Forecasting the price of gold

Hossein Hassani[a], Emmanuel Sirimal Silva[a], Rangan Gupta[b,*] and Mawuli K. Segnon[c]

[a]*Statistical Research Centre, The Business School, Bournemouth, University of Bournemouth, UK*
[b]*Department of Economics, University of Pretoria, Pretoria, South Africa* [c]*Department of Economics, Christian-Albrechts-University Kiel, Kiel, 24118, Germany*

This article seeks to evaluate the appropriateness of a variety of existing forecasting techniques (17 methods) at providing accurate and statistically significant forecasts for gold price. We report the results from the nine most competitive techniques. Special consideration is given to the ability of these techniques to provide forecasts which outperforms the random walk (RW) as we noticed that certain multivariate models (which included prices of silver, platinum, palladium and rhodium, besides gold) were also unable to outperform the RW in this case. Interestingly, the results show that none of the forecasting techniques are able to outperform the RW at horizons of 1 and 9 steps ahead, and on average, the exponential smooth-ing model is seen providing the best forecasts in terms of the lowest root mean squared error over the 24-month forecasting horizons. Moreover, we find that the univariate models used in this article are able to outperform the Bayesian autoregression and Bayesian vector autoregressive models, with exponential smoothing reporting statistically significant results in comparison with the former models, and classical autoregressive and the vector autoregressive models in most cases.

Keywords: ARIMA; ETS; TBATS; ARFIMA; AR; VAR; BAR; BVAR; random walk; gold; forecast; multivariate; univariate

JEL Classification: C22; C53

I. Introduction

Gold serves several functions in the world economy, and its link with financial and macroeconomic variables is well established (Pierdzioch *et al.*, 2014a, b). It has a monetary value and is sought after by central banks to be part of their international reserves which fulfil many purposes (Gupta *et al.*, 2014). It has industrial uses and can be transformed into jewellery. In modern finance, it is used as a hedge against inflation and as a safe haven during crises. Gold has also other distinguished characteristics. Its supply is accumulated over the years, and its global annual physical production can be as small as 2%

*Corresponding author. E-mail: rangan.gupta@up.ac.za

of total supply; thereby in contrast to other commodities, its annual production may not sway its price as other factors do. Moreover, unlike prices of stocks and bonds, the gold price depends on future supply and demand, and thus, it is forward-looking.

It is clear that given the significance of gold in the modern world, the ability to provide accurate forecasts into the future price of gold will be of utmost importance. Moreover, there are benefits from finding the right model that forecasts the gold price more accurately than others. Out-of-sample forecasting offers informational availability advantage for monetary policy-makers, hedge fund managers and international portfolio managers, which can be used in gauging future inflation, estimating demand for jewellery, discerning investment in precious metals and other commodities and assessing the future movement of the dollar exchange rate.[1]

Figure 1 shows the time series for gold price, which is exploited in this study. In general, it appears to portray an exponential growth over time, and a first look at the figure shows signs of two major shocks post-1980 and post-2010 which create structural breaks in the time series. In this article, we aim to evaluate the use of a variety of forecasting models representing both parametric and nonparametric techniques for obtaining accurate forecasts for the price of gold.

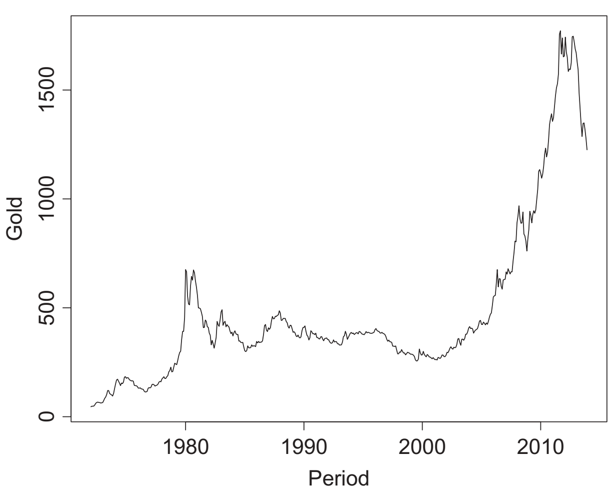

Fig. 1. Time series for gold price (January 1972–December 2013)

Whilst there exist various metrics which are used for comparing between two different out-of-sample forecasts, in this article, we rely on the root mean square error (RMSE) and the ratio of the RMSE criterion. As the RMSE criterion is in the process of gaining its popularity, we find it imperative to briefly describe this measure at the outset so that the reader has a clear understanding of the results reported in this article. In the very recent past, the RMSE criterion has been adopted as a popular measure in a range of forecasting studies (see for example, Hassani *et al.*, 2009, 2013a, b, 2015; Altavilla and De Grauwe, 2010; Silva, 2013; Silva and Hassani, 2015). As an example, we present the ratio of the RMSE between Model A and Model B:

$$\begin{aligned}\text{RRMSE} &= \frac{ModelA}{ModelB} \\ &= \frac{\left(\sum_{i=1}^{N}\left(\hat{y}_{T+h,i}-y_{T+h,i}\right)^2\right)^{1/2}}{\left(\sum_{i=1}^{N}\left(\tilde{y}_{T+h,i}-y_{T+h,i}\right)^2\right)^{1/2}}\end{aligned} \tag{1}$$

where $\hat{y}_{T+h}$ is the h-step ahead forecast obtained by Model A, $\tilde{y}_{T+h}$ is the h-step ahead forecast from Model B, y_t is the actual values, and N is the number of the forecasts. If $\frac{ModelA}{ModelB}$ is less than 1, then Model A outperforms Model B by $1-\frac{ModelA}{ModelB}$ per cent.

This article evaluates the application of 17 different forecasting techniques over 24 forecasting horizons from 1 month ahead up until 24 months ahead. This enables capturing both short- and long-run effectiveness of a given forecasting model at accurately predicting the future price of gold. The models

[1] Based on the suggestions of an anonymous referee, to test for the hedging characteristics of gold, we conducted constant parameter (Johansen, 1991) and time-varying parameter (Bierens and Martins, 2010) tests of cointegration between gold price, consumer price index and S&P500, first in a bivariate set-up between gold price and consumer price index, and gold price and S&P500, and then in a trivariate set-up involving all the variables. In a bivariate set-up, the constant parameter test of cointegration cannot detect any cointegration, suggesting that gold does not serve as a hedge for inflation and equity risks. However, realizing the existence of structural breaks and nonlinearities in the relationship between the price level and gold price and S&P500 and gold price, when we conducted the time-varying cointegration, we found overwhelming evidence of cointegration, thus validating that gold is a hedge against inflationary and equity risks. In addition, in the three-variable set-up, whilst the constant parameter cointegration detected one cointegrating vector, i.e. two common trends, the time-varying cointegration test detected two cointegrating vectors, implying one common trend, and thus again providing strong evidence in favour of gold serving as a hedge of inflation and stock prices.

evaluated in this study include an autoregressive model, an optimized autoregressive integrated moving average (ARIMA) model, exponential smoothing (ETS), trigonometric ETS state space model with Box–Cox transformation, ARMA errors, trend and seasonal components (TBATS), fractionalized ARIMA model (ARFIMA), vector autoregression (VAR), five variations of the Bayesian autoregression (BAR) models and five variations of the Bayesian VAR (BVAR) models. Note that the VAR-type models include the price of silver, platinum, palladium and rhodium as well, besides gold. We also ensure to compare each out-of-sample forecasting result along with a random walk (RW) model. The results from the nine most competitive models, selected based on the average lowest RMSE, are reported in this article.[2]

Note that unlike the existing literature (see for example, Pierdzioch *et al.*, 2014a, b; Aye *et al.*, forthcoming, and references cited therein) on forecasting the price of gold, which analyses the role of financial and macroeconomic variables in predicting gold price, we primarily concentrate on univariate approaches. This is important since gold is considered to be a leading indicator for inflation and growth (Stock and Watson, 2003), and hence, designing univariate models allows us to obtain forecasts of a leading indicator independent of other economic variables.[3] Further, the univariate approach also relieves us of the problem of choosing macroeconomic and financial variables that defines the world economic condition, given that gold is a globally traded asset. Having said this, we also use multivariate classical and BVAR models that include prices of four (silver, platinum, palladium and rhodium) other precious metals and compare with our univariate approaches. Finally to note is that, using our econometric approaches, which can handle nonstationarity of the data and hence, we forecast the price of gold in levels and not the gold returns as was done in the literature (Shafiee and Topal, 2010).[4] This is important since it is the forecast of future gold price trends, and not returns, which assists in mining companies to mitigate risk and uncertainty in gold price fluctuations and to carry out hedging, future investment and evaluation decisions. The results indicate that none of the forecasting models evaluated in this article are able to outperform the RW at the forecasting horizons of 1 and 9 steps ahead, whilst overall on average the ETS model illustrated the best out-of-sample forecasting performance by reporting the lowest average RMSE across all forecasting horizons with the optimized ARIMA model and TBATS reporting the second and third best performances, respectively. Interestingly, out of the reported models, the VAR model was the worst performer with the highest average RMSE.

The remainder of this article is organized as follows. Section II describes the methodology underlying the various forecasting techniques, whilst Section III is dedicated to an analysis of the data. Section IV reports the empirical results, and the article concludes in Section V.

II. Methodology

Forecasting models

Random walk. We use the RW as a benchmark, as it is a widely accepted practice that a forecasting technique which is recommended for a particular forecast should at least be more accurate than a RW (http://robjhyndman.com/hyndsight/benchmarks/). In brief, today's value for gold is forecasted to be tomorrow's value for gold.

[2] Note that the forecasting results from the eight models which are not reported in this article are available upon request.
[3] As suggested by an anonymous referee, we carried out a set of forecasting exercises for industrial production and the consumer price index separately in bivariate VAR models, with the other variable being gold, over the period 1972:01–2013:12. We observed that in case of the price level, the VAR model with gold always outperformed the RW model for horizons 1 to 24 months ahead (over the out-of-sample period of 1999:04–2013:12, with the Bai and Perron (2003) test determining the first structural break in 1999:03 for the price-level equation in a VAR(6) model), whilst for output the gains are observed till horizons 1 to 10 steps ahead (over the out-of-sample period of 1980:08–2013:12, with the Bai and Perron (2003) test determining the first structural break in 1980:07 for the output equation in a VAR(7) model). When we considered an out-of-sample period covering the crisis period of 2008:01–2013:12, we found that the VAR model with gold outperformed the RW model till horizon 8 for output and till horizon 6 and from 15 steps ahead onwards for CPI. Clearly then, whilst gold serves as a leading indicator for the output and price level, its predictive ability deteriorated during the crisis period. Further details on these results are available upon request from the authors.
[4] Shafiee and Topal (2010) address this issue of nonstationary gold prices by proposing a model that has three components: a long-term trend reversion component, a diffusion component and a jump or dip component.

Vector autoregression (VAR). The VAR model, though 'atheoretical' is particularly useful for forecasting purposes. An unrestricted VAR model, as suggested by Sims (1980), can be written as follows:

$$y_t = C + A(L)y_t + \epsilon_t \tag{2}$$

where $y\colon (n \times 1)$ is vector of variables (gold, silver, platinum, palladium and rhodium) being forecasted; $A(L)\colon (n \times n)$ is polynomial matrix in the backshift operator L with lag length p, i.e. $A(L) = A_1L + A_2L^2 + \ldots + A_pL^p$; $C\colon (n \times 1)$ is vector of constant terms, and $\varepsilon\colon (n \times 1)$ is vector of white-noise error terms.

The VAR model uses equal lag length for all the variables of the model. One drawback of VAR models is that many parameters are needed to be estimated, some of which may be insignificant. This problem of overparameterization, resulting in multicollinearity and loss of degrees of freedom, leads to inefficient estimates and large out-of-sample forecasting errors. One solution, often adapted, is simply to exclude the insignificant lags based on statistical tests. Another approach is to use near VAR, which specifies an unequal number of lags for the different equations.

However, an alternative approach to overcome this overparameterization, as described in Litterman (1981), Doan *et al.* (1984), Todd (1984), Litterman (1986)and Spencer (1993), is to use a BVAR model. Instead of eliminating longer lags, the Bayesian method imposes restrictions on these coefficients by assuming that these are more likely to be near 0 than the coefficient on shorter lags. However, if there are strong effects from less important variables, the data can override this assumption. The restrictions are imposed by specifying normal prior distributions with zero means and small SDs for all coefficients with the SD decreasing as the lags increases. The exception to this is, however, the coefficient on the first own lag of a variable, which has a mean of unity. Note that Litterman (1981) used a diffuse prior for the constant. This is popularly referred to as the 'Minnesota prior' due to its development at the University of Minnesota and the Federal Reserve Bank at Minneapolis.

The SD of the distribution of the prior for lag m of variable j in equation i for all i, j and m, defined as $S(i,j,m)$, can be specified as follows:

$$S(i,j,m) = [w \times g(m) \times f(i,j)]\frac{\sigma_i}{\sigma_j} \tag{3}$$

with $f(i,j) = 1$, if $i = j$ and k_{ij} otherwise, with $(0 \le k_{ij} \le 1)$, $g(m) = m^{-d}$, $d > 0$. Note σ_i is the SE of the univariate autoregression for variable i. The ratio $\frac{\sigma_i}{\sigma_j}$ scales the variables so as to account for differences in the units of measurement and, hence, causes specification of the prior without consideration of the magnitudes of the variables. The term w indicates the overall tightness and is also the SD on the first own lag, with the prior getting tighter as we reduce the value. The parameter $g(m)$ measures the tightness on lag m with respect to lag 1 and is assumed to have a harmonic shape with a decay factor of d, increasing which tightens the prior on increasing lags. The parameter $f(i,j)$ represents the tightness of variable j in equation i relative to variable i, and by increasing the interaction, i.e. the value of k_{ij}, we can loosen the prior. Following the extant literature on BVAR models, we look at the following combinations of w and d: (0.3, 0.5), (0.2, 1.0), (0.1, 1.0), (0.2, 2.0) and (0.1, 2.0), with k_{ij} set at 0.5. Univariate versions of the BVAR models, which we call BAR models, are estimated for the same values of w and d as above, but with k_{ij} set at 0.001, since a small interaction value basically reduces the multivariate model to its corresponding univariate version.

The BVAR model is estimated using Theil's (1971) mixed estimation technique, which involves supplementing the data with prior information on the distribution of the coefficients. In an artificial way, the number of observations and degrees of freedom are increased by one, for each restriction imposed on the parameter estimates. The loss of degrees of freedom due to overparameterization associated with a VAR model is, therefore, not a concern in the BVAR model. Further note that one major advantage of the BVAR and BAR models is that we can use nonstationary data for its estimation. Sims *et al.* (1990) indicate that with the Bayesian approach entirely based on the likelihood function, the associated inferences do not require special treatment for nonstationarity, since the likelihood function exhibits the same Gaussian shape regardless of the presence of nonstationarity.

Autoregressive integrated moving average (ARIMA). The optimal ARIMA model that is referred to as automatic ARIMA is provided through the forecast package for the R software. A more detailed description of the algorithm underlying automatic ARIMA can be found in Hyndman and Khandakar (2008). The number of differences, d, and the determination of its value are based on Kwiatkowski–Phillips–Schmidt–Shin (KPSS) (Kwiatkowski *et al.*, 1992) unit root test. Thereafter, the algorithm minimizes the Akaike information criterion (AIC) of the following form to determine the values of p and q.

$$AIC = -2\log(L) + 2(p + q + P + Q + k) \tag{4}$$

where $k = 1$ if $c \neq 0$ and 0 otherwise, and L represents the maximum likelihood of the fitted model.

The optimal model is chosen to be the model which represents the smallest AIC from the following options: ARIMA (2, d, 2), ARIMA (0, d, 0), ARIMA (1, d,0) and ARIMA (0, d, 1). The decision on the inclusion or exclusion of the constant c depends on the value of d.

In what follows, we provide a brief expansion to the above summary, and in doing so, we mainly follow Hyndman and Athanasopoulos (2013). A nonseasonal ARIMA model may be written as:

$$\begin{aligned}&(1 - \varphi_1 B - \ldots \varphi_p B^p)(1 - B)^d y_t \\ &\quad = c + (1 + \varphi_1 B + \ldots + \varphi_q B^q)e_t\end{aligned} \tag{5}$$

or

$$\begin{aligned}&(1 - \varphi_1 B - \ldots \varphi_p B^p)(1 - B)^d (y_t - \mu t^d/d!) \\ &\quad = (1 + \varphi_1 B + \ldots + \varphi_q B^q)e_t\end{aligned} \tag{6}$$

where μ represents the mean of $(1 - B)^d(y_t$, $c = \mu(1 - \phi_1 - \ldots - \phi_p)$, and B is the backshift operator. In R, the inclusion of a constant in a nonstationary ARIMA model is equivalent to inducing a polynomial trend of order d in the forecast function. It should be noted that when d = 0, μ is the mean of y_t.

According to Hyndman and Khandakar (2008), the seasonal ARIMA model can be expressed as:

$$\Phi(B^m)\,\phi(B)(1 - B^m)^D(1 - B)^d y_t = c + \Theta(B^m)\theta(B)\epsilon_t \tag{7}$$

where $\Phi(z)$ and $\Theta(z)$ are the polynomials of orders P and Q, and ϵ_t is white noise. If $c \neq 0$, there is an implied polynomial of order $d + D$ in the forecast function.

The process for obtaining point forecasts using the R software is concisely presented in Hyndman and Athanasopoulos (2013) as follows. First, the relevant ARIMA equation is expanded (i.e. nonseasonal or seasonal) so that y_t is on the left hand side with all other terms on the right. Second, we rewrite the ARIMA equation by replacing t with $T + h$, and finally, on the right hand side of this equation, we replace future observations by their forecasts, future errors by 0 and past errors by the corresponding residuals. Then, using the forecasting horizon h = 1 month ahead, we calculate all the forecasts for that horizon.

Exponential smoothing (ETS). The ETS technique incorporates the foundations of ETS and is made available through the forecast package for the R software. ETS overcomes a limitation found in earlier ETS models which did not provide a method for easy calculation of prediction intervals (Makridakis *et al.* 1998). The ETS model from the forecast package considers the error, trend and seasonal components along with over 30 possible options for choosing the best ETS model via optimization of initial values and parameters using the MLE and selecting the best model based on the AIC. A detailed description of ETS can be found in Hyndman and Athanasopoulos (2013).

Figure 2 summarizes in table format the several ETS formulas that are evaluated in the forecast package to select the best model to fit the data. Note that in this figure, ell_t denotes the series level at time t, b_t denotes the slope, s_t denotes the seasonal component of the series, and m denotes the number of seasons in a year; α, β, γ and ϕ are smoothing parameters, $\phi_h = \phi + \phi_2 + \ldots + \phi^h$, and $h_m^+ = [(h-1) mod m] + 1$ (Hyndman and Athanasopoulos, 2013).

Fractionalized ARIMA model (ARFIMA). We rely on the ARFIMA modelling process provided through the forecast package in R. Once again, the

Trend	Seasonal N	Seasonal A	Seasonal M
N	$\hat{y}_{t+h\|t} = \ell_t$ $\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t + s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)\ell_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = \ell_t s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)\ell_{t-1}$ $s_t = \gamma(y_t/\ell_{t-1}) + (1-\gamma)s_{t-m}$
A	$\hat{y}_{t+h\|t} = \ell_t + hb_t$ $\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t + hb_t + s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = (\ell_t + hb_t)s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t/(\ell_{t-1} + b_{t-1})) + (1-\gamma)s_{t-m}$
A_{d}	$\hat{y}_{t+h\|t} = \ell_t + \phi_h b_t$ $\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t + \phi_h b_t + s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1} - \phi b_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = (\ell_t + \phi_h b_t)s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t/(\ell_{t-1} + \phi b_{t-1})) + (1-\gamma)s_{t-m}$
M	$\hat{y}_{t+h\|t} = \ell_t b_t^h$ $\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}b_{t-1}$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*)b_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t b_t^h + s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1}$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1}b_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = \ell_t b_t^h s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1}$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t/(\ell_{t-1}b_{t-1})) + (1-\gamma)s_{t-m}$
M_{d}	$\hat{y}_{t+h\|t} = \ell_t b_t^{\phi_h}$ $\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}b_{t-1}^{\phi}$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*)b_{t-1}^{\phi}$	$\hat{y}_{t+h\|t} = \ell_t b_t^{\phi_h} + s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1}^{\phi}$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*)b_{t-1}^{\phi}$ $s_t = \gamma(y_t - \ell_{t-1}b_{t-1}^{\phi}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = \ell_t b_t^{\phi_h} s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1}^{\phi}$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*)b_{t-1}^{\phi}$ $s_t = \gamma(y_t/(\ell_{t-1}b_{t-1}^{\phi})) + (1-\gamma)s_{t-m}$

Fig. 2. Formulae for recursive calculations and point forecasts (Hyndman and Athanasopoulos, 2013)

modelling algorithm automatically estimates and selects the p and q for an ARFIMA(p,d,q) model based on the Hyndman and Khandakar (2008) algorithm, whilst d and parameters are selected based on the Haslett and Raftery (1989) algorithm.

ETS state space model with Box–Cox transformation, ARMA errors and trend and seasonal components (TBATS). The TBATS model is an ETS state space model with Box–Cox transformation, ARMA error correction and trend and seasonal components. The result is a technique which is aimed at providing accurate forecasts for time series with complex seasonality. A detailed description of the TBATS model can be found in De Livera *et al.* (2011) and is therefore not reproduced here.

III. Data

The data used in this study relate to the prices of gold, silver, platinum, palladium and rhodium. The data were obtained from www.kitco.com and represent monthly observations from January 1972 to December 2013. Note that though we are primarily interested in the price of gold, data for which are available from 1968, since we also estimate multivariate models which involve the prices of the other four precious metals, we start in January 1972 – the starting date from which price data on all the five precious metals are available. The end point of the sample is also purely driven by data availability at the time of writing this article (Table S1). In this article, we rely on a statistical test, namely the Bai and Perron (2003) sequential and repartition tests of multiple structural breaks, to determine the cut-off points for training and testing the forecasting models in order to ensure that the resulting output is robust, in the sense that we ensure that all our structural breaks fall in the out-of-sample part of the data, over which our models are recursively estimated to accommodate for changes in the parameter estimates of our forecasting methods. We evaluate out-of-sample forecasts for horizons of h = 1 step, up to h = 24 steps ahead, and thereby enable capturing and evaluating both short- and long-run forecasting abilities of the given forecasting models.

Table 1 presents some descriptives for these data to help understand the structure of the data and the time series. As the main interest of this article is the price of gold, we concentrate on describing these data whilst the descriptives for the other metals have been reported for those interested in comparing between these. The first observation is that based on the Shapiro–Wilk test, we are able to conclude with 99% confidence that the gold data are not normally distributed. We also considered the Kolmogorov–Smirnov test for normality which reported similar

Table 1. Descriptives for metals (January 1972–December 2013)

Series	Mean	Med.	SD	CV	Skew.	Shapiro–Wilk(p)	ADF
Gold	477.10	377.10	373.75	78.35	1.93	<0.01*	0.93†
Silver	891.80	553.00	760.56	85.28	2.17	<0.01*	–1.53†
Platinum	613.70	430.60	455.39	74.20	1.36	<0.01*	–0.32†
Palladium	230.30	141.80	201.09	87.32	1.58	<0.01*	–0.30†
Rhodium	1388.00	838.50	1581.75	113.96	2.62	<0.01*	–2.46†

Notes: * indicates data are not normally distributed based on a Shapiro–Wilk test at $p = 0.01$.
† indicates a nonstationary time series based on the augmented Dickey–Fuller (ADF) test at $p = 0.01$.

results. This in turn suggests that the median is a more accurate measure of the central tendency in this data-set and accordingly leads us to conclude that during this time period the median gold price of 377.10 is more reliable than the reported mean gold price of 477.10. This is because the positive skewness in these gold data has implications on the accuracy of the mean with outliers artificially increasing the mean value. A first look at Fig. 1 suggested that this time series has unit root problems, and this has been confirmed via all standard unit root tests. The results indicated that the gold series does indeed experience a unit root problem at a p-value of 0.01. We also report the coefficient of variation (CV) statistic for this metal in order to enable comparing the variation in gold during this time period with other metals. Accordingly, it is clear that in comparison with the other metals, platinum has illustrated the least variation during this time period, whilst rhodium has illustrated the highest variation as indicated by the highest CV. In comparison with the average, the SD in gold data appears to be considerably high, and this is likely to be influenced by the two major structural breaks visible in Fig. 1.

Given that the unit root tests confirmed a statistically significant unit root problem in the gold data, and as Fig. 1 suggests at least two clear structural breaks in the gold series, we apply the Bai and Perron (2003) test for break points on the gold equation of the VAR model, with lag lengths being chosen by the Akaike information criteria (AIC). The results from this test are reported in Table 2. First and foremost, the Bai and Perron (2003) test enables us to identify that there are five structural breaks affecting the gold time series as opposed to the two major breaks expected when analysing Fig. 1 with a naked eye. Second, based on the Bai and Perron (2003) test, it is suggested that the first structural break has occurred in April 1980 and also that there have been four other breaks during April 1984, August 2001, September 2007 and December 2007. Interestingly, this test is unable to capture the structural break which is visible post-2010, but this is more due to the fact that we lose 15% of the observations from both ends of the sample.[5] In any event, this is not a concern for us, since the models are recursively estimated over April 1980 till December 2013, with the in-sample observation being from January 1972 to March 1980.

Table 2. Break points in gold price time series

Break point	Time
Gold	1980(4), 1984(4), 2001(8), 2007(9), 2007(12)

IV. Empirical Results

Table 3 reports the RMSE for out-of-sample forecasting results. We begin our analysis of the results by comparing between the RMSE values. First and foremost, it is pertinent to point out that no single model is able to provide the best forecast for the gold price at all horizons. However, based on the lowest average RMSE, we can conclude that the ETS model is best for forecasting the gold price, should we be interested in relying on a single model over the selected forecasting horizons.

Interestingly, at the horizon of $h = 1$ and 9 steps ahead, the RW model is seen outperforming all other models, whilst the TBATS model provides the second best forecast in terms of the lowest

[5] Similar break dates were also obtained for the autoregressive equation of gold.

Table 3. RMSE for out-of-sample forecasts for gold

Horizon	RW	AR	VAR	BAR	BVAR	ARIMA	ETS	ARFIMA	TBATS
1	**28.17**	30.41	63.57	28.47	28.82	29.52	28.99	32.01	28.23
2	43.17	49.53	126.30	44.51	45.52	**24.97**	27.02	37.54	43.11
3	54.31	60.83	152.83	56.77	58.41	52.60	**48.05**	53.40	60.38
4	64.81	73.36	171.27	69.10	71.43	54.22	**52.21**	68.05	60.44
5	73.73	85.87	209.26	79.97	82.29	59.18	**57.46**	81.45	62.97
6	82.88	99.33	229.40	91.39	93.37	70.42	**66.68**	92.41	77.11
7	91.58	112.96	265.35	103.06	104.45	77.81	**70.55**	83.32	88.59
8	99.70	126.24	241.40	114.36	115.30	72.98	**71.40**	94.73	87.37
9	**106.77**	139.79	402.32	124.45	124.76	334.92	278.16	314.46	308.04
10	112.41	152.90	629.60	132.76	132.33	150.98	**95.11**	234.50	157.70
11	117.94	165.06	933.01	140.88	139.42	100.92	91.83	124.03	**90.67**
12	123.99	177.19	1165.96	149.39	147.82	97.12	**92.64**	140.62	103.42
13	130.08	191.56	1271.82	158.01	157.12	**83.98**	92.95	131.60	89.66
14	135.86	206.79	1492.50	166.33	165.79	109.71	**88.73**	129.20	128.12
15	141.39	221.91	972.61	174.33	173.86	92.78	**78.64**	150.91	92.26
16	147.18	238.32	2981.31	182.75	182.46	**90.79**	99.97	148.71	100.28
17	154.13	257.91	3697.32	193.69	194.39	108.57	**87.74**	158.95	99.09
18	161.05	278.95	7603.48	204.93	207.18	112.95	112.23	166.65	**108.86**
19	167.81	301.58	8396.89	216.40	220.59	137.12	148.92	172.20	**117.96**
20	174.09	326.08	11 332.61	227.66	235.71	135.24	155.47	170.00	**110.40**
21	179.78	351.72	14 051.50	236.98	249.95	94.46	103.86	174.93	**88.49**
22	185.41	380.46	11 727.67	245.81	265.40	118.71	**104.57**	175.23	110.23
23	190.65	411.11	31 921.26	251.97	282.26	122.46	**106.31**	188.17	124.49
24	196.23	447.99	21 536.77	259.67	304.31	141.60	**137.35**	190.52	138.91
Average	123.46	203.66	5065.67	152.23	157.62	103.08	**95.70**	138.07	103.20
Score	2	0	0	0	0	3	**14**	0	5

Notes: Score indicates the number of times a model is able to outperform all other models. Bold values indicate the minimum RMSE.

error at 1 step ahead and the BAR model provides the second best forecast at 9 steps ahead. The optimal ARIMA model is seen providing the best forecasts at horizons of 2, 13 and 16 steps ahead, whilst TBATS is seen outperforming the rest of the models at $h = 11$, 18, 19, 20 and 21 months ahead. The ETS model however captures the top position in terms of providing the best forecasts for a majority of the forecasting horizons with a score of 14 out of 24, where score indicates the number of times that a model reports the lowest RMSE (in comparison with the others) over the 24 forecasting horizons. In this case, it is also noteworthy that BAR, BVAR and ARFIMA models report a score of 0.[6] Moreover, based on the average RMSE, it is clear that the only two models which can provide a competitive forecast for gold price next to ETS are the optimal ARIMA and TBATS models.

Table 4 reports the relative RMSE (RRMSE) results for the gold price forecasts. As the ETS model was seen reporting the lowest average forecasting error, we now consider ETS as a benchmark and evaluate the forecasting performance of the other techniques in relation to that of ETS. Here, we also test the out-of-sample forecasting errors for statistical significance based on Harvey *et al.*'s (1997) modified Diebold–Mariano (DM) test.

Based on the suggestions of an anonymous referee, we also apply the superior predictive ability (SPA) test of Hansen (2005) using the RMSE as the

[6] Though we estimate five BAR and five BVAR models, we found that the best model corresponds to $w = 0.1$ and $d = 2.0$, implying the most tight-priored Bayesian models. For the BVAR models, this also highlights the fact that lagged information from the four other metal prices does not necessarily add any forecasting gains for the gold price. This is also vindicated by the fact that the BAR model outperforms the BVAR model.

Table 4. RRMSE for out-of-sample forecasts for gold

Horizon	$\frac{ETS}{RW}$	$\frac{ETS}{AR}$	$\frac{ETS}{VAR}$	$\frac{ETS}{BAR}$	$\frac{ETS}{BVAR}$	$\frac{ETS}{ARIMA}$	$\frac{ETS}{ARFIMA}$	$\frac{ETS}{TBATS}$
1	1.03	0.95*	0.46*	1.02*	1.01*	0.98	0.91	1.03
2	0.63*	0.55*	0.21[b]	0.61*	0.59*	1.08	0.72	0.63[b]
3	0.88	0.79[§]	0.31	0.85	0.82	0.91	0.90	0.80[§]
4	0.81[b]	0.71[b]	0.30	0.76[b]	0.73[b]	0.96	0.77	0.86
5	0.78[b]	0.67[b]	0.27	0.72[b]	0.70*	0.97	0.71	0.91
6	0.80	0.67[b]	0.29	0.73[§]	0.71*	0.95	0.72	0.86
7	0.77[§]	0.62[b]	0.27	0.68[b]	0.68[b]	0.91	0.85	0.80[b]
8	0.72	0.57[b]	0.30	0.62[b]	0.62[b]	0.98	0.75	0.82[§]
9	2.61	1.99	0.69	2.24	2.23	0.83[b]	0.88[§]	0.90[§]
10	0.85	0.62[§]	0.15	0.72	0.72	0.63	0.41	0.60
11	0.78	0.56[b]	0.10*	0.65*	0.66[b]	0.91[§]	0.74	1.01
12	0.75	0.52[b]	0.08*	0.62[b]	0.63[b]	0.95	0.66	0.90
13	0.71	0.49[§]	0.07*	0.59	0.59[b]	1.11	0.71	1.04
14	0.65	0.43[§]	0.06*	0.53[b]	0.54[b]	0.81	0.69	0.69
15	0.56	0.35[§]	0.08*	0.45[b]	0.45[b]	0.85	0.52	0.85
16	0.68	0.42[§]	0.03*	0.55[§]	0.55[§]	1.10	0.67	1.00
17	0.57	0.34[§]	0.02*	0.45[§]	0.45[b]	0.81[§]	0.55	0.89
18	0.70	0.40[§]	0.01*	0.55[§]	0.54[§]	0.99	0.67	1.03
19	0.89	0.49	0.02*	0.69[b]	0.68	1.09	0.86	1.26
20	0.89	0.48[§]	0.01*	0.68[b]	0.66	1.15	0.91	1.41
21	0.58	0.30[§]	0.01*	0.44[b]	0.42[b]	1.10	0.59	1.17
22	0.56	0.27[§]	0.01*	0.43[§]	0.39[b]	0.88	0.60	0.95
23	0.56	0.26[§]	0.00*	0.42[§]	0.38[b]	0.87	0.56	0.85
24	0.70	0.31[§]	0.01*	0.53[b]	0.45[b]	0.97	0.72	0.99
Average	0.78	0.57	0.16	0.63	0.61	0.93	0.69	0.93
Sig. Score	4	22	16	20	19	3	1	5

Notes: * indicates results are statistically significant based on a modified DM test at $p = 0.01$. [b] indicates results are statistically significant based on a modified DM test at $p = 0.05$. [§] indicates results are statistically significant based on a modified DM test at $p = 0.10$. Sig. Score indicates the number of times ETS has outperformed the other models with statistically significant results.

loss function. The advantage of the SPA test is that it allows to compare the forecasting performance of one model with that of all its competitors. We test the null hypothesis that the benchmark model cannot be outperformed by the competitive models. The results are reported in Table 5. We find that based on the RMSE, forecasts from the ETS model dominate forecasts from RW, AR, VAR, BAR, BVAR, ARIMA, ARFIMA and TBATS.

Overall, based on the RRMSE criterion, we can conclude that on average, for forecasting gold prices across the 24 horizons, the ETS model provides a forecast that is 22%, 43%, 84%, 37%, 39%, 7%, 31% and 7% better than forecasts from RW, AR, VAR, BAR, BVAR, ARIMA, ARFIMA and TBATS models. If we consider the RRMSE values in relation to the DM test results, then we can see some interesting observations. First, after the $h = 10$ months ahead step, none of the RRMSE results between ETS-RW, ETS-ARFIMA and ETS-TBATS are found to be statistically significant. Moreover, the Sig. Score criterion reported in Table 4 shows that the superior out-of-sample forecasting RMSE results obtained by ETS in relation to RW, ARIMA, ARFIMA and TBATS models are less reliable owing to its considerably low statistical significance across the 24 horizons. In contrast, the DM test on the RRMSE suggests that there is more statistical reliability behind the conclusion that ETS does provide a more accurate and statistically significant forecast for gold price in comparison with AR, VAR, BAR and BVAR models as these report comparatively high significant scores.

Table 5. SPA test results

Horizon	RW	AR	VAR	BAR	BVAR	ARIMA	ETS	ARFIMA	TBATS
	Benchmark models								
1	0.001	0.000	**1.000**	0.000	0.000	0.000	0.000	0.001	0.000
2	0.009	0.003	0.040	0.003	0.005	**0.527**	0.473	0.083	0.043
3	0.208	0.060	0.040	0.143	0.156	0.372	**0.971**	0.315	0.045
4	0.031	0.007	0.042	0.013	0.009	0.361	**0.735**	0.080	0.072
5	0.009	0.004	0.033	0.007	0.002	0.204	**0.796**	0.038	0.105
6	0.046	0.010	0.042	0.018	0.006	0.128	**1.000**	0.052	0.035
7	0.022	0.006	0.025	0.010	0.008	0.079	**1.000**	0.125	0.003
8	0.066	0.005	0.017	0.015	0.013	0.296	**0.704**	0.175	0.003
9	**1.000**	0.045	0.050	0.093	0.096	0.000	0.001	0.000	0.000
10	0.114	0.037	0.048	0.076	0.079	0.001	**1.000**	0.001	0.004
11	0.009	0.002	0.068	0.001	0.000	0.031	0.427	0.057	**0.573**
12	0.012	0.002	0.066	0.001	0.000	0.142	**1.000**	0.041	0.090
13	0.008	0.002	0.061	0.009	0.003	**0.804**	0.404	0.024	0.524
14	0.010	0.004	0.091	0.010	0.003	0.125	**0.909**	0.064	0.035
15	0.002	0.001	0.091	0.005	0.001	0.031	**0.909**	0.006	0.015
16	0.001	0.001	0.057	0.002	0.000	1.000	**0.294**	0.048	0.247
17	0.001	0.001	0.164	0.001	0.000	0.016	**0.836**	0.010	0.005
18	0.001	0.001	0.108	0.002	0.000	0.449	0.814	0.017	**0.818**
19	0.001	0.001	0.065	0.008	0.003	0.160	0.199	0.012	**1.000**
20	0.000	0.001	0.091	0.005	0.001	0.086	0.138	0.014	**1.000**
21	0.000	0.001	0.080	0.004	0.001	0.268	0.276	0.004	**0.983**
22	0.000	0.003	0.091	0.004	0.001	0.034	**0.909**	0.016	0.110
23	0.001	0.002	0.074	0.004	0.002	0.012	**1.000**	0.018	0.001
24	0.000	0.002	0.162	0.004	0.005	0.346	**0.923**	0.039	0.743

Notes: The values in the table are the p-values of the SPA test of Hansen (2005) using the MSE as the loss function. The null hypothesis is that the benchmark model is not outperformed by the other competitive models. A high p-value indicates that the null hypothesis is not rejected. Bold values indicate the highest p-value across models and horizons.

This result is strongly vindicated by the SPA test, which shows that the ETS dominates all other models.[7,8]

V. Conclusions

This article begins with a concise introduction relating to the importance of gold in the world economy. The foremost aim was to evaluate the use of 17 different parametric and nonparametric time series analysis and forecasting techniques, including both univariate and multivariate models for forecasting the price of gold. Prior to forecasting the gold price, the article considers a wide range of crucial statistical tests. The first of which tests the cointegration between gold price, consumer price index and S&P500 validates that gold is a hedge against inflationary and equity risks. In addition, the gold data undergo tests for break points, normality and unit

[7] Based on the suggestions of an anonymous referee, we also conducted the model confidence set (MCS) test of Hansen *et al.* (2011). The MCS test ranked the BVAR model for horizon 1, the RW model for horizons 2 to 8, 10 and 22–23, and the VAR model as the best model across the remaining horizons (9, 11–21 and 24 steps ahead). This result did not make sense to us in the light of the fact that ETS outperformed all the other models in 14 out of the 24 horizons. Given this, we chose not to report the MCS test results in the main text, but are available upon request from the authors.

[8] In addition to the evidence of inflation and equity risks hedging capabilities of inflation provided in Footnote 1, we also analysed the mean loss based on value at risk (MVaR) measure obtained from the out-of-sample forecasts for the ETS – our best forecasting model – and the RW model. We observed that, in primarily shorter horizons (1–9 and 13 being the exception), the RW performs better as it has a lower MVaR value. However, at horizons 11 and 12 and from 14 onwards, the ETS produces lower values and hence is better suited for portfolio allocation. On average, the ETS is better based on the MVaR measure. Recall that, in financial mathematics and financial risk management, VaR is a widely used measure of the risk of loss on a specific portfolio of financial assets.

root problems, which are then discussed prior to embarking on the main forecasting exercise. The out-of-sample forecasts are evaluated using the RMSE and RRMSE criterions, whilst the statistical significance of all forecasts is ascertained via the modified DM test (Harvey *et al.*, 1997) and Hansen's SPA test (Hansen, 2005).

Using monthly gold price data, we consider 24 different forecasting horizons which cover both the long run and short run. In text, we report and compare the results obtained from the nine most accurate forecasting models which include RW, AR, VAR, BAR, BVAR, ARIMA, ARFIMA, TBATS and ETS. Our results are interesting; first we find that no single model (out of the 17 models) is able to provide the most accurate forecast of gold price across both the short run and long run. Second, we see the forecasts from univariate models successfully outperforming those from multivariate BAR and BVAR models which were developed for forecasting the gold price. Third, we find forecasts from a univariate ETS model outperforming AR, VAR, BAR and BVAR forecasts with statistically significant results in most cases. Fourth, whilst forecasts from ETS are seen outperforming forecasts from RW, ARIMA, ARFIMA and TBATS models based on the RMSE criterion, when tested for statistical significance (DM), the ETS model reported the lowest significance score, suggesting that the results relating to the performance of ETS in comparison with these particular models could be attributable to chance occurrences. Based on the DM test, there is more confidence on the outcomes which show that ETS forecasts outperform forecasts from AR, VAR, BAR and BVAR models owing to very high number of statistically significant outcomes (as reported by the Sig. Score). Should one be interested in relying on a single model that can provide the most accurate forecast for gold price across 24 horizons, then ETS tops the list of contenders in comparison with the 17 models evaluated in this study owing to its lowest average RMSE. It is noteworthy that the aforementioned conclusion is strongly vindicated by the SPA test that finds ETS forecasts dominating all other models. Finally, based on the RRMSE criterion, we can conclude that the ETS model provides out-of-sample forecasts for gold which are 22%, 43%, 84%, 37%, 39%, 7%, 31% and 7% better than forecasts from RW, AR, VAR, BAR, BVAR, ARIMA, ARFIMA and TBATS models.

The findings reported through this article help forecasters of gold price in choosing the most appropriate model (from those evaluated here) based on the forecasting horizon which is of interest, or selecting one model which can provide the best average forecasts for gold price in the short run and long run (i.e. ETS). Moreover, the results suggest that when forecasting the price of gold, there is scope for accurate and reliable forecasts from univariate models as opposed to complex multivariate models. As discussed in Footnote 3, an additional study alongside the main analysis also found evidence which suggests that whilst gold does serve as a leading indicator for output and price level, its predictive ability has deteriorated during the crisis periods in the past. This result indicates that in the future, during recessions, decision-makers and analysts should be cautious when considering gold price as a main indicator for price level and output. Future research should consider evaluating and comparing ETS forecasts for gold price alongside other univariate and multivariate models such as singular spectrum analysis, neural networks and multivariate singular spectrum analysis.

Acknowledgements

We would like to thank an anonymous referee for many helpful comments. However, any remaining errors are solely ours.

Disclosure Statement

No potential conflict of interest was reported by the authors.

Supplemental Data

Supplemental data for this article is available as a separate file)

References

Altavilla, C. and De Grauwe, P. (2010) Forecasting and combining competing models of exchange rate determination, *Applied Economics*, **42**, 3455–80.

Aye, G. C., Gupta, R., Hammoudeh, S. *et al.* (forthcoming) Forecasting the price of gold using dynamic model averaging, *International Review of Financial Analysis*.

Bai, J. and Perron, P. (2003) Computation and analysis of multiple structural change models, *Journal* *of Applied Econometrics*, **18**, 1–22. doi:10.1002/jae.659

Bierens, H. J. and Martins, L. (2010) Time varying cointegration, *Econometric Theory*, **26**, 1453–90. doi:10.1017/S0266466609990648

De Livera, A. M., Hyndman, R. J. and Snyder, R. D. (2011) Forecasting time series with complex seasonal patterns using exponential smoothing, *Journal of the American Statistical Association*, **106**, 1513–27. doi:10.1198/jasa.2011.tm09771

Doan, T. A., Litterman, R. B. and Sims, C. A. (1984) Forecasting and conditional projection using realistic prior distributions, *Econometric Reviews*, **3**, 1–100. doi:10.1080/07474938408800053

Gupta, R., Hammoudeh, S., Kim, W. J. *et al.* (2014) Forecasting China's foreign exchange reserves using dynamic model averaging: the roles of macroeconomic fundamentals, financial stress and economic uncertainty, *The North American Journal of Economics and Finance*, **28**, 170–89. doi:10.1016/j.najef.2014.02.003

Hansen, P. R. (2005) A test for superior predictive ability, *Journal of Business and Economic Statistics*, **23**, 365–80. doi:10.1198/073500105000000063

Hansen, P. R., Peter, R., Lunde, A. *et al.* (2011) The model confidence set, *Econometrica*, **79**, 453–97. doi:10.3982/ECTA5771

Harvey, D. I., Leybourne, S. J. and Newbold, P. (1997) Testing the equality of prediction mean squared errors, *International Journal of Forecasting*, **13**, 281–91. doi:10.1016/S0169-2070(96)00719-4

Haslett, J. and Raftery, A. E. (1989) Space-time modelling with long-memory dependence: assessing Ireland's wind power resource (with discussion), *Applied Statistics*, **38**, 1–50. doi:10.2307/2347679

Hassani, H., Heravi, S. and Zhigljavsky, A. (2009) Forecasting European industrial production with singular spectrum analysis, *International Journal of Forecasting*, **25**, 103–18. doi:10.1016/j.ijforecast.2008.09.007

Hassani, H., Heravi, S. and Zhigljavsky, A. (2013a) Forecasting UK industrial production with multivariate singular spectrum analysis, *Journal of Forecasting*, **32**, 395–408. doi:10.1002/for.2244

Hassani, H., Soofi, A. S. and Zhigljavsky, A. (2013b) Predicting inflation dynamics with singular spectrum analysis, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **176**, 743–60. doi:10.1111/j.1467-985X.2012.01061.x

Hassani, H., Webster, A., Silva, E. S. *et al.* (2015) Forecasting U.S. tourist arrivals using optimal singular spectrum analysis, *Tourism Management*, **46**, 322–35. doi:10.1016/j.tourman.2014.07.004

Hyndman, R. J. and Athanasopoulos, G. (2013) *Forecasting: Principles and Practice*, OTexts, Australia. Available at OTexts.com/fpp

Hyndman, R. J. and Khandakar, Y. (2008) Automatic time series forecasting: the forecast package for R, *Journal of Statistical Software*, **27**, 1–22.

Johansen, S. (1991) Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, *Econometrica*, **59**, 1551–80. doi:10.2307/2938278

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. *et al.* (1992) Testing the null hypothesis of stationarity against the alternative of a unit root, *Journal of Econometrics*, **54**, 159–78.

Litterman, R. B. (1981). A Bayesian procedure for forecasting with vector autoregressions, Working Paper, Federal Reserve Bank of Minneapolis, Minneapolis, MN.

Litterman, R. B. (1986) Forecasting with Bayesian vector autoregressions: five years of experience, *Journal of Business and Economic Statistics*, **4**, 25–38.

Makridakis, S. G., Wheelwright, S. C. and Hyndman, R. J. (1998) *Forecasting: Methods and Applications*, 3rd edn, Wiley, New York.

Pierdzioch, C., Risse, M. and Rohloff, S. (2014a) On the efficiency of the gold market: results of a real-time forecasting approach, *International Review of Financial Analysis*, **32**, 95–108. doi:10.1016/j.irfa.2014.01.012

Pierdzioch, C., Risse, M. and Rohloff, S. (2014b) The international business cycle and gold-price fluctuations, *The Quarterly Review of Economics and Finance*, **54**, 292–305. doi:10.1016/j.qref.2014.01.002

Shafiee, S. and Topal, E. (2010) An overview of global gold market and gold price forecasting, *Resources Policy*, **35**, 178–89. doi:10.1016/j.resourpol.2010.05.004

Silva, E. S. (2013) A combination forecast for energy-related CO_2 emissions in the United States, *International Journal of Energy and Statistics*, **1**, 269–79. doi:10.1142/S2335680413500191

Silva, E. S. and Hassani, H. (2015) On the use of singular spectrum analysis for forecasting U.S. trade before, during and after the 2008 recession, *International Economics*, doi:10.1016/j.inteco.2014.11.003

Sims, C. A. (1980) Macroeconomics and reality, *Econometrica*, **48**, 1–48. doi:10.2307/1912017

Sims, C. A., Stock, J. H. and Watson, M. W. (1990) Inference in linear time series models with some unit roots, *Econometrica*, **58**, 113–44. doi:10.2307/2938337

Spencer, D. E. (1993) Developing a Bayesian vector autoregression forecasting model, *International Journal of Forecasting*, **9**, 407–21. doi:10.1016/0169-2070(93)90034-K

Stock, J. H. and Watson, M. W. (2003) Forecasting output and inflation: the role of asset prices, *Journal of Economic Literature*, **41**, 788–829. doi:10.1257/jel.41.3.788

Theil, H. (1971) *Principles of Econometrics*, John Wiley, New York, NY.

Todd, R. M. (1984). Improving economic forecasting with Bayesian vector autoregression. *Quarterly Review*, Federal Reserve Bank of Minneapolis, Fall, 18–29.