Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Stochastic Processes and their Applications 119 (2009) 2249–2276
stochastic processes and their applications
www.elsevier.com/locate/spa

Microstructure noise in the continuous case: The pre-averaging approach ☆

Jean Jacod[a,*], Yingying Li[b], Per A. Mykland[b], Mark Podolskij[c], Mathias Vetter[d]

[a] *Institut de Mathématiques de Jussieu, CNRS UMR 7586 and Université Pierre et Marie Curie, 175 rue du Chevaleret, 75013 Paris, France*
[b] *Department of Statistics, The University of Chicago, Chicago, IL 60637, USA*
[c] *CREATES, School of Economics and Management, University of Aarhus, Building 1322, DK-8000 Aarhus C, Danmark, Germany*
[d] *Ruhr-Universität Bochum, Fakultät für Mathematik, 44780 Bochum, Germany*

Received 14 December 2007; received in revised form 8 August 2008; accepted 3 November 2008
Available online 24 November 2008

Abstract

This paper presents a generalized pre-averaging approach for estimating the integrated volatility, in the presence of noise. This approach also provides consistent estimators of other powers of volatility — in particular, it gives feasible ways to consistently estimate the asymptotic variance of the estimator of the integrated volatility. We show that our approach, which possesses an intuitive transparency, can generate rate optimal estimators (with convergence rate $n^{-1/4}$).

MSC: primary 60G44; 62M09; 62M10; secondary 60G42; 62G20

Keywords: Consistency; Continuity; Discrete observation; Itô process; Leverage effect; Pre-averaging; Quarticity; Realized volatility; Stable convergence

---

☆ Financial support from the Stevanovich Center for Financial Mathematics at the University of Chicago, from the National Science Foundation under grants DMS 06-04758 and SES 06-31605, from CREATES funded by the Danish Research Foundation, and from Deutsche Forschungsgemeinschaft through SFB 475, is gratefully acknowledged.
* Corresponding author.
E-mail addresses: jean.jacod@upmc.fr (J. Jacod), yyli@uchicago.edu (Y. Li), mykland@galton.uchicago.edu (P.A. Mykland), mark.podolskij@math.ethz.ch (M. Podolskij), mathias.vetter@rub.de (M. Vetter).

0304-4149/$ - see front matter
doi:10.1016/j.spa.2008.11.004

1. Introduction

The recent years have seen a revolution in the statistics of high frequency data. On the one hand, such data is increasingly available and needs to be analyzed. This is particularly the case for market prices of stocks, currencies, and other financial instruments. On the other hand, the technology for the analysis of such data has grown rapidly. The emblematic problem is the question of how to estimate daily volatility for financial prices (in stochastic process terms, the quadratic variation of log prices).

The early theory was developed in the context of stochastic calculus, before the financial application was apparent. The sum of squared increments of the process was shown to be consistent for the quadratic variation in [1]. A limit theory was then developed in [2,3], and later in [4].

Meanwhile, these concepts were introduced to econometrics in [5–7]. A limit theory was developed in [8,9]. Further early econometric literature includes, in particular, [10–17]. The setting of confidence intervals using bootstrapping has been considered by Goncalves and Meddahi [18] and Kalnina and Linton [19].

The direct application to data of results from stochastic calculus have, however, run into the problem of microstructure. No-arbitrage based characterizations of securities prices (as in [20]) suggest that these must normally be semimartingales. Econometric evidence, however, suggests that there is additional noise in the prices. This goes back to Roll [21] and Hasbrouck [22]. In the nonparametric setting, the deviation from semimartigales is most clearly seen through the signature plots of Andersen et al. [10], see also the discussion in [23].

Statistical and econometric research has for this reason gravitated towards the concept that the price (and log price) semimartingale is latent rather than observed. Research goes back to the work on rounding by Jacod [24] and Delattre and Jacod [25]. Additive noise is studied in [26], and a consistent estimator in the nonparametric setting is found in [27]. Issues of bias-variance tradeoff are discussed in [28]. In the nonparametric case, rate optimal estimators are found in [29–31]. A development for low frequency data is given in [32].

There are currently three main approaches to estimation in the nonparametric case: linear combination of realized volatilities obtained by *subsampling* [27,29], and linear combination of *autocovariances* [31]. The purpose of this paper is to give more insight to the third approach of *pre-averaging*, which was introduced in Podolskij and Vetter [30] for i.i.d. noise and for non overlapping intervals. The idea is as follows. We suppose that the (say) log securities price X_t is a continuous semimartingale (of the form (2.1)). The observations are recorded prices at transaction times $t_i = i\Delta_n$, and what is observed is not X_{t_i}, but rather Z_{t_i}, given by

$$Z_{t_i} = X_{t_i} + \epsilon_{t_i}. \tag{1.1}$$

The noise ϵ_{t_i} can be independent of the X process, or have a more complex structure, involving for example some rounding. The idea is now that if one averages K of these Z_{t_i}'s, one is closer to the latent process. Define $\breve{Z}_{t_i}$ as the average of $Z_{t_{i+j}}$, $j = 0, \dots, K-1$. The variance of the noise in $\breve{Z}_i$ is now reduced by a factor of about $1/K$. If one calculates the realized volatility on the basis of $\breve{Z}_0, \breve{Z}_{t_1}, \breve{Z}_{t_2}, \dots$, the estimate is therefore closer to being based on the true underlying semimartingale. The scheme is particularly appealing since it is obviously robust to a wide variety of structures of the noise ϵ.

The paper provides a way of implementing this idea. There are several issues that have to be tackled in the process. First of all, the pre-averaging brings in a particular dependence structure