Available online at www.sciencedirect.com
ScienceDirect
Journal of Financial Markets 15 (2012) 1–28
Journal of FINANCIAL MARKETS
www.elsevier.com/locate/finmar

The information content of a limit order book: The case of an FX market ☆

Roman Kozhan*, Mark Salmon

Warwick Business School, University of Warwick, Scarman Road, Coventry, CV4 7AL, UK

Available online 4 August 2011

Abstract

In this paper we examine the question of whether knowledge of the information contained in a limit order book helps to provide economic value in a simple trading scheme. Given the greater information content of the order book, over simple price information, it might naturally be expected that the order book would dominate. Using Dollar Sterling tick data, we find that despite the in-sample statistical significance of variables describing the structure of the limit order book in explaining tick-by-tick returns, they do not consistently add significant economic value out-of-sample. We show this using a simple linear model to determine trading activity, as well as a model-free genetic algorithm based on price, order flow, and order book information. We also find that the profitability of all trading rules based on genetic algorithms dropped substantially in 2008 compared to 2003 data.

JEL classification: D81; F31; C53

Keywords: Profitability; Limit order book; High-frequency data; Algorithmic trading

1. Introduction

One important issue in recent market microstructure research has been whether knowledge of the structure of the limit order book is informative regarding future price movements.

☆We are grateful to an anonymous referee and Bruce Lehmann (the editor), whose insightful comments greatly improved the paper. We also thank Martin Evans, Matthijs Fleischer, Thierry Foucault, Lawrence Harris, Rich Lyons, Chris Neely, Carol Osler, Richard Payne, Dagfinn Rime, Lucio Sarno, Nick Webber and Paul Weller for their helpful comments and suggestions. We remain responsible for all errors.
*Corresponding author. Tel.: +44 24 7652 4118.
E-mail addresses: Roman.Kozhan@wbs.ac.uk (R. Kozhan), Mark.Salmon@wbs.ac.uk (M. Salmon).

1386-4181/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.finmar.2011.07.002

There is a growing body of theoretical work suggesting that limit orders imply the predictability of short-term asset returns (see Handa and Schwartz, 1996, 2003; Harris, 1998; Parlour, 1998; Foucault, 1999; Rosu, 2010 among others). This is in contrast with earlier papers that implied that informed traders would only use market orders (see Glosten, 1994; Rock, 1996; Seppi, 1997). This debate has also been carried out empirically by Harris and Hasbrouck (1996), Kavajecz (1999), Harris and Panchapagesan (2005), Cao, Hansch, and Wang (2009), and Hellström and Simonsen (2009), all of whom demonstrated that asset returns can be explained by limit order book information, such as depth and order flow. However, these studies have failed to demonstrate that the predictability of returns can be exploited in economic terms. In this paper we go beyond statistical significance and consider the economic value of limit order book information in an FX market.

We address this question by explicitly constructing trading strategies based on full limit order book and price information in the FX market. These strategies only use historical information in order to ensure that trading can be implemented in "real time" and focus on the economic value of ex ante predictability in out-of-sample prediction exercises.

Such an approach requires the explicit specification of a forecasting model, which might be misspecified and so spuriously unable to exploit the available information. In order to address this issue, we construct trading strategies in a model-free way by employing a genetic algorithm. Genetic algorithms serve as a systematic search mechanism for the best trading rule from amongst a huge universe of potential rules given the particular information set and have been successfully applied in a number of financial applications, most notably by Dworman, Kimbrough, and Laing (1996), Chen and Yeh (1997a, 1997b), Neely, Weller, and Ditmar (1997), Allen and Karjalainen (1999), Neely and Weller (2001), Dempster and Jones (2001), Chen, Duffy, and Yeh (1999), Arifovic (1996). Rather than adopting a single specific forecasting model, the genetic algorithm searches from a very large set for that trading rule which exploits the information most profitably. We then test if this approach generates significantly higher returns when new information constructed from the limit order book is included alongside price information.

It is important to recognize the theoretical and practical coherence offered by using genetic algorithms. A number of authors, since Leitch and Tanner (1991), have argued that the use of purely statistical criteria to evaluate forecasts and trading strategies is inappropriate (e.g., Satchell and Timmermann, 1995; Granger and Pesaran, 2000; Pesaran and Skouras, 2002; Granger and Machina, 2006). The issue turns on the appropriate loss function and whereas many statistical evaluation criteria are based on a quadratic loss, practical criteria are more likely to be based on the utility derived from profits. Critically, from our point of view, the genetic algorithm constructs trading rules using the same loss function as is used to evaluate the out-of-sample performance of the trading strategy, unlike a linear regression model where a statistical quadratic loss is used in estimation.

Another important factor that needs to be considered when testing the profitability of trading strategies in "real time" is transaction costs.[1] We analyze the performance of our trading rules on the basis of the best bid and ask prices using tick-by-tick data and so explicitly take into account transaction costs as measured by the bid–ask spread. This allows us to test if predictable components in exchange rate returns are economically exploitable net of transaction costs.

[1] Neely and Weller (2003) for instance emphasize the critical role of transaction costs and inconsistences between the data used by practitioners and in academic simulations.