

# The Risk Premium of Gold\*

Duc Binh Benno Nguyen<sup>†</sup>      Marcel Prokopczuk<sup>†,‡</sup>

Chardin Wese Simen<sup>‡</sup>

June 6, 2018

## Abstract

This paper examines the properties of the gold risk premium. We estimate a parsimonious model for the gold risk premium and uncover important time variations in the dynamics of the risk premium. We also estimate the risk premia of the stock and bond markets and investigate their co-movements. The results show that the co-movements of expected gold returns with stock and bond expected returns are positive, while co-movements of realized returns are zero or negative on average. This results holds not only during normal market periods, but also in times of market stress. Furthermore, we find no significant co-movement of expected and realized returns of gold with inflation.

**JEL classification:** G01, G10, G11, Q02

**Keywords:** Jump Risk; Tail Risk; Safe Haven; Hedge; Gold

---

\*We thank Kees Koedijk (the editor), an anonymous reviewer, Darien Huang (discussant) as well as participants at the meeting of the Commodity and Energy Markets Association (2017) for their comments and suggestions. Contact: [nguyen@fmt.uni-hannover.de](mailto:nguyen@fmt.uni-hannover.de) (D.B.B Nguyen), [prokopczuk@fmt.uni-hannover.de](mailto:prokopczuk@fmt.uni-hannover.de) (M. Prokopczuk) and [c.wese-simen@icmacentre.ac.uk](mailto:c.wese-simen@icmacentre.ac.uk) (C. Wese Simen).

<sup>†</sup>School of Economics and Management, Leibniz University Hannover, Koenigsworther Platz 1, 30167 Hannover, Germany.

<sup>‡</sup>ICMA Centre, Henley Business School, University of Reading, Reading, RG6 6BA, U.K.

# I Introduction

In this paper, we first show that the return of gold is predictable both in-sample and out-of-sample using a parsimonious forecasting model. In a second step, we study the co-movements of the expected return of gold, i.e. the gold risk premium, with those of stock and bond markets as well as with inflation. We find the expected gold returns to be positively related with the expected returns of stock and bond markets. In contrast, the co-movements of realized returns are insignificant or even negative. The same analysis with inflation reveals that both expected and realized gold returns do not co-move with expected and realized inflation, respectively.

Gold is often considered as a store of value. The media often claim that gold is a hedge and safe haven asset and the recent literature has empirically tested this claim ([Capie et al., 2005](#); [Baur & McDermott, 2010](#); [Ciner et al., 2013](#); [Reboredo, 2013](#)). Typically, these studies use realized returns and compute covariances or other dependence measures. As such, they focus on an ex-post setting and only answer the question whether gold and other assets realized returns co-moved. However, from an asset-pricing perspective, it is also important to understand whether the expected returns co-move in a similar fashion.

We contribute to the literature in at least two ways. First, we analyze the time-varying risk premium of gold by providing evidence of predictability. Second, we 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.

We differentiate in our discussion between hedges and safe havens in the spirit of [Baur & McDermott \(2010\)](#). The risk premium of gold is a hedge for another asset if their risk premia are uncorrelated or negatively correlated in general, while it serves as a safe haven asset for another asset when the risk premia of the two assets are uncorrelated or

negatively correlated in times of market stress.

Our empirical analysis shows that the jump tail premium and the variance risk premium of gold are strong predictors for the gold risk premium, with high explanatory power both in-sample and out-of-sample and for all horizons investigated, varying from one month to two years. The adjusted  $R^2$  and out-of-sample  $R^2$  reach values of 13.39% and 12.44% at the one-year horizon, respectively. We then investigate the relationship of gold and equity expected returns, relying on linear regression models. Our equity risk premium model relies on the two most predominant predictors: the dividend yield and the equity variance risk premium. We find that results based on the co-movement of realized returns and that of expected returns are markedly different not only in terms of magnitude but also in terms of sign. The results are similar for bonds. Relying on the framework of [Cochrane & Piazzesi \(2005\)](#) to model the bond risk premium, we find that gold's risk premium is not a hedge or a safe haven for the bond risk premium. The results based on realized returns are reversed. The relationship between gold and inflation is different. Both the expected and realized returns of gold are unrelated to expected and realized inflation, respectively.

Gold markets have been analyzed in several existing studies. [Capie et al. \(2005\)](#) investigate whether the realized return of gold acts as an exchange rate hedge for Sterling–US-Dollar and Yen–US-Dollar exchange rates. They find a negative relationship over more than thirty years (January 1971 to February 2004) of investigation. Their results are based on autoregressive lagged regressions including changes in the gold log-price and exchange rates. [Baur & McDermott \(2010\)](#) test whether gold is a safe haven against the stocks of major emerging and developing countries using realized returns for the period from 1979 to 2009. Gold returns are regressed on stock returns whereby they

differentiate between “normal” returns and extreme returns defined by empirical quantiles of the return distribution. They find that gold acts both as a hedge and a safe haven for major European stock markets and the U.S. but not for Australia, Canada, Japan and large emerging markets. [Baur & Loeffler \(2015\)](#) point out that the demand for gold can serve as a proxy for risk aversion and demonstrate that gold demand has predictive power for equity returns. [Reboredo \(2013\)](#) shows that gold can act as a hedge against U.S. Dollar movements and as a safe haven in periods of financial distress using weekly data in the period from January 2000 until September 2012. The author uses different copulas in order to model the dependence structure. [Ciner et al. \(2013\)](#) examine dynamic conditional correlation (DCC) GARCH models for crude oil, gold, currency, bond and stock markets using daily data from the U.S. and the U.K. Their results indicate that gold performs as a safe haven for exchange rates and bonds while crude oil acts as a safe haven only for bonds.<sup>1</sup>

Our paper is also related to the literature on safe haven assets in general, and in particular to safe haven currencies. Using high-frequency data, [Rinaldo & Soederlind \(2010\)](#) analyze the safe haven properties of major currencies and find that the Swiss Franc and the Japanese Yen act as safe haven against the U.S. stock market. [Habib & Stracca \(2012\)](#) study a large panel of 52 currencies and explore the characteristics of safe haven currencies. They conclude that the explanatory power of most variables they consider is weak but net foreign asset positions can be linked to the safe haven status. The safe haven properties of the Swiss Franc and the Japanese Yen are confirmed by [De Bock & de Carvalho Filho \(2015\)](#) in the context of high-volatility (VIX) periods. [Hossfeld & MacDonald \(2015\)](#) argue that the fact that a currency appreciates in times of market stress is not sufficient to be regarded as safe haven. If the reason for the appreciation is

---

<sup>1</sup>For further literature on gold, we refer to a very comprehensive survey by [O’Connor et al. \(2015\)](#).

simply the unwinding of carry trade positions, a currency should not be considered a safe haven. Empirically, they find that the Japanese Yen does indeed often serve as a funding currency for carry trades and therefore does not qualify as a true safe haven, whereas the Swiss Franc does. Similar to the existing literature on gold, the above mentioned studies analyze the safe haven properties of assets using realized returns. Our paper is different in that our analysis focuses on expected returns, i.e. risk premia.

The rest of the paper is organized as follows. Section II presents our data set. Section III presents our risk premium model for gold and Sections IV and V analyze the hedge and safe haven performance of gold for the stock and bond markets, respectively. Gold's role as an inflation hedge is investigated in Section VI. Section VII reports robustness tests and Section VIII concludes. The online appendix provides some additional tables.

## II Data & Prediction Variables

### A Data

The data used for our subsequent analyses come from various sources. Our primary data set consists of end-of-the-day futures for gold traded on the New York Mercantile Exchange/New York Commodities Exchange (NYMEX/COMEX). These are obtained from the Commodity Research Bureau (CRB). End-of-the-day futures for the S&P 500 index traded on the Chicago Mercantile Exchange (CME) are also obtained from the CRB. Futures contracts have expiration dates and hence cannot be tracked continuously. At each point of time we consider the two nearest contracts. For the computation of the returns, we follow [Diewald et al. \(2015\)](#) and differentiate between normal returns and

roll-over returns. More specifically, we compute gold futures returns as follows:

$$r_{t+1}^{normal} = \log \frac{F_{t+1}^{(1)}}{F_t^{(1)}} \qquad r_{t+1}^{roll} = \log \frac{F_{t+1}^{(1)}}{F_t^{(2)}} \qquad (1)$$

where  $r_{t+1}^{roll}$  denotes the return at time  $t + 1$  on a business day immediately after the expiration day and  $r_{t+1}^{normal}$  denotes the return on any other business day.  $F_t^{(1)}$  and  $F_t^{(2)}$  refer to the first nearby contract and the second nearby contract at time  $t$ , respectively.<sup>2</sup> This approach ensures that all returns are “real” returns, i.e. they are based on two consecutive prices of the same contract.

Options data for gold are obtained from the CRB and contain information on the strike price, maturity and settlement price. The options on futures contracts are traded on NYMEX/COMEX. Implied volatilities are calculated using binomial trees. We also use options data for the S&P 500 index. These consist of closing bid and ask quotes, strike prices, maturities and implied volatilities of options traded on the Chicago Board of Options Exchange (CBOE) and are obtained from Optionmetrics.

Our analysis covers the period from January 1996 until February 2015 leading to a total of 4825 trading days.<sup>3</sup> Our gold data set comprises American options. For short maturity, deep out-of-the-money (OTM) options, the difference between European and American options is negligible, so we rely on the original prices.<sup>4</sup>

---

<sup>2</sup>We also consider alternative rolling dates such as the end of the first or second month prior to delivery month in order to avoid irregular price behavior and obtain qualitatively similar results (Szymanowska et al., 2014). The relevant tables are available upon request. Following Gorton & Rouwenhorst (2006), Gorton et al. (2013) and Bhardwaj et al. (2015), the return of the futures price  $r_t$  is defined as an excess return without subtracting any proxy for the risk-free rate. For our predictive regressions, we also consider futures returns in excess of the one-month treasury bill, which is obtained from Kenneth R. French’s data library, leading to qualitatively similar results.

<sup>3</sup>The earliest available date for the gold options is 1989, but we start our analysis at a slightly later point since the data in Optionmetrics start only in 1996.

<sup>4</sup>Bakshi et al. (2003) argue that the early-exercise premium of OTM options can be ignored and hence the usage of American options barely changes the results. Further, Barone-Adesi & Whaley (1987) argue that the early-exercise premium is negligible for OTM options with a time-to-maturity less than 100 days. We conduct robustness tests supporting our choice of using the original options data.

## B Predictor Variables

### 1 Macroeconomic Variables

Macroeconomic variables such as employment rates, federal funds rates, industrial production, inflation or treasury bill rates are potential predictors of stock market movements (Geske & Roll, 1983; Thorbecke, 1997; Rapach et al., 2005; Chen, 2009). They affect future consumption and investment opportunities and consequently also stock returns, as outlined by the consumption capital asset-pricing model (CCAPM). Changes in interest rates are related to discounted cashflows on the one hand and represent monetary policy on the other, which both impact stock prices and returns. The macroeconomic variables have been shown to be related to gold returns as well. Sherman (1983), Fortune (1988), Jaffe (1989), Mahdavi & Zhou (1997), Ghosh et al. (2004) and Blose (2010), among others, investigate the relationship between gold prices and inflation. The impact of macroeconomic news announcements on gold prices has been analyzed by Christie-David et al. (2000) and Cai et al. (2001), especially news on the inflation and employment rate. We further include the oil price and the U.S. Dollar index as macroeconomic predictor variables for the gold premium (Capie et al., 2005; Levin et al., 2006; Tully & Lucey, 2007; Pukthuanthong & Roll, 2011; Baur, 2013; Reboredo, 2013):

- **Dollar:** The U.S. Dollar index is a real trade weighted index obtained from the Federal Reserve Bank of St. Louis (FRED) and presents a weighted average of the foreign exchange value against the currencies of major U.S. trading partners.<sup>5</sup> Our predictor variable is defined as changes in the U.S. Dollar index.
- **Employment, Federal funds rate and industrial production:** (Empl., FFR, IP) are employment rates, federal funds rates and industrial production obtained

---

<sup>5</sup>For details, we refer to the FRED website: <https://fred.stlouisfed.org/series/TWEXBPA>.

from FRED and the Board of Governors of the Federal Reserve System (FED). All time series are filtered by the Hodrick–Prescott filter ( $\lambda = 129,600$ ) following [Bloom \(2009\)](#).

- **Inflation:** This is defined as the change in the Consumer Price Index (CPI) and obtained from the Bureau of Labor Statistics.
- **Oil price changes:** (OIL) We include monthly changes in the nominal price of oil (West Texas Intermediate) obtained from FRED.
- **Treasury bill rates:** (Tbill) are the three-month treasury bill rates obtained from FRED.

## 2 Equity Market Related Variables

The dividend yield and earnings price ratio measure the stock price relative to fundamentals and are the most popular equity premium predictors ([Rozeff, 1984](#); [Campbell & Shiller, 1988](#); [Fama & French, 1988](#); [Hodrick, 1992](#); [Kothari & Shanken, 1997](#); [Lamont, 1998](#); [Lewellen, 2004](#)).

- **Dividend yield and earnings price ratio:** The dividend yield, which is defined as the difference between the log of dividends and the log of past prices  $\log(D/P)$  and the earnings price ratio, which is defined as the difference between the log of earnings and the log of prices  $\log(E/P)$ , measure the stock price relative to fundamentals and are the most popular equity premium predictors ([Rozeff, 1984](#); [Campbell & Shiller, 1988](#); [Fama & French, 1988](#); [Hodrick, 1992](#); [Kothari & Shanken, 1997](#); [Lamont, 1998](#); [Lewellen, 2004](#)).

[Kojien et al. \(2018\)](#) show that the carry of a commodity futures represents the ex-



pected returns of a commodity futures if nothing changes in the market. Motivated by this insight, we consider the basis of gold futures as a potential predictor. We consider three different definitions of the basis:

- **Basis:** [Fama & French \(1987\)](#) define the monthly basis  $B_t^{FF}$  as the normalized difference between the cash and futures price  $B_t^{FF} = \frac{F_{t,T} - S(t)}{S(t)}$  for one-, three-, six- and twelve-month maturities.

[Gorton & Rouwenhorst \(2006\)](#) and [Bhardwaj et al. \(2015\)](#) calculate the basis as the normalized difference between the first and second nearest futures contracts, which presents the slope of the futures curve:  $B_t^{GR} = \frac{F_{t,1} - F_{t,2}}{F_{t,1}} \frac{365}{t_2 - t_1}$ , where  $F_{t,1}$  and  $F_{t,2}$  are the two contracts closest to maturity with the relevant time to maturities  $t_1$  and  $t_2$ .

[Yang \(2013\)](#) defines the monthly basis  $B_t^Y$  as the normalized log difference between the one-month and twelve-month contract:  $B_t^Y = \frac{\log(F_{t,T_1}) - \log(F_{t,T_{12}})}{T_{12} - T_1}$  where  $F_{t,T_1}$  and  $F_{t,T_{12}}$  are the one-month and twelve-month futures prices, respectively.

### 3 Uncertainty and Tail Risk

Another source of fluctuations in gold and stock returns are changes in economic uncertainty ([Bansal & Yaron, 2004](#); [Bekaert et al., 2009](#)). Various methods have been introduced in order to capture uncertainty. Stock market volatility can be viewed as a measure of economic uncertainty, which has been represented by either the stock market variance ([French et al., 1987](#)) or the implied variance ([Bloom, 2009](#)). There is a growing literature investigating the predictability of the (equity) risk premium using the difference between the two, the variance risk premium, which proxies the aggregate degree of risk aversion in the market ([Bollerslev et al., 2009, 2014](#); [Bekaert & Hoerova, 2014](#)). Lastly, recent studies address the ability of rare disaster event models to capture the (equity)

risk premium (Gabaix, 2012; Wachter, 2013). We rely on the jump risk premium following Bollerslev & Todorov (2011b) and Bollerslev et al. (2015), which has been shown to amount for a large fraction (two-thirds) of the future excess return:

- **Left and right jump tail premia (LJP, RJP):** The calculation of the jump tail premia closely follows the approach of Bollerslev & Todorov (2011a) and Bollerslev et al. (2015). The jump risk premium is defined as:

$$JP_t(k) = \frac{1}{\tau} \left[ \mathbb{E}_t^{\mathbb{Q}} \left( \int_t^{t+\tau} \int_{|x|>k} x \nu_s^{\mathbb{Q}}(dx) ds - \mathbb{E}_t^{\mathbb{P}} \left( \int_t^{t+\tau} \int_{|x|>k} x \nu_s^{\mathbb{P}}(dx) ds \right) \right) \right] \quad (2)$$

We denote the left and right risk-neutral components of the jump tail premia as  $LJP^{\mathbb{Q}}$  and  $RJP^{\mathbb{Q}}$ , which are given by:

$$LJP^{\mathbb{Q}} = \int_t^{t+\tau} \int_{x<k} x \nu_s^{\mathbb{Q}}(dx) ds \quad (3)$$

$$RJP^{\mathbb{Q}} = \int_t^{t+\tau} \int_{x>k} x \nu_s^{\mathbb{Q}}(dx) ds \quad (4)$$

where the jump intensity process  $\nu_s^{\mathbb{Q}}(dx)$  is a function of a level shift parameter  $\phi_t^{\pm}$  and a tail decay parameter  $\alpha_t^{\pm}$ , which allow for time-varying and asymmetric dynamics for the left and the right tail:

$$\nu_t^{\mathbb{Q}}(dx) = \left[ \phi_t^- e^{-\alpha_t^- |x|} \mathbf{1}_{x<0} + \phi_t^+ e^{-\alpha_t^+ |x|} \mathbf{1}_{x>0} \right] dx. \quad (5)$$

The left and right tail measures are estimated in a two-step procedure where the tails are extrapolated from the short maturity and deep OTM options. Applying extreme

value theory leads to the following approximations, see [Bollerslev & Todorov \(2014\)](#):

$$\frac{e^{r_{t,\tau}} O_{t,\tau}(k)}{F_{t,\tau}} \approx \frac{\tau \phi_t^\pm e^{k(1 \mp \alpha_t^\pm)}}{\alpha_t^\pm (\alpha_t^\pm \mp 1)} \quad (6)$$

$$1 \pm \alpha^\pm \approx \frac{\log(O_{t,\tau}(k_{t,i})) - \log(O_{t,\tau}(k_{t,i-1}))}{k_{t,i} - k_{t,i-1}} \quad (7)$$

where  $O_{t,\tau}(k)$  denotes the price of an option with maturity  $\tau$  and log-moneyness  $k$  at time  $t$ .  $F_{t,\tau}$  is the corresponding futures price and  $r_{t,\tau}$  is the risk-free rate over the same horizon. The two parameters completely describe the jump intensity process resulting in the first moment of the jump intensity, i.e. for the time interval from  $t$  to  $t + \tau$ :

$$RJP_{[t,t+\tau]}^{\mathbb{Q}} = \tau \phi_t^+ e^{-a_t^+ k_t} \frac{[a_t^+ k_t + 1]}{(a_t^+)^2} \quad (8)$$

$$LJP_{[t,t+\tau]}^{\mathbb{Q}} = \tau \phi_t^- e^{-a_t^- k_t} \frac{[-(a_t^- k_t + 1)]}{(a_t^-)^2} \quad (9)$$

In unreported results, we show that the physical components of the jump tail premia are dwarfed by their risk-neutral counterparts and hence we focus on the risk-neutral components as a proxy for the jump tail premia.

- **Model-free implied variance (MFIV):** For the S&P 500 we proxy the model-free implied variance by the VIX obtained from Optionmetrics. For gold, we rely on the methodology proposed by [Bakshi et al. \(2003\)](#). The annualized model-free implied variance can be described as:

$$MFIV = \frac{1}{\tau} \int_{F_{t,\tau}}^{\infty} \frac{2(1 - \ln(\frac{K}{F_{t,\tau}}))}{K^2} C(t, \tau, K) dK + \int_0^{F_{t,\tau}} \frac{2(1 + \ln(\frac{K}{F_{t,\tau}}))}{K^2} P(t, \tau, K) dK \quad (10)$$

where  $F_{t,\tau}$  is the price of a futures contract at time  $t$  with time to maturity  $\tau$ .

$C(t, \tau, K)$  and  $P(t, \tau, K)$  denote the European call and put option prices at time  $t$  with strike  $K$  and time to maturity  $\tau$ , respectively.

- **Stock variance (Stock Var.):** We include the monthly stock variance which is given by the sum of squared daily returns in that month. The same procedure is applied to gold returns in order to obtain gold realized variances. We multiply the monthly realized variance by 12, in order to annualized the estimates.
- **Variance risk premium (VRP):** The monthly variance risk premium is defined as the difference between the annualized implied variance and the realized variance ([Bollerslev et al., 2009](#)).

### III Gold Risk Premium Prediction Model

#### A Risk Premium Prediction

Our first objective is to analyze whether the gold risk premium is time-varying. To find the best model, we include a variable only if it is a statistically significant regressor and is able to increase the explanatory power when added to the model. The variables we consider include financial, macroeconomic and option implied measures and are described in Section II.B. In summary, we use the following 18 predictor variables: gold basis, dividend yield, trade weighted U.S. Dollar index, earnings price ratio, employment rates, federal funds rate, implied variance of gold, industrial production, inflation, left and right jump risk premium (gold and stock market), oil price changes, stock market variance, treasury bill rates and variance risk premium (gold and stock market). We estimate the

following regression model for the gold futures return:

$$r_{t+h} = a_h + b_h X_t + \epsilon_{t+h} \quad (11)$$

where  $r_{t+h}$  is the continuously compounded excess futures return over the horizon  $h$ ,  $X_t$  presents one or more of the introduced predictor variables at time  $t$  and  $\epsilon$  is the error term. In order to account for the overlapping observations, we use [Newey & West \(1987\)](#) standard errors with lags equal to the return horizon expressed in months. In addition, we compute the more conservative [Hodrick \(1992\)](#) standard errors. We focus our discussion on the estimated slope coefficients and their statistical significance and the forecast accuracy of the regressions as measured by the corresponding adjusted  $R^2$ .

Table 1 summarizes the significance of the individual explanatory variables in simple regressions from the one-month to the two-year horizon. Even though the  $VRP^{S\&P}$  seems to show relatively good forecasting performance in simple regressions for gold futures excess returns, the variable is insignificant in multiple regressions and hence is excluded from the model. Investigating all predictor variables, we find that the best model for the gold risk premium includes the left jump risk premium ( $LJP$ ) of gold and the gold variance risk premium ( $VRP$ ) as explanatory variables. The final model is:

$$r_{t+h}^{Gold} = a_h + b_{1,h} LJP_t^{Gold} + b_{2,h} VRP_t^{Gold} + \epsilon_{t+h} \quad (12)$$

Panel A of Table 2 presents the results from the multiple regressions for the horizons from one month to two years.

The table shows that both the  $LJP$  and the  $VRP$  of gold are statistically significant predictors of futures excess returns for all horizons. The gold  $VRP$  is positively related

to future returns while  $LJP$  is positively related as well. The left jump tail premium is the difference between the risk neutral and the physical expectations of the left tail. Since the physical part of the jump tail premium is dwarfed by the risk-neutral component, the left jump tail premium can be expressed as  $LJP^Q$ . Hence more demand for downside protection (via put options) is associated with higher future excess returns. Our results are consistent with this intuition.

When relying on [Hodrick \(1992\)](#) standard errors, at least one explanatory variable is statistically significant and both coefficients for three of the six horizons, while the Wald test rejects the null of joint insignificance of the predictors for all horizons. The explanatory power in terms of adj.  $R^2$  varies from 4.65% to 15.14%.

We find that the contribution of individual predictor variables to the explanatory power depends on the return horizon. The time-series of the individual t-statistics from both simple and multiple regressions as well as the corresponding adjusted  $R^2$  are illustrated in Figure 1. The  $VRP$  (dotted line) shows generally larger t-statistics than the  $LJP$  (solid line) while both are statistically significant throughout all horizons, for both simple and multiple regressions. While both predictors contribute equally to the relatively high adj.  $R^2$  for short horizons, the additional explanatory power from the  $VRP$  when added to the  $LJP$  is much lower for longer horizons.<sup>6</sup>

In summary, the predictors in our prediction models have a both statistically and economically significant impact on future excess futures returns in the gold market. While their contribution to the explanatory power of the predictors depends on the prediction

---

<sup>6</sup>Our findings for the  $VRP$  are consistent with the literature. [Bollerslev et al. \(2009\)](#) find that the explanatory power of the  $VRP$  for the U.S. peaks at the horizon between three and six months and generally tapers off for longer return horizons, which is in line with the implications from their theoretical model. Further, they show that there is a positive relationship between the  $VRP$  and future expected returns, which is also consistent with our results. [Bollerslev et al. \(2014\)](#) extend these patterns to major economies including Belgium, France, Germany, Japan, the Netherlands, Switzerland and the U.K. These studies focus on the stock market while we focus on the gold market in this section. In Section IV we show similar results for the equity market as well.

horizon, they jointly ensure a generally increasing pattern for longer horizons.

## B Out-of-sample Prediction

Having investigated the in-sample predictability, we now turn to an out-of-sample setting. As argued by [Welch & Goyal \(2008\)](#), it is not sufficient to only investigate in-sample tests since most of the predictors are unable to consistently forecast the excess returns out-of-sample. Most of their examined models underperform the recursive mean model out-of-sample. Similar to them, we use the recursive mean as a benchmark for our models. The historical mean is simply given by:

$$\bar{r}_{t+h} = \frac{1}{N} \sum_{j=1}^t r_j \quad (13)$$

using  $N$  return observations until  $t$ . Following [Campbell & Thompson \(2008\)](#), we evaluate our models using the out-of-sample  $R^2$  which compares the mean squared prediction errors (MSPE) for the predictive model and the historical mean model, and is given by:

$$R_{OOS}^2 = 1 - \frac{\sum_{t=s}^T (r_{t+1} - \hat{r}_{t+1})^2}{\sum_{t=s}^T (r_{t+1} - \bar{r}_{t+1})^2} \quad (14)$$

where  $\hat{r}_{t+1}$  stands for the out-of-sample forecast obtained from the model in Equation (12) using the data until  $t$  and  $s$  is the break point splitting the whole sample for the out-of-sample analysis. Positive values for  $R_{OOS}^2$  indicate that the predictor outperforms the historical mean model in terms of the MSPE. We formally test whether our models significantly outperform the historical mean model using the [Clark & West \(2007\)](#)-adjusted test, i.e. testing the null of  $R_{OOS}^2 \leq 0$ . Under the null hypothesis, the MSPE-adjusted

test statistic of [Clark & West \(2007\)](#) follows a standard normal distribution. Defining

$$f_{t+1} = (r_{t+1} - \bar{r}_{t+1})^2 - [(r_{t+1} - \hat{r}_{t+1})^2 - (\bar{r}_{t+1} - \hat{r}_{t+1})^2] \quad (15)$$

and regressing  $f_{t+1}$  on a constant, i.e.  $f_{t+1} = \alpha + \epsilon_{t+1}$ , the MSPE-adjusted test statistic is equal to the t-statistic of the constant.

Panel B of Table 2 reports the results for the out-of-sample predictability analysis using five years of monthly observations for the initial estimation. Our prediction model shows good out-of-sample forecasting performance across all horizons, where it is able to outperform the historical mean model ( $R_{OOS}^2 > 0$ ). The higher performance relative to the historical mean is statistically significant for five of the six horizons. The  $R_{OOS}^2$  reaches values as high as 12.83% at the two-year horizon. In accordance with our previous results, our model is not only able to predict excess futures returns in-sample but is also able to beat the historical mean model out-of-sample. As such, as a first major result, we provide evidence that the excess return of gold is time-varying and predictable.

## IV Gold and the Stock Market

In this section we investigate the relationship of the gold and equity markets. In particular, we analyze the co-movements of the expected and realized returns of gold and stocks.

### A Equity Premium Prediction Model

In order to study the co-movement between the gold and the equity risk premia, we first also need to obtain predictions for the latter. We follow the same approach as for the gold



risk premium and consider the same variables as discussed in Section III. Table 3 shows the significance of the individual explanatory variables in simple regressions from the one-month to the two-year horizon. Even though  $RJP^{Gold}$  seems to show relatively good forecasting performance in simple regressions for the S&P 500 futures excess returns, it is insignificant in multiple regressions and hence is excluded from the model. The best model for the equity risk premium includes the dividend yield and the S&P 500 variance risk premium:

$$r_{t+h}^{S\&P} = a_h + b_{1,h} \log(D/P)_t + b_{2,h} VRP_t^{S\&P} + \epsilon_{t+h} \quad (16)$$

The results for the predictability regressions using this model are reported in Panel A of Table 4. We find that all coefficients are statistically significant at the 5% level or lower according to [Newey & West \(1987\)](#) standard errors and the signs make sense economically, just as for the gold market. A higher equity  $VRP$  leads to higher future returns. The  $VRP$  can be interpreted as a measure of aggregate economic uncertainty and the positive sign is consistent with the results of [Bollerslev et al. \(2009\)](#) and [Bollerslev et al. \(2014\)](#). [Bloom \(2009\)](#) shows that higher uncertainty impacts the aggregate real economy by lowering industrial production and employment rates, which again influences asset prices. The positive sign of the dividend yield slope coefficient is consistent with the literature. As argued by [Lewellen \(2004\)](#), the ratios should positively impact expected returns. This positive relationship is prescribed by a present value model ([Campbell & Shiller, 1988](#)). Our (best) model is able to explain 19.46% of the variation in expected returns at the six-month horizon.<sup>7</sup> The Wald test of joint significance rejects the null in favor of the

---

<sup>7</sup>This high adjusted  $R^2$  is comparable to the 21.39% of [Bollerslev et al. \(2015\)](#), who include the left jump tail variation and the dividend yield as predictors for the period from 1996 until 2013.

prediction model. Looking at the more conservative Hodrick (1992) standard errors, the slope coefficients both remain statistically significant for four of the six horizons, similar to the results for gold. For the nine- and twelve-month horizon, the  $VRP$  is not significant but the Wald statistic indicates the joint significance of both predictors with all values being above 10. We find a generally increasing pattern for the adjusted  $R^2$ , starting with 11.58% at the one-month horizon and reaching values as high as 41.64% for the two-year horizon.<sup>8</sup>

Turning next to the term structures of t-statistics and adj.  $R^2$  in Figure 2, we find that the t-statistics of the  $VRP$  (dotted line) are generally higher for short horizons up to nine months while the t-statistics of the  $\log(D/P)$  dominate for longer horizons. This is true for both simple and multiple regressions. In addition, the explanatory power is slightly higher when relying on the  $VRP$  for short horizons while it almost vanishes for longer horizons. This is manifested in the third plot showing the adj.  $R^2$ . There is a large increase for short-term horizons, where both the  $\log(D/P)$  and the  $VRP$  contribute significantly to a high explanatory power while the explanatory power mainly comes from the  $\log(D/P)$  for long horizons.

Panel B of Table 4 demonstrates that the equity risk premium prediction model shows good out-of-sample forecasting performance across all horizons. It is able to outperform the historical mean model ( $R_{OOS}^2 > 0$ ). The higher performance relative to the historical mean is statistically significant for all horizons. The  $R_{OOS}^2$  reaches values as high as 9.42% for the S&P 500.<sup>9</sup> Just as for the gold market, our equity risk premium model is able

---

<sup>8</sup>Our model delivers a higher explanatory power than proposed models of Welch & Goyal (2008), Kelly & Jiang (2014) and Bollerslev et al. (2015) for the one-month and the one-year horizons. For the one-year horizon, the authors find an adjusted  $R^2$  of 16.98% ( $LJV$  and continuous  $VRP$ ), 13.81% (kitchen sink regression) and 13.80% (tail risk and dividend yield), respectively, compared to our  $R^2$  of 24.84%.

<sup>9</sup>For comparison, Welch & Goyal (2008) and Kelly & Jiang (2014) find  $R_{OOS}^2$  of 0.2% and 0.3% for the one-month horizon and 2.04% and 4.5% for the one-year horizon when predicting the equity premium. The best performing models of Welch & Goyal (2008) rely on the Term Spread ( $tms$ ) and the

not only to predict excess futures returns in-sample but is also able to beat the historical mean model out-of-sample.

## B Co-movements of Gold with the Equity Market

We test the performance of gold as a hedge or safe haven asset following the approach of [Baur & McDermott \(2010\)](#). However, we rely on expected returns rather than realized returns as [Baur & McDermott \(2010\)](#) do. Thus, we analyze whether expected gold returns positively or negatively co-vary with expected equity returns. The model differentiates between co-movements on average and in times of extreme market movements. More formally, we jointly estimate the following regressions using the maximum likelihood method:

$$\hat{r}_t^{Gold} = a + b_t \hat{r}_t^{Stock} + \epsilon_t \quad (17)$$

$$b_t = c_0 + c_1 D(\hat{r}^{Stock} q_{10}) \quad (18)$$

$$h_t = \omega + \alpha \epsilon_{t-1}^2 + \beta h_{t-1} \quad (19)$$

where Equation (17) models the relation of the expected returns (premia) and  $\epsilon_t$  is the error term.<sup>10</sup> The slope coefficient  $b_t$  is a dynamic process and depends on  $c_0$  and  $c_1$ , the parameters of interest.  $D(\hat{r}^{Stock} q_{10})$  is a dummy variable which captures extreme stock market movements and equals one if the expected return  $\hat{r}^{Stock}$  falls below the 10% quantile of the distribution. Equation (19) presents a GARCH(1,1) model and allows for heteroskedasticity.

The regressions are based on conditional estimates of the expected returns and hence

---

Percent Equity Issuing (*eqis*) while [Kelly & Jiang \(2014\)](#) rely on their tail risk estimate  $\lambda$ .

<sup>10</sup>Normality is assumed for the error term. Our conclusions remain qualitatively similar when assuming a t-distribution.

shed light on the co-movements of the expected returns of gold and stocks. For comparison, we also re-estimate the Equations (17)-(19) using realized returns.<sup>11</sup>

The parameters of interest ( $c_0$  and  $c_1$ ) indicate whether gold serves as a hedge and/or a safe haven, i.e. how expected returns of gold and stocks co-move. If  $c_0$  is zero (negative and statistically significant) and  $c_1$  is not positive, exceeding the value of  $c_0$ , gold is a weak (strong) hedge. If both parameters are non-positive (and statistically significant), gold acts as a weak (strong) safe haven.

The results of our analysis are reported in Table 5. We focus on the following four time horizons: one-month, six-month, one-year and two-year, which include horizons of short-, mid- and long-term investors, respectively. In Table 5, columns (1) and (2) report the coefficients estimated from Equations (17)-(19) relying on the expected returns, while columns (3) and (4) show the coefficients estimated using the realized returns as dependent variables. The statistical significance of the coefficients is obtained from the Wald test statistics which are reported below the coefficients.<sup>12</sup>

One can observe that for the expected gold returns, none of the above criteria for a hedge or safe haven asset holds. At the one-month horizon, the coefficient  $c_0$  is positive and statistically significant and hence gold and stock market expected returns move in the same direction. For longer horizons the hedge coefficient  $c_0$  is negative and even statistically significant at the two-year horizon but is dwarfed by the co-movement during times of tumult in the stock market. The crisis coefficient ( $c_0 + c_1$ ) is positive and statistically significant for horizons longer than one month.

Turning next to the results based on the realized returns, we find that gold acts as

---

<sup>11</sup>We only include the 10% quantile (and exclude the 5% and 1% quantiles) as a proxy for extreme movements since our sample is much smaller, with a sample size of 230 observations. Figure 3 plots the time series of realized and expected returns on the gold and equity markets for the one-, six-, twelve- and twenty-four-month horizons.

<sup>12</sup>Again, we employ [Newey & West \(1987\)](#) standard errors.

both a weak hedge and weak safe haven for all horizons. All coefficients are statistically insignificant but all hedge coefficients are negative and of smaller absolute magnitude than the crisis coefficient. The results are similar to those of [Baur & McDermott \(2010\)](#), who show that realized gold returns serve as both a hedge and a safe haven for the U.S. stock market for the period from March 1979 until March 2009. The performance of gold as a weak or strong hedge/safe haven depends on the frequency (daily, weekly, monthly). The findings of [Baur & Lucey \(2010\)](#) also suggest that the realized returns of gold act as a hedge and safe haven for the U.S. stock market. In summary, the high co-movement between expected gold and stock market returns during times of stock market tumult offsets the expected hedging ability of gold.

## V Gold and Bond Risk Premia

### A Bond Premium Prediction Model

Next, we investigate the co-movement of gold with bonds. Again, we separately analyze the co-movements of expected and realized returns. We rely on forecasting regressions of bond excess returns on forward rates in order to obtain an estimate for the bond risk premium following [Cochrane & Piazzesi \(2005\)](#):

$$r_{n,t+1}^{Bond} = \beta_{n,0} + \beta_{n,1}y_{1,t} + \beta_{n,2}f_{2,t} + \dots + \beta_{n,5}f_{5,t} + \epsilon_{n,t+1} \quad (20)$$

where  $r_{n,t+1}^{Bond}$  is the holding period excess return from buying an  $n$ -year bond at time  $t$  and selling it as an  $n - 1$ -year bond at time  $t + 1$ ,  $y_{1,t}$  is the yield at time  $t$  and  $f_{n,t}$  is the forward at time  $t$  for loans between time  $t + n - 1$  and  $t + n$ . We also estimate a restricted specification in the two-step procedure. In the first step, the average bond return across

the different maturities is regressed on the forward rates:

$$\bar{r}_{n,t+1}^{Bond} = \gamma_{n,0} + \gamma_{n,1}y_{1,t} + \gamma_{n,2}f_{2,t} + \dots + \gamma_{n,5}f_{5,t} + \epsilon_{n,t+1} \quad (21)$$

In a second step, a single-factor  $b_n$  is estimated:

$$r_{n,t+1}^{Bond} = b_n(\gamma^T f_t) + \epsilon_{n,t+1} \quad (22)$$

$$\gamma^T f_t = \gamma_{n,0} + \gamma_{n,1}y_{1,t} + \gamma_{n,2}f_{2,t} + \dots + \gamma_{n,5}f_{5,t} \quad (23)$$

Cochrane & Piazzesi (2005) show that the linear combination of forward rates  $\gamma^T f_t$  is a state variable for the expected returns of all maturities, while the restriction has only a minor impact on the forecasting performance. We obtain monthly bond yields with maturities from one year to five years from the Board of Governors of the Federal Reserve System.<sup>13</sup> Since our data consist of monthly bond data with maturities varying from one to five years, we can only conduct the analysis for one-year bond excess returns, just as Cochrane & Piazzesi (2005).

The results are summarized in Table 6. We find that the adj.  $R^2$  values are similar for both the restricted and unrestricted model varying between 11.70% and 22.39% and 12.96% and 21.52%, respectively.<sup>14</sup> We find that the loadings  $b_n$  of expected returns on the forecasting factor  $\gamma^T f$  are statistically significant and are increasing in maturity. We

---

<sup>13</sup>Website: <https://www.federalreserve.gov/pubs/feds/2006/200628/200628abs.html>. Unlike the data sets of Fama & Bliss (1987) or McCulloch & Kwon (1993), the data are available at a daily frequency and include estimates out to thirty-year maturities. For our analysis, we work with the coarser monthly frequency, where monthly observations are obtained as either the end-of-month observation or the mean of daily observations within that month. The results are qualitatively similar for both specifications.

<sup>14</sup>The explanatory power is somewhat lower than those of Cochrane & Piazzesi (2005) or Kessler & Scherer (2009), but neither of these studies includes the recent financial crisis. The magnitudes of our adj.  $R^2$  are similar to Dahlquist & Hasseltoft (2013), who include the financial crisis and investigate the period from January 1975 to December 2009. They find adj.  $R^2$  values between 20% and 24%. When excluding the financial crisis, we also find much higher adj.  $R^2$ , indicating that times of market tumult have an important impact on the predictability of bond excess returns.

apply the [Newey & West \(1987\)](#) correction with 18 lags following [Cochrane & Piazzesi \(2005\)](#). The coefficients implied by the restricted model for each maturity  $n$  and the slope coefficients of the unrestricted model are displayed in Figure 4 in the top and bottom panels, respectively. We find that the parameters are very similar for both models and hence the single factor of the restricted model is able to mimic the unrestricted model. The coefficients do not follow a tent shape for either model and none of the maturities, which is consistent with the more recent results of [Kessler & Scherer \(2009\)](#).<sup>15</sup> Even though there is no clear pattern of the coefficients, we find that they are statistically significant overall, which supports the strong link between forward rates and bond excess returns. For the unrestricted model, we rely on Wald tests using also [Newey & West \(1987\)](#) 18 lags correction ([Cochrane & Piazzesi, 2005](#)). The null of zero coefficients can be rejected for all models except for the short-maturity bonds ( $n=2$ ). When using the long-maturity bonds ( $n=4,5$ ), the null test statistic  $\chi^2$  is even higher than the 1% critical value 15. All in all, our empirical findings suggest that we can be confident about our model(s) and we work with the estimates of expected bond risk premia as proxied by the fitted values of either the restricted or unrestricted model.

## B Co-movements of Gold with the Bond Market

We test the ability of gold as a hedge or safe haven against bond risk premia in the same manner as for the stock market:

$$\hat{r}_t^{Gold} = a + b_t \hat{r}_t^{Bond} + \epsilon_t \quad (24)$$

$$b_t = c_0 + c_1 D(\hat{r}^{Bond} q_{10}) \quad (25)$$

---

<sup>15</sup>The authors show that the tent shape is only found in certain time frames rather than being a consistent pattern. Their finding is supported by data from both Datastream and CRSP ([Fama & Bliss, 1987](#), data).

$$h_t = \omega + \alpha \epsilon_{t-1}^2 + \beta h_{t-1} \quad (26)$$

The results are reported Table 7. Overall, they are quite similar to those for the stock market. The results are both qualitatively similar for the restricted and unrestricted model and all maturities and in the following we focus our discussion on the restricted model. For expected returns, the hedge coefficient is negative and statistically insignificant, indicating that gold might serve as a hedge for bonds. But the positive 10% coefficient associated with the dummy variable, which is highly statistically significant, shows high co-movement of expected bond and gold risk returns during times of (bond) market stress and offsets the overall hedging performance of gold.

The analysis of realized returns suggests that gold acts as both a weak hedge and safe haven. Both the hedge and 10% coefficients are slightly above zero but statistically insignificant. Again, the results are similar across all bond maturities. The results are in line with [Baur & Lucey \(2010\)](#), who apply a similar methodology to realized returns in order to investigate the relationship between gold and bonds for the period from November 1995 until November 2005. They also show that both the hedge and crisis coefficients are statistically insignificant.

In summary, we show that the high positive co-movement of expected returns during times of (bond) market stress offsets the hedging property of gold. Analyzing realized returns, we show that gold serves as both a hedge and safe haven against bonds.

## VI Gold as an Inflation Hedge

The findings concerning gold as an inflation hedge in the literature are mixed. [Chua & Woodward \(1982\)](#) find that gold is an inflation hedge for the U.S. and not for other



major countries, but consider only the period from 1975 until 1980. [Batten et al. \(2014\)](#) investigate the dynamic inflation-beta of gold for the period from 1985 until 2012 and find that the relationship is time-varying. Before the 1990s, the beta is generally positive and quite high, reaching values above 2.5. Throughout the 1990s, they show evidence of very small, close to zero, inflation-betas, and then a significant increase in the 2000s.<sup>16</sup>

We want to explore the extent to which expected gold returns co-move with expected inflation. To do so, we follow the approach of [Chua & Woodward \(1982\)](#) and estimate the following regression:

$$\hat{r}_t^{Gold} = \alpha + \beta \hat{I}_t + \epsilon_t \quad (27)$$

where  $\hat{I}_t$  is the expected inflation rate at time  $t$  and  $\epsilon_t$  is the error term. If the slope coefficient  $\beta$  is positive and statistically significant, expected gold returns positively co-move with expected inflation, i.e. when there is an increase in expected inflation, there is a contemporaneous increase in the expected return of gold. We also repeat the analysis, but replace the expected return and the expected inflation with the realized excess returns  $r_t^{Gold}$  and the actual inflation rate, respectively. We focus on the same horizons  $h$  as in the stock market analysis: one month, six months, one year and two years.

Similar to our analysis for the stock market, we first need to obtain an estimate of the expected inflation. [Ang et al. \(2007\)](#) compare 39 forecasting models, and show that the time-series of inflation rate can be well described by time-series models such as Autoregressive (AR) models, Random Walk (RW) models or Autoregressive Moving Average (ARMA) models. We follow their advice and rely on an ARMA(1,1) model and AR models. The order of the AR-order  $p$  is chosen according to the Bayesian information

---

<sup>16</sup>For more research on gold and inflation, we refer to the literature survey of [Blose \(2010\)](#).

criterion (BIC).<sup>17</sup> We evaluate the forecasting performance of the three models for the horizons from one month to two years by comparing the Root Mean Squared Error (RMSE). Each month, we estimate the models using all the observations available until that month and obtain forecasts of the inflation over the next  $h$  months. The initial estimation uses the first 60 observations. We then compare the expected inflation over the  $h$  months with the realized inflation over the  $h$  months and compute the RMSE.

Table 8 reports the results. One can observe in Panel A that both the AR and ARMA(1,1) models outperform the historical mean for all horizons except for the twenty-four-month horizon. The AR model shows the lowest overall RMSE. Only at the three-month horizon does the ARMA model show a slightly smaller RMSE. Panel B reports out-of-sample  $R^2$  and the relevant p-values following [Clark & West \(2007\)](#) and [Campbell & Thompson \(2008\)](#), where the AR model is the benchmark model. The results support the choice of the AR model, since none of the models is able to outperform the the AR model for all horizons. Only the ARMA model is able to beat the AR model at the three-month horizon but the outperformance is statistically insignificant. For our hedging analysis in the following we thus rely on the AR model.

After computing the expected inflation as the forecast of the AR model:

$$I_{t+1} = \phi_0 + \sum_{i=1}^p \phi_i I_{t+1-i} + \epsilon_{t+1} \quad (28)$$

we regress the expected gold return on the former as in Equation (27) and report the coefficients in Table 9. We find that both expected and realized gold returns are unrelated to expected and realized inflation, respectively. The latter is in contrast to the

---

<sup>17</sup>[Ang et al. \(2007\)](#) show that expected inflation obtained from surveys is a strong competitor to the time-series models. We consider most of the competing models as advocated by the authors. Details of this analysis are reported in Section VII.D.

finding of [Chua & Woodward \(1982\)](#). When comparing our estimates to theirs, one can notice that our estimate is still positive, which indicates that gold has served as inflation hedge. However, the magnitude is substantially lower than the one reported by [Chua & Woodward \(1982\)](#) which makes it no longer possible anymore to reject the null hypothesis that it is equal to zero.

## VII Robustness

We provide additional evidence in favor of our prediction models by obtaining p-values of both slope coefficients and  $R^2$  (in-sample and out-of-sample) with a parametric bootstrap. Further, we acknowledge the potential issues of errors-in-variables and finite sample bias for our empirical analysis. The former is relevant since various regressions rely on estimated values as explanatory variables. The potential finite sample bias is related to our relatively short sample period from 1996 until 2015, which leads to 230 monthly observations. In the following, we present robustness tests which mitigate these potential issues. We focus on our main results for the stock market. Lastly, we show results for competing models for the inflation rate, following [Ang et al. \(2007\)](#).

### A Statistical Inferences of the Prediction Model

We follow [Welch & Goyal \(2008\)](#) and apply a parametric bootstrap in order to obtain the statistical significance of our OLS coefficients in Equations (12) and (16). The data generating process under the null is assumed to be:

$$r_{t+h} = a_h + u_{1,t+h} \tag{29}$$

$$X_{t+1} = \alpha + \beta X_t + u_{2,t+h} \tag{30}$$

where  $X_t$  includes  $LJP$  and  $VRP^{Gold}$  for gold and  $\log(D/P)$  and  $VRP^{S\&P}$  for the S&P 500. The data generating process under the alternative is given by:

$$r_{t+h} = a_h + b_h X_t + u_{1,t+h} \quad (31)$$

$$X_{t+1} = \alpha + \beta X_t + u_{2,t+h} \quad (32)$$

By allowing for an autoregressive structure for the predictors we control for the potential [Stambaugh \(1999\)](#) bias. We obtain pseudo time-series for both the returns and predictor time series under the null by drawing with replacement from the residuals simultaneously. This procedure thus preserves the cross-correlation structure of the residuals in the predictive regression and the two autoregressive models. We then compute and store the t-statistics of the coefficients, in-sample adjusted  $R^2$ , out-of-sample  $R^2_{OOS}$  and the MSPE-adjusted test statistic related to Equations (12) and (16). We repeat this process 5,000 times, which gives us empirical distributions for the test statistics and the  $R^2$ . After ordering the distribution for each statistic, critical values and p-values are obtained by the quantiles.

The results for the in-sample and out-of-sample analyses are reported in Panels A and B of Table 10 of the Online Appendix, respectively. We find that the p-values of all slope coefficients are all statistically significant for both the gold and S&P 500 prediction models, just as in our main analysis (when relying on [Newey & West, 1987](#), standard errors). The  $LJP$  is significant at the 10% level at the one-month horizon while the  $VRP$  is significant at the 5% level for the one-, six-, nine- and twenty-four-month horizons. The remaining p-values are all below 1%. For the S&P 500, the coefficients are all statistically significant at the 1% level.<sup>18</sup> The p-values for the in-sample adj.  $R^2$

---

<sup>18</sup>The results are consistent with our main findings when relying on [Newey & West \(1987\)](#) and [Hodrick](#)

are all smaller than 0.001 for the S&P 500 as well. For gold, the p-value is 2.18% for the one-month horizon and smaller than 1% for the remaining values.

The bootstrapped p-values for the out-of-sample  $R_{OOS}$  and the MSFE-adjusted test statistic also confirm the results in our main analysis. The MSFE-statistics show that our prediction model performs better than the historical mean model for all horizons at a significance level of 1%. Only for the S&P 500 and the one-month horizon is the statistical significance at the 5% level. The  $R_{OOS}$  are statistically significant at the 5% level or lower for both gold and the S&P 500 for all horizons.

We thus verify the performance of our prediction models concerning both the statistical significance of the predictors and the explanatory power (in-sample and out-of-sample) by relying on bootstrapping methods instead of corrections for heteroskedasticity and autocorrelation (Newey & West, 1987; Hodrick, 1992).

## B Errors-in-Variables

We account for the possible errors-in-variables (EIV) problem since our expected premia in Equation (17) and (18) are estimates obtained from linear regressions. The standard econometric approach to deal with the EIV problem is the use of instrumental variables (Greene, 1998; Christensen & Prabhala, 1998). Christensen & Prabhala (1998) propose using lagged observations as an instrument. Algebraically, we estimate the following equations:

$$\hat{r}_t^{Gold} = a + b_t \hat{r}_t^{Stock} + \epsilon_t \quad (33)$$

$$b_t = c_0 + c_1 D(\hat{r}^{Stock} q_{10}) \quad (34)$$

---

(1992) standard errors in Tables 2 and 4. The *VRP* slope coefficient shows the lowest statistical significance at the one-, nine- and twenty-four-month horizon for gold as well, while the S&P 500 coefficients generally show higher t-statistics than those of gold.

$$h_t = \omega + \alpha \epsilon_{t-1}^2 + \beta h_{t-1} \quad (35)$$

$$\hat{r}_t^{Stock} = \beta_0 + \beta_1 \hat{r}_{t-1}^{Stock} + \eta_t \quad (36)$$

where  $\eta_t$  denotes the measurement error which is uncorrelated with  $\hat{r}_t^{Stock}$ . In the first-stage regression, Equation (36), the expected equity premium  $\hat{r}_t^{Stock}$  is regressed on an instrument, its lagged observation  $\hat{r}_{t-1}^{Stock}$ . Fitted values from this regression then replace the expected equity premium  $\hat{r}_t^{Stock}$  in the second-stage regression in Equation (33).<sup>19</sup>

Table 11 of the Online Appendix reports the IV estimates in the second-stage regression. The coefficients are of slightly higher magnitudes than in Table 5 but the conclusions remain the same. The hedge coefficient is negative for horizons of six months and more, while there is statistically significant positive co-movement during times of stock market tumult.

## C Finite Sample Bias

In a two-step approach we investigate the robustness of our hedge and safe haven results against finite sample bias, as discussed in the literature. The use of Monte Carlo or bootstrap simulations is documented in recent studies and for various applications. [Nelson & Kim \(1993\)](#) rely on annual returns from 1872 until 1927 for stock return predictability regressions, and argue that the biases should be accounted for. [Mark \(1995\)](#) accounts for small-sample biases in his multiple-period regressions of exchange rates by relying on bootstrap distributions under the null. [Bekaert et al. \(1997\)](#) examine the expectations hypothesis of the term structure of interest rates and show evidence of extreme bias in

---

<sup>19</sup>We also conducted the analysis with the exclusion of Equation (35) and the relevant least squares (OLS) estimation as in [Baur & Lucey \(2010\)](#), which leads to qualitatively similar results. By doing so we reduce the number of parameters to be estimated from six to three compared to our sample size of 230. We further investigate the potential finite sample bias in Section VII.C.

the small-sample distribution of their regression-based tests.

In the first step we quantify the small-sample bias. In a second step we obtain critical values for our test statistics from a bootstrap approach, which does not rely on asymptotic results that may not be valid for finite samples.

First, we conduct a residual resampling bootstrap approach. For this purpose residuals are estimated from Equation (17). Block-bootstraps of the dependent variable are then generated by sampling from the residuals with replacement, which are then added to the fitted values from Equation (17). This leads to the same number of observations as in the initial model.<sup>20</sup> The coefficients of interest are then estimated from Equation (17) using the simulated data. We repeat this procedure 5,000 times.<sup>21</sup> The small-sample bias of a coefficient is estimated as the difference between the original coefficient estimate and the average across the 5,000 simulated coefficients.

In a second step, new residuals are computed from the bias-corrected coefficients. New dependent variables under the null hypotheses are obtained by sampling the residuals. The original regression model in Equation (17) is then estimated again in order to obtain the Wald statistics. The procedure is repeated 5,000 times, which leads to a distribution of the statistics. From the percentiles of the distribution of simulated test statistics we obtain the critical values and p-values and conclude on the statistical significance of  $c_0$  and the sum  $c_0 + c_1$ .

We present the results of the two steps in Panels A and B of Table 12 of the Online Appendix. The results suggest that our main conclusions are generally robust to potential finite sample bias. The absolute bias in coefficient estimates is negligible and varies

---

<sup>20</sup>We follow [Hall et al. \(1995\)](#) using a block length of  $n^{1/3}$ , where  $n$  is the total sample size. We also consider non-block bootstraps, leading to qualitatively similar results.

<sup>21</sup>[Efron & Tibshirani \(1986\)](#), [Kho \(1996\)](#) and [Kosowski et al. \(2006\)](#) show by means of different applications that their results are not sensitive for repetitions larger than 500–1,000. By the choice of 5,000 replications we strike the right balance between our computational capacity and sufficient repetitions.

between 0.01 and 2.14 percentage points, which should not overturn our results on the hedging and safe haven performance of gold. In Panel B, we report the bias-corrected coefficients, and show results for the finite sample distributions of the test statistics. The results are qualitatively similar to the results when relying on asymptotic critical values for the tests. Overall, the first coefficient  $c_0$  speaks in favor of gold as a hedge but the high and statistically significant co-movement during the crisis offsets the hedging ability. In conclusion, it is unlikely that finite sample bias and distortions significantly affect our main results.

## D Modeling Inflation

In our main analysis we rely on time-series models for the expected inflation. Our choice is supported by the findings of [Ang et al. \(2007\)](#) but is also motivated by the available data frequency of potential explanatory variables. The competing non-time-series models of the authors use data at a coarser frequency and hence monthly forecasts of the inflation cannot be estimated. More specifically, their analysis focuses at the quarterly and yearly horizon and one-year-ahead inflation forecasts.

In this section, we investigate the forecasting performance of alternative inflation models. We include most of the models investigated by [Ang et al. \(2007\)](#) relying on quarterly data.<sup>22</sup> We also focus on one-year-ahead inflation forecasts, which is mainly due to the non-availability of alternative forecast horizons for the survey data.

---

<sup>22</sup>We exclude the random walk on annual inflation (AORW) and the models based on the Livingston survey (LIV1, LIV2, LIV3), since these are of yearly and semi-yearly frequency, respectively. Further we exclude regime-switching models and the empirical term structure model and the term structure model suggested by [Ang et al. \(2008\)](#). When estimating a regime-switching model for the inflation rate in a short sample with 76 quarterly observations, the algorithm fails to converge. We exclude models which include the Bernanke–Boivin–Eliasz *FAC* measure since the data are only available until the end of 2001. For the term structure data, we rely on the same data set as for our bond analysis in Section V. Lastly, the [Stock & Watson \(1989\)](#) experimental leading indices were discontinued. Following the advice of the authors, we rely on the “most direct successor”, the Chicago Fed National Activity Index (CFNAI and CFNAIMA3), obtained from the Federal Reserve Bank of St. Louis.



Similar to our previous analyses we obtain out-of-sample forecasts using the different models, where the initial estimation takes into account the first five years of observations. We also compare the out-of-sample forecasting performance with respect to the AR model, following [Stock & Watson \(1989\)](#):

$$I_{t,t+4} = \lambda \hat{I}_t^{AR} + (1 - \lambda) \hat{I}_t^X + \epsilon_{t,t+4} \quad (37)$$

where  $\hat{I}_t^{AR}$  is the forecast of the inflation over the next year from the AR time-series model,  $\hat{I}_t^X$  is the forecast from an alternative model and  $\epsilon_{t,t+4}$  is the error term associated with the combined forecast. If  $\lambda = 1$ , then the forecasting model  $X$  does not add anything to the forecast from the AR time-series benchmark. If  $\lambda = 0$ , then forecasts from the AR model add nothing to the alternative model. We correct the standard errors of the coefficients due to the overlapping observations using 4 lags and the procedure of [Newey & West \(1987\)](#).<sup>23</sup>

Table 13 of the Online Appendix reports the results. In accordance with our main results, we find that the AR model shows relatively strong out-of-sample forecasting performance in the means of RMSE. Only 3 out of the 29 models (PC1, PC6 and PC7) show slightly smaller RMSE with ratios of 0.9870, 0.9701 and 0.9990, respectively. Nonetheless, the additional information added by these models is not statistically significant, where the coefficient  $1 - \lambda$  in Equation (37) varies between 0.06 and 0.48.<sup>24</sup>

We repeat our regression analysis, which tests whether gold serves as an inflation hedge using the alternative inflation models and at the quarterly horizon. Table 14 of the Online Appendix reports the results. From an ex-post point of view, inflation does not

---

<sup>23</sup>Using [Hodrick \(1992\)](#) standard errors yields qualitatively similar results.

<sup>24</sup>In unreported results we find that the coefficient  $\lambda$  in Equation (37) is close to 1 or higher in most cases and statistically significant in 21 of the 29 cases.

serve as a hedge at the quarterly horizon, which shows that our main results are robust against the choice of frequency. From an ex-ante point of view, we find that expected gold returns and expected inflation also do not co-move. The coefficient is insignificant when relying on the AR model or PC1 and even negative and statistically significant when relying on PC6 or PC7. All in all, this subsection supports our main results: gold does not serve as a hedge for inflation. Even though the coefficients (and the significance) differ when using alternative models, the conclusion remains the same.

## VIII Conclusion

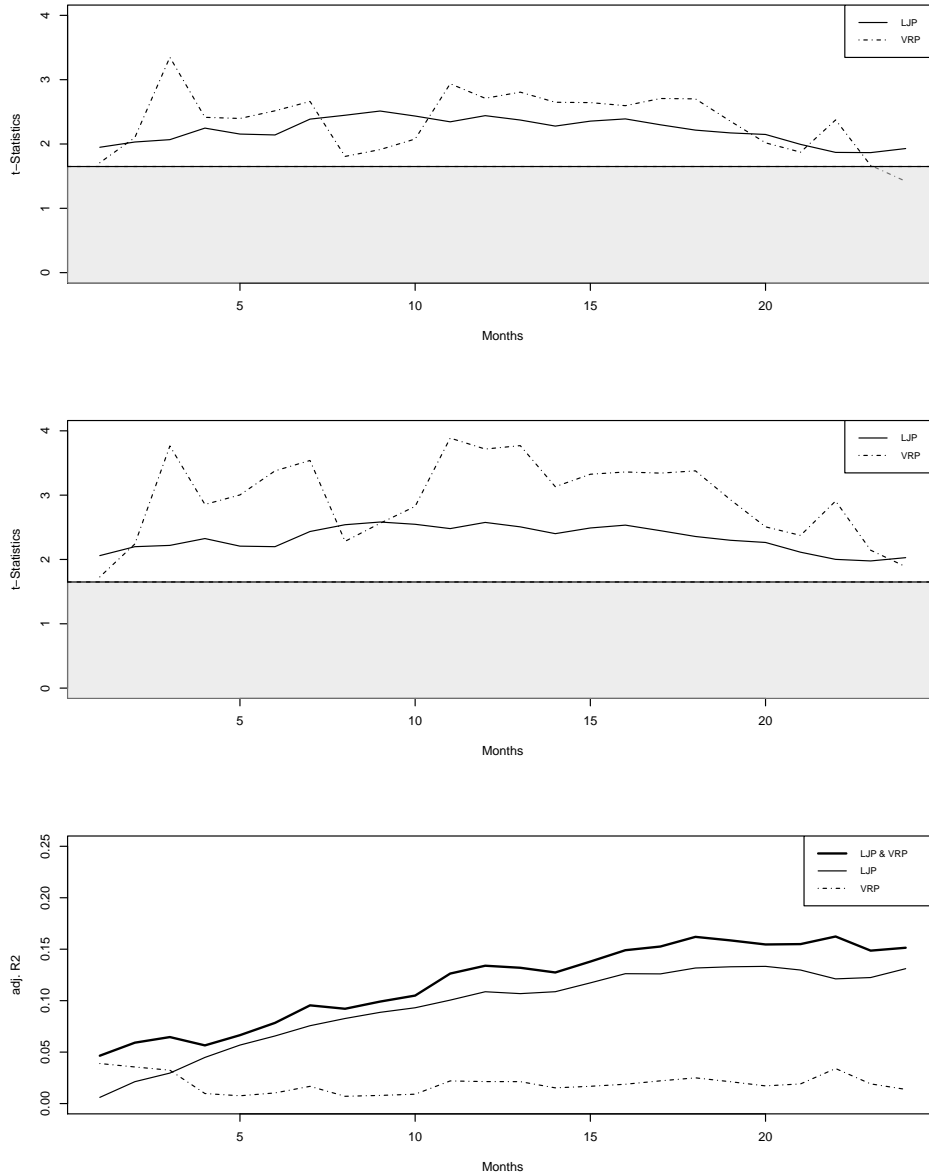
This paper provides new evidence of gold as a hedge and safe haven asset for the stock market, the bond market, and inflation based on expected returns. In the first step, we present a model for the gold risk premium both in-sample and out-of-sample. Thus, our first major result is that the future return of gold is predictable.

Based on the conditional risk premium estimate, i.e. the expected returns, we analyze the safe haven and hedging properties of gold. We apply state of the art models in order to estimate expected stock and bond risk premia as well as expected inflation. We find that the co-movements of expected gold returns with both, expected stock and bond market returns, are positive, while the co-movements of realized returns are zero or negative. This result holds not only during normal market periods, but also in times of market stress. Moreover, we find that co-movement of gold returns with inflation is positive, irrespectively whether we analyze expected or realized returns.

Lastly, we want to briefly discuss future research questions. As reviewed in the Introduction, the literature on safe haven currencies has also focused on realized values. Similar to our study, it might also be useful to analyze whether the safe haven status

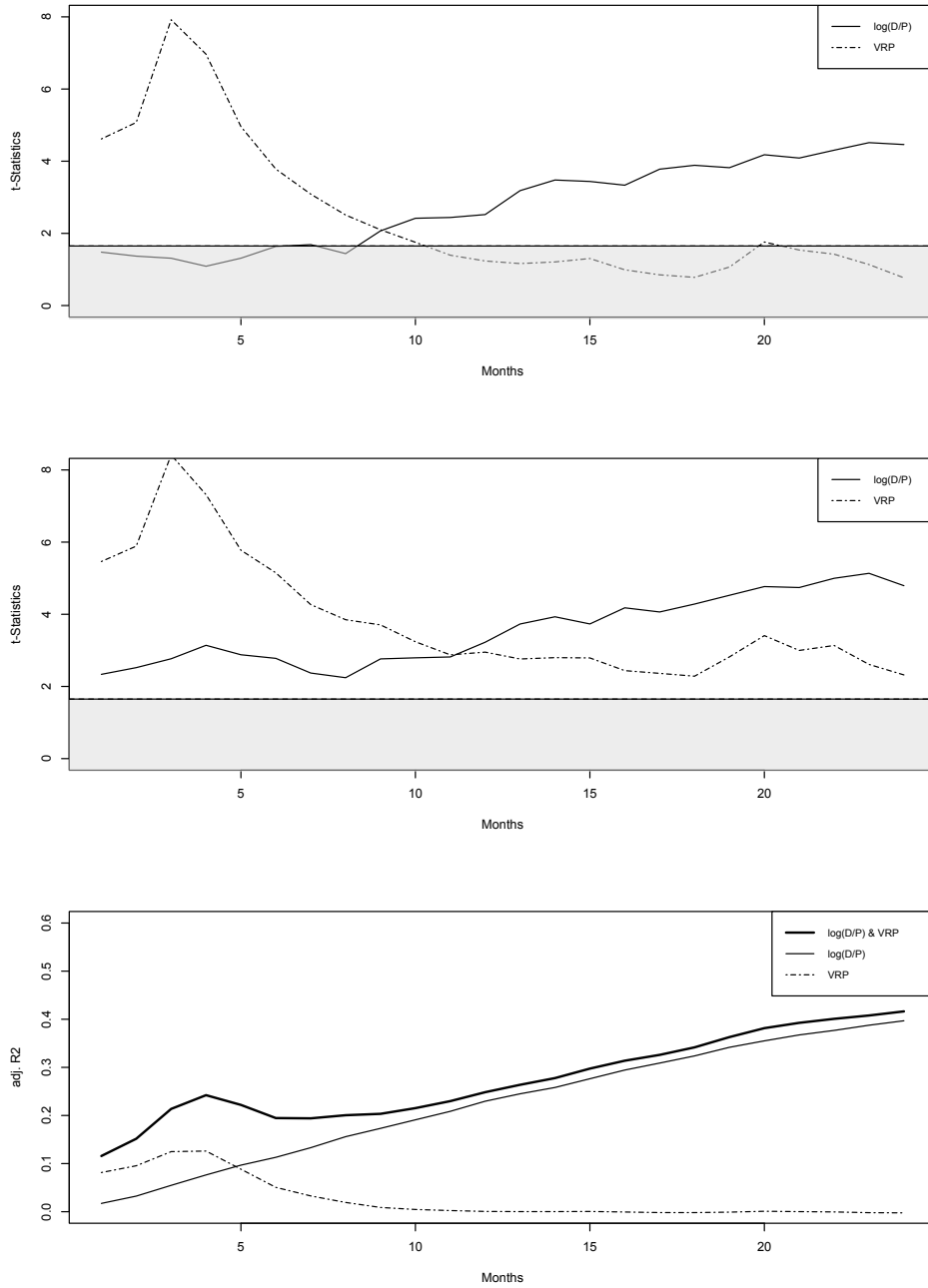
of specific currencies holds when analyzing expectations of returns. Furthermore, it will be interesting to compare whether gold or some safe haven currency, such as the Swiss Franc, act as a “safer” haven for equity investors.

Figure 1: Predictability Regressions: Gold



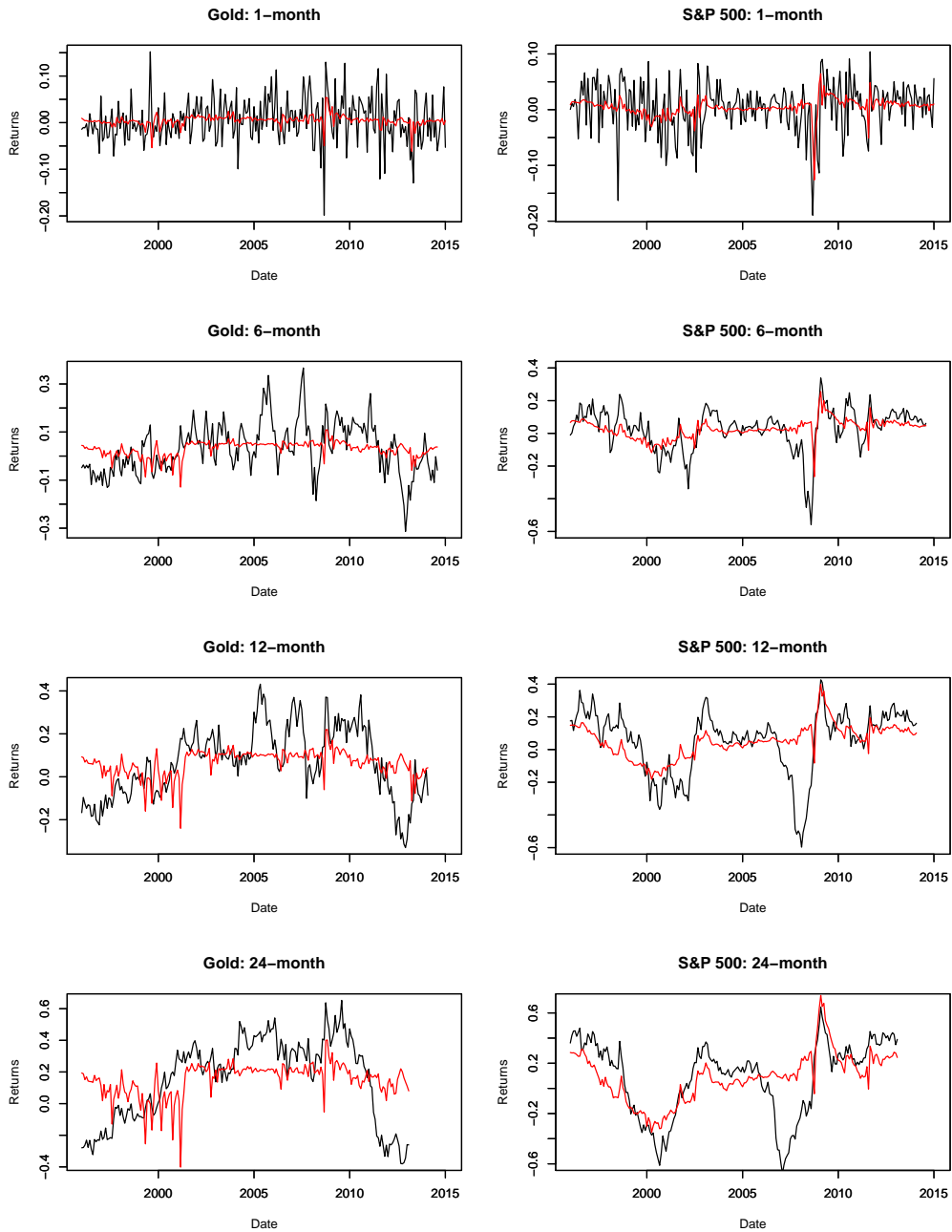
This figure presents [Newey & West \(1987\)](#) t-statistics from the return predictability regressions for the gold futures returns. The independent variables are the *LJP* (solid line) and the *VRP* (dotted line). The first (second) panel reports t-statistics from simple (multiple) return predictability regressions. The shaded areas indicate statistical significance at the 10% level. The third panel shows the corresponding adj.  $R^2$  for the simple regression (solid and dotted lines) and the multiple regression (bold solid line).

Figure 2: Predictability Regressions: S&P 500



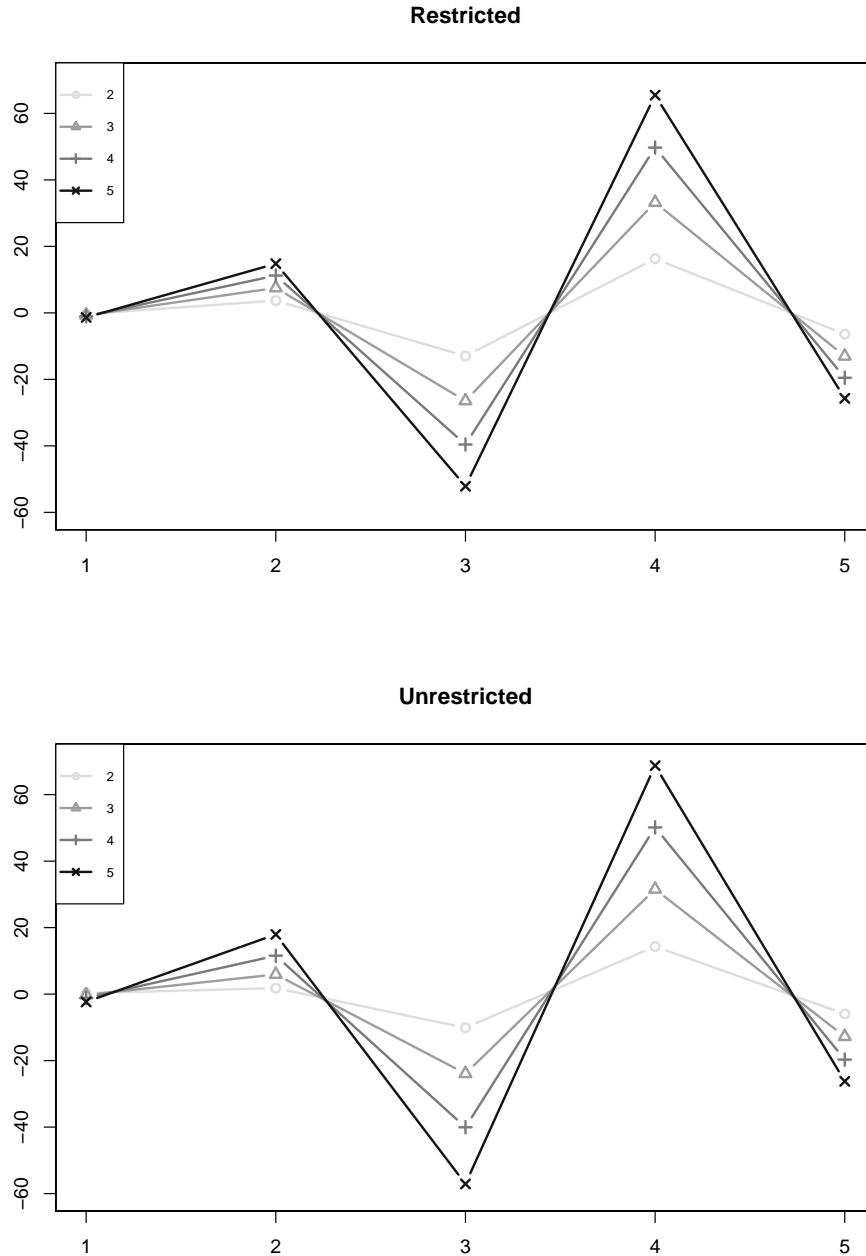
This figure presents [Newey & West \(1987\)](#) t-statistics from the return predictability regressions for the S&P 500 futures returns. The independent variables are the  $\log(D/P)$  (solid line) and the  $VRP$  (dotted line). The first (second) panel reports t-statistics from simple (multiple) return predictability regressions. The shaded areas indicate statistical significance at the 10% level. The third panel shows the corresponding adj.  $R^2$  for the simple regression (solid and dotted lines) and the multiple regression (bold solid line).

Figure 3: Expected Returns vs. Realized Returns



This figure plots the realized futures excess returns (black) against the expected premium, i.e. expected returns (fitted values in red) of gold and the S&P 500. The first, second, third and fourth panel report results for the 1-month, 6-month, 12-month and 24-month prediction horizon, respectively.

Figure 4: Regression Coefficients of Bond Excess Returns



This figure plots the estimates of  $\beta$  from the unrestricted regressions of bond excess returns and restricted estimates  $b\gamma^T$  in the top and bottom panel, respectively. The numbers in the legend indicate the maturity of the bonds, which is used as the dependent variable, while the numbers on the horizontal axis are the maturity of the independent variables (forward rates).

Table 1: Comparison of Predictors for the Gold Risk Premium

This table compares the significance of the introduced predictors. We regress gold futures excess returns on a constant and the explanatory variables in simple regressions for horizons from one month to two years. The check marks indicate whether the explanatory variable is statistically significant (at the 5% level).

	1-month	3-month	6-month	9-month	12-month	24-month
Basis						
Dollar						
Empl.						
FFR						
Inflation	✓					
IP						
log(D/P)						
log(E/P)						
<i>LJP</i> (Gold)		✓	✓	✓	✓	
<i>LJP</i> (S&P)						
LTR						
MFIV (Gold)						✓
Oil						
<i>RJP</i> (Gold)						
<i>RJP</i> (S&P)						
Stock Var.						✓
Tbill						
<i>VRP</i> (Gold)		✓	✓		✓	
<i>VRP</i> (S&P)		✓				✓



Table 2: Predictive Regressions: Gold

This table presents the results for monthly predictive regressions for the period from 1996 until 2015 in Panel A. The investigated predictors are the  $LJP$  and the  $VRP$  of gold. The  $LJP$  is calculated on the basis of  $k = 5\sigma_{ATM,t}$ , where  $\sigma_{ATM,t}$  stands for the at-the-money (ATM) volatility. The dependent variables are the gold futures excess returns. Robust Newey & West (1987) standard errors are reported in parentheses below using lags equal to the return horizon expressed in months. We also report Hodrick (1992) standard errors in square brackets for the slope coefficients. Stars indicate significance of the estimates: \* significant at  $p < 0.10$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$  according to the Newey & West (1987) standard errors. The last row reports Wald test statistics for the joint significance of the predictor variables using Newey & West (1987) standard errors in parentheses and Hodrick (1992) standard errors in square brackets. Panel B presents results for monthly out-of-sample predictive regressions for horizons from one month to two years for gold. The investigated predictors are the  $LJP$  and the  $VRP$  of gold. The  $LJP$  is calculated on the basis of  $k = 5\sigma_{ATM,t}$ , where  $\sigma_{ATM,t}$  stands for the ATM volatility. The dependent variables are the gold futures excess returns. We rely on expanding rolling windows and include five years of data for the initial regression. To obtain statistical significance we conduct a Clark & West (2007) MSPE test. The null hypothesis is the recursive mean model outperforming the predictive model, i.e.  $R_{OOS} \leq 0$ . The p-values are reported in braces below. Stars indicate significance of the estimates: \* significant at  $p < 0.10$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

	1-month	3-month	6-month	9-month	12-month	24-month
<i>Panel A: In-Sample</i>						
Intercept	0.0050 (0.0038)	0.0206* (0.0113)	0.0479* (0.0257)	0.0749** (0.0349)	0.0988** (0.0414)	0.2068** (0.0894)
$LJP$	0.7278** (0.3533) [0.3838]	1.9513** (0.8795) [1.0505]	3.8673** (1.7587) [2.0674]	5.6609** (2.1925) [2.9472]	7.4067** (2.8755) [3.7316]	13.2254** (6.5209) [6.9681]
$VRP$	0.4548* (0.2631) [0.2781]	0.6511*** (0.1730) [0.3608]	0.6012*** (0.1780) [0.4928]	0.7115** (0.2777) [0.5355]	1.1931*** (0.3210) [0.6745]	1.9257* (1.0202) [1.0198]
adj. $R^2$	0.0465	0.0646	0.0784	0.0991	0.1339	0.1514
Wald	(5.6427) [6.1591]	(19.2089) [7.2024]	(22.8497) [4.9038]	(15.3677) [5.3547]	(26.1202) [6.6457]	(9.4064) [8.1298]
<i>Panel B: Out-of-Sample</i>						
Gold	0.0253 {0.1026}	0.0597*** {0.0093}	0.0721** {0.0114}	0.0878*** {0.0055}	0.1244*** {0.0003}	0.1283*** {0.0004}

Table 3: Comparison of Predictors for the Equity Premium

This table compares the significance of the introduced predictors. We regress S&P 500 futures excess returns on the explanatory variables in simple regressions for horizons from one month to two years. The check marks indicate whether the explanatory variable is statistically significant (at the 5% level).

	1-month	3-month	6-month	9-month	12-month	24-month
Basis		✓	✓			
Dollar						
Empl.	✓					✓
FFR						✓
Inflation						
IP		✓	✓			
log(D/P)				✓	✓	✓
log(E/P)						
<i>LJP</i> (Gold)						
<i>LJP</i> (S&P)						
LTR						
MFIV (Gold)						
Oil						
<i>RJP</i> (Gold)		✓	✓	✓	✓	✓
<i>RJP</i> (S&P)						
Stock Var.						
Tbill						
<i>VRP</i> (Gold)						
<i>VRP</i> (S&P)	✓	✓	✓	✓		

Table 4: Predictive Regressions: Equity Premium

This table presents the results for monthly predictive regressions for the period from 1996 until 2015 in Panel A. The investigated predictors are the  $VRP$  of the S&P 500 and the dividend yield. The dependent variables are the S&P 500 futures excess returns. Robust Newey & West (1987) standard errors are reported in parentheses below using lags equal to the return horizon expressed in months. We also report Hodrick (1992) standard errors in square brackets for the slope coefficients. Stars indicate significance of the estimates: \* significant at  $p < 0.10$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$  according to the Newey & West (1987) standard errors. The last row reports Wald test statistics for the joint significance of the predictor variables using Newey & West (1987) standard errors in parentheses and Hodrick (1992) standard errors in square brackets. Panel B presents results for monthly out-of-sample predictive regressions for horizons from one month to two years for the S&P 500. The investigated predictors are the  $VRP$  of the S&P 500 and the dividend yield. The dependent variables are the S&P 500 futures excess returns. We rely on expanding rolling windows and include five years of data for the initial regression. To obtain statistical significance we conduct a Clark & West (2007) MSPE test. The null hypothesis is the recursive mean model outperforming the predictive model, i.e.  $R_{OOS} \leq 0$ . The p-values are reported in braces below. Stars indicate significance of the estimates: \* significant at  $p < 0.10$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

	1-month	3-month	6-month	9-month	12-month	24-month
<i>Panel A: In-Sample</i>						
Intercept	0.1651** (0.0716)	0.4714** (0.1720)	0.9200** (0.3374)	1.3635** (0.5053)	1.8231** (0.5795)	3.5464*** (0.7395)
log(D/P)	0.0411** (0.0176) [0.0174]	0.1165** (0.0421) [0.0503]	0.2248** (0.0809) [0.0933]	0.3307** (0.1197) [0.1344]	0.4415** (0.1369) [0.1753]	0.8597*** (0.1794) [0.3255]
$VRP$	0.4105*** (0.0752) [0.1637]	0.9432*** (0.1121) [0.2876]	1.0419*** (0.2024) [0.4832]	0.8332*** (0.2248) [0.5366]	0.7997** (0.2710) [0.5765]	1.1937** (0.5146) [0.6933]
adj. $R^2$	0.1158	0.2136	0.1946	0.2034	0.2484	0.4164
Wald	(36.6227) [12.9924]	(86.7176) [16.5738]	(51.6379) [16.4435]	(26.4963) [11.9341]	(23.1149) [12.1866]	(23.6173) [16.3726]
<i>Panel B: Out-of-Sample</i>						
S&P 500	0.0646** {0.0306}	0.1544** {0.0147}	0.0603** {0.0105}	0.0312* {0.0787}	0.0501* {0.0979}	0.0942* {0.0624}

Table 5: Hedge and Safe Haven

This table presents the estimation results for the role of gold as a hedge and safe haven relying on expected returns (columns (2)-(3)) or realized returns (columns (4)-(5)). Negative coefficients in columns (2) and (4) indicate that gold is a hedge against the stock market while zero (negative) coefficients in columns (3) and (5) indicate that gold is a weak (strong) safe haven. We report Wald test statistics for the significance of the coefficients below. Stars indicate significance of the estimates: \* significant at  $p < 0.10$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$ .

	Expected Returns		Realized Returns	
	Hedge	10%	Hedge	10%
1-month	0.2432**	-0.1469	-0.0118	0.1187
	6.0221	0.9250	0.0192	0.9868
6-month	-0.1561	0.2430***	-0.2003	0.0131
	2.6767	13.7804	1.1012	0.0244
12-month	-0.1405	1.2301***	-0.1978	0.0033
	0.6117	16.2719	2.0724	0.0018
24-month	-0.1891***	1.1917***	-0.1155	-0.0730
	19.2115	32.2851	0.2957	0.5571

Table 6: Bond Single-Factor Model

This table presents the estimation results for the regressions of one-year excess bond returns on forward rates. Panels A and B report results for the restricted and unrestricted model, respectively. The significance of the coefficients  $b_n$  and the Wald test statistics are based on [Newey & West \(1987\)](#) corrected standard errors with 18 lags. Stars indicate significance of the estimates: \* significant at  $p < 0.10$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

	$Bond_{n=2}$	$Bond_{n=3}$	$Bond_{n=4}$	$Bond_{n=5}$
<i>Panel A: Restricted Model</i>				
$b_n$	0.3957** (0.1401)	0.8069*** (0.2250)	1.2072*** (0.2788)	1.5902*** (0.3146)
$R^2$	0.1211	0.1489	0.1876	0.2274
adj. $R^2$	0.1170	0.1450	0.1839	0.2239
<i>Panel B: Unrestricted Model</i>				
Wald	6.6953	10.5108	16.8167	24.5810
$R^2$	0.1497	0.1531	0.1879	0.2333
adj. $R^2$	0.1296	0.1331	0.1687	0.2152

Table 7: Gold as a Hedge for Bonds

This table presents the estimation results for the role of gold as a hedge and safe haven relying on expected returns (columns (2)-(5)) or realized returns (columns (6)-(7)) both for the one-year horizon. All results are based on the one-year horizon relying on two- to five-year zero bonds. Negative coefficients in columns (2), (4) and (6) indicate that gold is a hedge against the bond market while zero (negative) coefficients in columns (3), (5) and (7) indicate that gold is a weak (strong) safe haven. We report Wald test statistics for the significance of the coefficients below. Stars indicate significance of the estimates: \* significant at  $p < 0.10$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$ .

	Expected Returns				Realized Returns	
	Restricted	10%	Unrestricted	10%	Hedge	10%
<i>Bond</i> <sub><i>n</i>=2</sub>	Hedge	0.2113***	Hedge	0.1858***	0.0089	0.0151
		17.7382	1.7380	33.2884	0.2293	0.0907
<i>Bond</i> <sub><i>n</i>=3</sub>	Hedge	0.1036***	Hedge	0.1084***	0.0061	0.0097
		17.7384	0.1023	16.0364	0.2089	0.0982
<i>Bond</i> <sub><i>n</i>=4</sub>	Hedge	0.0692***	Hedge	0.0732***	0.0043	0.0091
		17.7384	0.1029	16.3356	0.4134	0.3251
<i>Bond</i> <sub><i>n</i>=5</sub>	Hedge	0.0526***	Hedge	0.0511***	0.0044	0.0068
		17.7387	0.1561	14.0398	0.7025	0.3260

Table 8: Predictive Regressions: Inflation

This table presents the results for monthly predictive regressions for the period from 1996 until 2015. Forecasts for the next one, three, six, nine, twelve and twenty-four months are obtained from expanding window estimation, where the initial estimation takes into account the first sixty observations, using the following models: Autoregressive (AR) model, Random Walk (RW), Autoregressive Moving average (ARMA) and Historical Mean (HM). We report the Root Mean Squared Error (RMSE) for each model and horizon in Panel A. We also report out-of-sample  $R^2$  and the relevant p-values following [Clark & West \(2007\)](#) and [Campbell & Thompson \(2008\)](#) in Panel B.

	1-month	3-month	6-month	9-month	12-month	24-month
<i>Panel A: RMSE</i>						
AR	0.0385	0.0309	0.0215	0.0169	0.0141	0.0092
RW	0.0479	0.0496	0.0471	0.0465	0.0454	0.0439
ARMA	0.0386	0.0305	0.0217	0.0171	0.0143	0.0093
HM	0.0442	0.0316	0.0222	0.0174	0.0145	0.0092
<i>Panel B: Out-of-Sample <math>R^2</math></i>						
RW	-0.5522	-1.5731	-3.7991	-6.5316	-9.3570	-21.8170
	0.9174	0.8333	0.6279	0.5974	0.6604	0.6272
ARMA	-0.0040	0.0252	-0.0158	-0.0230	-0.0301	-0.0203
	0.7964	0.1675	0.5031	0.6471	0.7896	0.7114
HM	-0.3180	-0.0433	-0.0620	-0.0522	-0.0541	-0.0073
	0.6891	0.9226	0.8231	0.8234	0.8179	0.7900

Table 9: Gold as an Inflation Hedge

This table presents the estimation results for the role of gold as a hedge against inflation relying on expected returns (columns (2)-(3)) or realized returns (columns (4)-(5)). We report [Newey & West \(1987\)](#) robust standard errors in parentheses below. Stars indicate significance of the estimates: \* significant at  $p < 0.10$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

	Expected Returns		Realized Returns	
	Intercept	$\beta$	Intercept	$\beta$
1-month	0.0086*	-1.1832	0.0022	1.4309
	(0.0047)	(2.0942)	(0.0045)	(1.6856)
6-month	0.0042	1.0848	0.0009	2.2005
	(0.0041)	(1.9923)	(0.0059)	(2.3333)
12-month	0.0022	2.2535	-0.0026	4.2211
	(0.0081)	(3.8844)	(0.0074)	(3.0123)
24-month	0.0106	-1.3830	-0.0033	4.9833
	(0.0089)	(4.3364)	(0.0194)	(8.3049)



## References

- Ang, A., Bekaert, G., & Wei, M. (2007). Do macro variables, asset markets, or surveys forecast inflation better? *Journal of Monetary Economics*, *54*(4), 1163–1212.
- Ang, A., Bekaert, G., & Wei, M. (2008). The term structure of real rates and expected inflation. *Journal of Finance*, *63*(2), 797–849.
- Bakshi, G., Kapadia, N., & Madan, D. (2003). Stock return characteristics, skew laws, and the differential pricing of individual equity options. *Review of Financial Studies*, *16*(1), 101–143.
- Bansal, R., & Yaron, A. (2004). Risks for the long run: A potential resolution of asset pricing puzzles. *Journal of Finance*, *59*(4), 1481–1509.
- Barone-Adesi, G., & Whaley, R. E. (1987). Efficient analytic approximation of American option values. *Journal of Finance*, *42*(2), 301–320.
- Batten, J. A., Ciner, C., & Lucey, B. M. (2014). On the economic determinants of the gold–inflation relation. *Resources Policy*, *41*, 101–108.
- Baur, D. G. (2013). Gold-fundamental drivers and asset allocation. *University of Technology Sydney, Working Paper*.
- Baur, D. G., & Loeffler, G. (2015). Predicting the equity premium with the demand for gold coins and bars. *Finance Research Letters*, *13*, 172–178.
- Baur, D. G., & Lucey, B. M. (2010). Is gold a hedge or a safe haven? An analysis of stocks, bonds and gold. *Financial Review*, *45*(2), 217–229.

- Baur, D. G., & McDermott, T. K. (2010). Is gold a safe haven? International evidence. *Journal of Banking & Finance*, *34*(8), 1886–1898.
- Bekaert, G., Engstrom, E., & Xing, Y. (2009). Risk, uncertainty, and asset prices. *Journal of Financial Economics*, *91*(1), 59–82.
- Bekaert, G., Hodrick, R. J., & Marshall, D. A. (1997). On biases in tests of the expectations hypothesis of the term structure of interest rates. *Journal of Financial Economics*, *44*(3), 309–348.
- Bekaert, G., & Hoerova, M. (2014). The VIX, the variance premium and stock market volatility. *Journal of Econometrics*, *183*(2), 181–192.
- Bhardwaj, G., Gorton, G., & Rouwenhorst, G. (2015). Facts and fantasies about commodity futures ten years later. *Yale International Center for Finance, New Haven, CT, ICF Working Paper No. 15-18*.
- Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica*, *77*(3), 623–685.
- Blose, L. E. (2010). Gold prices, cost of carry, and expected inflation. *Journal of Economics and Business*, *62*(1), 35–47.
- Bollerslev, T., Marrone, J., Xu, L., & Zhou, H. (2014). Stock return predictability and variance risk premia: Statistical inference and international evidence. *Journal of Financial and Quantitative Analysis*, *49*(3), 633–661.
- Bollerslev, T., Tauchen, G., & Zhou, H. (2009). Expected stock returns and variance risk premia. *Review of Financial Studies*, *22*(11), 4463–4492.
- Bollerslev, T., & Todorov, V. (2011a). Estimation of jump tails. *Econometrica*, *79*(6), 1727–1783.

- Bollerslev, T., & Todorov, V. (2011b). Tails, fears, and risk premia. *Journal of Finance*, *66*(6), 2165–2211.
- Bollerslev, T., & Todorov, V. (2014). Time-varying jump tails. *Journal of Econometrics*, *183*(2), 168–180.
- Bollerslev, T., Todorov, V., & Xu, L. (2015). Tail risk premia and return predictability. *Journal of Financial Economics*, *118*(1), 113–134.
- Cai, J., Cheung, Y.-L., & Wong, M. (2001). What moves the gold market? *Journal of Futures Markets*, *21*(3), 257–278.
- Campbell, J. Y., & Shiller, R. J. (1988). Stock prices, earnings, and expected dividends. *Journal of Finance*, *43*(3), 661–676.
- Campbell, J. Y., & Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies*, *21*(4), 1509–1531.
- Capie, F., Mills, T. C., & Wood, G. (2005). Gold as a hedge against the dollar. *Journal of International Financial Markets, Institutions and Money*, *15*(4), 343–352.
- Chen, S.-S. (2009). Predicting the bear stock market: Macroeconomic variables as leading indicators. *Journal of Banking & Finance*, *33*(2), 211–223.
- Christensen, B. J., & Prabhala, N. R. (1998). The relation between implied and realized volatility. *Journal of Financial Economics*, *50*(2), 125–150.
- Christie-David, R., Chaudhry, M., & Koch, T. W. (2000). Do macroeconomics news releases affect gold and silver prices? *Journal of Economics and Business*, *52*(5), 405–421.

- Chua, J., & Woodward, R. S. (1982). Gold as an inflation hedge: A comparative study of six major industrial countries. *Journal of Business Finance & Accounting*, 9(2), 191–197.
- Ciner, C., Gurdgiev, C., & Lucey, B. M. (2013). Hedges and safe havens: An examination of stocks, bonds, gold, oil and exchange rates. *International Review of Financial Analysis*, 29, 202–211.
- Clark, T. E., & West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 138(1), 291–311.
- Cochrane, J. H., & Piazzesi, M. (2005). Bond risk premia. *American Economic Review*, 95(1), 138–160.
- Dahlquist, M., & Hasseltoft, H. (2013). International bond risk premia. *Journal of International Economics*, 90(1), 17–32.
- De Bock, R., & de Carvalho Filho, I. (2015). The behavior of currencies during risk-off episodes. *Journal of International Money and Finance*, 53, 218–234.
- Diewald, L., Prokopczuk, M., & Wese Simen, C. (2015). Time-variations in commodity price jumps. *Journal of Empirical Finance*, 31, 72–84.
- Efron, B., & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, 1(1), 54–75.
- Fama, E. F., & Bliss, R. R. (1987). The information in long-maturity forward rates. *American Economic Review*, 77(4), 680–692.
- Fama, E. F., & French, K. R. (1987). Commodity futures prices: Some evidence on

- forecast power, premiums, and the theory of storage. *Journal of Business*, 60(1), 55–73.
- Fama, E. F., & French, K. R. (1988). Dividend yields and expected stock returns. *Journal of Financial Economics*, 22(1), 3–25.
- Fortune, J. N. (1988). The inflation rate of the price of gold, expected prices and interest rates. *Journal of Macroeconomics*, 9(1), 71–82.
- French, K. R., Schwert, G. W., & Stambaugh, R. F. (1987). Expected stock returns and volatility. *Journal of Financial Economics*, 19(1), 3–29.
- Gabaix, X. (2012). Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance. *Quarterly Journal of Economics*, 127(2), 645–700.
- Geske, R., & Roll, R. (1983). The fiscal and monetary linkage between stock returns and inflation. *Journal of Finance*, 38(1), 1–33.
- Ghosh, D., Levin, E. J., Macmillan, P., & Wright, R. E. (2004). Gold as an inflation hedge? *Studies in Economics and Finance*, 22(1), 1–25.
- Gorton, G., & Rouwenhorst, K. G. (2006). Facts and fantasies about commodity futures. *Financial Analysts Journal*, 62(2), 47–68.
- Gorton, G. B., Hayashi, F., & Rouwenhorst, K. G. (2013). The fundamentals of commodity futures returns. *Review of Finance*, 17(1), 35–105.
- Greene, W. H. (1998). *Econometric Analysis (3rd ed.)*. New York: Prentice Hall.
- Habib, M. M., & Stracca, L. (2012). Getting beyond carry trade: What makes a safe haven currency? *Journal of International Economics*, 87, 50–64.

- Hall, P., Horowitz, J. L., & Jing, B.-Y. (1995). On blocking rules for the bootstrap with dependent data. *Biometrika*, *82*(3), 561–574.
- Hodrick, R. J. (1992). Dividend yields and expected stock returns: Alternative procedures for inference and measurement. *Review of Financial Studies*, *5*(3), 357–386.
- Hossfeld, O., & MacDonald, R. (2015). Carry funding and safe haven currencies: A threshold regression approach. *Journal of International Money and Finance*, *59*, 185–202.
- Jaffe, J. F. (1989). Gold and gold stocks as investments for institutional portfolios. *Financial Analysts Journal*, *45*(2), 53–59.
- Kelly, B., & Jiang, H. (2014). Tail risk and asset prices. *Review of Financial Studies*, *27*(10), 2841–2871.
- Kessler, S., & Scherer, B. (2009). Varying risk premia in international bond markets. *Journal of Banking & Finance*, *33*(8), 1361–1375.
- Kho, B.-C. (1996). Time-varying risk premia, volatility, and technical trading rule profits: Evidence from foreign currency futures markets. *Journal of Financial Economics*, *41*(2), 249–290.
- Koijen, R., Moskowitz, T. J., Pedersen, L. H., & Vrugt, E. B. (2018). Carry. *Journal of Financial Economics*, *127*, 197–225.
- Kosowski, R., Timmermann, A., Wermers, R., & White, H. (2006). Can mutual fund stars really pick stocks? New evidence from a bootstrap analysis. *Journal of Finance*, *61*(6), 2551–2595.

- Kothari, S. P., & Shanken, J. (1997). Book-to-market, dividend yield, and expected market returns: A time-series analysis. *Journal of Financial Economics*, 44(2), 169–203.
- Lamont, O. (1998). Earnings and expected returns. *Journal of Finance*, 53(5), 1563–1587.
- Levin, E. J., Montagnoli, A., & Wright, R. (2006). Short-run and long-run determinants of the price of gold. *London, World Gold Council. Research Study No. 32*.
- Lewellen, J. (2004). Predicting returns with financial ratios. *Journal of Financial Economics*, 74(2), 209–235.
- Mahdavi, S., & Zhou, S. (1997). Gold and commodity prices as leading indicators of inflation: Tests of long-run relationship and predictive performance. *Journal of Economics and Business*, 49(5), 475–489.
- Mark, N. C. (1995). Exchange rates and fundamentals: Evidence on long-horizon predictability. *American Economic Review*, 85(1), 201.
- McCulloch, J. H., & Kwon, H.-C. (1993). US term structure data, 1947-1991. *Working Paper No. 93-6*.
- Nelson, C. R., & Kim, M. J. (1993). Predictable stock returns: The role of small sample bias. *Journal of Finance*, 48(2), 641–661.
- Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3), 703–708.
- O'Connor, F. A., Lucey, B. M., Batten, J. A., & Baur, D. G. (2015). The financial

- economics of gold – A survey. *International Review of Financial Analysis*, 41, 186–205.
- Pukthuanthong, K., & Roll, R. (2011). Gold and the dollar (and the euro, pound, and yen). *Journal of Banking & Finance*, 35(8), 2070–2083.
- Ranaldo, A., & Soederlind, P. (2010). Safe haven currencies. *Review of Finance*, 14, 385–407.
- Rapach, D. E., Wohar, M. E., & Rangvid, J. (2005). Macro variables and international stock return predictability. *International Journal of Forecasting*, 21(1), 137–166.
- Reboredo, J. C. (2013). Is gold a safe haven or a hedge for the US dollar? Implications for risk management. *Journal of Banking & Finance*, 37(8), 2665–2676.
- Rozeff, M. S. (1984). Dividend yields are equity risk premiums. *Journal of Portfolio Management*, 11(1), 68–75.
- Sherman, E. J. (1983). A gold pricing model. *Journal of Portfolio Management*, 9(3), 68–70.
- Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics*, 54(3), 375–421.
- Stock, J. H., & Watson, M. W. (1989). New indexes of coincident and leading economic indicators. *NBER Macroeconomics Annual*, 4, 351–394.
- Szymanowska, M., Roon, F., Nijman, T., & Goorbergh, R. (2014). An anatomy of commodity futures risk premia. *Journal of Finance*, 69(1), 453–482.



- Thorbecke, W. (1997). On stock market returns and monetary policy. *Journal of Finance*, *52*(2), 635–654.
- Tully, E., & Lucey, B. M. (2007). A power GARCH examination of the gold market. *Research in International Business and Finance*, *21*(2), 316–325.
- Wachter, J. A. (2013). Can time-varying risk of rare disasters explain aggregate stock market volatility? *Journal of Finance*, *68*(3), 987–1035.
- Welch, I., & Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, *21*(4), 1455–1508.
- Yang, F. (2013). Investment shocks and the commodity basis spread. *Journal of Financial Economics*, *110*(1), 164–184.

# The Risk Premium of Gold

Online Appendix

Table 10: Statistical Inference of the Prediction Model

This table presents the results accounting for potential [Stambaugh \(1999\)](#) biases. In Panel A, we report the bootstrapped p-values for the slope coefficients and the in-sample adj.  $R^2$ . In Panel B, we report the bootstrapped p-values for the out-of-sample  $R^2_{OOS}$  and the MSPE-adjusted test statistic. The bootstrap procedure is repeated 5,000 times and the p-values are obtained from the empirical distributions of the statistics.

	Gold			S&P 500		
<i>Panel A: In-Sample</i>						
	$LJP$	$VRP$	adj. $R^2$	$\log(D/P)$	$VRP$	adj. $R^2$
1-month	0.0932	0.0288	0.0218	0.0042	0.0010	0.0008
3-month	0.0086	0.0018	0.0006	0.0000	0.0000	0.0000
6-month	0.0020	0.0106	0.0008	0.0000	0.0006	0.0000
9-month	0.0000	0.0174	0.0000	0.0000	0.0030	0.0000
12-month	0.0000	0.0010	0.0000	0.0000	0.0078	0.0000
24-month	0.0000	0.0266	0.0000	0.0000	0.0058	0.0000
<i>Panel B: Out-of-Sample:</i>						
	$R^2_{OOS}$	MSFE		$R^2_{OOS}$	MSFE	
1-month	0.0034	0.0056		0.0160	0.0370	
3-month	0.0004	0.0000		0.0022	0.0014	
6-month	0.0130	0.0000		0.0006	0.0000	
9-month	0.0218	0.0000		0.0002	0.0000	
12-month	0.0120	0.0000		0.0000	0.0000	
24-month	0.0000	0.0000		0.0000	0.0000	

Table 11: Hedge and Safe Haven: Instrumental Variable Regression

This table presents the estimation results for the role of gold as a hedge and safe haven relying on expected returns and instrumental variables. Negative coefficients in column (2) indicate that gold is a hedge against the stock market while zero (negative) coefficients in column (3) indicate that gold is a weak (strong) safe haven. We report Wald test statistics for the significance of the coefficients below. Stars indicate significance of the estimates: \* significant at  $p < 0.10$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

	Hedge	10%
1-month	0.3861	0.2698
	1.0310	0.5493
6-month	-0.2106	0.3936***
	2.1282	9.9802
12-month	-0.1497	1.3669***
	0.5387	14.9951
24-month	-0.1999***	1.2804***
	18.8311	31.7491

Table 12: Finite Sample Bias

This table presents the results accounting for potential finite sample biases for the coefficient  $c_0$  and the sum  $c_0 + c_1$ . In Panel A, we report the results for the bias in coefficient estimates in percentage points. The bias is computed as the difference between the initial coefficient estimates and the mean of the coefficients obtained from a block-bootstrap of the dependent variable with 5,000 repetitions. Panel B reports the results for the hypothesis tests of the coefficients with bootstrapped critical values/p-values. Bias-corrected coefficient estimates are used to simulate the dependent variables under the null. We repeat this 5,000 times and obtain distributions of the Wald test statistics. We report the bias-corrected coefficients and the bootstrapped p-values below. Stars indicate significance of the estimates: \* significant at  $p < 0.10$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

	Hedge	10%
<i>Panel A: Coefficient Bias</i>		
1-month	0.0827	0.0056
6-month	0.3773	-0.9272
12-month	0.4727	-2.1364
24-month	0.1022	-1.0532
<i>Panel B: Finite Sample Distributions</i>		
1-month	0.2433*	-0.1469
	0.0512	0.4854
6-month	-0.1485	0.2338**
	0.2034	0.0180
12-month	-0.1344	1.2087**
	0.5333	0.0301
24-month	-0.1833***	1.1811***
	0.0052	0.0066

Table 13: Forecasting Annual Inflation

This table presents the results for the forecast of annual inflation at a quarterly frequency. The column labeled “Relative” reports the ratio of the RMSE relative to the AR model. The column labeled “ $1 - \lambda$ ” reports the coefficient from Equation (37), where [Newey & West \(1987\)](#) corrected standard errors and p-values are given in columns “NW SE” and “NW p”, respectively. The abbreviations for the different models are as in [Ang et al. \(2007\)](#).

	RMSE	Relative	$1 - \lambda$	NW SE	NW p
ARMA	0.1089	1.1005	-0.5563	1.1043	0.6162
AR	0.0990	1.0000			
RW	0.2610	2.6377	-0.0557	0.0390	0.1586
PC1	0.0977	0.9870	0.0580	0.3384	0.8646
PC2	0.1008	1.0182	0.1099	0.2541	0.6670
PC3	0.1036	1.0465	-0.1057	0.3238	0.7452
PC4	0.1002	1.0128	-1.7387	0.5905	0.0046
PC5	0.1090	1.1014	0.0634	0.2905	0.8279
PC6	0.0960	0.9701	0.4765	0.3479	0.1760
PC7	0.0989	0.9990	0.3051	0.4122	0.4620
PC9	0.1013	1.0240	0.0824	0.2374	0.7297
PC10	0.1042	1.0531	-0.0915	0.2774	0.7427
TS1	0.1039	1.0500	-0.2700	0.2466	0.2780
TS2	0.1121	1.1326	-0.1543	0.3263	0.6380
TS3	0.1208	1.2203	-0.3963	0.4459	0.3777
TS4	0.1082	1.0931	-0.3036	0.3302	0.3617
TS5	0.1149	1.1613	-0.0407	0.3785	0.9148
TS6	0.1004	1.0150	0.2717	0.2992	0.3676
TS7	0.1041	1.0520	0.1461	0.3267	0.6564
TS9	0.0992	1.0024	-0.1185	0.4161	0.7768
TS10	0.1081	1.0920	-0.1039	0.2130	0.6275
TS11	0.1076	1.0875	-0.0861	0.1337	0.5222
VAR	0.1157	1.1695	0.0183	0.1210	0.8800
SPF1	0.0998	1.0080	0.2553	0.5217	0.6263
SPF2	0.1038	1.0486	0.0684	0.6711	0.9191
SPF3	0.1062	1.0735	-0.0324	0.3889	0.9339
MICH1	0.1364	1.3786	-0.6637	0.3966	0.0992
MICH2	0.1032	1.0427	-0.4310	0.6081	0.4813
MICH3	0.1081	1.0920	-0.6198	0.6302	0.3295

Table 14: Gold as an Inflation Hedge - Alternative Models

This table presents the estimation results for the role of gold as a hedge against inflation relying on expected returns (columns (2)-(3)) or realized returns (columns (4)-(5)). The models used to obtain the expected inflation rate is reported are the AR model and three Phillips curve models (PC1, PC6 and PC7). We report [Newey & West \(1987\)](#) robust standard errors in parentheses below. Stars indicate significance of the estimates: \* significant at  $p < 0.10$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

	Expected Returns		Realized Returns	
	Intercept	$\beta$	Intercept	$\beta$
AR	0.0200 (0.0280)	-0.7347 (1.1995)	-0.0094 (0.0134)	0.5102 (0.4098)
PC1	0.0016 (0.0127)	0.0430 (0.5184)		
PC6	0.0132*** (0.0049)	-0.4575** (0.2210)		
PC7	0.0176*** (0.0061)	-0.6412** (0.2795)		