Statistica Sinica **35** (2025), 1-19
doi:https://doi.org/10.5705/ss.202023.0020

VOLATILITY ANALYSIS WITH HIGH-FREQUENCY AND LOW-FREQUENCY HISTORICAL DATA, AND OPTIONS-IMPLIED INFORMATION

Huiling Yuan[1], Kexin Lu*[2] and Guodong Li[2]

[1]*East China Normal University and* [2]*The University of Hong Kong*

Abstract: In order to better forecast the volatility of underlying securities, this paper proposes a new model to integrate three important information sources, high-frequency and low-frequency historical observations, and options data, and the extracted options-implied information, rather than the option prices themselves, is used as an exogenous variable for simplicity. The quasi-maximum likelihood estimation is considered, and its asymptotic properties, as well as a hypothesis test, are further established. The proposed model is applied to the S&P 500 index as the underlying security, and the implied volatility index (VIX) is chosen as the options-implied information. Its superior performance can be observed in terms of out-of-sample forecast against many existing models.

Key words and phrases: Exogenous variable, prediction performance, quasi-maximum likelihood estimation, simple volatility model.

1. Introduction

Volatility analysis has significant importance in modern finance and economics. A natural information source of the volatility is the historical data of a security, and they can be further classified into high- and low-frequency ones. Specifically, high-frequency data pertains to intra-day prices, such as tick-by-tick, 1-second and 5-minute data. Scholars commonly employ continuous-time Itô processes for high-frequency data, and the corresponding estimation can also be established. Examples include two-time scale realized volatility (TSRV) in Zhang, Mykland and Aït-Sahalia (2005), multi-scale realized volatility (MSRV) in Zhang (2006), kernel realized volatility (KRV) in Barndorff-Nielsen et al. (2009), pre-averaging realized volatility (PRV) in Jacod et al. (2009) and quasi-maximum likelihood estimation (QMLE) in Xiu (2010). On the other hand, low-frequency data refer to observed prices of a security over daily or longer time intervals, and many volatility models have been developed and evaluated for them; see, e.g., the autoregressive conditional heteroskedasticity (ARCH) in Engle (1982) and the generalized autoregressive conditional heteroskedasticity (GARCH) in Bollerslev (1986).

*Corresponding author. E-mail: neithen@connect.hku.hk