Forecasting Excess Returns of the Gold Market: Can We Learn from Stock Market Predictions?*

Hubert Dichtl[#]

dichtl research & consulting GmbH
Am Bahnhof 7, 65812 Bad Soden am Taunus, Germany

First Version: July 2015
This Version: July 2019

Abstract

As some recent studies have shown empirically, future gold price fluctuations are especially difficult to forecast. Against this background, this study evaluates the forecasting power of three approaches that have been applied successfully in a stock market prediction context: 1) technical indicators, 2) diffusion indices, and 3) economically motivated restrictions in predictive regressions. The results are evaluated using statistical and economic evaluation criteria over the entire data sample, as well as separately for expansive and recessive business cycles. We observe that none of the three prediction techniques leads to better forecasts of gold excess returns. The forecast power of fundamental predictor variables is not only highly regime-dependent, but also dependent on the selected economic evaluation criterion. Future gold forecast studies should address these issues.

JEL Classification: G11, G12, G14, G17

Keywords: Gold excess return prediction, fundamental factors, technical factors, diffusion indices, predictive regression models, restrictions, business cycles

---

* I am grateful to Wolfgang Drobetz for helpful comments and suggestions.
# Corresponding author: Tel.: +49-6196-97299-00; E-mail: dichtl@dichtl-rc.de.
Address: Dichtl research & consulting GmbH, Am Bahnhof 7, 65812 Bad Soden am Taunus, Germany.

1. Introduction

The degree of predictability of the gold market is not only interesting from a market efficiency perspective, but also relevant for private and institutional investors as well as asset management companies. Compared to equity, bond, and currency markets, the academic research on gold was for a long time negligible (Lucey, 2011). However, O'Connor et al. (2015) provides evidence that the number of peer-reviewed publications on gold and the predictability of gold returns has increased significantly in recent years.

However, there are major concerns with respect to existing studies. As our literature review in section 2 below shows, there are many forecasting studies which evaluate the predictability of gold returns only with statistical evaluation criteria (e.g., the mean squared forecast error), but not with economic evaluation criteria (e.g., the return and risk of a simulated trading strategy). However, as Leitch and Tanner (1991) and Cenesizogolu and Timmermann (2012) demonstrate, there may be a large discrepancy between both kind of measures. A good forecast accuracy in terms of statistical measures – like for example the mean squared forecast error – does not necessarily imply attractive trading gains. The few available studies evaluating the predictability of gold returns with economic evaluation criteria show sobering results. Their prediction accuracy is mostly not sufficient to beat simple benchmark strategies, e.g., a passive buy-and-hold strategy (Pierdzioch et al., 2014a, 2014b, 2015a).

Given this empirical evidence, we apply three different forecast approaches, which have been successfully used in the field of stock market predictions: technical indicators, diffusion indices, and economically motivated restrictions in predictive regressions. In terms of stock market forecasts, there is some evidence that all three prediction approaches lead to an improvement in terms of statistical as well as economical evaluation criteria (Ludvigson and Ng, 2007; Campbell and Thompson, 2008; Neely et al., 2014). As the behavior of various prediction approaches seemingly depends on the market state (Neely et al., 2014), we explicitly measure the statistical and economical prediction success conditional on expansive and recessive market states. Most importantly, there are several plausible explanations why we can also expect an improvement of our gold market predictions, and not only stock market predictions, with all these forecasting approaches.

In order to provide successful forecasts with technical indicators, gold prices must exhibit a positive time series momentum. In their comprehensive study, Moskowitz et al. (2012) detect a persistent and significant time-series momentum in the prices of fifty-eight liquid instruments of equity index, currency, bond, and commodity markets. A simple momentum-based trading strategy implemented with gold futures provides a statistically significant positive (gross) Sharpe ratio. Hurst et al. (2013) provide an overview over various rational and behavioral explanations how trends emerge in financial markets and why time series momentum exists.

Diffusion indices represent instruments that can conveniently track the main co-movements in a large set of potential predictor variables. The diffusion indices can be consistently estimated by principal components, so that the co-movements in the predictor variables are mainly illustrated by fluctuations in a relatively small number of factors, which are then applied as regressors in the predictive regression model. While all predictor variables potentially contain valuable forecast information, they may also contain some noise. The principal components enable a separation of the information content of all predictor variables into an “important common fluctuations” component and a noise component. Therefore, it seems obvious that better forecasting results can be expected when implementing the predictive regressions with the factor structure of the potential predictor variables instead of the variables themselves. While the principal component approach has been successfully applied in predicting excess returns of stock markets (Ludvigson and Ng, 2007) and bond markets (Ludvigson and Ng, 2009), there is no reason why its application should be restricted to these markets.

The consideration of economic restrictions to improve forecasts has become popular with Campbell and Thompson's (2008) equity risk premium prediction study. However, already Vrugt et al. (2007) considered economically meaningful restrictions in the context of commodity return predictions and report positive effects in terms of trading success. In their real-time forecasting approach, they only take into account those predictive regression models at each forecasting date where the regression coefficient exhibits the theoretically expected sign.

Given these empirical findings, we hypothesize that all these approaches also have the potential to improve gold market forecasts. In contrast to existing gold prediction studies, we evaluate the economic value of the forecasts not only within a simple market timing strategy (i.e., investing 100% in the gold market, or alternatively in the cash market), but also based on a portfolio model. In order to gain further insights, the predictive ability of the models is evaluated separately for expansive and recessive business cycles. As a benchmark model, our study uses the historical mean of excess gold market returns. While this simply benchmark at first may not seem very powerful, it has proven a real challenge to the field of stock market predictions (Welch and Goyal, 2008; Rapach and Zhou, 2013).

The main finding of our study is that none of the three tested concepts, technical indicators, diffusion indices, or regression restrictions, leads to consistent improvements in gold return predictions. An explanation could be that the structure of the gold market is different from that of the stock market (see section 7). However, our study provides strong evidence that some fundamental variables are more suited to forecasting gold excess returns in expansive business cycles, while others exhibit stronger predictive power in recessive business cycles. We also observe that the forecast power of specific prediction approaches depends on the economic evaluation criterion being considered (e.g., certainty equivalent, Sharpe ratio, or hit rate). In this way, our study provides valuable hints for future research regarding the forecasting of gold returns, e.g., the application of regime-dependent forecast methods, or the use of classification-based prediction approaches when the hit rate is the preferred evaluation criterion.

The remainder of this paper is structured as follows. Section 2 provides a review of the relevant gold prediction literature, while section 3 presents the predictive regression models based on fundamental and macroeconomic factors. Section 4 describes the additionally verified forecasting approaches, namely technical indicators, diffusion indices, and economically motivated regression restrictions. The statistical and economic evaluation criteria are presented in section 5. Section 6 outlines the design of the study and provides empirical results, while section 7 discusses potential explanations for the results. Section 8 concludes, and offers implications for future academic work and the asset management industry.

2. Literature review

Compared to the overwhelming amount of general literature on stock market predictions (or equity risk premium predictions), there are few extant studies that explore the predictability of gold market returns (e.g., Lucey, 2011; Baur et al., 2016). For gold market predictions, we differentiate between two strands of literature: 1) forecasting approaches based on publicly available fundamental and macroeconomic data, and 2) approaches that assess only historical gold prices (i.e., time series models or technical indicator models).

2.1 Predictions with fundamental and macroeconomic variables

In their thorough study, Pierdzioch et al. (2014a) analyze the predictability of monthly excess gold market returns (spot gold fixing prices from the London Bullion Market) within a comprehensive real-time forecasting approach. They analyze whether publicly available information about a large set of fundamental and macroeconomic variables (inflation rate, exchange rate changes, oil price changes, stock market returns, term spread, corporate bond spread, lagged returns of gold prices) help forecast out-of-sample monthly excess returns in order to invest in gold. Their implemented real-time forecasting approach accounts for the possibility that the optimal forecasting model may change over time (Pesaran and Timmermann, 1995).

With their “thin” modeling approach, Pierdzioch et al. (2014a) select the most promising forecast model for each single monthly prediction by means of various model selection criteria (e.g., the Akaike or the Hannan-Quinn information criterion). They also implement a “thick” modeling approach, where they combine all forecasts based on different combination methods (e.g., Rapach et al., 2010). In order to judge the forecast quality of their prediction models for the 1997-2012 out-of-sample period, they set up a simple trading rule (with and without transaction costs), and compare the results with a buy-and-hold strategy. They conclude that the gold market is informationally efficient with respect to the predictor variables considered in their study.[1]

Due to the positive results in Cooper and Priestley (2009) in terms of stock and bond market predictions, Pierdzioch et al. (2014b) study whether the international business cycle, as measured in terms of the output gaps of G7 countries, has out-of-sample predictive power for excess gold returns. They find some evidence of predictive power for gold price fluctuations. But a simple trading rule built on real-time out-of-sample forecasts does not lead to superior performance over a buy-and-hold strategy after accounting for transaction costs.

Based on the dynamic model averaging framework proposed by Raftery et al. (2010), Baur et al. (2016) apply this forecasting approach – together with a dynamic model selection approach – to predict gold returns over one, three, and twelve months (the gold price is the 3pm London fixing price, denominated in U.S. dollars (USD)). Their findings show that the dynamic model

[1] Due to these results, the combination of forecasts is not considered in our study.

averaging framework improves forecasts compared to other frameworks, and provides evidence for the time variation of gold price predictors.

Aye et al. (2015) apply the same prediction approach to forecast gold prices. However, in contrast to Baur et al. (2016), they do not apply the possible predictor variables directly, instead aggregating them using a recursive principal component analysis to six global factors (business cycle, inflation rate, interest rate, commodity, exchange rate, and stock price).[2] In this study, the dynamic model selection approach provides the highest prediction quality across all forecast horizons (one, three, six, nine, and twelve months), while the exchange rate factor exhibits the strongest predictive power. Unfortunately, Aye et al. (2015) measure prediction quality only with the mean squared forecast error and the sum of log predictive likelihoods, not economic evaluation criteria. It thus remains unclear whether the higher prediction quality in terms of statistical criteria can be profitably exploited within an active investment strategy.

Pierdzioch et al. (2015a) apply a boosting approach in a real-time setup to forecast gold price fluctuations.[3] In order to ensure comparability with their earlier work (Pierdzioch et al., 2014a), they use the same data used in this study, and forecast fluctuations in excess of the short-term interest rate. The three predictor variables included most often in the optimal forecasting model are lagged excess gold returns, the inflation rate, and the corporate bond spread. Pierdzioch et al.'s (2015a) results show that the performance measures implied by an active trading rule dominate the corresponding values of the buy-and-hold strategy, but only for small transaction costs. However,

[2] They also consider the Kansas City Fed's financial stress index and the U.S. economic policy uncertainty index (both variables are used directly and are not part of their principal component analysis).

[3] See Bühlmann (2006) and Bühlmann and Hothorn (2007) for a description of the boosting approach. Berge (2014) uses this prediction technique to forecast exchange rate movements in real time.

an additionally conducted bootstrap simulation reveals that the differences between the performance measures of the trading rule and the buy-and-hold strategy are not significant even when zero transaction costs are assumed.

In Pierdzioch et al. (2015b), gold returns are forecasted with a real-time quantile regression approach. Within this approach (Koenker and Hallock, 2001), they consider that a forecaster may have an asymmetric loss function where over- and underestimates of the same size are weighted differently. Based on this asymmetric loss function, they evaluate their forecasts with an out-of-sample R^2 statistic similar to that proposed in Campbell and Thompson (2008). Pierdzioch et al. (2015b) ultimately show that their approach outperforms forecasts implied by an autoregressive benchmark model in terms of out-of-sample R^2 when the loss function implies that underestimations are more costly than overestimations (of the same size).

Malliaris and Malliaris (2015) conduct a decision tree analysis to predict the direction of daily gold price movements (up or down). Their forecasts are based on equity returns (S&P 500 index), equity volatility (VIX), oil prices, the Cleveland Financial Stress Indicator, and the Euro. Due to their extraordinarily positive results (correct direction forecasts ranging from 85.9% to 95.9%), however, it is necessary to conduct further robustness tests on this innovative prediction methodology.[4]

With their quantile-boosting approach to forecasting gold returns, Pierdzioch et al. (2016) combine the advantages of quantile regression techniques and boosting techniques. For optimistic investors who incur higher losses for an underprediction than an overprediction, Pierdzioch et al.

[4] The same holds for Parisi et al.'s (2008) study, which predicts gold price changes with neural network models. Despite their 60% level of correct direction forecasts (clearly lower than that documented in Malliaris and Malliaris, 2015), further research using this innovative prediction technique also seems warranted.

(2016) find several situations where a forecasting-based trading strategy provides higher terminal wealth and a higher Sharpe ratio than the buy-and-hold. This result holds after transaction costs.

Gupta et al. (2016) also use a quantile predictive regression approach to analyze whether terror attacks predict gold returns. They find that terror attacks have predictive power for the lower and particularly the upper quantiles of the conditional distribution of gold returns. However, because they evaluate their forecasts using the same out-of-sample R^2 as Pierdzioch et al. (2015b), it is not clear whether the prediction accuracy is sufficient to generate higher economic profits than a simple buy-and-hold strategy.

Using data from 54 countries, Sharma (2016) tests whether consumer price index predicts gold price returns. Her predictability test is based on the flexible generalized least squares estimator proposed by Westerlund and Narayan (2015). This study provides limited evidence that consumer price index predicts gold price returns in in-sample tests. However, the conducted out-of-sample test. reveal relatively strong evidence that consumer price index predicts gold returns.

With more than 140 years of data, Prokopczuk et al. (2018) comprehensively analyze the predictability of returns and volatilities of 30 different commodity spot markets. They follow the literature on stock return predictability and use 16 predictive variables that are usually considered to have predictive power for stock returns. In terms of the out-of-sample predictability of the gold market, this study provides evidence for a better out-of-sample predictability of yearly excess returns compared to monthly excess returns. The default yield spread is the explanatory variable with the highest out-of-sample predictability of the 1-year gold excess returns in a univariate regression setup. While the model selection approach provides overall poor results, for the prediction of the yearly gold excess returns they show a slightly improvement in terms of the out-of-sample R^2 measure.

Nguyen et al. (2019) examine the predictability of the gold risk premium and analyze the question of how expected gold returns co-move with the expected returns of stock and bond markets, as well as with expected inflation. They find that the best prediction model for the gold risk premium is a parsimonious regression model with the jump risk premium (Bollerslev et al., 2015) and the variance risk premium (Bollerslev et al., 2009) of gold as the two explanatory variables. Both variables show strong predictability in-sample and out-of-sample and for all horizons investigated, varying from one month to two years.

2.2 Predictions based on historical gold prices

Marshall et al. (2008) analyze the profitability of technical indicator-based market timing strategies (see Sullivan et al., 1999) applied in fifteen major commodity futures markets (including gold). Their analysis is based on daily log returns over the January 1984-December 2005 period. Just as in most other futures markets, Marshall et al. (2008) fail to find a statistically significant outperformance for gold when they account for data snooping using White's (2000) reality check method.

Szakmary et al. (2010) implement trend-following trading strategies in twenty-eight commodity future markets based on moving averages and the channel indicator. They report positive results after transaction costs in at least twenty of the twenty-eight markets. All parameterizations of the moving average and the channel strategy provide a positive mean net return for the gold market (which is statistically significant in most cases).

In their comprehensive study of time series momentum, Moskowitz et al. (2012) detect a persistent and significant momentum in the time series of fifty-eight liquid instruments of equity index, currency, commodity, and bond markets (and thus also in gold futures). They find that the strongest relationship exists between a security's next month excess return and the lagged twelve-month return. A simple momentum-based trading strategy implemented with gold futures provides a positive (gross) Sharpe ratio that is statistically significantly different from zero at a 5% level.

Beside some other precious metal markets, Hassani et al. (2015) forecast gold prices with an autoregressive model, an optimized autoregressive integrated moving average (ARIMA) model, exponential smoothing (ETS), a trigonometric ETS state space model with Box-Cox transformation, ARMA errors, trend and seasonal components (TBATS), a fractionalized ARIMA model (ARFIMA), vector autoregression (VAR), Bayesian autoregression (BAR) models, and Bayesian VAR models (BVAR). Over all forecast horizons (ranging from one to twenty-four months), the exponential smoothing model provides on average the best forecasts in terms of root mean squared errors. Interestingly, over the one-month forecast horizon, no forecasting technique was able to outperform the random walk model.

3. Predictive regressions with fundamental and macroeconomic factors

In the domain of stock market predictions, the application of a simple bivariate regression model seems to be fairly standard (e.g., Goyal and Welch, 2003; Welch and Goyal, 2008; Campbell and Thompson, 2008; Neely et al., 2014; among many others). Besides the ability to generate quantitative predictions, this approach has the advantage of being able to assess the impact of a specific factor on the forecast variable by means of various statistical measures (e.g., R^2, *t*-statistic). The simple bivariate predictive regression model is defined as:

(1a) $$r_{t+1} = \alpha_i + \beta_i x_{i,t} + \varepsilon_{i,t+1}$$ (1b) $$\hat{r}_{t+1} = \hat{\alpha}_i + \hat{\beta}_i x_{i,t}$$

In a gold forecasting context, r_{t+1} in Equation (1a) represents the log return on the gold price in excess to the log risk-free rate from period t to t+1 (e.g., Pierdzioch et al., 2014a, 2014b). $x_{i,t}$ is a predictor variable, and α_i and β_i are regression parameters that can be estimated using an OLS method. $\varepsilon_{i,t+1}$ labels the regression residuum. Once the regression parameters are estimated, they can be used together with an observed value of the predictor variable to forecast the excess return (Equation (1b)).

While there are no established forecasting models for the price of gold, several factors can significantly influence it. The factors are derived from the properties gold is generally associated with, i.e., an inflation hedge, a currency hedge, a safe haven, and an investment diversifier (portfolio protection) (Baur, 2013a; Erb and Harvey, 2013; Baur et al., 2016).[5]

3.1 Gold and inflation

One of the most widely discussed properties of gold is its (potential) ability to hedge against inflation. The main argument for this property is based on the money-like status of gold. In contrast to a fiat currency (like the USD or the Euro), it is not possible to increase the supply of gold immediately. Gold has a limited stock and a relatively inelastic supply in the short run, because increasing production can take a great deal of time (Feldstein, 1980; O'Connor et al., 2015). Various studies support these arguments by showing empirically proven positive relationships between gold price fluctuations and the inflation rate (e.g., Worthington and Pahlavani, 2007; Bampinas and Panagiotidis, 2015); other studies view this relationship as more or less important (e.g., Sjaastad, 2008; Blose, 2010; Baur, 2011; Erb and Harvey, 2013).

3.2 Gold and currencies

Following O'Connor et al. (2015), it has been frequently argued that the USD is one, if not the primary, driver of gold prices. The basis for this argument is that gold is traded primarily in

[5] While a "hedge asset" can be defined as an asset that is uncorrelated or negatively correlated with another asset or portfolio, a "safe haven asset" must be uncorrelated or negatively correlated during times of market stress or turmoil (Baur and Lucey, 2010). Baur (2013a) further distinguishes a "safe haven property" from an "investment diversification (portfolio protection) property" in terms of timing. Demand for a safe haven occurs during or shortly after a crisis or a crash; demand for investment diversification or portfolio protection occurs before a crisis or a crash.

dollars. A weaker USD (as measured by its trade-weighted exchange rate) makes gold cheaper for other nations to purchase, thereby increasing demand. This leads to rising gold prices, which explains the negative relationship to the USD. Several studies provide evidence for this negative relationship (e.g., Capie et al., 2005; Tully and Lucey, 2007; Pukthuanthong and Roll, 2011; Baur, 2011; Erb and Harvey, 2013; Reboredo, 2013). However, despite these empirical findings, it remains unclear whether gold is really a "good" currency hedge.[6] Aye et al. (2015) find that the USD seems to be a better predictor of gold price fluctuations than other variables.

3.3 Gold and interest rates

In contrast to various other economic variables, the link between gold and interest rates is not as clear as it appears at first glance (Baur, 2013a; O'Connor et al., 2015). As per Koutsoyiannis (1983) and Fortune (1987), gold and interest rates are related due to an asset substitution relationship. They argue that increases in expected interest rates should encourage gold owners to shift from gold to interest-bearing assets, because gold does not provide cash flow benefits. This is also the reason investors should be discouraged from making new purchases of gold. The logical consequence of this argument is a negative relationship between gold price fluctuations and interest rates (e.g., Koutsoyiannis, 1983; Blose, 2010).

In contrast, Abken (1980) sees the link between gold and inflation as the real driver of the gold-interest rate relationship. He argues that an increase in expected inflation will drive up nominal interest rates by a similar level (see also Blose, 2010). In his equilibrium reflections, he also posits that gold investors will demand compensation for holding a non-interest-bearing asset class

[6] For example, in Erb and Harvey's (2013) regression analysis, all coefficients show a predictably statistically significant negative relationship between gold returns and various exchange rate returns. However, Erb and Harvey (2013) emphasize that "the average beta coefficient is significantly different from zero but also significantly different from -1.0." Corresponding dollar exchange rate moves can only be partially compensated for with gold. Capie et al. (2005) find that gold has served as a hedge against fluctuations in the foreign exchange value of the dollar, but to what degree seems highly dependent on unpredictable political attitudes and events.

that equals the interest rate, resulting in a similar rate of gold price appreciation (see also Blose, 2010).

Thus, interest rates and gold price fluctuations should move in the same direction. In order to explore this issue, Baur (2013a) suggests the application of real interest rates (i.e., the difference between nominal interest rates and the inflation rate), where both effects are combined.[7] Although the direction of the relationships among short-term interest rates (e.g., one- or three-month T-bills), bonds, and term spread variables and gold price fluctuations is unclear, the opposite is true for the default yield spread (i.e., the difference between BAA- and AAA-rated corporate bond yields) and the default return spread (i.e., the difference between long-term corporate bond and long-term government bond returns).[8] Various studies view these variables as business cycle indicators (e.g., Fama and French, 1989; Chen, 1991), or as economic and financial crisis indicators (Hartmann et al., 2008). Therefore, we expect a positive (negative) relationship between the default yield spread (default return spread) and gold price fluctuations (Prokopczuk et al., 2018).

3.4 Gold and stock markets

From the perspective of a stock market investor, gold achieves two significant accomplishments: 1) it provides a safe haven during stock market crises, and 2) it serves as an investment diversifier within a portfolio. As a result, some studies have provided evidence of a negative relationship between gold price fluctuations and stock market returns (e.g., Baur and Lucey, 2010; Baur and McDermott, 2010); others find a positive relationship between gold returns and stock market volatility (e.g., Hillier et al., 2006). With respect to its property as a business cycle indicator (Chen, 1991), Vrugt et al. (2007) also consider the (annualized) dividend yield on the S&P 500 in

[7] Baur (2013a) emphasizes that a macroeconomic regime in which the nominal interest rate is below the inflation rate (i.e., an environment of negative real interest rates) can be expected to exert an extraordinarily strong influence on gold price fluctuations.

[8] See Welch and Goyal (2008, p. 1459).

their commodity prediction models. In contrast to the relationship between gold returns and interest rates, the direction between the various stock market variables and gold returns seems quite clear.

3.5 Gold and oil prices

The relationship between gold and oil prices would appear to be of great economic interest (e.g., Pierdzioch et al., 2014a, 2015a; Aye et al., 2015). The price of oil is assumed to be an indicator of geopolitical risk (Pierdzioch et al., 2014a), as well as a harbinger of the business cycle in many developed countries (Hamilton, 2009). Due to its safe haven property during times of economic and financial market turmoil, as well as its inflation-driving property (O'Connor et al., 2015), a positive link between oil price developments and the price of gold is expected.

Based on the preceding discussions about the relationships between gold price fluctuations and various fundamental and macroeconomic variables, Exhibit 1 lists the factors considered in the fundamental predictive regression models.[9]

[9] Due to the limited availability of historical oil price information, we do not consider oil prices in the analysis. The USD variable represents the Broad Trade Weighted U.S. Dollar Index from the Federal Reserve Bank of St. Louis (https://research.stlouisfed.org/fred2). All other data come from Welch and Goyal's (2008) dataset (see also the data descriptions therein, and in Neely et al., 2014). The data can be retrieved from Amit Goyal's webpage at http://www.hec.unil.ch/agoyal/.

Exhibit 1: Fundamental and macroeconomic predictor variables

Variable	Description	Publication Lag	Sign
INFL	Inflation calculated from the Consumer Price Index for all urban consumers	1 month	+
USD	Exchange rate as the continuously compounded year-on-year change in the trade-weighted effective nominal U.S. exchange rate	1 month	-
TBL	Interest rate on a three-month Treasury bill	no	+/-
LTY	Long-term government bond yield	no	+/-
LTR	Return on long-term government bonds	no	+/-
TMS	Term spread computed as long-term yield minus Treasury bill rate	no	+/-
DFY	Default yield spread, computed as the difference between Moody's BAA- and AAA-rated corporate bond yields	no	+
DFR	Default return spread, computed as the long-term corporate bond return minus the long-term government bond return	no	-
ERP	Equity risk premium, calculated as the difference between the log return on the S&P 500 index (including all dividends) and the log return on a risk-free bill	no	-
DY	Dividend yield (log), calculated as the log of a twelve-month moving sum of dividends paid on the S&P 500 index, minus the log of lagged stock prices (S&P 500 index)	no	-
RVOL	Volatility of the equity risk premium, based on a twelve-month moving standard deviation estimator (Mele, 2007)	no	+

Notes: This table lists all the fundamental factors used in the predictive regressions, including their descriptions. The table also contains information about potentially considered publication lags. The "Sign" column provides information about the expected relationship between the fundamental factor and the gold excess return. The "+" sign labels a positive expected relationship, and the "-" sign a negative one. "+/-" indicates that the expected relationship can be either positive or negative (see section 3).

Beside the description of the fundamental input variables, Exhibit 1 provides further information about the considered publication lags as well as the expected impact direction of the variables (i.e., the signs of the corresponding regression coefficients). While our list of used fundamental and macroeconomic-based predictor variables is representative (e.g., Prokopczuk et al., 2018), we are aware that it is not a complete list of all potential forecasting variables. Several others have been applied (with more or less success) to forecast future gold price fluctuations.[10] As Baur et al.

[10] For example, Aye et al. (2015) additionally consider the Kansas City Fed's financial stress index as well as the U.S. economic policy uncertainty index as predictive variables in their study. Due to the empirically verified co-movements of commodities (Pindyck and Rotemberg, 1990; Baur, 2011), other commodity markets or broad indices are also considered as predictive factors for future gold price fluctuations (e.g., Aye et al., 2015). Some recent studies have also considered the output gap of major countries as predictive variables for gold price fluctuations, albeit with relatively

(2016) emphasize, a contemporaneous and thus non-predictive relationship between gold prices and particular determinants over a specific period does not guarantee successful out-of-sample predictability. Hence, whether the potential factors of gold price fluctuations discussed here are also useful in a forecasting framework is of great interest.

4. Further promising prediction approaches

This section presents some prediction approaches that have been successfully applied in a stock market prediction context: 1) technical indicators (e.g., Neely et al., 2014), 2) diffusion indices (e.g., Ludvigson and Ng, 2007; Rapach and Zhou, 2013; Neely et al., 2014), and 3) economically motivated restrictions (Campbell and Thompson, 2008; Rapach and Zhou, 2013).

4.1 Regressions with technical indicators

The predictor variable in the regression model defined in Equations (1a) and (1b) need not be a fundamental or macroeconomic factor. It can also be a technical indicator. Following some stock market prediction studies (e.g., Neely et al., 2014; Baetje and Menkhoff, 2016; Hammerschmid and Lohre, 2015), we implement predictive regressions based on technical indicators with the moving average indicator and the momentum indicator. The moving average trading rule is defined as:

(2a) $$S_{i,t} = \begin{cases} 1 & if\ \ MA_{s,t} \geq MA_{l,t} \\ 0 & if\ \ MA_{s,t} < MA_{l,t} \end{cases} \qquad with\ s < l$$

where

(2b) $$MA_{j,t} = (1/j) \sum_{i=0}^{j-1} P_{t-i} \qquad for\ j = s, l$$

negligible success (e.g., Pierdzioch et al., 2014b; Aye et al., 2015). Nguyen et al. (2019) identify the jump risk premium of gold and the gold variance premium as successful predictors for the gold risk premia.

The moving average indicator is based on the moving averages on the gold price index (P_t), as defined in Equation (2b). According to Equation (2a), we would invest in the gold index if the short moving average $\left(MA_{s,t}\right)$ is above the long moving average $\left(MA_{l,t}\right)$. Otherwise, an allocation to risk-free bills is taken $\left(S_{i,t} = 1 \text{ or } S_{i,t} = 0\right)$.

Following certain stock market prediction studies (Neely et al., 2014; Baetje and Menkhoff, 2016; Hammerschmid and Lohre, 2015), the short index for the moving average is set to $s = 1,2,3$, and the long index to $l = 9,12$, resulting in six moving average strategies labeled as MA(*s*-*l*).[11]

The time series momentum indicator is calculated as the difference between the actual price (P_t) and the *m*-month lagged price (P_{t-m}), as follows:

$$(3) \qquad S_{i,t} = \begin{cases} 1 & if \;\; P_t \geq P_{t-m} \\ 0 & if \;\; P_t < P_{t-m} \end{cases}$$

We would thus invest in the gold index if the time series momentum is positive $\left(S_{i,t} = 1\right)$. Otherwise, an allocation to risk-free bills is taken $\left(S_{i,t} = 0\right)$. We follow Neely et al. (2014), and set $m = 9\ ,12$. The resulting momentum strategies are labeled as MOM(9) and MOM(12).

4.2 Regressions with diffusion indices

Some equity premium forecast studies substitute for the predictor variables in their regressions with diffusion indices, and report positive results (e.g., Ludvigson and Ng, 2007; Rapach and Zhou, 2013; Neely et al., 2014). Diffusion indices represent instruments that can conveniently track the main co-movements in a large set of potential return predictors. The diffusion index approach is grounded on following latent factor model (Rapach and Zhou, 2013):

[11] This parameterization is similar to that in Szakmary et al. (2010). They parameterize their moving average strategies (applied in commodity futures markets) with $s = 1,2$ and $l = 6,12$ months.

(4) $x_{i,t} = \lambda_i' f_t + e_{i,t}$ $\quad (i = 1, \ldots, K)$,

where $x_{i,t}$ is a demeaned potential predictor variable, f_t is a q-vector of latent factors, λ_i is a q-vector of factor loadings, and $e_{i,t}$ is a zero-mean disturbance term. A "strict" factor model is based on the assumption of contemporaneously and serially uncorrelated disturbance terms, but a limited degree of both is allowed in an "approximate" factor model (e.g., Stock and Watson, 2002; Bai, 2003). The latent factors can be consistently estimated by principal components, so that the co-movements in the predictor variables are mainly illustrated by fluctuations in the relatively small number of factors $(q \ll K)$. These factors are then applied as regressors in the predictive regression model (see Equations (1a) and (1b)):

(5a) $r_{t+1} = \alpha_{DI} + \beta_{DI}' f_t + \varepsilon_{t+1}$ $\quad$ (5b) $\hat{r}_{t+1}^{DI} = \hat{\alpha}_{DI,t} + \hat{\beta}_{DI,t}' \hat{f}_{t,t}$.

β_{DI} in Equation (5a) labels a vector of slope coefficients with length q. In Equation (5b), $\hat{f}_{t,t}$ represents the principal component estimate of f_t based on data available through t. $\hat{\alpha}_{DI,t}$ and $\hat{\beta}_{DI,t}$ are OLS estimates of α_{DI} and β_{DI} from regressing $\{r_j\}_{j=2}^{t}$ on a constant, and $\{\hat{f}_{j,t}\}_{j=1}^{t-1}$.

While all K predictors $x_{i,t}$ $(i = 1, \ldots, K)$ potentially contain valuable information for forecasting r_{t+1}, they may also contain some noise. The latent factor model in Equation (4) enables a separation of the information content of all *K* predictor variables into an "important common fluctuations" component (f_t) and a noise component $(e_{i,t})$. Therefore, it seems obvious that better forecasting results can be expected when implementing the predictive regressions with the factor structure of the *K* potential predictor variables (instead of the variables themselves).

To apply this predictive approach, the number of latent variables (q) must be specified. Rapach and Zhou (2013) advise keeping q relatively small in a predictive context, in order to avoid an overparameterized forecasting model. Due to the various documented positive results of this predictive approach in terms of equity premium predictions, it is also applied here.[12] To avoid the problem of overparameterization, the predictive regressions in the baseline simulations are implemented using one factor (e.g., Rapach and Zhou, 2013) and two-factor models are additionally tested in the robustness tests.

4.3 Regressions with economically motivated restrictions

Due to the poor results of predictive regression models in an equity risk premium prediction context, Campbell and Thompson (2008) demonstrate that including some (weak) economically meaningful constraints can significantly enhance forecast quality. They take two types of restrictions into account. First, they set the estimated regression coefficient to zero when it does not exhibit the theoretically expected sign (i.e., $\hat{\beta}_i = 0$ in the predictive regression model (1b)). In this way, the regression constant (i.e., $\hat{\alpha}_i$ in the predictive regression model (1b)) predicts the excess return. Second, because they expect positive risk premiums for assets with positive volatility, they also set the forecast to zero if the predictive regression model (1b) predicts a negative excess return (i.e., if $\hat{r}_{t+1} < 0$).

The consideration of economic restrictions became popular with Campbell and Thompson's (2008) equity risk premium study. However, others have also incorporated economic theory into empirical models. For example, in a bond yield prediction context, Ang and Piazzesi (2003) show that the forecasting performance of their vector autoregressive models improves when they impose non-arbitrage restrictions. Vrugt et al. (2007) consider economically meaningful restrictions in the context of commodity return predictions, and report positive effects. Their real-time forecasting

[12] Aye et al. (2015) apply the diffusion index approach in a context of gold price predictions. However, it remains unclear whether their reported positive results are attributable to the application of the diffusion index or their sophisticated dynamic model averaging method.

approach only takes into account the predictive regression models at each forecasting date that exhibit the theoretically expected sign.

Pettenuzzo et al. (2014) also report positive results when considering economic restrictions in their equity premium prediction. Besides the non-negativity equity premium restriction, they impose additional bounds on the conditional Sharpe ratio. Due to the various positive documented results, we also analyze the effect of Campbell and Thompson's (2008) proposed restrictions on the results of the predictive regressions.

5. Forecast evaluation criteria

Leitch and Tanner (1991) convincingly demonstrate that prediction accuracy as measured with a mean squared error (or a similar statistical measure) implies nothing about the economic success potential of a forecast model. For this reason, both statistical and economic evaluation criteria are applied in this study.

5.1 Statistical evaluation criteria

Within the in-sample evaluation of a simple bivariate regression model, it is common to examine the sign and magnitude of the regression coefficient, the corresponding *t*-statistic, and the coefficient of determination (R^2). In order to obtain further insight into the relative strength of gold returns during expansive and recessive business cycles, Neely et al. (2014) propose the "following intuitive versions of the conventional R^2 statistic":

(6) $$R_c^2 = 1 - \frac{\sum_{t=1}^{T} I_t^c \hat{\varepsilon}_{i,t}^2}{\sum_{t=1}^{T} I_t^c (r_t - \bar{r})^2} \qquad \text{for } c = \text{EXP, REC}.$$

The indicator variable I_t^{EXP} (I_t^{REC}) takes the value unity when month *t* is classified as an "expansive" ("recessive") business cycle, and 0 otherwise. $\hat{\varepsilon}_{i,t}$ is the fitted residual based on the full sample estimates of the predictive regression model, $\bar{r}$ represents the full sample mean, and T labels the number of observations in the full sample. Following Rapach and Zhou (2013), as well as Neely et al. (2014), we apply National Bureau of Economic Research (NBER)-dated business cycle expansions and recessions. In contrast to the full sample R^2 statistic, the R^2_{EXP} and R^2_{REC} statistics can be negative.

In order to evaluate the predictive regression models out-of-sample (from $t = s, \dots, T$), we apply the commonly used mean squared forecast error (*MSFE*), as well as the out-of-sample R^2 proposed by Campbell and Thompson (2008):

(7a) $$R^2_{OS} = 1 - \frac{\sum_{t=s}^{T}(r_t - \hat{r}_t)^2}{\sum_{t=s}^{T}(r_t - \bar{r}_t)^2}$$ (7b) $$R^2_{OS} = 1 - \frac{MSFE_i}{MSFE_0}.$$

In Equation (7a), $\hat{r}_t$ represents the fitted value form of a predictive regression model estimated through period $s - 1$, and $\bar{r}_t$ is the historical average return also estimated through period $s - 1$. Alternatively, R^2_{OS} can be formulated in terms of *MSFE* values in Equation (7b), where $MSFE_i$ denotes the *MSFE* of prediction model i, and $MSFE_0$ is the *MSFE* of the historical mean (Rapach and Zhou, 2013).

When $R^2_{OS} > 0$, the forecast of the predictive regression model is more accurate than the historical average in terms of *MSFE* $(MSFE_i < MSFE_0)$. Analogously to the in-sample R^2 measure, the out-of-sample R^2 can also be calculated separately for expansive and recessive business cycles. The question natural arises whether the detected improvement in the predictive regression model is statistically significant. Formally, we test $H_0\!: MSFE_0 \leq MSFE_i$ against $H_A\!: MSFE_0 > MSFE_i$, or, alternatively, $H_0\!: R^2_{OS} \leq 0$ against $H_A\!: R^2_{OS} > 0$.

Diebold and Mariano (1995) and West (1996) provide an appropriate test statistic that is asymptotically standard normally distributed. However, it has a non-standard asymptotic distribution when forecasts from nested models are compared, as is done here. If the null hypothesis ($\beta_i = 0$) holds in the predictive regression model, then the forecast model reduces to the historical mean, the benchmark model used in our study. Fortunately, Clark and West (2007) provide an adjusted test statistic (*MSFE-adjusted*) that is suitable for comparing forecasts from nested models. Their proposed test statistic exhibits an asymptotic distribution that is well approximated by the standard normal, and can be calculated in two steps.

First, the $\widehat{f_t}$ values are computed in the out-of-sample period as:

(8) $$\hat{f}_t = (r_t - \bar{r}_t)^2 - [(r_t - \hat{r}_t)^2 - (\bar{r}_t - \hat{r}_t)^2] \qquad \text{for } t = s, \dots, T,$$

where r_t denotes the realizations, $\hat{r}_t$ are the forecasts, and $\bar{r}_t$ are the forecasts of the benchmark model (the historical average, which also represents the nested model if $\beta_i = 0$). Second, the computed $\hat{f}_t$ values are regressed on a constant. The resulting *t*-statistic for a zero coefficient then represents the test statistic of interest. The null hypothesis is rejected at a (one-sided) 10% level if the test statistic is greater than +1.282, or, alternatively, at a 5% level if the test statistic exceeds +1.645 (Clark and West, 2007).

As Neely et al. (2014) show, the decomposition of the *MSFE*, as proposed in Theil (1971), can also provide valuable insights:

(9) $$MSFE = \left(\bar{\hat{r}} - \bar{r}\right)^2 + \left(\sigma_{\hat{r}} - \rho_{\hat{r},r}\sigma_r\right)^2 + \left(1 - \rho_{\hat{r},r}^2\right)\sigma_r^2$$

It is straightforward to show that the second and third summands in Equation (9) correspond to the forecast error variance (i.e., $Var(\hat{r} - r)$). In this way, the *MSFE* is decomposed into the squared bias (systematic forecast error) and the error variance (unsystematic forecast error).

5.2 *Economic evaluation criteria*

Leitch and Tanner (1991) find only a weak relationship between statistical evaluation criteria (e.g., the *MSFE*) and forecast profitability. They note only one criterion, directional accuracy (e.g., the proportion of times the sign of excess returns is correctly predicted), that is significantly correlated with forecasts. For this reason, this evaluation criterion is also reported here.

Pesaran and Timmermann (1992) provide a market timing test statistic that is based on forecast directional accuracy. By using this test statistic, a one-sided test of no market timing skills (null hypothesis) versus the alternative of market timing skills can be conducted. The beginning point for their test is the series of real and predicted excess returns (i.e., r_t and $\hat{r}_t$, respectively), each with length *n*. The asymptotically $N(0,1)$ distributed test statistic S_n is defined as:

$$S_n = \frac{\hat{P} - \hat{P}_*}{\{\widehat{var}(\hat{P}) - \widehat{var}(\hat{P}_*)\}^{1/2}}, \tag{10}$$

where $\hat{P} = n^{-1}\sum_{t=1}^{n} I(r_t \hat{r}_t)$, $\hat{P}_* = \hat{P}_r \hat{P}_{\hat{r}} + (1 - \hat{P}_r)(1 - \hat{P}_{\hat{r}})$, $\hat{P}_r = n^{-1}\sum_{t=1}^{n} I(r_t)$, and $\hat{P}_{\hat{r}} = n^{-1}\sum_{t=1}^{n} I(\hat{r}_t)$. $I(\cdot)$ denotes the indicator function, which is defined as $I(\cdot) = \begin{cases} 1, & if \cdot > 0 \\ 0, & otherwise \end{cases}$, and the variance terms in Equation (10) are defined as:

$\widehat{var}(\hat{P}) = n^{-1}\hat{P}_*(1 - \hat{P}_*)$ and $\widehat{var}(\hat{P}_*) = n^{-1}(2\hat{P}_r - 1)^2 \hat{P}_{\hat{r}}(1 - \hat{P}_{\hat{r}}) + n^{-1}(2\hat{P}_{\hat{r}} - 1)^2 \hat{P}_r(1 - \hat{P}_r) + 4n^{-2}\hat{P}_r \hat{P}_{\hat{r}}(1 - \hat{P}_r)(1 - \hat{P}_{\hat{r}})$.

Based on the direction forecasts, a simple switching strategy is often used in the finance literature where a risky asset is held during periods when its returns are expected to outperform those from holding risk-free bills (i.e., the predicted excess return of the risky asset is positive). If the opposite holds, an allocation to risk-free bills would be taken instead (Pesaran and Timmermann, 1995). This simple market timing strategy has been applied in various gold forecasting studies to measure forecast ability in terms of economic profits (e.g., Pierdzioch et al., 2014a, 2014b, 2015a).

However, risk is not considered in this simple trading strategy, so it implicitly assumes risk-neutral investors. We thus follow various stock market prediction studies here, and evaluate the forecast ability of the models with a utility-based metric. Within this approach, risk aversion is incorporated into the asset allocation decision. The beginning point is a mean-variance investor with a relative risk aversion γ, who allocates his portfolio between gold and risk-free bills based on the predictive regression forecast of the excess return (Equation (1b)). At the end of t, the investor optimally allocates the following proportion of this portfolio to gold during month $t+1$ (e.g., Campbell and Thompson, 2008):

$$(11) \quad w_t = \left(\frac{1}{\gamma}\right)\left(\frac{\hat{r}_{t+1}}{\hat{\sigma}_{t+1}^2}\right),$$

where $\hat{r}_{t+1}$ is the predicted simple gold excess return, and $\hat{\sigma}_{t+1}^2$ is the forecast of its variance.[13]

With an allocation of $(1-w_t)$ into risk-free bills, the portfolio return in $t+1$ is given by:

$$(12) \quad R_{P,t+1} = w_t\left(r_{t+1} + r_{t+1}^f\right) + (1-w_t)r_{t+1}^f,$$

where r_{t+1} labels the excess return of gold over the risk-free rate r_{t+1}^f.With the mean ($\hat{\mu}_P$) and the variance ($\hat{\sigma}_P^2$) of the portfolio returns over the forecast evaluation period, the certainty equivalent is then given by:

[13] This asset allocation exercise is usually implemented with simple (instead of log) returns, so that the portfolio return is given by the sum of the portfolio weights multiplied by asset returns (e.g., Rapach and Zhou, 2013; Neely et al., 2014). Various empirical studies estimate the variance by using the sample variance computed from a rolling or recursive window of historical returns (e.g., Campbell and Thompson, 2008; Neely et al., 2014). However, other (more sophisticated) variance estimators are also possible.

(13) $CE_P = \hat{\mu}_P - \frac{1}{2}\gamma\hat{\sigma}_P^2.$

The certainty equivalent (*CE*) can be interpreted as the risk-free rate when an investor is indifferent to the risky portfolio. Substituting the predictive regression forecasts of gold excess returns in Equation (11) with the corresponding historical mean estimate, the *CE* can also be calculated for this benchmark strategy. The *CE* gain (*ΔCE*) is simply the difference between the predictive regression's *CE* and the historical average *CE*. After multiplying this difference by 1,200, it can be interpreted as the annual percentage portfolio management fee that an investor is willing to pay for the predictive regression forecasts instead of the historical average forecasts (Neely et al., 2014; Campbell and Thompson, 2008).

Because the relationship between the *MSFE* and the utility gains seems weak (Cenesizoglu and Timmermann, 2012), this measure is also reported here. We follow Neely et al. (2014), and set the relative risk aversion coefficient to five ($\gamma = 5$), prevent short sales and allow maximum leverage of 50% (i.e., $0 \leq w_t \leq$1.5). In addition to the *CE* gain (*ΔCE*), we report the monthly Sharpe ratio, defined as the mean of the portfolio excess returns divided by their standard deviations (Sharpe, 1994). In contrast to the *CE* measure, this does not depend on a (investor-specific) relative risk aversion coefficient. All the economic evaluation criteria discussed in this section can also be computed separately to account for differences between expansive and recessive business cycles.

6. Empirical results

6.1 Data

Due to the monthly availability of many fundamental and macroeconomic factors, we conduct our study with monthly data. Our available dataset comprises data from December 1975 through December 2014. Therefore, the in-sample analysis is performed for the January 1976

(1976:01) to December 2014 (2014:12) period. After considering an initial estimation period for the fundamental predictive regression models, the out-of-sample period begins in 1991:01 (and ends in 2014:12).[14]

We use the end-of-month spot gold fixing prices from the London Bullion Market (3:00 PM, London time) in USD.[15] Due to the critical importance of this marketplace, this gold price has been intensively researched in many studies (e.g., Blose, 2010; Capie et al., 2005; Baur et al., 2016; Pierdzioch, 2014a, 2014b, 2015a).[16] Following Pierdzioch et al. (2014a, 2014b, 2015a), we forecast the continuously compounded monthly excess returns of gold over the risk-free rate. As risk-free rate we use the Treasury bill rate provided in Welch and Goyal's (2008) dataset.

As outlined in section 5.2, the forecasted monthly excess returns can be immediately transformed into corresponding trading signals (Pesaran and Timmermann, 1995, p. 1218). In order to properly assess forecasting ability in different market environments, the NBER-dated business cycle expansions and recessions data are applied.[17]

Exhibit 2 reports descriptive statistics for the monthly gold excess returns, as well as the results of some simple weak-form market efficiency tests (i.e., autocorrelation tests and runs tests).[18] The corresponding values are reported for the entire data sample (1976:01-2014:12) and the out-of-sample period (1991:01-2014:12).

[14] All computations are coded with the free R programming environment (R Core Team, 2015).

[15] The gold price series comes from the Federal Reserve Bank of St. Louis (https://research.stlouisfed.org/fred2).

[16] In terms of turnover, the London OTC market is one of the two major markets for gold (the other is the COMEX in New York) (see the comparison of marketplaces by O'Connor et al., 2015).

[17] These data are publicly available at http://www.nber.org.

[18] Several studies analyze the weak-form information efficiency of gold prices (see the literature review in O'Connor et al., 2015, and Charles et al., 2015). However, these tests provide no further information about potential profitability, so they are not explored in detail here.

Exhibit 2: Descriptive statistics and simple tests for weak-form market efficiency

Panel A: Descriptive Statistics							
Period	Mean	Std.	Min	Max	Skewness	Kurtosis	Jarque-Bera
1976:01 – 2014:12	0.05	5.58	-26.39	23.33	-0.03	3.45	226.11***
1991:01 – 2014:12	0.16	4.58	-19.19	15.60	-0.07	1.54	26.90***
Panel B: Autocorrelations							
Period	AC(1)	Q(1)	Q(3)	Q(6)	Q(12)	Q(24)	Q(36)
1976:01 – 2014:12	-0.05	1.11	1.69	10.21	25.96***	36.61**	58.55***
1991:01 – 2014:12	-0.01	4.42**	6.84*	12.72**	24.31**	44.47***	62.05***
Panel C: Runs Test							
Period	Cutoff: Mean	Cutoff: Median	Cutoff: Zero				
1976:01 – 2014:12	-1.19	-0.83	-1.20				
1991:01 – 2014:12	0.13	-0.12	-0.11				

Notes: This table reports descriptive statistics (panel A), as well as results of simple tests for weak-form market efficiency (panels B and C) for monthly gold excess returns over the entire January 1976-December 2014 sample period and the January 1991-December 2014 out-of-sample period. All calculations are based on monthly log excess returns. The mean return and standard deviation in panel A are not annualized. Kurtosis is adjusted so that the normal distribution exhibits a kurtosis of zero. The Jarque-Bera test statistic tests the null hypothesis that the gold excess returns are normally distributed. AC(1) in panel B denotes autocorrelation in the excess returns at lag 1. The statistical significance of autocorrelation is tested using the Ljung-Box test. Q(·) denotes the Q-test statistic that tests for an autocorrelation up to the lag shown in brackets. Panel C gives the test statistics of the runs test, where the definition of a run is based on three different cutoff values: mean, median, and zero. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

The values in panel A show that the mean monthly gold excess return from 1976 through 2014 is remarkably lower than for the 1991-2014 period (0.05 versus 0.16). This observation is attributable to the substantial and persistent increase in gold prices around 2000 (Pierdzioch et al., 2015b). The monthly excess returns in both time periods exhibit skewness of nearly zero, as well as positive excess kurtosis. The normal distribution assumption must be rejected in both cases at a 1% significance level.

Panel B gives the autocorrelations at a lag of 1 (AC(1)), and the results of the Ljung-Box test for higher-order autocorrelations (Q(·)). Regardless of the statistical significance of the autocorrelations, in both periods, all autocorrelations up to lag 36 are small in magnitude (no absolute values higher than 0.2).

The runs test applied in both of the same periods, however, provides a different picture. This test detects no statistically significant run in either of the test periods. The result holds regardless of whether the mean, median, or zero is used as the cutoff value for defining runs. Apart from the heterogenous results of both weak-form market efficiency tests, none of both tests can answer the question, whether the (potentially) detected inefficiencies can be successfully exploited within a realistic investment strategy.

6.2 In-sample analysis

In the next step an in-sample analysis is used to verify the impact direction and explanatory power of the fundamental variables and technical indicators. The corresponding bivariate regression models are estimated from 1976:01 to 2014:12. Exhibit 3 reports the slope coefficients, the Newey-West (1987)-corrected *t*-statistics, and the coefficient of determination (R^2) calculated over the entire data sample and separately for NBER-dated expansion (R^2_{EXP}) and recession cycles (R^2_{REC}). [19]

[19] The significance of regression coefficients is frequently determined using a two-sided hypothesis test. But Inoue and Kilian (2004) recommend a one-sided test in a predictive regression context when theory suggests the sign of the slope coefficient (see also Neely et al., 2014). Note that the sign of some potential influencing factors of gold price fluctuations are clearly theoretically founded, but the opposite holds for some of the other factors (see Exhibit 1). As a result, we do not provide statistical significance levels for the *t*-statistics listed in Exhibit 3, although their determination is straightforward. For a one-sided hypothesis test, the critical values are ±1.28 (10% level), ±1.65 (5% level), and ±2.33 (1% level). The corresponding values for a two-sided test are ±1.65 (10% level), ±1.96 (5% level), and ±2.58 (1% level).

Exhibit 3: In-sample predictive regression results

Predictor	Slope Coefficient	R^2 (%)	R^2_{EXP} (%)	R^2_{REC} (%)	Predictor	Slope Coefficient	R^2 (%)	R^2_{EXP} (%)	R^2_{REC} (%)
Panel A: Regressions with Fundamental Variables					Panel B: Regressions with Technical Indicators				
INFL	0.14 [0.11]	0.01	0.08	-0.23	MA(1-9)	0.14 [0.27]	0.01	0.04	-0.06
USD	-0.08 [-2.58]	1.00	1.85	-1.76	MA(1-12)	0.49 [0.87]	0.15	0.25	-0.21
TBL	-0.15 [-1.42]	0.90	0.72	1.46	MA(2-9)	0.08 [0.16]	0.00	0.00	0.03
LTY	-0.21 [-2.01]	1.11	1.09	1.16	MA(2-12)	0.06 [0.12]	0.00	0.00	0.02
LTR	0.02 [0.30]	0.02	-0.02	0.16	MA(3-9)	0.00 [0.00]	0.00	0.00	0.00
TMS	0.09 [0.33]	0.06	-0.06	0.43	MA(3-12)	-0.12 [-0.21]	0.01	0.02	-0.03
DFY	0.12 [0.18]	0.01	-0.07	0.27	MOM(9)	0.15 [0.29]	0.01	-0.01	0.10
DFR	0.45 [2.08]	1.38	-0.67	8.05	MOM(12)	-0.16 [-0.27]	0.01	0.09	-0.23
ERP	0.02 [0.38]	0.03	0.05	-0.07					
DY	-0.14 [-0.21]	0.01	-0.02	0.13					
RVOL	-0.63 [-0.15]	0.00	0.02	-0.07					
Panel C: Regressions with Diffusion Indices									
$\hat{I}_1^{FUND}$	0.22 [1.03]	0.52	0.47	0.70					
$\hat{I}_1^{TECH}$	-0.02 [-0.17]	0.01	0.00	0.03					
$\hat{I}_1^{ALL}$	0.03 [0.30]	0.02	0.01	0.02					

Notes: This table reports the results of the in-sample predictive regressions with fundamental variables (panel A), technical indicators (panel B), and diffusion indices (panel C). The "Slope Coefficient" column lists the regression coefficients of the predictive variables as well as their Newey-West (1987)-adjusted *t*-statistics. Statistical significance levels are not provided, as it is not obvious whether a one- or two-sided hypothesis test would be most appropriate (see the discussion in section 6.2). For a one-sided hypothesis test, the critical values are ±1.28 (10% level), ±1.65 (5% level), and ±2.33 (1% level). The corresponding values for a two-sided test are ±1.65 (10% level), ±1.96 (5% level), and ±2.58 (1% level). The table also contains the in-sample R^2 measure, as well as the R^2 values calculated separately for expansive and recessive business cycles (section 5.1).

Panel A in Exhibit 3 gives the results for the regressions with fundamental variables. The variables with the highest R^2s are USD (U.S. exchange rate), TBL (T-bill rate), LTY (long-term government bond yield), and DFR (default return spread), all around 1.0. While the slope coefficient of the USD exhibits the theoretically expected negative sign, its influence, with a coefficient of -0.08, is only minor (albeit statistically significant). The business cycle-specific R^2s clearly show that the influence in the expansion cycles (R^2_{EXP} of 1.85) is much more pronounced than in the recession cycles (R^2_{REC} of -1.76).

Both variables of interest, TBL and LTY, exhibit a negative relationship with the gold risk premium. With regression coefficients of -0.15 (TBL) and -0.21 (LTY), both factors also exert a stronger impact on the gold risk premium than USD. Furthermore, both models exhibit comparatively stronger explanatory power during negative business cycles (R^2_{REC} of 1.46 for TBL and 1.16 for LTY).

We observe the highest R^2 (1.38) for the “default return spread” (DFR) variable. Its explanatory power is extremely pronounced during recessive business cycles (R^2_{REC} of 8.05 and R^2_{Exp} of -0.67). While the regression coefficient is statistically significant (coefficient of 0.45 with a *t*-statistic of 2.08), the positive sign is different from what would theoretically be expected (see section 3). Against this background, the out-of-sample predictive ability of this variable is particularly interesting. All other fundamental variables exhibit fairly low in-sample predictive accuracy, as indicated by their *t*-statistics and R^2s.

Panel B in Exhibit 3 reports the regression results for both technical indicators, “moving average” and “momentum,” with their specific parameterizations. We observe remarkably low coefficients of determination for all technical indicators. This holds for the R^2 values over the entire data sample, as well as for those in expansive and recessive business cycles. Furthermore, all regression coefficients are close to zero and far beyond statistically significant (independent of whether a one-sided or two-sided hypothesis test is considered). These results clearly contradict those in Neely et al. (2014), where the same technical indicators are successfully used to predict (and explain) the next period’s stock market risk premium.[20]

[20] See Table 2 in Neely et al. (2014, p. 1776). In their in-sample analysis, all moving average and momentum indicators are statistically significant at 1%, 5%, or at least 10% levels. A potential explanation for this different result is given in section 7.

The in-sample results for the diffusion index approach are in panel C of Exhibit 3. Separate models for the fundamental variables, technical indicators, and both variable groups together are estimated. We evaluate the adjusted R^2s, and report the estimation results for all three diffusion index models with one principal component (Neely et al., 2014).[21] Note that the diffusion index model based on fundamental variables exhibits the highest R^2s, and the highest regression coefficient (by level) with the highest *t*-statistic (albeit not statistically significant). Also in a diffusion index model context, the technical indicators seem to offer no predictive power (at least in-sample). Compared to the best fundamental single-factor models (panel A), the fundamental diffusion index model is clearly inferior. These results also contradict evidence provided in Neely et al. (2014) regarding equity risk premium predictions. But, while this in-sample analysis reveals interesting insights, it is not a suitable substitute for a rigorous out-of-sample analysis.

6.3 Out-of-sample analysis

The out-of-sample analysis is based on an expanding window approach. This means that the initial estimation period from 1976:01 to 1990:12 (to predict the gold excess return in 1991:01) is expanded each month with new available data (the prediction for 1991:02 then depends on parameters estimated from 1976:01 to 1991:01, and so on).

[21] All diffusion index models are estimated up to three principal components. For the fundamental-based and technical indicator-based models, the one-factor model exhibits the highest adjusted R^2. While the first principal component explains 31.5% of the variation in our fundamental variables, it's explanatory power is 82.7% for the technical variables (taking both variable groups together, the first principal component explains 36,5% of the variance). Out-of-sample predictive regression models are implemented in the baseline simulations using only one factor in order to avoid an overparameterized forecasting model (see Rapach and Zhou, 2013). Therefore, we report in-sample results for the one-factor models for all diffusion index models.

Exhibit 4: Out-of-sample forecasting results (based on statistical evaluation criteria)

Predictor	MSFE	R^2_{OS} (%)	MSFE-Adjusted	R^2_{OS} EXP (%)	R^2_{OS} REC (%)	$(\bar{e})^2$	Rem. Term
Panel A: Bivariate Predictive Regression Results with Fundamental Variables							
INFL	21.42	-2.02	-0.99	-1.57	-3.83	0.31	21.11
USD	21.00	-0.03	0.90	1.41	-5.83	0.00	21.00
TBL	21.11	-0.55	1.01	-0.40	-1.12	0.15	20.96
LTY	21.64	-3.07	0.74	-3.46	-1.50	0.50	21.13
LTR	21.08	-0.42	-1.33	-0.36	-0.63	0.09	20.99
TMS	21.03	-0.16	-0.46	-0.16	-0.14	0.08	20.95
DFY	21.30	-1.47	-1.38	-1.16	-2.74	0.01	21.29
DFR	20.63	1.74	1.61*	1.38	3.18	0.07	20.56
ERP	21.25	-1.22	-1.25	-1.01	-2.08	0.08	21.17
DY	21.25	-1.21	0.33	-1.26	-1.02	0.47	20.78
RVOL	21.17	-0.84	-0.16	-0.47	-2.30	0.07	21.10
Panel B: Bivariate Predictive Regression Results with Technical Indicators							
MA(1-9)	21.18	-0.87	-1.10	-0.73	-1.47	0.06	21.11
MA(1-12)	21.20	-0.99	-0.72	-0.80	-1.76	0.06	21.14
MA(2-9)	21.07	-0.38	-1.57	-0.30	-0.69	0.07	21.00
MA(2-12)	21.10	-0.50	-1.71	-0.40	-0.93	0.07	21.03
MA(3-9)	21.08	-0.40	-2.00	-0.40	-0.44	0.08	21.00
MA(3-12)	21.10	-0.51	-1.91	-0.50	-0.52	0.08	21.02
MOM(9)	21.07	-0.35	-0.87	-0.23	-0.82	0.08	20.99
MOM(12)	21.08	-0.40	-1.14	-0.27	-0.91	0.08	21.00
Panel C: Predictive Regression Results with Diffusion Indices							
$\hat{I}_1^{FUND}$	21.06	-0.30	0.54	-0.10	-1.11	0.02	21.04
$\hat{I}_1^{TECH}$	21.12	-0.60	-1.31	-0.51	-0.98	0.07	21.05
$\hat{I}_1^{ALL}$	21.18	-0.87	-1.25	-0.80	-1.14	0.05	21.13

Notes: This table reports the out-of-sample results in terms of statistical evaluation criteria. The results of the predictive regressions with fundamental variables are in panel A, the results with technical indicators are in panel B, and the results with diffusion indices are in panel C. The "MSFE" column contains the mean squared forecast errors, and the "R^2_{OS} (%)" column shows the out-of-sample R^2 statistic (section 5.1). The "MSFE-Adjusted" column reports the test statistic of the one-sided hypothesis test proposed in Clark and West (2007) to test whether the R^2_{OS} value is statistically significantly positive (section 5.1). * (**) indicates statistical significance at a 10% (5%) level. The fifth and sixth columns contain the out-of-sample R^2_{OS} values calculated separately for expansive and recessive business cycles. The last both columns document the results of the error decomposition as proposed in Theil (1971). The seventh column reports the squared bias, and the eighth column gives the remainder term (section 5.1).

Exhibit 4 reports the out-of-sample forecasting results obtained using the statistical evaluation criteria discussed in section 5.1. In panel A, we observe that the predictive regression model based on the default return spread exhibits the lowest *MSFE* of all the fundamental regression models. The statistically significant R^2 of 1.74 (significant at a 10% level) indicates further that this forecast model clearly dominates the historical average forecast as our benchmark (also calculated on an expanding window basis). The dominance is observable for both expansive (R^2_{EXP} of 1.38), and recessive (R^2_{REC} of 3.18) business cycles. As Campbell and Thompson (2008) demonstrate, a monthly out-of-sample R^2 of 0.5% may still be economically significant.

All other fundamental predictive regressions exhibit negative R^2s, indicating lower predictive accuracy than the historical average in terms of the *MSFE*. Only the USD model provides a positive R^2 in expansive phases (R^2_{EXP} of 1.41), but the poor predictive accuracy in recessive phases (R^2_{REC} of -5.83) leads to a slightly negative R^2 over the entire out-of-sample period. While some fundamental predictive regression models provide higher squared biases than others (i.e., INFL, LTY, and DY), the remainder term (representing the variance of the error) is the main source for the *MSFE*.

Panel B in Exhibit 4 gives the corresponding values for the predictive regression models with technical indicators. The negative R^2s show that all predictive models provide poorer prediction quality than the historical average in terms of the *MSFE*. This evidence holds not only over the entire out-of-sample period, but also for all expansive and recessive business cycles. Just as with the fundamental prediction models, the main source for the *MSFE* is not the squared bias, but the variance of the forecast error (the remainder term).

The out-of-sample forecast results for the predictive regressions based on diffusion indices are given in panel C. All three diffusion index models exhibit negative R^2s. This holds for the entire data sample, as well as separately for the expansive and recessive business cycles. Thus, to summarize, neither the technical indicator regression approach nor the diffusion index approach leads to an improvement over the simple bivariate predictive regression models with fundamental variables.

In a next step, the effects of economically motivated restrictions on the performance of fundamental-based predictive regressions are analyzed. We only consider those where the impact direction of the economic factor on the gold excess return is obvious (see Exhibit 1 and the description in section 3).

Exhibit 5: Economically motivated restrictions: Impact on out-of-sample R^2s

	R^2_{OS} Overall (in %)		R^2_{OS} Expansion (in %)		R^2_{OS} Recession (in %)	
Predictor	Coefficient Restriction	Forecast Restriction	Coefficient Restriction	Forecast Restriction	Coefficient Restriction	Forecast Restriction
INFL (+)	-2.04	0.26	-1.59	0.32	-3.83	0.02
USD (-)	-0.03	0.59	1.41	1.48	-5.83	-3.04
DFY (+)	-3.43	-0.35	-3.62	-0.51	-2.66	0.27
DFR (-)	-0.03	0.91	-0.03	0.56	-0.02	2.34
ERP (-)	-0.03	0.01	-0.03	-0.02	0.00	0.14
DY (-)	-60.87	0.22	-69.24	0.19	-27.19	0.32
RVOL (+)	-2.92	0.33	-3.13	0.21	-2.10	0.81

Notes: This table illustrates the implications of coefficient restrictions and forecast restrictions on the R^2_{OS} measure (section 4.3). The R^2_{OS} statistic is calculated over the entire out-of-sample period, as well as separately for expansive and recessive business cycles. This analysis is conducted only for fundamental predictor variables where the expected relationship is clearly positive or negative (see section 3 and Exhibit 1).

Exhibit 5 reports the R^2s when restrictions on the predictive regression coefficients or forecasts are considered (see section 4.3). The values in the second column show that the consideration of coefficient forecasts generally implies a deterioration in R^2s. This holds for the DFY forecasts (R^2_{OS} of -3.43, compared to -1.47), the DFR forecasts (R^2_{OS} of -0.03 compared to +1.74), the RVOL forecasts (R^2_{OS} of -2.92 compared to -0.84), and especially for the DY forecasts (R^2_{OS} of -60.87 compared to -1.21). For the DFR and DY forecasts, the deterioration in R^2_{OS} is observable over expansive and recessive business cycles. Only for the ERP model can one observe an improvement in R^2 after considering coefficient restrictions (R^2_{OS} of -0.03, compared to -1.22). This improvement again holds for both expansive and recessive cycles.

The third column in Exhibit 5 lists the R^2s when the forecasted monthly gold excess return is restricted to be positive (positive risk premium). The introduction of forecast restrictions leads to improved R^2s (with only one exception, DFR). For example, for the INFL model, R^2_{OS} increases from -2.02 to +0.26, and for the USD model it goes from -0.03 to +0.59. For both, the improvement is observable in both expansive and recessive cycles.

Due to the unusually loose link between statistical and economic evaluation criteria (Leitch and Tanner, 1991; Cenesizoglu and Timmermann, 2012), next the prediction models are evaluated by using economic evaluation criteria. Within the simulated investment strategies, turnover-dependent transaction costs of 50 basis points (bp) are considered.

Exhibit 6: Portfolio performance and market timing potential

	Overall			Expansion			Recession		
Predictor	ΔCE	Sharpe	Hit Rate	ΔCE	Sharpe	Hit Rate	ΔCE	Sharpe	Hit Rate
HA	2.89	0.02	0.53	3.03	0.03	0.54*	1.58	-0.06	0.45
Panel A: Bivariate Predictive Regression Results with Fundamental Variables									
INFL	-0.85	-0.04	0.51	-0.62	-0.03	0.53	-2.96	-0.06	0.34
USD	-1.28	0.05	0.48	0.30	0.08	0.49	-16.05	-0.08	0.41
TBL	-1.89	0.03	0.52	-1.24	0.04	0.52	-8.11	-0.04	0.52
LTY	-5.75	0.00	0.51	-5.15	0.01	0.51	-11.70	-0.02	0.52
LTR	-0.52	-0.06	0.53	-0.45	-0.04	0.55*	-1.05	-0.18	0.34
TMS	-0.09	0.00	0.48	-0.13	0.01	0.48	0.28	-0.06	0.45
DFY	-1.96	-0.08	0.49	-2.15	-0.07	0.49	-0.23	-0.12	0.41
DFR	0.12	0.04	0.50	-0.25	0.01	0.50	3.36	0.15	0.48
ERP	-1.86	-0.18	0.49	-1.93	-0.18	0.49	-1.25	-0.16	0.45
DY	0.03	0.02	0.55**	-0.01	0.03	0.55**	0.42	-0.02	0.55
RVOL	-0.26	0.03	0.52	-0.44	0.03	0.52	1.27	0.07	0.48
Panel B: Bivariate Predictive Regression Results with Technical Indicators									
MA(1-9)	-1.04	-0.07	0.50	-1.09	-0.06	0.51	-0.69	-0.35	0.38
MA(1-12)	-0.97	-0.05	0.49	-1.04	-0.05	0.50	-0.48	-0.28	0.41
MA(2-9)	-0.40	-0.05	0.50	-0.35	-0.02	0.51	-0.86	-0.42	0.34
MA(2-12)	-0.56	-0.06	0.50	-0.50	-0.04	0.51	-1.13	-0.41	0.34
MA(3-9)	-0.40	-0.05	0.50	-0.40	-0.03	0.51	-0.42	-0.18	0.48
MA(3-12)	-0.48	-0.05	0.48	-0.54	-0.04	0.49	-0.05	-0.15	0.38
MOM(9)	-0.24	-0.02	0.50	-0.36	-0.02	0.51	0.71	0.01	0.34
MOM(12)	-0.30	-0.03	0.50	-0.39	-0.03	0.52	0.52	-0.03	0.34
Panel C: Bivariate Predictive Regression Results with Diffusion Indices									
$\hat{I}_1^{FUND}$	-0.88	0.01	0.52	-0.54	0.02	0.53	-4.10	-0.10	0.45
$\hat{I}_1^{TECH}$	-0.61	-0.05	0.49	-0.68	-0.05	0.51	-0.03	-0.23	0.34
$\hat{I}_1^{ALL}$	-1.16	-0.07	0.48	-1.28	-0.07	0.49	-0.12	-0.22	0.41

Notes: This table reports the out-of-sample results in terms of economic evaluation criteria, which are calculated over the entire out-of-sample period as well as separately for expansive and recessive business cycles. For the historical average model (HA) as the benchmark prediction model, the absolute certainty equivalent value is reported in the

ΔCE column. In all other cases, the *ΔCE* column lists the certainty equivalent differences of the predictive regression models against the historical average benchmark model. These values can be interpreted as the annual percentage portfolio management fee that an investor would be willing to pay for the predictive regression forecasts instead of the historical average forecasts (section 5.2). The “Sharpe” column contains the monthly Sharpe ratio calculated as the mean of monthly excess returns divided by their standard deviations (Sharpe, 1994). All reported *ΔCE* values and Sharpe ratios are net of 0.5% transaction costs. The “Hit Rate” column reports the proportion of correct direction forecasts in percent. The hit rates are tested for statistical significance with the one-sided hypothesis test proposed by Pesaran and Timmermann (1992). * (**) indicates statistical significance at a 10% (5%) level (section 5.2). The corresponding results of the predictive regressions with fundamental variables are in panel A, the results with technical indicators are in panel B, and the results with diffusion indices are in panel C.

Exhibit 6 reports *ΔCE*, the monthly Sharpe ratio, and the hit rate. As outlined in section 5.2, *ΔCE* represents the annual percentage portfolio management fee an investor is willing to pay for the predictive regression forecast instead of the historical average forecast (note that absolute *CE* is reported for the historical average forecast).

Panel A shows that DY and DFR are the only two fundamental predictive regression models that provide slightly higher *CE*s than the historical average forecast (*ΔCE*s of 3 bp for DY and 12 bp for DFR). The USD model provides rather interesting results. With a *ΔCE* of -1.28, measured over the entire out-of-sample period, this model clearly performs worse than the historical average forecast. However, the same model exhibits a positive *ΔCE* of 0.30 when only expansive business cycles are considered. A negative *ΔCE* over the entire data sample can be attributed to the poor performance of the USD model during recessive business environments (*ΔCE* of -16.05).

A similar phenomenon is observed for the TMS, DFR, DY, and RVOL models. In contrast to the USD model, they exhibit positive *CE* gains against the historical average forecast in recessive business cycles, and negative gains in expansive cycles. Seemingly, the forecasting power of the specific fundamental variables depends strongly on the business cycle.

The same phenomenon plays out as well for the Sharpe ratio. The USD and DFR models feature a monthly Sharpe ratio that is twice as high as that of the historical average forecast over the whole data sample (0.05 versus 0.2 for USD, and 0.04 versus 0.02 for DFR). However, while the USD is particularly dominant in terms of Sharpe ratios during expansive market cycles (Sharpe ratio of 0.08), the DFR dominates during recessive market cycles (Sharpe ratio of 0.15). In terms of correct direction forecasts, we observe a clear superiority of the DY model, with a hit rate of 55%. This result is observable over all three evaluation periods, and is significant for the entire data sample and for the expansive cycle. While the historical average forecast also exhibits a hit rate of 54% in an expansive business cycle (which is statistically significant at a 10% level), all other fundamental models perform more poorly in terms of correct direction forecasts.

Panel B lists the results for the bivariate predictive regression models with technical indicators. Only the momentum models (MOM9 and MOM12) provide higher *CE*s than the historical average forecast in recessive market environments (*ΔCE*s of 71 bp for MOM(9) and 52 bp for MOM(12)). Except for MOM(9) in a recessive business environment, all other technical indicator models feature negative Sharpe ratios in all three evaluation periods. Moreover, with only one exception (the MOM(12) in an expansive business cycle), all hit rates of the technical indicator models are below 52%.

Panel C reports the results for the diffusion index models. All three models provide lower *CE* values in all three evaluation periods than the historical average forecast, as indicated by the negative *ΔCE*s. Only the $\hat{I}_1^{FUND}$ model exhibits a slightly positive Sharpe ratio over the entire data sample and the expansive business cycle, although these values are lower than those of the historical mean prediction. The same holds for the corresponding hit ratios of the $\hat{I}_1^{FUND}$ forecast model in both evaluation periods.

Overall, compared to the fundamental predictive regression results, those of the technical indicator-based and diffusion index models are rather sobering.

In a next step, the effects of coefficient restrictions on the performance of the fundamental-based predictive regressions are analyzed.[22] Again, we consider only fundamental bivariate predictive regressions where the impact direction of the economic factor on the gold excess return is obvious (see Exhibit 1).

[22] Considering forecast restrictions in the context of economic evaluation criteria is somewhat illogical. As demonstrated in Equation (11), substituting for a negative gold excess return forecast with zero implies a portfolio weight of zero (instead of a negative weight indicating a short position). However, short positions would be excluded anyway, due to the allocation restriction that the portfolio weights range from 0 to 1.5 (see the description in section 5.2). Therefore, a forecast restriction leads to the same results for ΔCE and the Sharpe ratio as those already documented in Exhibit 6. In terms of a hit rate evaluation, the restriction that the forecasted gold excess return be either positive or zero implies that the strategy in each month is invested in the gold market (and never in the risk-free rate). This assumes the same hit rate for each forecasting model, which is furthermore identical to the hit rate of a buy-and-hold strategy.

Exhibit 7: Economically motivated restrictions: Portfolio performance and market timing potential

	Overall		Expansion		Recession	
Predictor	No Restriction	Coefficient Restriction	No Restriction	Coefficient Restriction	No Restriction	Coefficient Restriction
			Panel A: ΔCE			
INFL (+)	-0.85	-0.85	-0.62	-0.62	-2.96	-2.96
USD (-)	-1.28	-1.28	0.30	0.30	-16.05	-16.05
DFY (+)	-1.96	-4.94	-2.15	-4.61	-0.23	-8.33
DFR (-)	0.12	-0.01	-0.25	-0.03	3.36	0.11
ERP (-)	-1.86	0.04	-1.93	0.03	-1.25	0.19
DY (-)	0.03	-4.93	-0.01	-5.01	0.42	-4.61
RVOL (+)	-0.26	-6.01	-0.44	-5.04	1.27	-15.30
			Panel B: Sharpe Ratio			
INFL (+)	-0.04	-0.04	-0.03	-0.03	-0.06	-0.06
USD (-)	0.05	0.05	0.08	0.08	-0.08	-0.08
DFY (+)	-0.08	-0.09	-0.07	-0.08	-0.12	-0.16
DFR (-)	0.04	0.01	0.01	0.03	0.15	-0.05
ERP (-)	-0.18	0.02	-0.18	0.04	-0.16	-0.05
DY (-)	0.02	0.03	0.03	0.03	-0.02	-0.01
RVOL (+)	0.03	0.01	0.03	0.01	0.07	0.00
			Panel C: Hit Rate			
INFL (+)	0.51	0.51	0.53	0.53	0.34	0.34
USD (-)	0.48	0.48	0.49	0.49	0.41	0.41
DFY (+)	0.49	0.51	0.49	0.51	0.52	0.41
DFR (-)	0.50	0.53	0.50	0.54*	0.45	0.48
ERP (-)	0.49	0.55*	0.49	0.56**	0.45	0.45
DY (-)	0.55**	0.51	0.55**	0.51	0.55	0.52
RVOL (+)	0.52	0.51	0.52	0.51	0.52	0.48

Notes: This table reports the implications of "coefficient restrictions" on the economic criteria ΔCE (panel A), the Sharpe ratio (panel B), and the hit rate (panel C). Panel A lists the certainty equivalent differences of the predictive regression models against the historical average benchmark model. These values can be interpreted as the annual percentage portfolio management fee that an investor would be willing to pay for the predictive regression forecast instead of the historical average forecast (section 5.2). Panel B contains the monthly Sharpe ratio, calculated as the mean of monthly excess returns divided by their standard deviations (Sharpe, 1994). All reported ΔCE values and Sharpe ratios are net of 0.5% transaction costs. Panel C reports the proportion of correct direction forecasts in percent (hit rate). The hit rates are tested for statistical significance with the one-sided hypothesis test proposed by Pesaran and Timmermann (1992). * (**) indicates statistical significance at a 10% (5%) level (section 5.2). All economic evaluation criteria are calculated over the entire out-of-sample period, as well as separately for expansive and recessive business cycles. This analysis is conducted only for those fundamental predictor variables where the expected relationship is clearly either positive or negative (see section 3 and Exhibit 1).

Panel A in Exhibit 7 reports the results in terms of the "certainty equivalent gain" (ΔCE) economic criterion, compared to the values of the corresponding unrestricted models. For the INFL and USD variables, including the coefficient restrictions is of no consequence. This is because the coefficients of both variables already exhibit the expected sign, so the restrictions are not binding.

Except for the ERP model, the coefficient restriction implies worse results in terms of ΔCE for all other models, with especially pronounced deterioration for DY and RVOL. In both cases, the introduction of coefficient restrictions leads to worse results in expansive and recessive cycles. Only the ERP model exhibits a different result. Within this model, we observe that coefficient restrictions improve ΔCE in both positive and negative business environments. The Sharpe ratios listed in panel B confirm that the introduction of coefficient restrictions does not lead to substantially higher economic profits. While the Sharpe ratios for ERP and DY are slightly higher under the restrictions, the opposite holds for DFY, DFR, and RVOL.

Evaluating the hit rates in panel C provides the same results. While the hit rate of the ERP model rises to a statistically significant 55% after consideration of coefficient restrictions, the statistically significant hit rate of the DY model now decreases from a statistically significant 55% to 51%.

Overall, neither technical indicators, diffusion indices, nor economically motivated restrictions can substantially improve gold market excess return predictions. The results contradict stock market predictions, where these forecasting methods clearly lead to higher predictive accuracy in terms of statistical and economic measures (e.g., Campbell and Thompson, 2008; Rapach and Zhou, 2013; Neely et al., 2014). However, the results provide strong evidence that some fundamental predictor variables perform better during expansive business cycles, while others outperform during recessive business cycles. Furthermore, the forecast power of fundamental predictor variables is apparently not only highly regime-dependent, but also dependent on the economic evaluation criteria being considered. While USD (DFR) generates the highest portfolio performance, as measured by ΔCE and the Sharpe ratio during expansive (recessive) business cycles, DY exhibits the strongest market timing potential as measured by the hit rate (see panel A in Exhibit 6).

6.4 Robustness tests and further insights

In the baseline analysis, an expanding (recursive) estimation window approach is implemented, which means that the estimation sample always begins in 1976:01, with further observations added as they become available. Alternatively, the parameters of the forecasting model can be estimated with a rolling window approach, where earlier observations are dropped when further observations become available. In this way, the estimation window always exhibits a fixed length. Compared to the expanding window approach, the rolling window approach has the disadvantage of higher outlier sensitivity, due to the more limited number of data that define a rolling estimation window. However, the rolling window approach has the advantage that the impact of structural changes in the link between gold excess returns and predictor variables tends to fade as the rolling estimation window moves forward in time (Rapach and Zhou, 2013; Pierdzioch et al., 2014a).

For this reason, a rolling window approach is also implemented in this study (the first estimation window covers 1976:01-1990:12, the second covers 1976:02-1991:01, and so on). This approach ultimately provides not better results than those documented for the expanding window approach.

In order to avoid overparameterized forecasting models, we follow Rapach and Zhou (2013), and implement the predictive regression models with only one diffusion index (the first principal component). As an additional robustness check, the first two principal components are also applied to the diffusion index models. However, these models do not increase forecast accuracy in terms of statistical or economic evaluation criteria. Moreover, the forecasting ability of the dividend-price ratio to the dividend yield variable (DY) is also verified. While the dividend yield sets dividends in relation to lagged stock prices, the dividend-price ratio sets them in relation to actual stock prices (Welch and Goyal, 2008). Both variables return nearly identical results.

In a further robustness test, we follow Hammerschmid and Lohre (2015), and implement more momentum models, for one-, three-, and six-month momentum lengths (i.e., MOM(1), MOM(3), and MOM(6)). The corresponding in- and out-of-sample results in terms of statistical and economic evaluation criteria are in Exhibit 8.

Exhibit 8: Forecasting results of additional momentum strategies

Panel A: In-Sample Forecast Results (statistical evaluation)				
Predictor	Slope Coefficient	R^2_{OS} (%)	R^2_{EXP} (%)	R^2_{REC} (%)
MOM(1)	0.53 [0.91]	0.22	0.19	0.33
MOM(3)	-0.27 [-0.41]	0.05	-0.04	0.37
MOM(6)	0.66 [1.19]	0.29	0.46	-0.26

Panel B: Out-of-Sample Forecast Results (statistical evaluation)							
Predictor	MSFE	R^2_{OS} (%)	MSFE-Adjusted	R^2_{EXP} (%)	R^2_{REC} (%)	$(\bar{e})^2$	Rem. Term
MOM(1)	21.33	-1.58	-0.70	-1.88	-0.36	0.06	21.27
MOM(3)	21.08	-0.40	-0.54	-0.51	0.06	0.09	20.98
MOM(6)	21.38	-1.86	-0.56	-1.57	-2.99	0.05	21.33

Panel C: Out-of-Sample Forecast Results (economic evaluation)									
	Overall			Expansion			Recession		
Predictor	ΔCE	Sharpe	Hit Rate	ΔCE	Sharpe	Hit Rate	ΔCE	Sharpe	Hit Rate
MOM(1)	-2.86	-0.14	0.50	-3.06	-0.14	0.50	-1.28	-0.08	0.52
MOM(3)	-1.32	-0.08	0.54*	-1.31	-0.08	0.53	-1.37	-0.07	0.62*
MOM(6)	-2.10	-0.06	0.48	-2.22	-0.06	0.48	-1.12	-0.28	0.41

Notes: This table reports the results for the three additionally implemented momentum models MOM(1), MOM(3), and MOM(6). Panel A (B) reports the results of an in-sample (out-of-sample) evaluation with statistical criteria. Panel C lists the results in terms of portfolio performance and market timing potential. For the corresponding descriptions, see the table notes in Exhibit 3 (description for panel A), Exhibit 4 (description for panel B), and Exhibit 6 (description for panel C).

Despite the fact that we observe a statistically significant hit rate of 54% for MOM(3), all other results generally confirm the unsatisfying predictive power of the MOM(9) and MOM(12) models. As outlined in section 4.1, the technical indicator models (moving average and momentum) provide 1/0 signals, which indicate a long position in the gold or cash markets (risk-free). These trading signals are transformed with a predictive regression model into quantitative forecasts of the gold excess return. This is necessary in order to verify the predictive ability with statistical evaluation criteria, and also with some portfolio-theoretic-founded economic criteria (i.e., a gain in *CE* or the Sharpe ratio).

In the baseline simulations, the quantitative predictions of the gold excess returns are transformed back into 1/0 signals in order to evaluate the hit rates. However, actual investment practice would be to evaluate the 1/0 signals of the technical indicators directly, with no transformations. Therefore, we also follow this procedure. All hit rates from the six moving average models and the five momentum models are below 50%. This implies that the main reason for the poor predictive performance of these models is the insignificant informational content of the variables, not the linearization due to their application within a linear predictive regression model.

In terms of the analysis regarding economically motivated restrictions, we follow Campbell and Thompson (2008) and Rapach and Zhou (2013), and apply coefficients and forecast restrictions together. The resulting R^2s are similar to those documented in Exhibit 5 (where either coefficient or forecast restrictions were applied solely).

The out-of-sample results documented in section 6.3 suggest that some fundamental predictor variables provide better forecast results during expansive business cycles, while others are superior in recessive business environments. For example, as Exhibit 6 shows, the USD model provides a good ΔCE value and a fairly high Sharpe ratio in expansive cycles. On the other hand, the DFR model dominates in terms of both these performance measures during recessive cycles. Exhibit 9 gives the results for an investment strategy that applies either the USD-based or DFR-based forecasting model depending on the market cycle.

Exhibit 9: Portfolio performance and market timing potential of a sample business cycle-dependent forecasting model

	Overall			Expansion			Recession		
Predictor	*ΔCE*	Sharpe	Hit Rate	*ΔCE*	Sharpe	Hit Rate	*ΔCE*	Sharpe	Hit Rate
USD / DFR	0.56	0.09	0.49	0.30	0.08	0.49	3.36	0.15	0.48
USD	-1.28	0.05	0.48	0.30	0.08	0.49	-16.05	-0.08	0.41
DFR	0.12	0.04	0.50	-0.25	0.01	0.50	3.36	0.15	0.48

Notes: This table reports the *ΔCE*, the Sharpe ratio, and the hit rate of a sample business cycle-dependent forecasting model. The model (labeled as USD/DFR) uses forecasts of the USD factor in expansive business cycles, and forecasts of the DFR factor in recessive cycles. Both models are also reported in the table for comparison (see the "USD" and "DFR" rows). The "*ΔCE*" column lists the CE differences of the predictive regression models against the historical average benchmark model. These values can be interpreted as the annual percentage portfolio management fee that an investor would be willing to pay for the predictive regression forecasts instead of the historical average forecasts (section 5.2). The "Sharpe" column contains the monthly Sharpe ratio calculated as the mean of monthly excess returns divided by their standard deviations (Sharpe, 1994). All reported *ΔCE* values and Sharpe ratios are net of 0.5% transaction costs. The "Hit Rate" column reports the proportion of correct direction forecasts in percentage. The hit rates are tested for statistical significance with the one-sided hypothesis test proposed by Pesaran and Timmermann (1992). * (**) indicates statistical significance at a 10% (5%) level (section 5.2). All economic evaluation criteria are calculated over the entire out-of-sample period as well as separately for expansive and recessive business cycles.

Such a regime-dependent investment strategy provides a *ΔCE* value of 56 bp and a Sharpe ratio of 0.09 over the entire data sample. Both values clearly dominate the corresponding USD and DFR model values. Note that the market state is ex ante not known, however, and therefore also forecast-dependent. This simple exercise demonstrates that a regime-dependent forecasting approach exhibits a reasonable potential to enhance forecast accuracy.

7. Potential explanations

The empirical results of this study show that technical indicators, diffusion indices, and economically motivated restrictions clearly do not provide superior forecasts of gold excess returns. This is in contrast to stock market predictions. This section attempts to provide some coherent explanations for this issue.

Trend-following strategies (like the moving average and momentum strategies implemented here) are only successful if markets tend to be trend-following. Hurst et al. (2013) give a rational explanation for why trends exist and how they emerge. Following their argumentation, a trend emerges from an initial underreaction to new information, combined with a subsequent overreaction that may extend the trend beyond the fundamental value. In addition to some other less significant reasons, certain empirically verified behavioral finance phenomena may also be responsible for the under- and overreaction effects (Hurst et al., 2013).

For example, research has shown that investors tend to "anchor" their views to historical data, and adjust their views only on a step-by-step basis. This anchoring behavior can cause prices to underreact to news (Barberis et al., 1998). The disposition effect (e.g., Shefrin and Statman, 1985; Frazzini, 2006) states that investors tend to sell winners too early in order to realize gains. This phenomenon creates downward price pressure, which causes upward price adjustments to new positive information to lose momentum, and also intensifies the underreaction effect. Once a price trend has begun, some empirically verified phenomena may also extend it beyond the fundamental value (overreaction). Due to the disposition effect, investors also tend to hold losing investments too long in order to avoid realizing losses. Logically, fewer willing sellers prevent prices from adjusting downward as fast as they could. Herding and feedback trading are some other reasons that may extend the trend beyond fundamental value and force an overreaction effect (Hurst et al., 2013). Therefore, in summary, strong evidence exists to explain why trends exist and how they emerge.

Because trend-following strategies are not typically as successful in the gold market as they are in stock markets, the trend-following behavior of both these markets can differ quite dramatically. This section will explore potential explanations for this issue by studying the factors that affect the supply and demand for gold. While the main motivation for a stock market investment is the prospect for positive returns (in the form of dividends and increased value), the reasons behind the demand for gold are manifold. In addition to investments, further demand comes from the jewelry industry, industrial production, and ETF purchasers. However, compared to jewelry production and investment demand, industrial uses and ETFs play only a minor role.

Jewelry represents by far the largest and most stable source of demand for gold, and the primary markets are in Asia, particularly in China and India (O'Connor et al., 2015). Baur (2013b) finds that the wedding season in India and the pre-Christmas season in many developed countries are important drivers of jewelry demand, and generally induce seasonal gold price patterns.

However, it is not only the demand side that can differ between the gold and stock markets. The supply side also can. Just as with demand, the supply side of the gold market is equally manifold. As shown in O'Connor et al. (2015), in addition to mine production and scrap (coming from individuals recycling old jewelry, and, to a lesser degree, electronics), central bank sales/purchases, as well as producer hedging, also heavily influence the supply of gold.

Overall, the completely different supply and demand structures of the gold market versus the stock market seem to be a major reason for the different trending behaviors of these markets. The result in our study is that trend-following strategies (e.g., those based on moving average and momentum indicators) do not work as successfully in gold markets as they do in stock markets. Further research could shed light on this issue.

However, for the other forecast approaches implemented here, it seems reasonable to look for methodological refinements and improvements. In contrast to the positive results reported in some stock market studies, the diffusion index approach applied in this study does not lead to superior forecast results. As in the corresponding stock market studies, all prediction variables are aggregated by means of a principal component approach. However, if the predictive power of the fundamental and macroeconomic variables really does depend on the business cycles (as the evidence in this study suggests), this approach cannot be superior.

Under these circumstances, it would be more rational to cluster all available fundamental and macroeconomic variables into two sets: One for expansive business cycles, and one for recessive ones. Based on these sets, separate diffusion indices for expansive and recessive cycles could be generated and applied in a context of business cycle-dependent forecasts.

The same argumentation holds for the coefficient restrictions. Given that this forecast approach also does not show anticipated positive results, it seems just as plausible that some variables would have a different impact on the forecast variable in expansive and recessive business cycles. This may be especially true for the interest variables, where the direction of the impact can be positive or negative (see section 3 and Exhibit 1).

Beside these more specific concerns, there may be some general concerns which could possibly also explain why we don’t observe a superior predictability like in the case of the equity premium prediction. The recent study of Nguyen et al. (2019) show that the jump risk premium and the variance risk premium of gold are seemingly two powerful predictor variables. The additional consideration of these both predictors could possibly improve the forecast results of the diffusion index approach and/or the coefficient restrictions. Furthermore, we constrained our research on the prediction of monthly gold excess returns. However, some recent research results provide evidence that more long term forecasts of the gold excess returns (e.g., 12 month predictions) are more successful (Prokopczuk et al., 2018; Nguyen et al., 2019). All of these issues warrant further research.

8. Concluding remarks and implications

This study analyzes whether some prediction methods that have been successfully applied in stock return forecasting are also suitable for predicting future gold price fluctuations. Specifically, we explore whether technical indicators, diffusion indices, and economically motivated restrictions for predictive regressions can lead to improvements in order to beat the historical mean return as benchmark model. The results show that these prediction techniques do not result in superior forecasts of gold excess returns. Thus, this study confirms the findings of other studies that have demonstrated that future gold price fluctuations are difficult to forecast (Pierdzioch et al., 2014a, 2014b; 2015a). However, I provide some new insights that could be exploited further in future research.

First, the results show that some fundamental variables are more suitable for forecasting future gold returns during expansive business cycles, while others exhibit higher predictive accuracy during recessive business cycles. Against this background, it seems reasonable to evaluate the predictive power of forecasting techniques that explicitly take regime shifts into account (Hamilton, 1989). While such techniques have been successfully applied in the context of stock market forecasts (Henkel et al., 2011; Hammerschmid and Lohre, 2015) or in the prediction of macroeconomic variables (Koop and Korobilis, 2014), the results here provide strong evidence that they could also be suitable for improving gold market forecasts.

Second, some studies document a loose relationship between statistical and economic evaluation criteria (Leitch and Tanner, 1991; Cenesizoglu and Timmermann, 2012). I observe the same phenomenon among the different economic evaluation criteria. For example, the forecast approaches that generate the highest portfolio performance (as measured by the gain in certainty equivalent or the Sharpe ratio) are not the same as those that provide the highest market timing potential (as measured by the hit rate). Therefore, future academic work, as well as practical real-world investment applications, should explicitly consider this issue. For directional forecasts (used for market timing strategies), Leung et al. (2000) provide evidence that specific classification methods (e.g., linear discriminant analysis, logit and probit models, probabilistic neural networks) lead to better forecasts than level estimation approaches. While these methods have already been successfully applied in a stock market prediction context, evaluating them in terms of gold return predictions would be instructive for future research in order to obtain better directional forecasts.

Finally, given some recent research results (Westerlund and Narayan, 2013; Prokopczuk et al., 2018; Nguyen et al., 2019), the additional consideration of further prediction variables (e.g., the gold futures price, jump risk premium, variance risk premium) as well as the additional implementation of more long-term prediction models also seem to be promising avenues for future research.

References

Abken, P.A. (1980): The Economics of Gold Price Movements, in: *Economic Review*, Federal Reserve Bank of Richmond, March/April, 3-12.

Ang, A., and Piazzesi, M. (2003): A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables, in: *Journal of Monetary Economics*, Vol. 50, 745-787.

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

Baetje, F., and Menkhoff, L. (2016): Equity Premium Prediction: Are Economic and Technical Indicators Unstable? Discussion Paper 1552, DIW Berlin.

Bai, J. (2003): Inferential Theory for Factor Models of Large Dimensions, in: *Econometrica*, Vol. 71, 135-171.

Bampinas, G., and Panagiotidis, T. (2015): Are gold and silver a hedge against inflation? A two century perspective, in: *International Review of Financial Analysis,* Vol. 41, 267-276.

Barberis, N., Shleifer, A., and Vishny, R. (1998): A model of investor sentiment, in: *Journal of Financial Economics*, Vol. 49, 307-343.

Baur, D.G. (2011): Explanatory mining for gold: Contrasting evidence from simple and multiple regressions, in: *Resources Policy*, Vol. 36, 265-275.

Baur, D.G. (2013a): Gold – Fundamental Drivers and Asset Allocation, Working paper, University of Technology, Sydney.

Baur, D.G. (2013b): The autumn effect of gold, in: *Research in International Business and Finance*, Vol. 27, 1-11.

Baur, D.G., Beckmann, J., and Czudaj, R. (2016): A melting pot – Gold price forecasts under model and parameter uncertainty, in: *International Review of Financial Analysis*, Vol. 48, 282-291.

Baur, D.G., and Lucey, B.M. (2010): Is Gold a Hedge or a Safe Haven? An Analysis of Stocks, Bonds and Gold, in: *Financial Review*, Vol. 45, 217-229.

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

Berge, T.J. (2014): Forecasting Disconnected Exchange Rates, in: *Journal of Applied Econometrics*, Vol. 29, 713-735.

Blose, L.E. (2010): Gold prices, cost of carry, and expected inflation, in: *Journal of Economics and Business*, Vol. 62, 35-47.

Bollerslev, T., Tauchen, G., and Zhou, H. (2009): Expected Stock Returns and Variance Risk Premia, in: *Review of Financial Studies*, Vol. 22, 4463-4492.

Bollerslev, T., Todorov, V., and Xu, L. (2015): Tail risk premia and return predictability, in: *Journal of Financial Economics*, Vol. 118, 113-134.

Bühlmann, P. (2006): Boosting for High-Dimensional Linear Models, in: *Annals of Statistics*, Vol. 34, 559-583.

Bühlmann, P., and Hothorn, T. (2007): Boosting Algorithms: Regularization, Prediction and Model Fitting, in: *Statistical Science*, Vol. 22, 477-505.

Campbell, J.Y., and Thompson, S.B. (2008): Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? in: *Review of Financial Studies*, Vol. 21, 1509-1531.

Capie, F., Mills, T.C., and Wood, G. (2005): Gold as a hedge against the dollar, in: *Journal of International Financial Markets, Institutions & Money*, Vol. 15, 343-352.

Cenesizoglu, T., and Timmermann, A. (2012): Do return prediction models add economic value? in: *Journal of Banking & Finance*, Vol. 36, 2974-2987.

Charles, A., Darné, O., and Kim, J.H. (2015): Will precious metals shine? A market efficiency perspective, in: *International Review of Financial Analysis*, Vol. 41, 284-291.

Chen, N.-F. (1991): Financial Investment Opportunities and the Macroeconomy, in: *Journal of Finance*, Vol. 46, 529-554.

Clark, T.E., and West, K.D. (2007): Approximately normal tests for equal predictive accuracy in nested models, in: *Journal of Econometrics*, Vol. 138, 291-311.

Cooper, I., and Priestley, R. (2009): Time-Varying Risk Premiums and the Output Gap, in: *Review of Financial Studies*, Vol. 22, 2801-2833.

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

Erb, C.B., and Harvey, C.R. (2013): The Golden Dilemma, in: *Financial Analysts Journal*, Vol. 69, 10-42.

Fama, E.F., and French, K.R. (1989): Business Conditions and Expected Returns on Stocks and Bonds, in: *Journal of Financial Economics*, Vol. 25, 23-49.

Feldstein, M. (1980): Inflation, tax rules, and the prices of land and gold, in: *Journal of Public Economics*, Vol. 14, 309-317.

Fortune, J.N. (1987): The Inflation Rate of the Price of Gold, Expected Prices and Interest Rates, in: *Journal of Macroeconomics*, Vol. 9, 71-82.

Frazzini, A. (2006): The Disposition Effect and Underreaction to News, in: *Journal of Finance*, Vol. 61, 2017-2046.

Goyal, A., and Welch, I. (2003): Predicting the Equity Premium with Dividend Ratios, in: *Management Science*, Vol. 49, 639-654.

Gupta, R., Majumdar, A., Pierdzioch, C., and Wohar, M.E. (2016): Do Terror Attacks Predict Gold Returns? Evidence From a Quantile-Predictive-Regression Approach, Working paper, University of Pretoria.

Hamilton, J.D. (1989): A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle, in: *Econometrica*, Vol. 57, 357-384.

Hamilton, J.D. (2009): Causes and Consequences of the Oil Shock of 2007-08, in: *Brookings Papers of Economic Activity*, Spring 2009, 215-283.

Hammerschmid, R., and Lohre, H. (2015): Regime Shifts and Stock Return Predictability, Working paper, University of Zurich.

Hartmann, D., Kempa, B., and Pierdzioch, C. (2008): Economic and financial crises and the predictability of U.S. stock returns, in: *Journal of Empirical Finance*, Vol. 15, 468-480.

Hassani, H., Silva, E.S., Gupta, R., and Segnon, M.K. (2015): Forecasting the price of gold, in: *Applied Economics*, Vol. 47, 4141-4152.

Henkel, S.J., Martin, J.S., and Nardari, F. (2011): Time-varying short-horizon predictability, in: *Journal of Financial Economics*, Vol. 99, 560-580.

Hillier, D., Draper, P., and Faff, R. (2006): Do Precious Metals Shine? An Investment Perspective, in: *Financial Analysts Journal*, Vol. 62, 98-106.

Hurst, B., Ooi, Y.H., and Pedersen, L.H. (2013): Demystifying Managed Futures, in: *Journal of Investment Management*, Vol. 11, 42-58.

Inoue, A., and Kilian, L. (2004): In-Sample or Out-of-Sample Tests of Predictability: Which One Should We Use? in: *Econometric Reviews*, Vol. 23, 371-402.

Koenker, R., and Hallock, K.F. (2001): Quantile Regression, in: *Journal of Economic Perspectives*, Vol. 15, 143-156.

Koop, G., and Korobilis, D. (2014): A new index of financial conditions, in: *European Economic Review*, Vol. 71, 101-116.

Koutsoyiannis, A. (1983): A short-run pricing model for a speculative asset, tested with data from the gold bullion market, in: *Applied Economics*, Vol. 15, 563-581.

Leitch, G., and Tanner, J.E. (1991): Economic Forecast Evaluation: Profits versus the Conventional Error Measures, in: *American Economic Review*, Vol. 81, 580-590.

Leung, M.T., Daouk, H., and Chen, A.-S. (2000): Forecasting stock indices: A comparison of classification and level estimation models, in: *International Journal of Forecasting*, Vol. 16, 173-190.

Lucey, B.M. (2011): What Do Academics Think They Know About Gold? in: *The Alchemist*, Issue 62, 12-14.

Ludvigson, S.C., and Ng, S. (2007): The empirical risk-return relation: A factor analysis approach, in: *Journal of Financial Economics*, Vol. 83, 171-222.

Ludvigson, S.C., and Ng, S. (2009): Makro Factors in Bond Risk Premia, in: *Review of Financial Studies*, Vol. 22, 5027-5067.

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

Marshall, B.R., Cahan, R.H., and Cahan, J.M. (2008): Can commodity futures be profitably traded with quantitative market timing strategies? in: *Journal of Banking & Finance*, Vol. 32, 1810-1819.

Mele, A. (2007): Asymmetric stock market volatility and the cyclical behavior of expected returns, in: *Journal of Financial Economics*, Vol. 86, 446-478.

Moskowitz, T.J., Ooi, Y.H., and Pedersen, L.H. (2012): Time series momentum, in: *Journal of Financial Economics*, Vol. 104, 228-250.

Neely, C.J., Rapach, D.E., Tu, J., and Zhou, G. (2014): Forecasting the Equity Risk Premium: The Role of Technical Indicators, in: *Management Science*, Vol. 60, 1772-1791.

Newey, W., and West, K. (1987): A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, in: *Econometrica*, Vol. 55, 703-708.

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

O'Connor, F.A., Lucey, B.M., Batten, J.A., and Baur, D.G. (2015): The financial economics of gold – A survey, in: *International Review of Financial Analysis*, Vol. 41, 186-205.

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

Pesaran, M.H., and Timmermann, A. (1992): A Simple Nonparametric Test of Predictive Performance, in: *Journal of Business & Economic Statistics*, Vol. 10, 461-465.

Pesaran, M.H., and Timmermann, A. (1995): Predictability of Stock Returns: Robustness and Economic Significance, in: *Journal of Finance*, Vol. 50, 1201-1228.

Pettenuzzo, D., Timmermann, A., and Valkanov, R. (2014): Forecasting stock returns under economic constraints, in: *Journal of Financial Economics*, Vol. 114, 517-553.

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

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

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

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

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

Pindyck, R.S., and Rotemberg, J.J. (1990): The Excess Co-Movement of Commodity Prices, in: *Economic Journal*, Vol. 100, 1173-1189.

Prokopczuk, J., Tharann, B., and Wese Simen, C. (2018): Predictability in Commodity Markets: Evidence from more than a Century, Working paper, Leibniz University Hannover/University of Reading.

Pukthuanthong, K., and Roll, R. (2011): Gold and the Dollar (and the Euro, Pound, and Yen), in: *Journal of Banking & Finance*, Vol. 35, 2070-2083.

R Core Team (2015): R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, URL http://www.R-project.org/.

Raftery, A.E., Karny, M., and Ettler, P. (2010): Online Prediction Under Model Uncertainty via Dynamic Model Averaging: Application to a Cold Rolling Mill, in: *Technometrics*, Vol. 52, 52-66.

Rapach, D.E., Strauss, J.K., and Zhou, G. (2010): Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy, in: *Review of Financial Studies*, Vol. 23, 821-862.

Rapach, D.E., and Zhou, G. (2013): Forecasting Stock Returns, in: *Handbook of Economic Forecasting*, G. Elliott and A. Timmermann, eds., Vol. 2A, Elsevier (Amsterdam), 328-383.

Reboredo, J.C. (2013): Is gold a safe haven or a hedge for the US dollar? Implications for risk management, in: *Journal of Banking & Finance*, Vol. 37, 2665-2676.

Sharma, S.S. (2016): Can consumer price index predict gold price returns?, in: *Economic Modelling*, Vol. 55, 269-278.

Sharpe, W.F. (1994): The Sharpe Ratio, in: *Journal of Portfolio Management*, Vol. 21, 49-58.

Shefrin, H., and Statman, M. (1985): The Disposition to Sell Winners Too Early and Ride Losers Too Long: Theory and Evidence, in: *Journal of Finance*, Vol. 40, 777-790.

Sjaastad, L.A. (2008): The price of gold and the exchange rates: Once again, in: *Resources Policy*, Vol. 33, 118-124.

Stock, J.H., and Watson, M.W. (2002): Forecasting Using Principal Components from a Large Number of Predictors, in: *Journal of the American Statistical Association*, Vol. 97, 1167-1179.

Sullivan, R., Timmermann, A., and White, H. (1999): Data-Snooping, Technical Trading Rule Performance, and the Bootstrap, in: *Journal of Finance*, Vol. 54, 1647-1691.

Szakmary, A.C., Shen, Q., and Sharma, S.C. (2010): Trend-following trading strategies in commodity futures: A re-examination, in: *Journal of Banking & Finance*, Vol. 34, 409-426.

Theil, H. (1971): *Applied Economic Forecasting*, North-Holland Publishing (Amsterdam).

Tully, E., and Lucey, B.M. (2007): A power GARCH examination of the gold market, in: *Research in International Business and Finance*, Vol. 21, 316-325.

Vrugt, E.B., Bauer, R., Molenaar, R., and Stelenkamp, T. (2007): Dynamic Commodity Timing Strategies, in: *Intelligent Commodity Investing: New Strategies and Practical Insights for Informed Decision Making*, H. Till and J. Eagleeye, eds., Risk Books (London), 419-438.

Welch, I., and Goyal, A. (2008): A Comprehensive Look at the Empirical Performance of Equity Premium Prediction, in: *Review of Financial Studies*, Vol. 21, 1455-1508.

West, K.D. (1996): Asymptotic Inference about Predictive Ability, in: *Econometrica*, Vol. 64, 1067-1084.

Westerlund, J., and Narayan, P. (2013): Testing the Efficient Market Hypothesis in Conditionally Heteroskedastic Futures Markets, in: *Journal of Futures Markets*, Vol. 33, 1024-1045.

Westerlund, J., and Narayan, P. (2015): Testing for Predictability in Conditionally Heteroskedastic Stock Returns, in: *Journal of Financial Econometrics*, Vol. 13, 342-375.

White, H. (2000): A Reality Check for Data Snooping, in: *Econometrica*, Vol. 68, 1097-1126.

Worthington, A.C., and Pahlavani, M. (2007): Gold investment as an inflationary hedge: Cointegration evidence with allowance for endogenous structural breaks, in: *Applied Financial Economics Letters*, Vol. 3, 259-262.