7 Econometric Analysis of High Frequency Data

Helmut Herwartz[1]

[1]Institut für Statistik und Ökonometrie, Christian Albrechts–Universität zu Kiel
herwartz@stat-econ.uni-kiel.de

Summary: Owing to enormous advances in data acquisition and processing technology the study of high (or ultra) frequency data has become an important area of econometrics. At least three avenues of econometric methods have been followed to analyze high frequency financial data: Models in tick time ignoring the time dimension of sampling, duration models specifying the time span between transactions and, finally, fixed time interval techniques. Starting from the strong assumption that quotes are irregularly generated from an underlying exogeneous arrival process, fixed interval models promise feasibility of familiar time series techniques. Moreover, fixed interval analysis is a natural means to investigate multivariate dynamics. In particular, models of price discovery are implemented in this venue of high frequency econometrics. Recently, a sound statistical theory of 'realized volatility' has been developed. In this framework high frequency log price changes are seen as a means to observe volatility at some lower frequency.

7.1 Introduction

With the enormous advances in computer technology, data acquisition, storage and processing has become feasible at higher and higher frequencies. In the extreme case of ultra high frequency financial data the analyst has access to numerous characteristics, called marks, of each transaction (price and quantity traded, corresponding bid and ask quotes etc.) and to the time of its occurrence, measured in seconds. As a consequence, numerous financial market microstructure hypotheses undergo empirical tests based on ultra frequency data. Typical issues in this vein of microstructure analysis are, for instance, the informational content of traded volumes for (future) prices (Karpoff, 1987), the relation between prices and clustering of transactions (Easley and O'Hara, 1992), or the significance of bid ask spreads as a means to identify the presence of informed traders in the market (Admati and Pfleiderer, 1988). From an econometric perspective such hypotheses naturally require an analysis of the marks in tick time, and, eventually motivate a duration model. The

methodology for the analysis of marked point processes as well as durations has experienced substantial progress since the introduction of Autoregressive Conditional Duration (ACD) models by Engle and Russell (1998). For a recent overview the reader may consult Engle and Russell (2005).

Another area of market microstructure modeling is information diffusion across markets trading the same asset or close substitutes. Then, it is of interest if independent price discovery (Schreiber and Schwartz, 1986) takes place in some major market or, alternatively, if the efficient price is determined over a cross section of interacting exchanges. Following Harris *et al.* (1995) or Hasbrouck (1995) price discovery is investigated by means of vector error correction models (VECM) mostly after converting transaction time to fixed time intervals of 1, 10 or 30 minutes, say. Although the latter conversion goes at the cost of loosing information on the trading intensity, it appears inevitable since the price quotations of interest are collected as a vector valued variable. Owing to irregular time spacing of quotes the statistical analysis of fixed interval data has to cope with methodological issues arising from the incidence of missing values. A condensed review over econometric approaches to model price discovery will be given in Section 7.2.

Apart from market microstructure modeling high frequency data have recently attracted large interest in econometrics as a means to estimate conditional volatility of asset prices at lower frequencies (Anderson *et al.*, henceforth, ABDL, 2001, 2003). Owing to its consistency for the process of conditional variances this estimator has particular appeal since it makes the latent volatility observable in the limit. A sound statistical theory on 'realized volatility' is now available making it a strong competitor to parametric approaches to modeling time varying second order moments. Section 7.3 will provide theoretical and empirical features of 'realized volatility'.

7.2 Price Discovery

A particular issue in empirical finance is the analysis of dynamic relationships between markets trading simultaneously a given security. Since cross sectional price differentials cannot persist, it is of interest, if the involved market places contribute jointly to the fundamental value of the asset or if particular markets lead the other. The process of incorporating new information into the efficient price has become popular as price discovery (Schreiber and Schwartz, 1986).

Starting from stylized features of high frequency data this section will first discuss the scope of fixed interval techniques. Then, the VECM is motivated as a means to address price discovery and a few empirical results are provided. A formal measure of a particular market's contribution to price discovery is derived along with a brief formalization of the VECM. Finally, parameter estimation in case of missing values is discussed.