



# South African stock return predictability in the context data mining: The role of financial variables and international stock returns <sup>☆</sup>

Rangan Gupta <sup>\*</sup>, Mampho P. Modise

Department of Economics, University of Pretoria, Pretoria, 0002, South Africa

## ARTICLE INFO

### Article history:

Accepted 21 December 2011

### JEL classification:

C22  
C52  
C53  
G12  
G14

### Keywords:

Stock return predictability  
Financial variables  
Nested models  
In-sample tests  
Out-of-sample tests  
Data mining  
General-to-specific model selection

## ABSTRACT

In this paper, we examine the predictive ability, both in-sample and the out-of-sample, for South African stock returns using a number of financial variables, based on monthly data with an in-sample period covering 1990:01 to 1996:12 and the out-of-sample period of 1997:01 to 2010:04. We use the *t*-statistic corresponding to the slope coefficient in a predictive regression model for in-sample predictions, while for the out-of-sample, the MSE-F and the ENC-NEW tests statistics with good power properties were utilised. To guard against data mining, a bootstrap procedure was employed for calculating the critical values of both the in-sample and out-of-sample test statistics. Furthermore, we use a procedure that combines in-sample general-to-specific model selection with out-of-sample tests of predictive ability to further analyse the predictive power of each financial variable. Our results show that, for the in-sample test statistic, only the stock returns for our major trading partners have predictive power at certain short and long run horizons. For the out-of-sample tests, the Treasury bill rate and the term spread together with the stock returns for our major trading partners show predictive power both at short and long run horizons. When accounting for data mining, the maximal out-of-sample test statistics become insignificant from 6-months onward suggesting that the evidence of the out-of-sample predictability at longer horizons is due to data mining. The general-to-specific model shows that valuation ratios contain very useful information that explains the behaviour of stock returns, despite their inability to predict stock return at any horizon. The model also highlights the role of multiple variables in predicting stock returns at medium- to long run horizons.

© 2011 Elsevier B.V. All rights reserved.

## 1. Introduction

The recent financial turmoil has once again highlighted the importance of accurate forecasting, especially when it involves predicting the path of leading indicators of the economy. There exists international evidence that asset prices, including stock prices, not only help in predicting output and inflation by acting as leading indicators (Stock and Watson, 2003), but also that there are major (asymmetric) spillovers from the stock markets to the real sector of the economy (for some recent evidence, refer to, Apergis and Miller, 2004, 2005a,b, 2006; Das et al., forthcoming; Lettau and Ludvigson, 2001, 2004; Lettau et al., 2002; Pavlidis et al., 2009; Rapach and Strauss, 2006, 2007, amongst others). Hence, obtaining accurate predictions of stock prices cannot be understated, since if predicted accurately, the forecasts not only paves a path for relevant policy decision in advance, but can also provide important information

for policy makers to appropriately design policies to avoid the impending crisis.

In a recent study, Gupta and Modise (2010), using monthly South African data for 1990:01–2009:10, examined the in-sample predictability of real stock prices based on valuation ratios, namely, price–dividend and price–earnings ratios. The authors could not detect either short-horizon or long-horizon predictability; that is, the hypothesis that the current value of a valuation ratio is uncorrelated with future stock price changes cannot be rejected at both short- and long-horizons based on bootstrapped critical values constructed from both linear and non-linear representations of the data. Gupta and Modise (2010), however, note that, future research should aim to investigate not only in-sample, but also out-of-sample predictability of real stock returns based on a wider set of financial variables, since it is possible for a variable to carry significant out-of-sample information even when it is not the case in-sample (Rapach and Wohar, 2006a; Rapach et al., 2005). In addition, Gupta and Modise (2010), following the recent work by Rapach et al. (2010), suggested the need to analyze the role played by stock returns of major trading partners of South Africa in explaining the future path of stock returns.

Against this backdrop, using a predictive regression framework, we aim to implement the above set of extensions suggested by Gupta and Modise (2010), and herein lies our contribution to the

<sup>☆</sup> We would like to thank two anonymous referees for many helpful comments. Any remaining errors are, however, solely ours.

<sup>\*</sup> Corresponding author. Tel.: +27 12 420 3460.

E-mail addresses: [Rangan.Gupta@up.ac.za](mailto:Rangan.Gupta@up.ac.za) (R. Gupta), [mamphomodise@yahoo.com](mailto:mamphomodise@yahoo.com) (M.P. Modise).

literature. To the best of our knowledge, this is the first study using South African data that looks at not only in-sample, but also out-of-sample forecasting ability of stock returns of South Africa's major trading partners, besides valuation ratios (Campbell and Shiller, 1998), term spread (Campbell, 1987), short-term interest rate (Ang and Bekaert, 2007), and payout ratio (Lamont, 1998). Since we are using quite a number of predictors, we avoid of data mining problems by computing appropriate critical values using a bootstrap procedure. Further, given that predictive regressions are essentially a bivariate approach, where the predictability of each of the potential predictors are tested individually, we use general-to-specific model selection in order to choose the best in-sample forecasting model, where we start with a model that includes all the financial variables. This approach allows us to incorporate information simultaneously from (possibly) multiple predictors, without suffering from the degrees of freedom problem. Thus, in essence, the predictive regression framework based on the general-to-specific approach could encompass the bivariate predictive regression model, if in case multiple predictors are chosen in the best forecasting model.

Following the extant literature, our stock price predictions are based on a predictive regression model, which essentially amounts to regressing the growth rate of real stock price (over various horizons) on a variable thought to be capable of explaining the future path of stock prices. Note that the predictive regression framework, despite its limitations discussed below in Section 2, continue to be the most widely used econometric model in examining stock return predictability. Recent innovations involving non-linearity, time-varying parameters, latent factors and Bayesian priors, amongst others, have recently been incorporated into the framework as well.<sup>1</sup> Based on data availability, our in-sample period covers the period from 1990:01 to 1996:12, while our out-of-sample period begins from 1997:01 to 2010:04. Note, the choice of the out-of-sample period is aimed to cover the effects of the East Asian crisis, the move to an inflation-targeting regime, the currency crisis in late 2001 and the recent financial turmoil. We assess in-sample predictability via the  $t$ -statistic corresponding to the slope coefficient in a predictive regression model. In order to test for out-of-sample predictability, we compare out-of-sample forecasts generated by a model of constant returns to forecasts generated by a model that utilizes a given financial variable using two recently developed powerful test statistics by Clark and McCracken (2001) and McCracken (2004). In addition, following the argument by Inoue and Kilian (2002) that both in-sample and out-of-sample tests are subject to potential data mining problems, we address issues of possible data mining by computing appropriate critical values for all the test statistics using data mining-robust bootstrap procedure. Finally, following Clark (2002) and Rapach et al. (2005), we first use general-to-specific model selection approach in order to choose the best forecasting model based on in-sample data, where we start with a model that includes all the variables. Using a recursive approach, all the variables that have insignificant  $t$ -statistics (less than 1.654) are excluded from the final model, as a result, the general-to-specific model that we use will only contain those variables that have significant  $t$ -statistics. The selected model, in turn, is used to compute forecasts over the out-of-sample period, again based on the Clark and McCracken (2001) and McCracken (2004) test statistics. As before, to guard against overfitting, we base our inferences on a data mining-robust bootstrap procedure.

Our results show that most of the financial variables in the vast literature show no in-sample predictive power on South Africa's stock returns. Only the stock returns for our major trading partners have relatively strong predictive power on stock returns at longer horizons. For the out-of-sample period only two extra financial variables show some predictive ability. The Treasury bill rate shows predictive

ability from three-months-ahead horizon, while the term spread has relatively weak predictive ability and it's only at a one-month-ahead horizon. Accounting for data mining, only the in-sample test remains significant at all horizons, while for the out-of-sample (from six-months-ahead horizon) both the MSE-F and the ENC-NEW test statistics lack predictive power. On the other hand, the model that combines general-to-specific model selection with out-of-sample test statistics shows interesting results. In all the horizons, at least one valuation ratio is included in the model specification. This may suggest that valuation ratios contain important information about stock return behaviour in South Africa, despite our earlier results showing no predictive ability in both in-sample and out-of-sample periods. Further, the model also tends to indicate predictability at medium to long-term horizons, even after accounting for data mining. The rest of the paper is organised as follows: Section 2 discusses the econometric; Section 3 outlines the data and the results obtained from the models; and Section 4 summarises our main findings and concludes.

## 2. Econometric methodology

### 2.1. In-sample predictability

Following extant literature, including Rapach and Wohar (2006a) and Campbell and Shiller (1998), amongst others, we used a predictive regression model to analyse the behaviour of the stock return in long horizon. The predictive regression takes the form,

$$y_{t+k} = \alpha + \beta \cdot x_t + \gamma \cdot y_t + \mu_{t+k} \quad (1)$$

where  $y_t$  is the real stock return to holding to holding stock from period  $t - 1$ ,  $y_{t+k}$  is the log real return to holding stock from period  $t$  to  $t + k$ ,  $x_t$  represents the fundamentals used in predicting future real stock returns and  $\mu_{t+k}$  is the error term. When  $\beta = 0$  then the variable  $x_t$  has no predictive power for future stock return (null hypothesis), while under the alternative hypothesis,  $x_t$  does have predictive power for future returns ( $\beta \neq 0$ ). Suppose we have observations for  $y_t$  and  $x_t$  for  $t = 1, \dots, T$ . This leaves us with  $T - k$  usable observations with which to estimate the in-sample predictive regression model. The predictive ability of  $x_t$  is typically assessed by examining the  $t$ -statistic corresponding to  $\beta$ , the OLS estimate of  $\beta$  in Eq. (1), together with the goodness-of-fit measure,  $R^2$ . We also normalise each of the predictors ( $x_t$ ) by its standard deviation to make it easier to compare the estimated  $\beta$  in the predictive regression, Eq. (1). This normalisation, however, has no effect on the in-sample and out-of-sample statistical inferences. Note that, the efficient markets hypothesis argues that the best predictor of the next period's stock price is the current stock price, since it contains all the information in the market. Thus, the rate of return on stocks should correspond to a white noise error term. So tests for in-sample and out-of-sample predictability based on other predictors using the predictive regression framework, allows us to search for violations of the efficient markets hypothesis.<sup>2</sup>

Although Eq. (1) is widely used, it poses potential problems when estimating future stock returns. The first problem is small-sample bias, as  $x_t$  is not an exogenous regressor in Eq. (1). Rapach and Wohar (2006a,b) show a case when  $k = 1$  to illustrate the biasness in  $\beta$ . Another potential problem emerges when  $k > 1$  in the predictive regression model, Eq. (1). The observations for the regression in Eq. (1) are overlapping when  $k > 1$  and thus induce serial correlation in the error term,  $\mu_{t+k}$ . To account for this, we use Newey and West (1987) standard errors, as these account for serial correlation

<sup>1</sup> The reader is referred to Rapach and Zhou (forthcoming) for an extensive survey in this regard.

<sup>2</sup> We also estimated and forecasted with a model of real stock returns regressed on just a constant, to capture a scenario of random walk with drift for real stock prices. Understandably, the forecasts performed poorly relative to the AR(1) model for stock returns, hence validating the choice of the AR(1) model as our benchmark. These results are available upon request from the authors.

Download English Version:

<https://daneshyari.com/en/article/5055050>

Download Persian Version:

<https://daneshyari.com/article/5055050>

[Daneshyari.com](https://daneshyari.com)