Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Empirical Finance 13 (2006) 231–247
Journal of EMPIRICAL FINANCE
www.elsevier.com/locate/jempfin

In-sample vs. out-of-sample tests of stock return predictability in the context of data mining

David E. Rapach [a,1], Mark E. Wohar [b,*]

[a] Department of Economics, Saint Louis University, 3674 Lindell Boulevard, Saint Louis, MO 63108-3397, United States
[b] Department of Economics, University of Nebraska at Omaha, RH-512K, Omaha, NE 68182-0286, United States

Accepted 19 August 2005
Available online 20 December 2005

Abstract

We undertake an extensive analysis of in-sample and out-of-sample tests of stock return predictability in an effort to better understand the nature of the empirical evidence on return predictability. We find that a number of financial variables appearing in the literature display both in-sample and out-of-sample predictive ability with respect to stock returns in annual data covering most of the twentieth century. In contrast to the extant literature, we demonstrate that there is little discrepancy between in-sample and out-of-sample test results once we employ out-of-sample tests with good power. While conventional wisdom holds that out-of-sample tests help guard against data mining, Inoue and Kilian [Inoue, A., Kilian, L., 2004. In-sample or out-of-sample tests of predictability: which one should we use? Econometric Reviews 23, 371–402.] recently argue that in-sample and out-of-sample tests are equally susceptible to data mining biases. Using a bootstrap procedure that explicitly accounts for data mining, we still find that certain financial variables display significant in-sample and out-of-sample predictive ability with respect to stock returns.

JEL classification: C22; C52; C53; G12
Keywords: Stock return predictability; Nested models; In-sample tests; Out-of-sample tests; Data mining

1. Introduction

There now exists a voluminous literature on the predictability of stock returns from past information. Interestingly, researchers have identified a large number of financial variables that

* Corresponding author. Tel.: +1 402 554 3712; fax: +1 402 554 2853.
E-mail addresses: rapachde@slu.edu (D.E. Rapach), mwohar@mail.unomaha.edu (M.E. Wohar).
[1] Tel.: +1 314 977 3601.

0927-5398/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.jempfin.2005.08.001

appear to predict future stock returns. These include the dividend-price ratio (Rozeff, 1984; Campbell and Shiller, 1988a; Fama and French, 1988; Hodrick, 1992), price-earnings ratio (Campbell and Shiller, 1988b, 1998), book-to-market ratio (Kothari and Shanken, 1997; Pontiff and Schall, 1998), market value-to-net worth ratio or "Fed q" (Smithers and Wright, 2000; Robertson and Wright, 2002), dividend-payout ratio (Lamont, 1998), term and default spreads on bonds (Campbell, 1987; Fama and French, 1989), short-term interest rate (Campbell, 1987; Hodrick, 1992; Ang and Bekaert, 2001), equity share in total new equity and debt issues (Baker and Wurgler, 2000), and consumption-wealth ratio (Lettau and Ludvigson, 2001).[2] The evidence for the predictability of stock returns comes primarily from in-sample predictive regression models. While there are important econometric difficulties relating to the lack of exogenous regressors and overlapping observations in predictive regression models (Mankiw and Shapiro, 1986; Stambaugh, 1986, 1999; Richardson and Stock, 1989; Nelson and Kim, 1993), Campbell (2000, p. 1523) nevertheless concludes, "Despite these difficulties, the evidence for predictability survives at reasonable if not overwhelming levels of statistical significance. Most financial economists appear to have accepted that aggregate returns do contain an important predictable component."

While Campbell (2000) is probably correct in his assessment, as noted above, the extant literature primarily relies on in-sample tests of stock return predictability. This raises concerns of data mining, also referred to as model overfitting or data snooping. Lo and MacKinlay (1990) and Foster et al. (1997) provide theoretical analyses of data mining in the context of return predictability. It is typically believed that out-of-sample tests provide a measure of protection against data mining, as statistical models are tested using out-of-sample observations that are not used in the estimation of the statistical model itself. It is interesting to note that the relatively few studies that employ out-of-sample tests of return predictability typically obtain negative results. For example, in an effort to guard against model overfitting, Bossaerts and Hillion (1999) use different model selection criteria to choose the best forecasting model of real stock returns for a number of industrialized countries over the postwar period. Testing for out-of-sample forecasting power by regressing actual returns on the forecasts from the best models, they find that the best forecasting models for the United States fail to have significant out-of-sample forecasting power for S&P 500 excess returns at the 1-month horizon over the 1990:06–1995:05 out-of-sample period. They thus conclude that there is no external validation of the best forecasting models. Goyal and Welch (2003) also employ out-of-sample tests. They examine the predictive ability of the dividend-price ratio for CRSP value-weighted annual excess returns over the 1926–2000 period. While they find evidence of in-sample predictability, a model that includes the dividend-price ratio exhibits little out-of-sample predictive ability compared to a model of constant returns according to the Diebold and Mariano (1995) and West (1996) statistic.[3] The negative results typically generated by out-of-sample tests suggest that the in-sample evidence of return predictability is spurious.

The disparities between in-sample and out-of-sample test results of return predictability in the literature make an overall assessment of return predictability difficult. In this paper, we undertake an extensive analysis of both in-sample and out-of-sample tests of stock return predictability in an effort to better understand the empirical evidence on return predictability. We

[2] The studies cited are representative and do not constitute an exhaustive list.

[3] Goyal and Welch (2003) use the standard normal as the limiting distribution for the Diebold and Mariano (1995) and West (1996) statistic. However, McCracken (2004) shows that this statistic has a non-standard limiting distribution when comparing forecasts from nested models, as Goyal and Welch (2003) do.