gold and stock index show that the predictions of the random forests do not detect recessions well. This is likely because the NBER recession indicator has low relative variable importance in every model. This leads to a persistent loss in returns during a crisis such as the Covid-19 crisis in 2020. Unfortunately, this also implies that the out-of-sample performance for the portfolios created from the forecasted gold and stock weights do not statistically outperform the benchmark portfolios. The weights reconstructed from the utility model change moderately over the test sample. While this is good for limiting transaction costs, the reconstructed weights do not adapt well to signals from macroeconomic features, especially in recessions. This causes the portfolio with the reconstructed weights to perform similarly to the benchmark portfolios.

To conclude, the portfolio rebalancing with the random forests of macroeconomic features used in this paper does not achieve significantly higher out-of-sample Sharpe ratios than simple benchmark portfolios such as the Markowitz (1952) portfolio. While the conclusions drawn from the variable importance and ALE plots for the random forests are interesting and confirm some other empirical findings, no significant benefits are realized using random forests to model optimal portfolio weights. Considering the computational efforts compared to a simple, equally-weighted portfolio, the random forests models should be significantly improved to provide investors with an excellent alternative to the benchmark portfolios.

The most important limitation of this research is the high correlation among the features in the random forests. This causes noisy estimates in variable importance and can increase overfitting in the model. While random forests are a better alternative to model these highly-correlated variables than simple regression trees, it might very well be the case that this causes bad out-of-sample performance. When the model cannot determine the most relevant variables in the dataset, i.g. the variable that decreases the MSE the most, single trees in the forest are likely to provide wrong estimates for the response variable. If this occurs often in the tree, it can worsen the performance of the entire forest.

Another limitation of the random forests and regression trees is that they tend to be not very good at extrapolating observations out-of-sample. Given the previously fitted data from the training sample, random forests predict values based on where you end up in the tree in the test sample. Therefore, it is more likely that the random forest predicts a value that it has visited before in the training sample or that is within the values observed at a certain leaf node in the training sample. This could cause the portfolio weights not to react as significantly to some of the signals from the macro features compared to the optimal portfolio weights. Consequently, this might explain why the random forests are less good at predicting sharper increases or decreases in the stock or gold weights.

Finally, the random forests in this paper use the MSE as splitting criteria. While this is most commonly done for random forests, it might not be the most suitable choice for portfolio weights. As transaction costs bound changes in portfolio weights, the weights tend to be relatively close to one another from period to period. On top, forecasting the portfolio weights results in small values – most often between zero and one – such that errors in forecasted values also tend to be relatively small. Squaring values smaller than one leads to even smaller values, making it hard for the random forests to detect which variables offer the best split and which variables significantly increase model performance and which do not.

Based on previous literature and other empirical applications, random forests are still promising in this line of research. It is, however, a must that the models are improved first. Therefore, I suggest the following for future research. First, an interesting suggestion would be to investigate whether the problems with high correlation among macro features can be solved. A possible solution could be to include a very limited amount of features in the sample and restrict the random choice for features to variables that are not correlated as much. This could offer trees in the forest that consist of features with relatively low correlation. Furthermore, `mtry` could be even larger, so only small fractions of the total set of features are included in each tree.

Unfortunately, the second limitation is harder to improve, as random forests inherently have this limitation. A possible suggestion could be to obtain a larger data set, such that leaves are visited more often and better estimates are provided for the response variables. Finally, the third limitation can be mitigated by adjusting the splitting criteria. Therefore, a suggestion for future research could be first to determine the optimal splitting criteria, e.g. Root Mean Squared Error, Mean Absolute Error or other criteria. Then, the optimal splitting criteria can possibly be used to provide better random forest models, such that better forecasting performance is obtained.

References

Al-Ameer, M., Hammad, W., Ismali, A. & Hamdan, A. (2018). The relationship of gold price with the stock market: The case of Frankfurt Stock Exchange. *International Journal of Energy Economics and Policy*, *8*(5), 357–371.

Ang, A. (2014). *Asset management: A systematic approach to factor investing.* Oxford University Press.

Ang, A. & Bekaert, G. (2004). How regimes affect asset allocation. *Financial Analysts Journal*, *60*(2), 86–99.

Apley, D. W. & Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *82*(4), 1059–1086.

Barsky, R., Epstein, C., Lafont-Mueller, A. & Yoo, Y. (2021). What drives gold prices? *Chicago Fed Letter*, *464*. Retrieved from `https://www.chicagofed.org/publications/chicago-fed-letter/2021/464`

Boucher, C. & Tokpavi, S. (2019). Stocks and bonds: Flight-to-safety for ever? *Journal of International Money and Finance*, *95*, 27–43.

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

Bryzgalova, S., Pelger, M. & Zhu, J. (2021). Forest through the trees: Building cross-sections of stock returns. *Working Paper*.

Carrizosa, E., Molero-Río, C. & Morales, D. R. (2021). Mathematical optimization in classification and regression trees. *TOP*, *29*, 5–33.

Chakrabarti, D. (2021). Parameter-free robust optimization for the maximum-sharpe portfolio problem. *European Journal of Operations Research*, *293*(1), 388–399.

Chicago Board Options Exchange. (2023). Historical Data for CBOE VIX Index and Other Volatility Indices. *CBOE - database*.

Chirwa, T. G. & Odhiambo, N. M. (2020). Determinants of gold price movements: An empirical investigation in the presence of multiple structural breaks. *Resources Policy*, *69*, 1–9.

Choudry, T., Hassan, S. S. & Shabi, S. (2015). Relationship between gold and stock markets during the global financial crisis: Evidence from nonlinear causality tests. *International Review of Financial Analysis*, *41*, 247–256.

Coulombe, P. G. (2020). To bag is to prune. *Working paper*.

Coulombe, P. G. (2021). The macroeconomy as random forest. *Working paper*.

Cutler, D. M., Poterba, J. M. & Summers, L. H. (1988). What moves stock prices? *Working Paper*.

De Nard, G., Ledoit, O. & Wolf, M. (2021). Factor models for portfolio selection in large dimensions: The good, the better and the ugly. *Journal of Financial Econometrics*, *19*(2), 236–257.

DeMiguel, V., Garlappi, L. & Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *Review of Financial Studies*, *22*(5), 1915–1953.

Detzel, A., Novy-Marx, R. & Velikov, M. (2023). Model comparison with transaction costs. *The Journal of Finance*, *78*(3), 1743–1775.

Federal Reserve Bank of St. Louis. (2023). Federal Reserve Economic Data. *Economic Research Federal Reserve Bank of St. Louis - database*. Retrieved from `https://fred.stlouisfed.org/`

Guidolin, M. (2011). Markov switching models in empirical finance. *Working paper*.

Ilmanen, A. (2003). Stock-bond correlations. *The Journal of Fixed Income*, *13*(2), 55–56.

Iuorio, J. (2023). How does gold perform with inflation, stagflation and recession? *Forbes*. Retrieved from `https://www.forbes.com/sites/cme-group/2023/06/02/how-does-gold-perform-with-inflation-stagflation-and-recession/`

Jagannathan, R. & Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, *58*(4), 1651–1683.

Kazak, E., Li, Y., Nolte, I. & Nolte, S. (2022). Direct portfolio weight estimator: Mitigating specification risk with realized utility. *Working paper*.

Ledoit, O. & Wolf, M. (2008). Robust performance hypothesis testing with the sharpe ratio. *Journal of Empirical Finance*, *15*, 850–859.

Loh, W.-Y. (2011). Classification and regression trees. *Wiley interdisciplinary reviews: data mining and knowledge discovery*, *1*(1), 14–23.

Loh, W.-Y. (2014). Fifty years of classification and regression trees. *International Statistical Review*, *82*(3), 329–348.

Maio, P. & Philip, D. (2013). Macro factors and the cross-section of stock returns. *Working paper*.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, *7*(1), 77–91.

National Bureau of Economic Research. (2023). NBER Recession Indicator - data. *NBER Recession Indicator - Data*. Retrieved from `https://www.nber.org/research/business-cycle-dating`

Neifar, M., Dhouib, S., Bouhamed, J., Ben Abdallah, F., Arous, I., Ben Braiek, F. & Mrabet, D. (2021). The impact of macroeconomic variables on stock market in United Kingdom. *Working paper*.

Oliveira, A. B. & Valls Pereira, P. L. (2018). Asset allocation with markovian regime switching: Efficient frontier and tangent portfolio with regime switching. *Working paper*.

Organisation for Economic Co-operation and Development. (2023). OECD Data. *OECD - database*. Retrieved from `https://data.oecd.org/`

Pellegrino, F. (2021). Factor-augmented tree ensembles. *Working paper*.

Pinelis, M. & Ruppert, D. (2022). Machine learning portfolio allocation. *The Journal of Finance and Data Science*, *8*, 35–54.

Refinitiv. (2023). Eikon Refinitiv Datastream. *Refinitiv Datastream - database*.

Rokach, L. (2016). Decision forest: Twenty years of research. *Information Fusion*, *27*, 111–125.

Samitas, A. & Armenatzoglou, A. (2014). Regression tree model versus markov regime switching: a comparison for electricity spot price modelling and forecasting. *Operational Research*, *14*, 319–340.

Sharpe, W. F. (1964). Captial asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, *19*(3), 425–442.

Sharpe, W. F. (2007). Expected utility asset allocation. *Financial Analysts Journal*, *63*(5), 18–30.

von Neumann, J. & Morgenstern, O. (1994). *Theory of games and economic behavior*. Princeton University Press.

A Macroeconomic variables and their transformations

Table A1. *Macroeconomic Variables and their Transformations as used in the Random Forests. The Frequency of the Data and the Databases from which the Data is obtained are also mentioned.*

Macroeconomic Variable	Transformation	Frequency	Database obtained from
Long-term interest rates	-	Monthly	OECD
Long-term interest rates forecasts	-	Quarterly	OECD
Short-term interest rates	-	Monthly	OECD
Short-term interest rates forecasts	-	Quarterly	OECD
Fed funds rate	-	Monthly	FED St. Louis
Yield ten-year treasury-bill	-	Monthly	FED St. Louis
Inflation (% annual growth)	-	Monthly	OECD
Inflation (% annual growth)	Demeaned	Monthly	OECD
Unemployment rate	-	Monthly	Refinitiv Datastream
Volatility (VIX)	-	Monthly	CBOE
Exchange rate EUR/USD	-	Monthly	Refinitiv Datastream
Exchange rate GBP/USD	-	Monthly	Refinitiv Datastream
Narrow money (M1)	-	Monthly	OECD
Narrow money (M1)	Monthly growth	Monthly	OECD
Narrow money (M1)	Annual growth	Monthly	OECD
Broad money (M3)	-	Monthly	OECD
Broad money (M3)	Monthly growth	Monthly	OECD
Broad money (M3)	Annual growth	Monthly	OECD
Gross Domestic Product (GDP) (level)	-	Quarterly	Refinitiv Datastream
Real GDP forecasts (annual growth)	-	Quarterly	OECD
Housing prices	-	Quarterly	OECD
Housing prices	Quarterly growth	Quarterly	OECD
Housing prices	Annual growth	Quarterly	OECD
NBER Recession Indicator	-	Monthly	St. Louis FED
Personal consumption expenditures	-	Quarterly	Refinitiv Datastream
Government consumption and investment	-	Quarterly	Refinitiv Datastream
Private domestic fixed investment	-	Quarterly	Refinitiv Datastream
Exports	-	Quarterly	Refinitiv Datastream
Imports	-	Quarterly	Refinitiv Datastream
Foreign reserve assets	-	Quarterly	Refinitv Datastream
Public debt	-	Quarterly	Refinitiv Datastream
Consume confidence index	-	Monthly	Refinitiv Datastream
Industrial production index	-	Monthly	Refinitiv Datastream
Unit labour costs	-	Quarterly	Refinitiv Datastream

The table continues on the next page

Macroeconomic Variable	Transformation	Frequency	Database obtained from
Capacity utilization rate	-	Monthly	Refinitiv Datastream
Housing authorized	-	Monthly	Refinitiv Datastream
New private housing units started	-	Monthly	Refinitiv Datastream
New private housing units authorized	-	Monthly	Refinitiv Datastream
Construction expenditures	-	Monthly	Refinitiv Datastream
Bankruptcy filings	-	Quarterly	Refinitiv Datastream
Goods and Services balance	-	Monthly	Refinitiv Datastream
Sales of new family houses	-	Monthly	Refinitiv Datastream

B Descriptive statistics

Table B1 shows the descriptive statistics for the macroeconomic variables included in the data.

Table B1. *Descriptive Statistics for the Macroeconomic Features Included in the Model over the period December 1989 to November 2022*

Variable	Name in dataset	Mean	St. Dev	Min.	Max.
Long-term interest rates	LongInterest	4.270	2.030	0.620	8.890
Long-term interest rates forecasts	LongInterestForecasts	4.270	2.023	0.650	8.703
Short-term interest rates	ShortInterest	2.920	2.376	0.090	8.420
Short-term interest rates forecasts	ShortInterestForecasts	2.913	2.371	0.100	8.437
Fed funds rate	FedFunds	4.394	2.384	0.050	19.100
Yield ten-year treasury bill	TreasuryYield10	2.716	2.037	0.550	15.840
Inflation	Inflation	3.321	1.630	-2.097	14.756
Inflation demeaned	InflationDemeaned	-0.699	1.630	-5.442	5.715
Unemployment rate	Unemployment	6.170	1.730	3.500	14.700
Volatility (VIX)	Volatility	19.740	7.797	9.450	68.510
Exchange rate EUR/USD	FXEUR	0.839	0.110	0.635	1.181
Exchange rate GBP/USD	FXGBP	0.643	0.081	0.480	0.887
Narrow money (M1)	M1 level	103.650	0.015	26.240	683.920
Narrow money (M1) Monthly	M1 monthly	0.0111	0.120	-0.033	2.388
Narrow money (M1) Annual	M1 annual	0.160	0.573	-0.054	3.612
Broad money (M3)	M3 level	69.760	40.149	26.170	180.180
Broad money (M3) Monthly	M3 monthly	0.005	0.005	-0.006	0.064
Broad money (M3) Annual	M3 annual	0.062	0.040	0.002	0.269
Gross Domestic Product (GDP) (level)	GDP	13622	5310	5754	26138
Real GDP forecasts (annual growth)	RealGDPgrowth	0.025	0.047	-0.299	0.353
Housing prices	Housing level	97.050	19.521	73.390	152.430
Housing prices Monthly	Housing monthly	0.005	0.013	-0.037	0.041
Housing prices Annual	Housing annual	0.022	0.048	-0.126	0.134
NBER Recession Indicator	NBERRecession	0.091	0.288	0.000	1.000
Personal consumption expenditures	PersonalConsumption	9146	3690	3654	17750
Government consumption and investment	GovConsumpInvest	2548	925	1180	4575
Private domestic fixed investment	PrivateFixedInvest	2331.100	932.076	940.100	4508.200
Exports	Exports	1551.400	713.480	515.400	3065.000
Imports	Imports	1993.200	929.575	598.200	4074.400

The table continues on the next page

Variable	Name in dataset	Mean	St. Dev	Min.	Max.
Foreign reserve assets	ForeignAssetsReserve	102866	41072	64222	253217
Public debt	PublicDebt	11805	7805	2953	31420
Consumer confidence index	ConsumerConfidence	95.180	26.530	25.300	144.700
Industrial production index	IndustrialProduction	89260	13.144	60.310	104.120
Unit labour costs	UnitLaborCosts	95.820	12.123	74.220	127.920
Capacity utilization rate	CapacityUtilizationRate	78.770	3.716	64.570	85.010
Housing authorized	HousingAuth	112.030	36.651	36.300	211.900
New private housing units started	PrivateHousingStarted	1320	393	478	2273
New private housing units authorized	PrivateHousingAuth	1345	408	513	2263
Construction expenditures	ConstructionExpenditures	937.600	342.474	424.800	1840.300
Bankruptcy filings	BankruptcyFilings	40392	15872	12748	73232
Goods and Services balance	GoodsServicesBalance	-35616	21277	-102536	-63
Sales of new family houses	SalesNewFamilyHouses	699.000	251.501	270.000	1389.000

Note. The numbers are rounded to three decimal places.

Table B2 shows the descriptive statistics for the portfolio data used in this research.

Table B2. *Descriptive Statistics of the Prices and Monthly Returns for the S&P GSCI Gold Index and the S&P 500 Stock Index over the period January 1980 to December 2022*

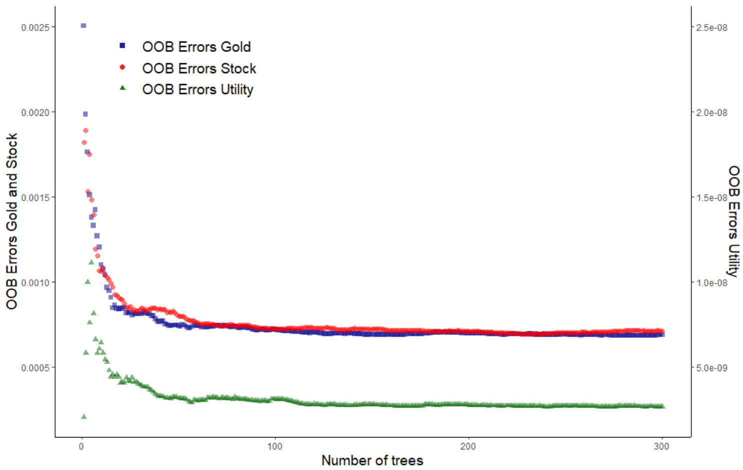
Variable	Mean	St. Dev.	Min.	Max.
S&P GSCI Gold Price	430.900	294.742	148.800	1156.400
S&P 500 Price	1197.200	1032.166	102.200	4796.600
S&P GSCI Gold Returns	0.004	0.051	-0.197	0.301
S&P 500 Returns	0.008	0.045	-0.219	0.157

Note. The numbers are rounded to three decimal places.

C Additional Figures

Figure C1 shows the OOB errors, where the OOB errors stabilize when 100 trees are grown in the forest.

Figure C1. The Out-Of-Bag (OOB) Errors for an Increasing Amount of Trees Grown in the Random Forests.



Note. The OOB errors for the model with gold weights are in **blue** on the left axis, the OOB errors for the model with stock weights are in **red** on the left axis, and the OOB errors for the utility model are in **green** on the right axis.

Figures C2, C3, and C4 show the variable importance for the three random forest models.

Figure C2. Variable Importance for the Macro Features from the Random Forest Model for the Gold Weights

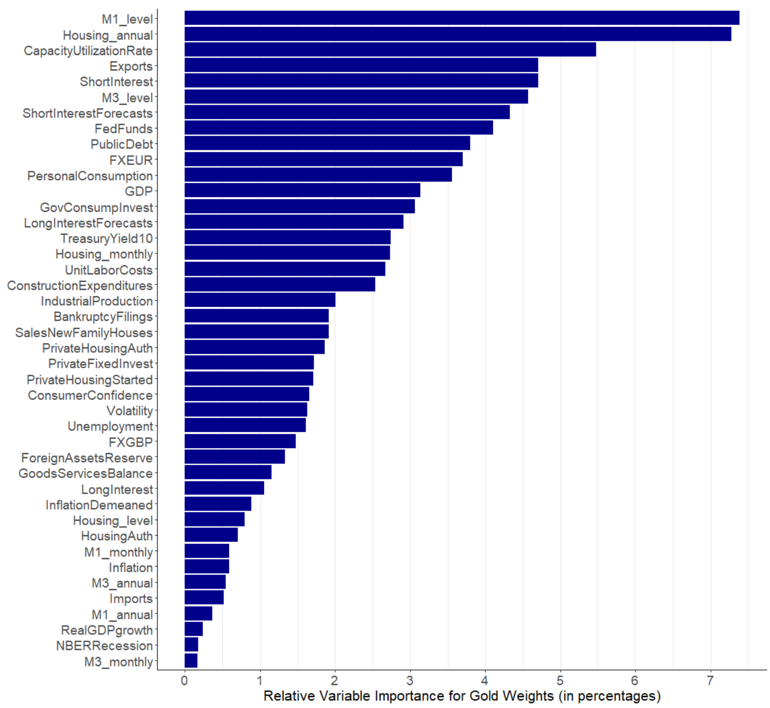


Figure C3. Variable Importance for the Macro Features from the Random Forest Model for the Stock Weights

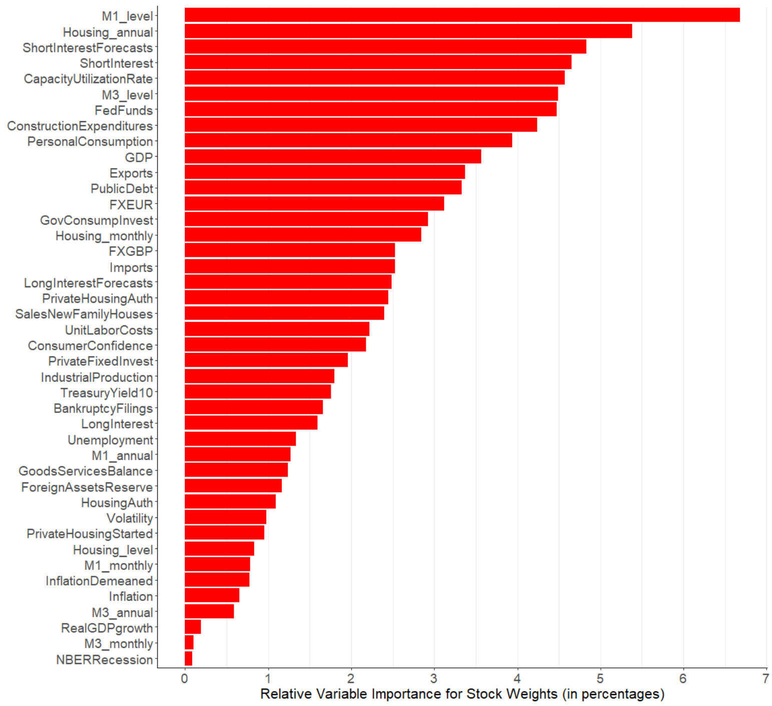
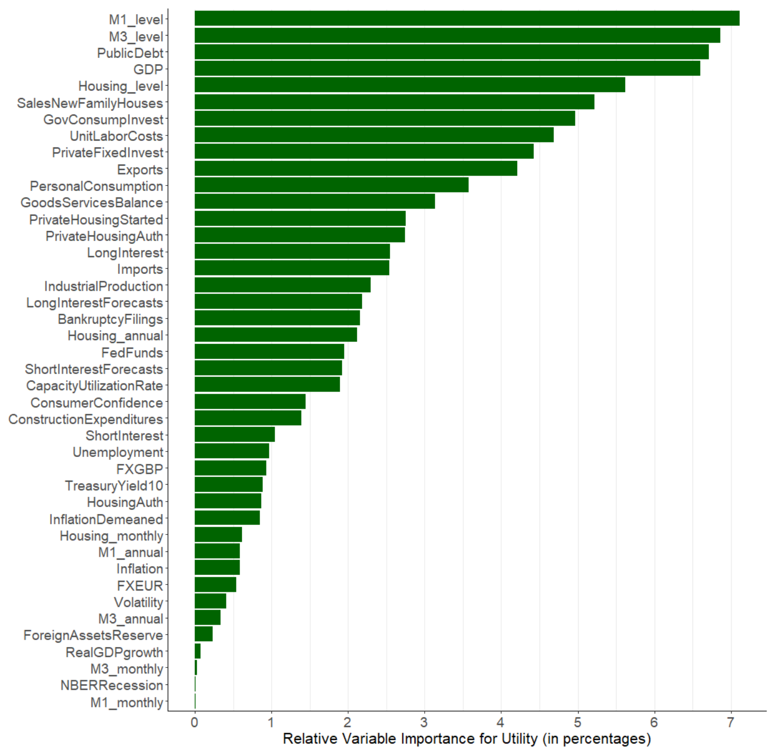
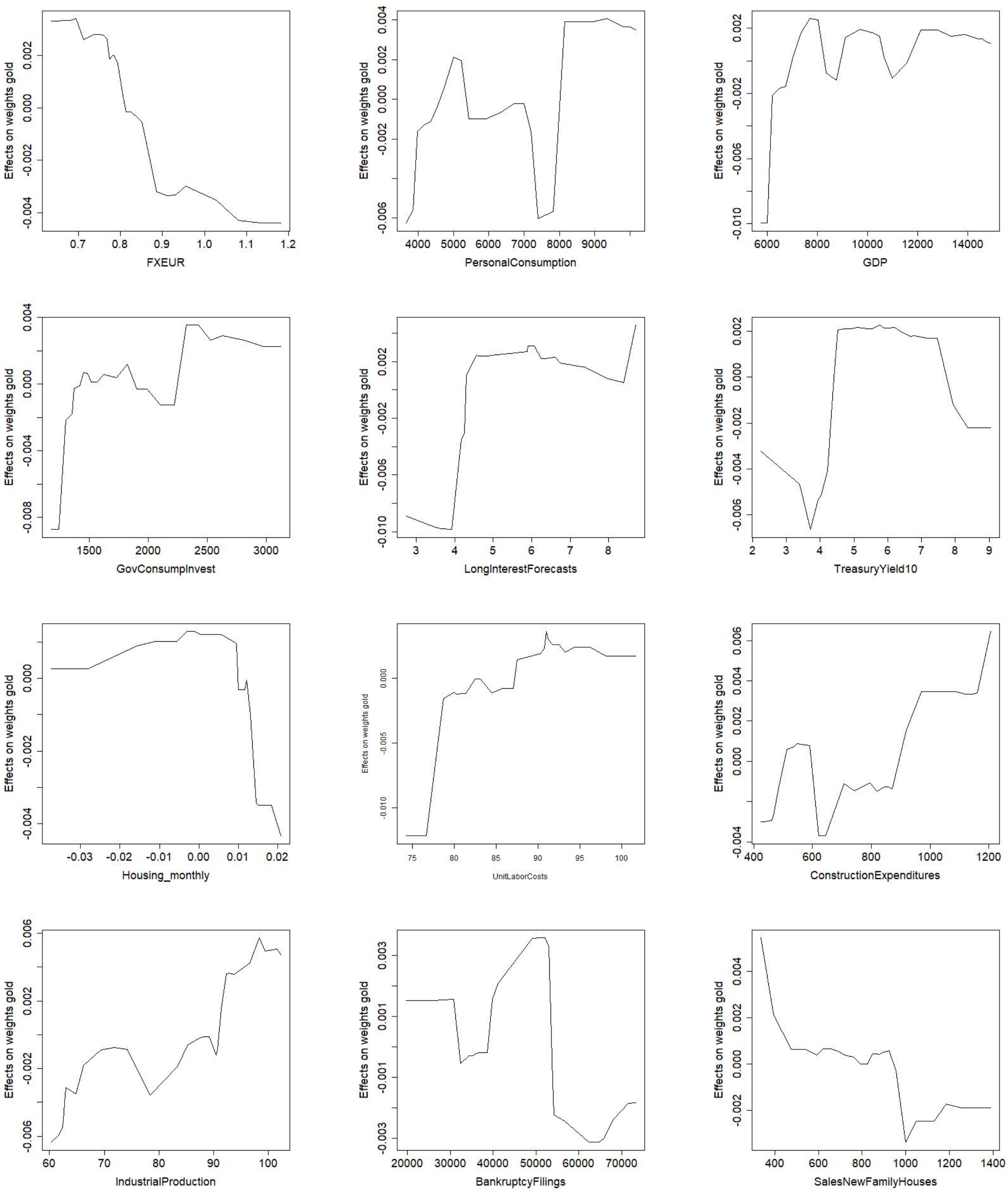


Figure C4. Variable Importance for the Macro Features from the Random Forest Model for Utility

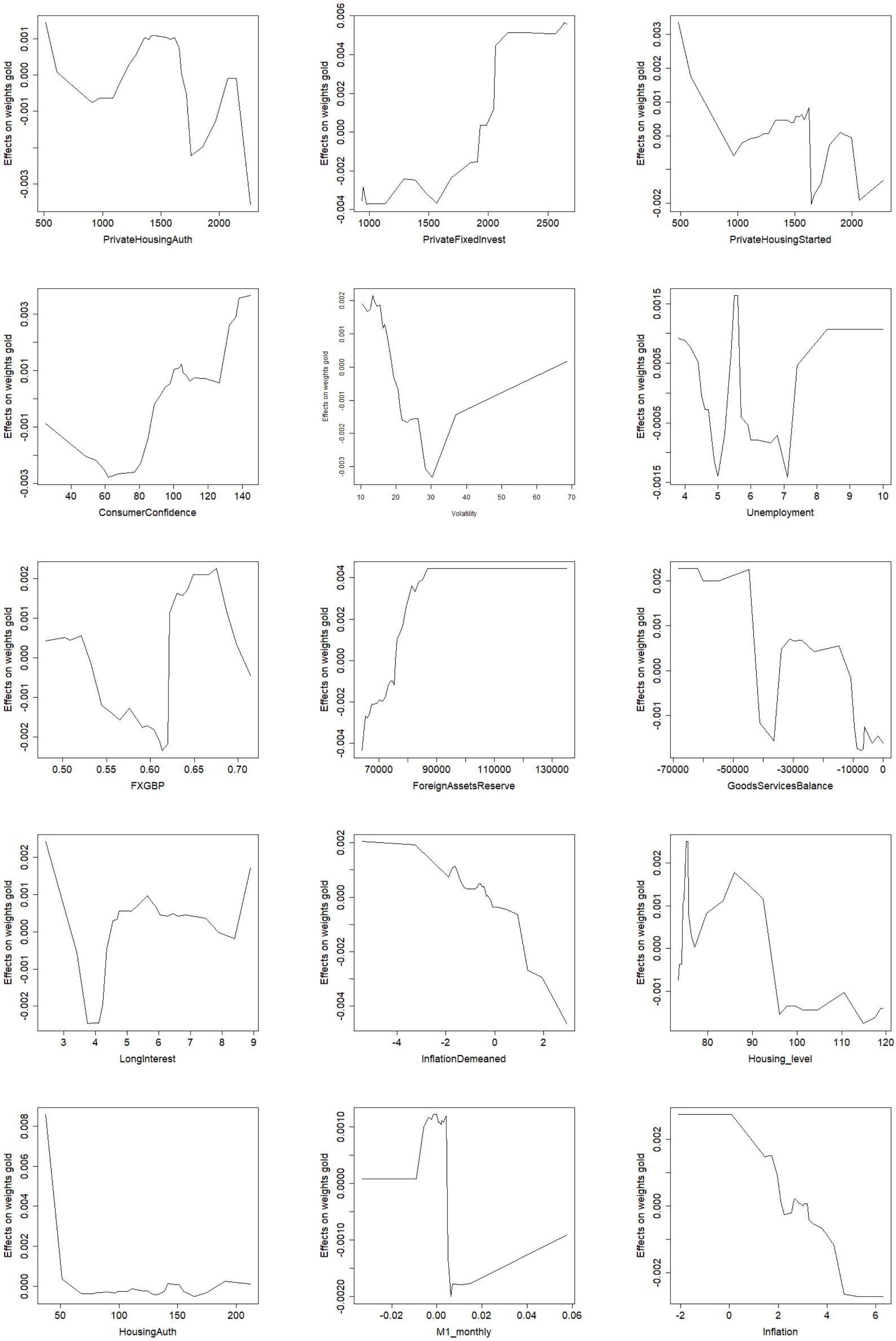


C.1 ALE Plots for the Gold, Stock, and Utility models

Figure C5. The ALE Plots for the Remaining 33 Features from the Random Forests Model for the Gold Weights in order of importance according to Figure C2.



The figure continues on the next two pages.


Effects on weights gold
PrivateHousingAuth
PrivateFixedInvest
PrivateHousingStarted
ConsumerConfidence
Volatility
Unemployment
FXGBP
ForeignAssetsReserve
GoodsServicesBalance
LongInterest
InflationDemeaned
Housing_level
HousingAuth
M1_monthly
Inflation

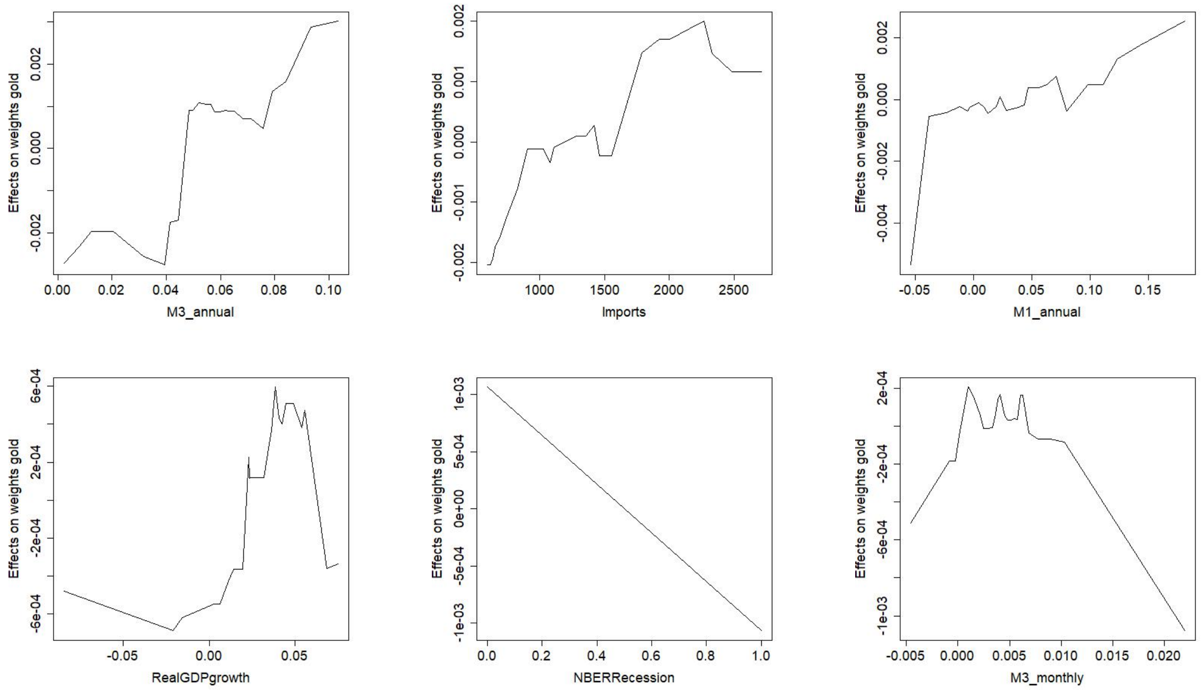
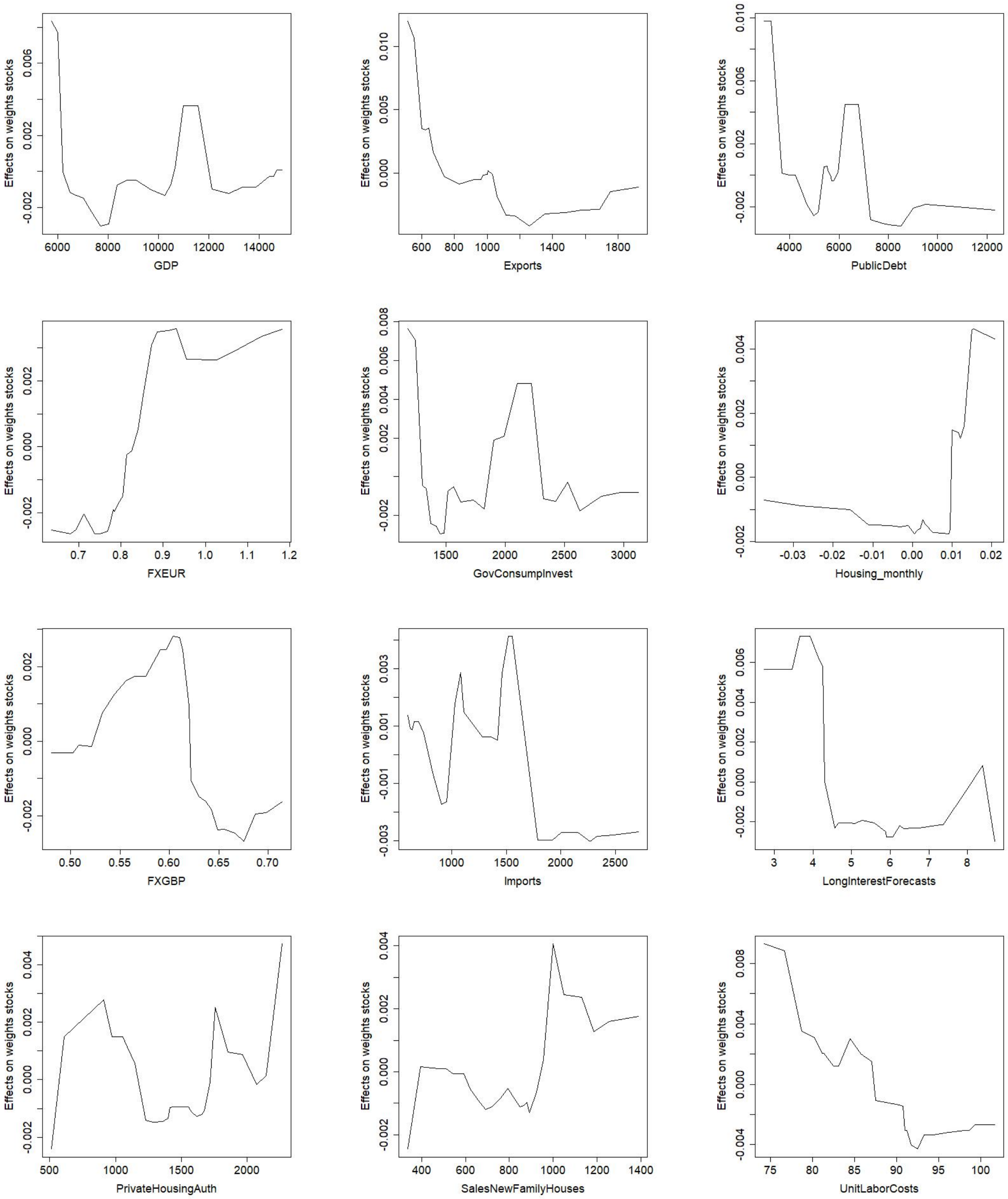
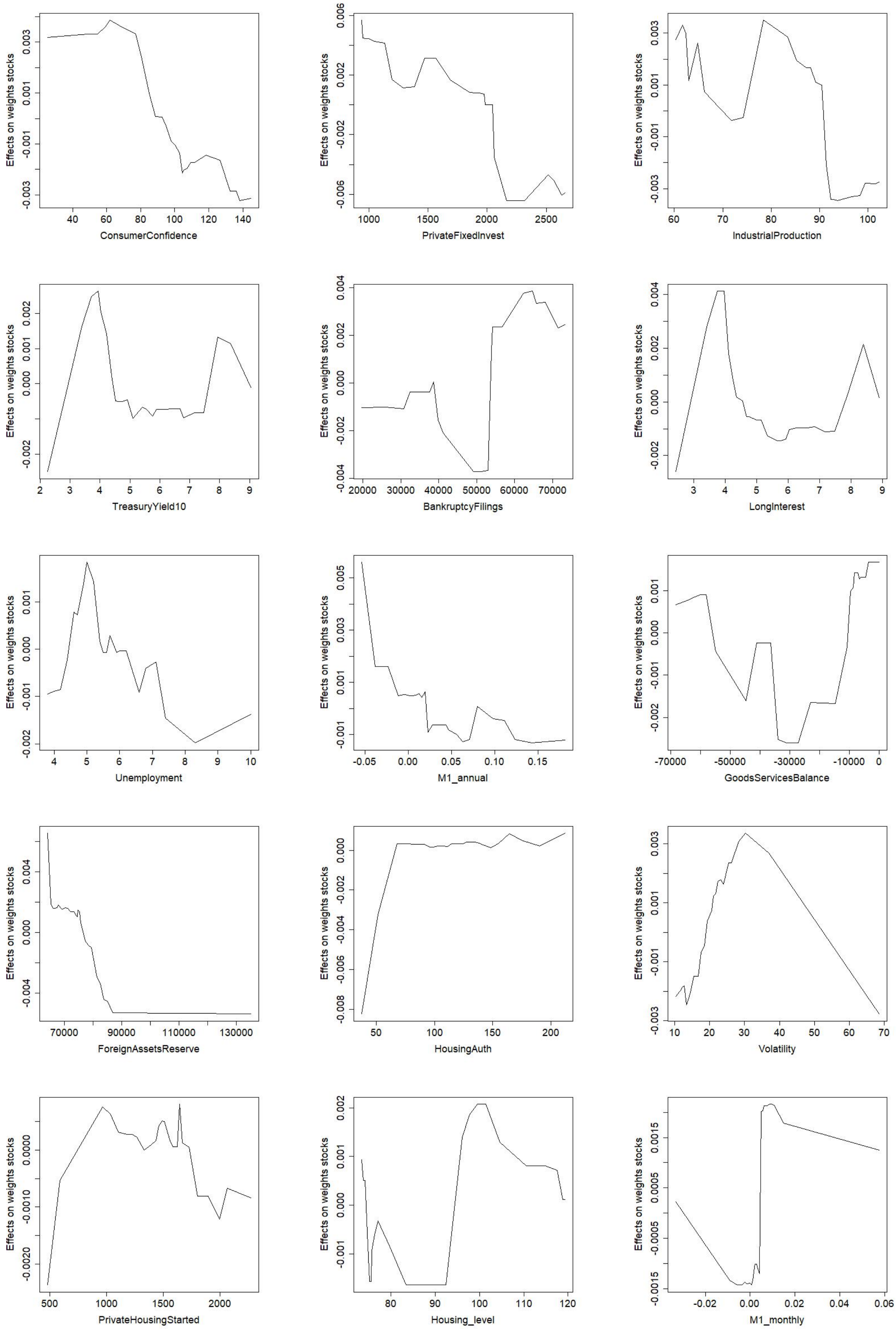

Effects on weights gold
M3_annual
Effects on weights gold
Imports
Effects on weights gold
M1_annual
Effects on weights gold
RealGDPgrowth
Effects on weights gold
NBERRecession
Effects on weights gold
M3_monthly

Figure C8. The ALE Plots for the 33 Remaining Features from the Random Forests Model for the Stock Weights in order of importance according to Figure C3.



The figure continues on the next two pages.


Effects on weights stocks
ConsumerConfidence
PrivateFixedInvest
IndustrialProduction
TreasuryYield10
BankruptcyFilings
LongInterest
Unemployment
M1_annual
GoodsServicesBalance
ForeignAssetsReserve
HousingAuth
Volatility
PrivateHousingStarted
Housing_level
M1_monthly

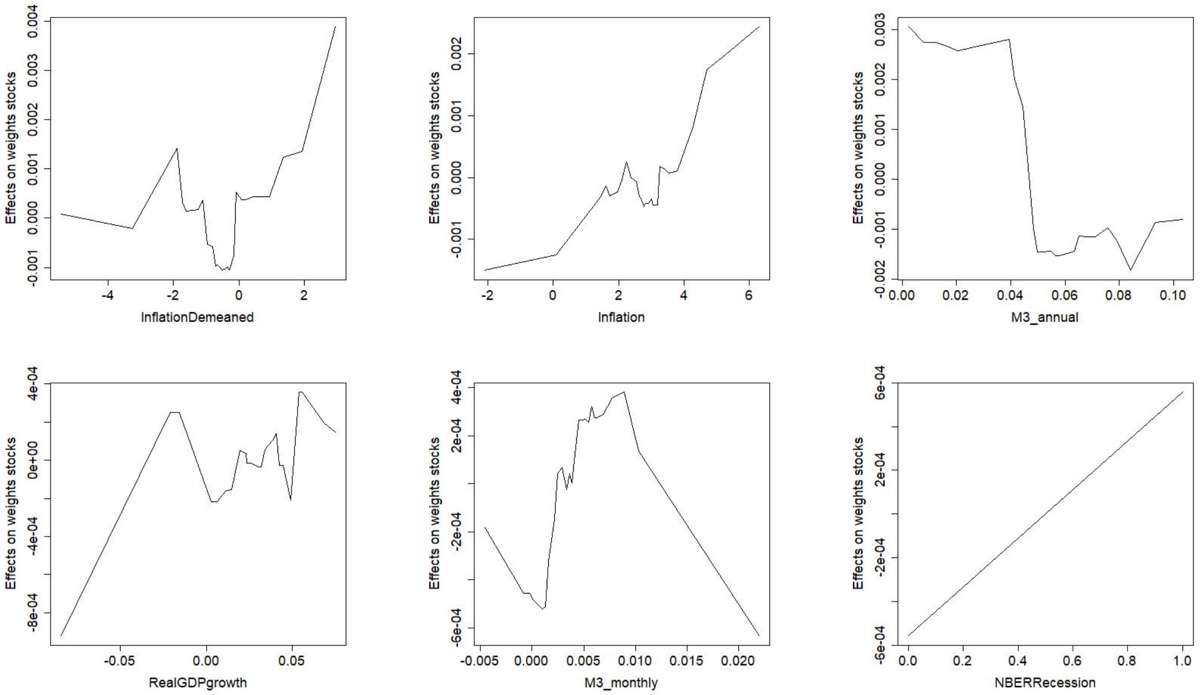
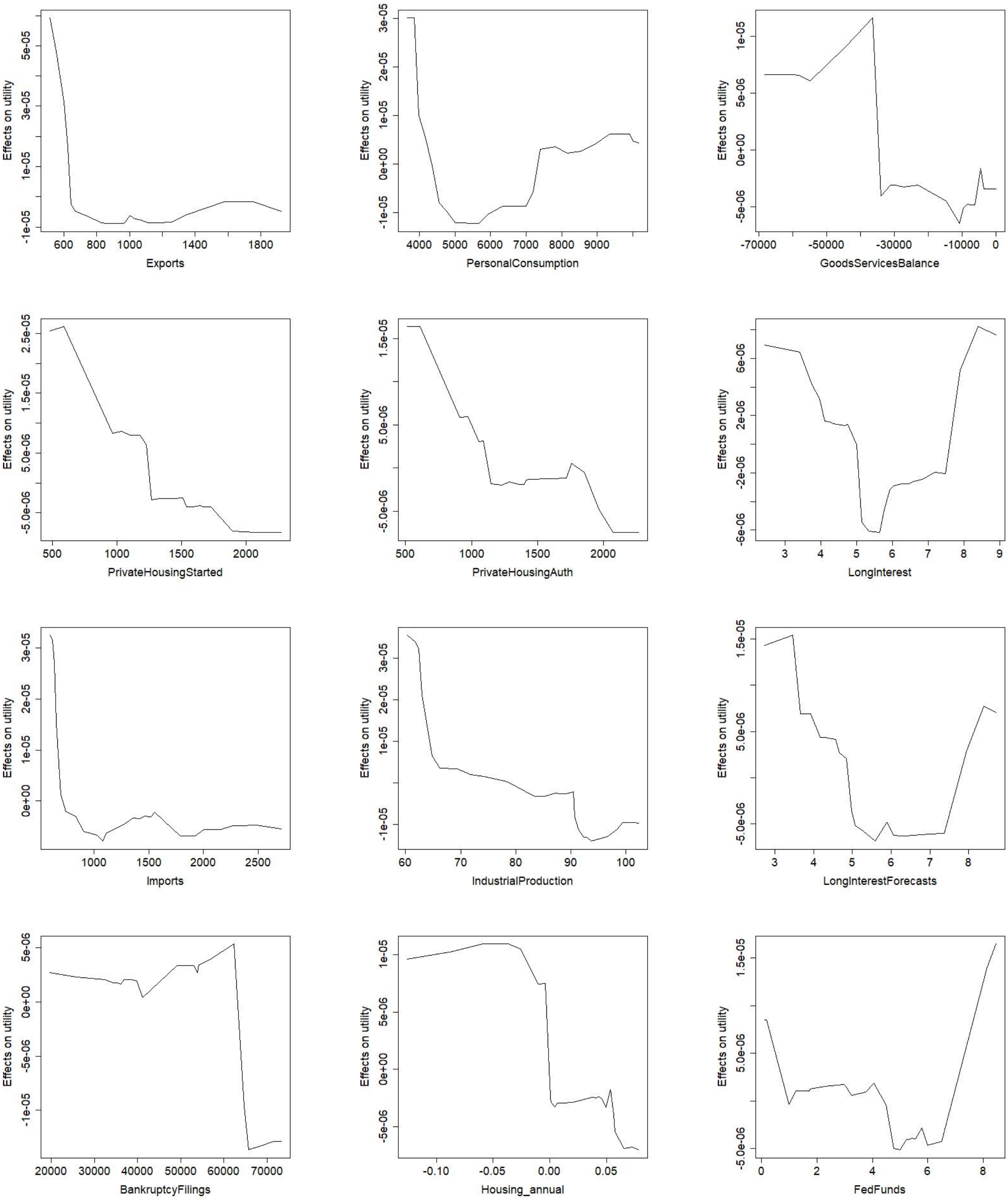
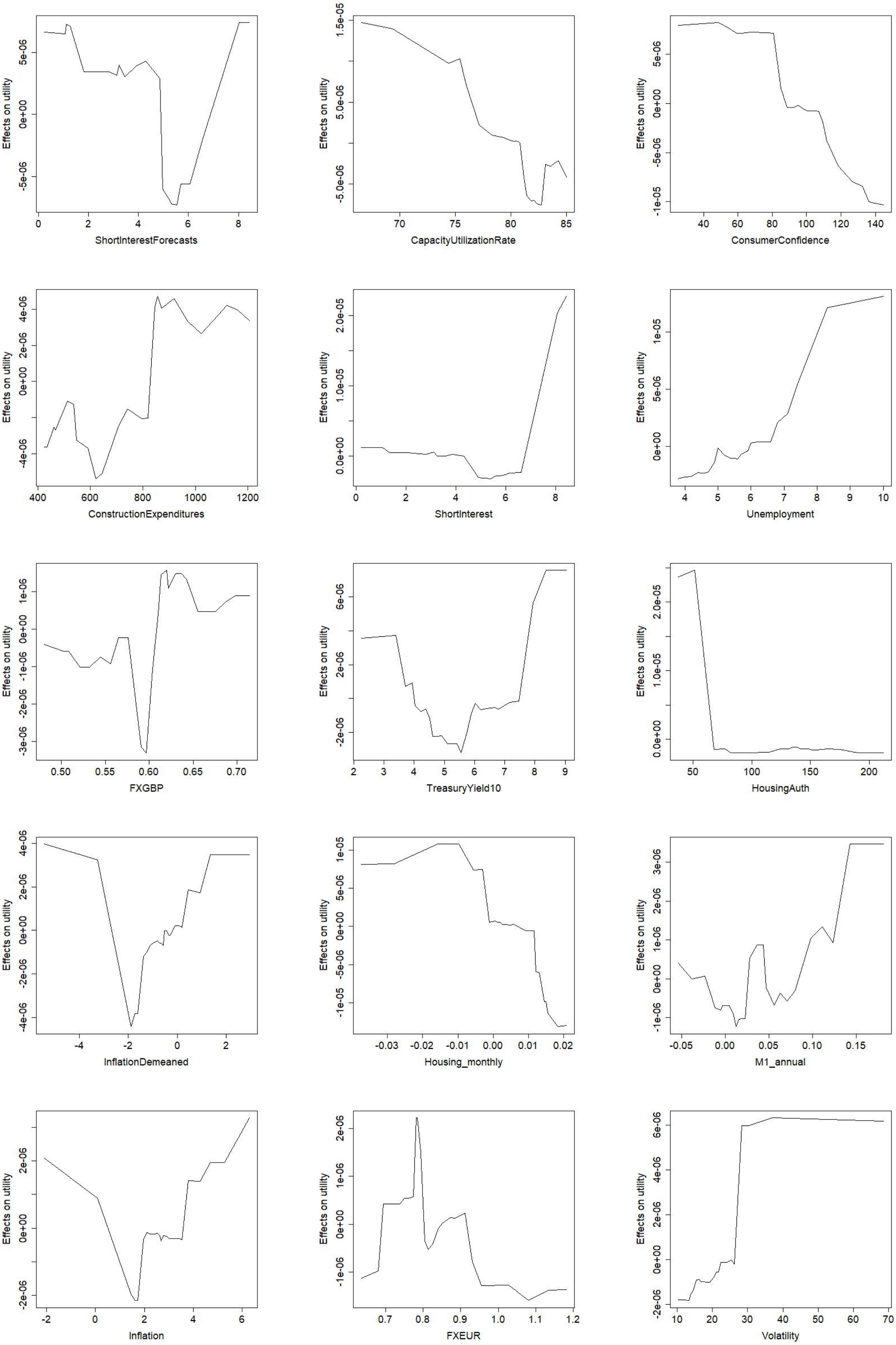
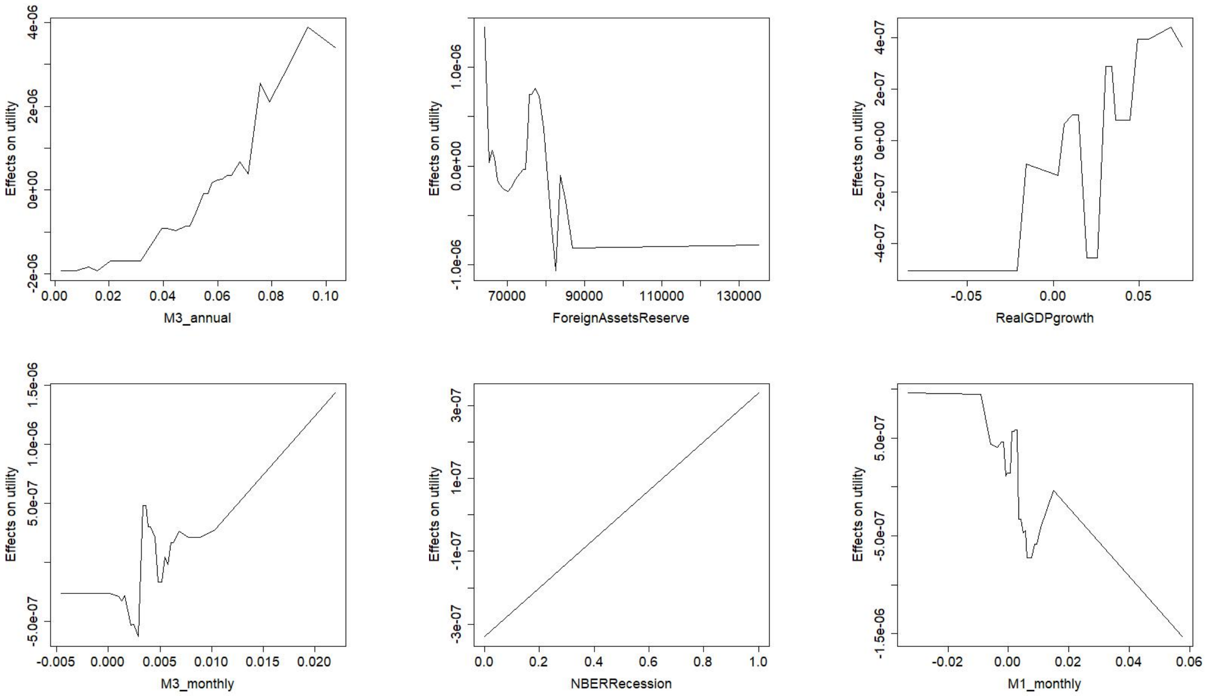

Effects on weights stocks
InflationDemeaned
Effects on weights stocks
Inflation
Effects on weights stocks
M3_annual
Effects on weights stocks
RealGDPgrowth
Effects on weights stocks
M3_monthly
Effects on weights stocks
NBERRecession

Figure C11. The ALE Plots for the 33 Remaining Features from the Random Forests Model for Utility in order of importance according to Figure C4.



The figure continues on the next two pages.


Effects on utility
5e-06
0e+00
-5e-06
0
2
4
6
8
ShortInterestForecasts
Effects on utility
1.5e-05
5.0e-06
-5.0e-06
70
75
80
85
CapacityUtilizationRate
Effects on utility
5e-06
0e+00
-5e-06
-1e-05
40
60
80
100
120
140
ConsumerConfidence
Effects on utility
4e-06
2e-06
0e+00
-4e-06
400
600
800
1000
1200
ConstructionExpenditures
Effects on utility
2.0e-05
1.0e-05
0.0e+00
0
2
4
6
8
ShortInterest
Effects on utility
1e-05
5e-06
0e+00
4
5
6
7
8
9
10
Unemployment
Effects on utility
1e-06
0e+00
-1e-06
-2e-06
-3e-06
0.50
0.55
0.60
0.65
0.70
FXGBP
Effects on utility
6e-06
2e-06
-2e-06
2
3
4
5
6
7
8
9
TreasuryYield10
Effects on utility
2.0e-05
1.0e-05
0.0e+00
50
100
150
200
HousingAuth
Effects on utility
4e-06
2e-06
0e+00
-2e-06
-4e-06
-4
-2
0
2
InflationDemeaned
Effects on utility
1e-05
5e-06
0e+00
-5e-06
-1e-05
-0.03
-0.02
-0.01
0.00
0.01
0.02
Housing_monthly
Effects on utility
3e-06
2e-06
1e-06
0e+00
-1e-06
-0.05
0.00
0.05
0.10
0.15
M1_annual
Effects on utility
2e-06
0e+00
-2e-06
-2
0
2
4
6
Inflation
Effects on utility
2e-06
1e-06
0e+00
-1e-06
0.7
0.8
0.9
1.0
1.1
1.2
FXEUR
Effects on utility
6e-06
4e-06
2e-06
0e+00
-2e-06
10
20
30
40
50
60
70
Volatility


Effects on utility
-2e-06
0e+00
2e-06
4e-06
0.00 0.02 0.04 0.06 0.08 0.10
M3_annual
Effects on utility
-1.0e-06
0.0e+00
1.0e-06
70000 90000 110000 130000
ForeignAssetsReserve
Effects on utility
-4e-07
-2e-07
0e+00
2e-07
4e-07
-0.05 0.00 0.05
RealGDPgrowth
Effects on utility
-5.0e-07
5.0e-07
1.0e-06
1.5e-06
-0.005 0.000 0.005 0.010 0.015 0.020
M3_monthly
Effects on utility
-3e-07
-1e-07
1e-07
3e-07
0.0 0.2 0.4 0.6 0.8 1.0
NBERRecession
Effects on utility
-1.5e-06
-5.0e-07
5.0e-07
-0.02 0.00 0.02 0.04 0.06
M1_monthly

D Additional Tables

Table D1 presents the variable importance for the Random Forest model with stock weights.

Table D1. *Variable Importance for the Random Forest Model with Gold Weights*

Variable	Relative Importance (in %)	Increase in MSE (in %)
M1 level	7.389	0.089
Housing annual	7.278	0.088
CapacityUtilizationRate	5.479	0.066
Exports	4.705	0.056
ShortInterest	4.703	0.057
M3 level	4.570	0.055
ShortInterestForecasts	4.327	0.052
FedFunds	4.105	0.050
PublicDebt	3.799	0.046
FXEUR	3.700	0.046
PersonalConsumption	3.555	0.043
GDP	3.138	0.038
GovConsumpInvest	3.068	0.037
LongInterestForecasts	2.908	0.035
TreasuryYield10	2.746	0.033
Housing monthly	2.730	0.033
UnitLaborCosts	2.670	0.032
ConstructionExpenditures	2.535	0.031
IndustrialProduction	2.007	0.024
BankruptcyFilings	1.920	0.023
SalesNewFamilyHouses	1.913	0.023
PrivateHousingAuth	1.860	0.022
PrivateFixedInvest	1.716	0.020
PrivateHousingStarted	1.711	0.020
ConsumerConfidence	1.654	0.020
Volatility	1.632	0.020
Unemployment	1.608	0.019
FXGBP	1.474	0.018
ForeignAssetsReserve	1.334	0.016

The table continues on the next page

Variable	Relative Importance (in %)	Increase in MSE (in %)
GoodsServicesBalance	1.150	0.014
LongInterest	1.055	0.013
InflationDemeaned	0.884	0.011
Housing level	0.792	0.010
HousingAuth	0.705	0.009
M1 monthly	0.588	0.007
Inflation	0.587	0.007
M3 annual	0.544	0.007
Imports	0.520	0.006
M1 annual	0.366	0.004
RealGDPgrowth	0.235	0.003
NBERRecession	0.173	0.002
M3 monthly	0.170	0.002

Note. Numbers are rounded to three decimal points. The Relative Importance is also shown in the graphs in Appendix C.

Table D2 presents the variable importance for the stock model.

Table D2. *Variable Importance for the Random Forest Model with Stock Weights*

Variable	Relative Importance (in %)	Increase in MSE (in %)
M1 level	6.681	0.076
Housing annual	5.382	0.062
ShortInterestForecasts	4.833	0.055
ShortInterest	4.654	0.053
CapacityUtilizationRate	4.569	0.053
M3 level	4.491	0.051
FedFunds	4.477	0.051
ConstructionExpenditures	4.239	0.048
PersonalConsumption	3.937	0.045
GDP	3.562	0.041
Exports	3.370	0.039
PublicDebt	3.333	0.038
FXEUR	3.116	0.036
GovConsumpInvest	2.924	0.033
Housing monthly	2.841	0.033
FXGBP	2.528	0.029
Imports	2.527	0.029

The table continues on the next page

Variable	Relative Importance (in %)	Increase in MSE (in %)
LongInterestForecasts	2.488	0.028
PrivateHousingAuth	2.444	0.028
SalesNewFamilyHouses	2.395	0.027
UnitLaborCosts	2.218	0.025
ConsumerConfidence	2.181	0.025
PrivateFixedInvest	1.961	0.022
IndustrialProduction	1.798	0.021
TreasuryYield10	1.758	0.020
BankruptcyFilings	1.656	0.019
LongInterest	1.592	0.018
Unemployment	1.336	0.015
M1 annual	1.266	0.014
GoodsServicesBalance	1.241	0.014
ForeignAssetsReserve	1.163	0.013
HousingAuth	1.091	0.012
Volatility	0.976	0.011
PrivateHousingStarted	0.955	0.011
Housing level	0.833	0.010
M1 monthly	0.780	0.009
InflationDemeaned	0.772	0.009
Inflation	0.653	0.007
M3 annual	0.593	0.007
RealGDPgrowth	0.195	0.002
M3 monthly	0.106	0.001
NBERRecession	0.087	0.001

Note. Numbers are rounded to three decimal points. The Relative Importance is also shown in the graphs in Appendix C.

Table D3 represents the variable importance for the utility model.

Table D3. *Variable Importance for the Random Forest Model with Utility*

Variable	Relative Importance (in %)	Increase in MSE (in %)
M1 level	7.117	9.745e-07
M3 level	6.867	9.403e-07
PublicDebt	6.711	9.189e-07
GDP	6.602	9.041e-07
Housing level	5.620	7.696e-07
SalesNewFamilyHouses	5.216	7.142e-07

The table continues on the next page

Variable	Relative Importance (in %)	Increase in MSE (in %)
GovConsumpInvest	4.964	6.797e-07
UnitLaborCosts	4.687	6.418e-07
PrivateFixedInvest	4.429	6.064e-07
Exports	4.208	5.762e-07
PersonalConsumption	3.575	4.895e-07
GoodsServicesBalance	3.139	4.299e-07
PrivateHousingStarted	2.750	3.765e-07
PrivateHousingAuth	2.741	3.754e-07
LongInterest	2.546	3.486e-07
Imports	2.533	3.469e-07
IndustrialProduction	2.293	3.140e-07
LongInterestForecasts	2.183	2.989e-07
BankruptcyFilings	2.153	2.948e-07
Housing annual	2.117	2.899e-07
FedFunds	1.948	2.668e-07
ShortInterestForecasts	1.924	2.635e-07
CapacityUtilizationRate	1.890	2.588e-07
ConsumerConfidence	1.441	1.974e-07
ConstructionExpenditures	1.388	1.900e-07
ShortInterest	1.047	1.434e-07
Unemployment	0.970	1.328e-07
FXGBP	0.930	1.274e-07
TreasuryYield10	0.883	1.209e-07
HousingAuth	0.869	1.190e-07
InflationDemeaned	0.852	1.166e-07
Housing monthly	0.611	8.371e-08
M1 annual	0.589	8.059e-08
Inflation	0.584	7.995e-08
FXEUR	0.539	7.387e-08
Volatility	0.410	5.617e-08
M3 annual	0.333	4.557e-08
ForeignAssetsReserve	0.228	3.119e-08
RealGDPgrowth	0.074	1.006e-08
M3 monthly	0.023	3.086e-09
NBERRecession	0.008	1.146e-09
M1 monthly	0.008	1.038e-09

Note. Numbers are rounded to three decimal points. The Relative Importance is also shown in the graphs in Appendix C.

E Programming code

First, I cleaned the portfolio and macroeconomic data in Excel, i.g. I calculated simple portfolio returns and organized the data. Afterwards, the datasets are loaded into `R` with the script ***"Main Script.R"***. This is the script in which almost all of the modelling for this research paper is done. The script ***"themed_apa_Annegien.R"*** ensures all plots this paper presents follow the APA guidelines. The script is mainly based on the function theme_apa() from the `jtools` package in R.

The script ***"ALE plot function.R"*** creates the ALE plots for the macro features in the dataset, which is obtained from the package `ALEplot` in `R`. Only minor modifications are done with the script, mainly to change the axis titles. Finally, the script ***"ledoit and wolf Sharpe ratio test.R"*** is the script that performs the Ledoit and Wolf (2008) test for the Sharpe ratios. The functions in this script are directly obtained from the codes from the paper presented by Michael Wolf on his website.