Available online at www.sciencedirect.com
ScienceDirect
International Journal of Forecasting 23 (2007) 427–445
www.elsevier.com/locate/ijforecast

Judgemental bootstrapping of technical traders in the bond market

Roy Batchelor *, Tai Yeong Kwan

Cass Business School, City of London, United Kingdom

Abstract

In many domains the decisions of experts are inferior to the decisions of statistical models of experts. The aim of this paper is to test this proposition in the financial markets, where genuine expertise is hard to find and the drivers of success are unclear. We exploit a unique database containing the recommended trading positions of technical analysts following the German bond market, and questionnaires revealing the technical indicators they used. The analysts have only average directional forecasting ability, but make consistent profits through superior market timing. Ordered-response models describing their positions in each market, driven by a subset of the technical indicators they claim to use, make even more profits. Models based on pooled data from several markets do better still. However, the pattern of model based trades is different from, and more risky than, the pattern of analyst trades. So it cannot be claimed that the models mimic the judgement process, or that the outcomes clearly dominate those from expert judgement.

JEL classification: E47; G1

Keywords: judgemental bootstrapping; Technical analysis; Forecasting

1. Methodology

In 1954 Paul Meehl (1954, reprinted 1996) published an influential study reviewing 20 pieces of research that compared decisions made by human experts with decisions indicated by the fitted values of simple linear statistical models parameterised on the experts' decisions. The applications were in fields as disparate as the diagnosis of schizophrenia, the probability of released prisoners reoffending, and the academic attainments of college students. In every case, the statistical model performed as well as, and generally better than, the human judges, in spite of the fact that the amount of information available to the human judges was usually greater than the limited number of quantitative inputs available to the statistical model. The logic of this "judgemental bootstrapping" procedure is that in many contexts quite simple models can remove the noise and inconsistency from human decisions, and this more than compensates for the lower information content of the models. Meehl's work stimulated a fierce and negative reaction from medical practitioners, and a summary and

* Corresponding author. HSBC Professor of Banking and Finance, Sir John Cass Business School, City of London, 106 Bunhill Row, London EC1Y 8TZ, United Kingdom. Tel.: +44 207 040 8733.
E-mail address: r.a.batchelor@city.ac.uk (R. Batchelor).

0169-2070/$ - see front matter © 2007 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.ijforecast.2007.05.007

rebuttal of their arguments is found in Grove and Meehl (1996). Subsequent studies have reinforced Meehl's findings. The meta-analysis by Grove, Zald, Lebow, Snits, and Nelson (2000) found 136 studies, including several in business and finance-related areas such as bankruptcy prediction and credit rating by banks. Of these studies, 64 show the statistical approach of weighted linear prediction to be superior, 64 show approximately the same outcome from human and statistical approaches, and only 8 favour the human judges.

This paper aims to extend the domain of judgemental bootstrapping further, by investigating whether simple models of the trading recommendations of technical analysts in bond futures markets can outperform the analysts themselves. Technical analysts believe that patterns in the time series of prices can be used to identify profitable trading opportunities in financial markets, in contrast to fundamental analysts, who look at news about interest rates, inflation, company earnings and other economic variables. For determining short term intra-day and day-to-day trading positions, technical analysis is overwhelmingly the more common decision support system used by traders (Lui & Mole, 1998; Taylor & Allen, 1992).

Our exercise promises to be interesting for five reasons. First, there is little published evidence on the validity of financial market trading systems built on the performance of experts. The many books, journals and articles on expert systems in finance are mostly concerned with finding complex nonlinear models to determine "ideal" trading positions. Second, technical analysts use a combination of quantifiable indicators, subjective pattern recognition tools, and the occasional piece of economic and business news, to support their decisions. Only the quantifiable indicators can be readily used as inputs to a high frequency statistical model, so in this field the human judges have a substantial information advantage. Third, the criterion of success in the markets is profit rather than the percentage of correct outcomes, the typical metric used in the studies surveyed by Grove et al. (2000). Because of the non-normal distribution of price movements in financial markets, profitability and accuracy are only weakly related, as was demonstrated in the study of interest rate futures trading by Leitch and Tanner (1991). Fourth, trading involves making judgements over time against a changing environment, whereas most models of judgement have been developed from more static cross-sectional "case-based" data. Finally, and again in contrast to, say, corporate bankruptcy events and loan defaults, price movements in financial markets do not have clear-cut drivers. On the contrary, a key prediction of the mainstream modern theory of finance is that futures price changes are made near-random by the actions of profit-motivated traders. Timmerman and Granger (2004) give a balanced review of reasons why prices might nonetheless be forecastable even in an efficient market.

A major barrier to the development of expert financial market trading systems is the paucity of objectively verifiable experts (as opposed to self-proclaimed experts, of whom there are many). Section 2 of this paper introduces the German bond futures markets, and the track records of two analysts who followed these markets in the years 2000–01. We conduct tests suggesting that these analysts show genuine expertise. Section 3 reviews the methods of technical analysis, and the results of a survey designed to establish what technical indicators these particular analysts use. In Section 4, we relate the analysts' recommended trading positions to a subset of relevant indicators using the ordered-response model of Aitchison and Silvey (1957). In the clinical literature this procedure of modelling experts by simple quasi-linear models is termed the "statistical" or "actuarial" approach. We have instead followed current practice in calling our procedure "judgemental bootstrapping", the terminology of Dawes (1971), Dawes and Corrigan (1974) and Armstrong (2001a). Even then, there is the potential for confusion with statistical bootstrap inference methods, compounded here by the fact that we use a resampling methodology when testing for analyst expertise.

Armstrong (2001a) sets out some criteria for best practice in the implementation of conventional judgemental bootstrap exercises. Benchmarking our data against these criteria may help us to clarify its likely strengths and limitations. The criteria relate to the choice of cases, experts and driver variables, and, at a more fundamental level, the choice of problem. The number of cases should ideally be large and disparate. With over 700 trading recommendations in very different trading environments (rising/stable/falling markets), our data satisfy this criterion. The number of experts should be "more than one", and experts should differ and have demonstrable expertise. We have only two experts, both from the same