Available online at www.sciencedirect.com
SCIENCE DIRECT®
Computational Statistics & Data Analysis 50 (2006) 1478 – 1495
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Stock and bond return predictability: the discrimination power of model selection criteria

Rosario Dell'Aquila[a,*], Elvezio Ronchetti[b]

[a] *Investment Research, Zürcher Kantonalbank, Postfach, CH-8010 Zürich, Switzerland and University of Lugano (Università della Svizzera italiana), Lugano, Switzerland*

[b] *Department of Econometrics, University of Geneva, Blv. Pont d'Arve 40, CH-1211 Geneva, Switzerland and University of Lugano (Università della Svizzera italiana), Lugano, Switzerland*

Received 8 December 2004; accepted 5 January 2005
Available online 26 January 2005

Abstract

The discrimination power of well-known model selection criteria is analyzed when the R-squared is low as in typical asset return predictability studies. It turns out that the discrimination power is low in this situation and this may explain, already in a simple i.i.d. setup, why often in-sample predictability, but no out-of-sample predictability is found. In particular it is possible to give another interpretation to the results of the well-cited Bossaerts and Hillion (Rev. Financial Stud. 12 (1999) 405–428) study. As a consequence, model selection criteria are put in a testing framework and a bootstrap-based procedure is proposed as a diagnostic tool to construct the class of models which are statistically indistinguishable from the best model chosen by a model selection criterion. In an empirical illustration the Pesaran and Timmerman (J. Finance 50 (1995) 1201–1228) results are reanalyzed and it turns out that in this case this class of models can be large. Finally it is shown that similar problems arise in a more hidden way in the context of recent model uncertainty studies using Bayesian model selection criteria.

Keywords: Model selection; Akaike; Schwarz; Stock return predictability; Forecasting; Asset pricing; Factor model; Risk model; Bootstrap; Bayesian model selection

* Corresponding author. Tel.: +41 01 2923404; fax: +41 01 2923212.
E-mail addresses: rosario.dellaquila@lu.unisi.ch (R. Dell'Aquila), elvezio.ronchetti@metri.unige.ch (E. Ronchetti).

0167-9473/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2005.01.001

1. Introduction

Recently, there has been growing evidence of stock and bond return predictability across different international stock and bond markets as well as over different time horizons; see for example Breen et al. (1990), Campbell (1987), Cochrane (1991), Brock et al. (1992), Fama and French (1989), Fama (1991) and Keim and Stambaugh (1986). While it is still open whether this may be due to time varying risk premia and/or market inefficiencies, the choice of the predictive variables is an empirical issue. In this context statistical model selection criteria such as the adjusted R^2, the Akaike criterion (Akaike, 1973), the Schwarz criterion (Schwarz, 1978), Mallow's C_p (Mallows, 1973) or Shao's cross-validation criterion (Shao, 1993) to cite only the most popular ones, are widely used and are very popular among practitioners. More recently Bayesian model selection techniques have been used as well.

However, in predictability studies the typical R^2 is very low, ranging from 1% to 10%. To our knowledge, no study has focused on the discrimination power of model selection criteria for low R^2 and on its consequences on asset return predictability studies. Notice that this is a different issue than testing for the null of no predictability when a preliminary search has been performed, see e.g. Foster et al. (1997) and White (2000).

The aim of this paper is to investigate the discrimination power of model selection criteria and try to assess their reliability. While the problems and solutions apply for model selection in every context, we focus our analysis on model selection of predictive models for asset returns, in particular stock returns, because in this case the R^2 is very low, the problems appear more clearly, and because of its relevance in finance. Similar issues arise when choosing a macroeconomic or fundamental factor model in an asset pricing or risk modelling context, where the R^2 is higher, but some of the variables have a low signal-to-noise ratio, or when choosing the lag order in a time series context.

We would like to stress that we do not have a priori a skeptical view on return predictability, rather the aim is to understand the usefulness of currently used frequentist or Bayesian model selection procedures in the context of low R^2. This is of particular importance in practice, since when real money is 'at risk' one is interested to know whether the method works in the specific context (e.g. low R^2 environment) and to which extent the claimed results are reliable, before implementing a more complex and possibly less transparent method.

To discuss our points, we limit ourselves to examples of well-known academic studies, where statistical model selection criteria are *explicitly* used to search over the whole set of models spanned by p potential predictive variables.

Pesaran and Timmerman (1995, 2000) for example use common model selection criteria like Akaike's Information Criterion and Schwarz Bayesian Criterion to determine a predictive model in each period and argue, that there is learning in the marketplace and that predictor performance in strategies improves, when one switches models over time based on formal model selection criteria. They show that this fact could have been exploited successfully in investment strategies in the past. Bossaerts and Hillion (1999) apply several popular model selection criteria to select forecasting models for excess stock returns. They report in-sample evidence of predictability across different international stock markets, but limited out-of-sample forecasting power when using the model selected in the in-sample