Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Infinite-order, long-memory heterogeneous autoregressive models

Eunju Hwang, Dong Wan Shin *

Institute of Mathematical Sciences, Ewha University, Republic of Korea
Department of Statistics, Ewha University, Seoul, Republic of Korea

ARTICLE INFO

Article history:
Received in revised form 31 July 2013
Accepted 13 August 2013
Available online xxxx

Keywords:
HAR-RV model
Least squares estimator
Asymptotic property
Prediction mean-squared error
Realized volatility

ABSTRACT

We develop an infinite-order extension of the HAR-RV model, denoted by HAR(∞). We show that the autocorrelation function of the model is algebraically decreasing and thus the model is a long-memory model if and only if the HAR coefficients decrease exponentially. For a finite sample, a prediction is made using coefficients estimated by ordinary least squares (OLS) fitting for a finite-order model, HAR(p), say. We show that the OLS estimator (OLSE) is consistent and asymptotically normal. The approximate one-step-ahead prediction mean-square error is derived. Analysis shows that the prediction error is mainly due to estimation of the HAR(p) coefficients rather than to errors made in approximating HAR(∞) by HAR(p). This result provides a theoretical justification for wide use of the HAR(3) model in predicting long-memory realized volatility. The theoretical result is confirmed by a finite-sample Monte Carlo experiment for a real data set.

1. Introduction

The volatility of financial data is an important component of financial markets in both practical applications and theoretical studies. The long-memory property of volatility, meaning that historical volatility has a persistent impact on future volatility, is important in investment decision-making. Thus, the time series model for financial data should reflect the long-memory property. While the FIGARCH and ARFIMA models can be used for long-memory empirical analysis, such fractional integration models lack a clear economic interpretation.

To describe the long memory of volatility, Corsi (2004, 2009) proposed an additive cascade model containing volatility components defined for different time periods, called the heterogeneous autoregressive model of realized volatility (HAR-RV), that has three heterogeneous volatility components. Inspired by the HARCH model of Müller et al. (1997) and Dacorogna et al. (1998), the HAR-RV (called HAR hereafter) model is consistent with the heterogeneous market hypothesis and with the asymmetric propagation of volatility between long- and short-time horizons, and has different volatility components generated by the actions of different types of market participant. Although the HAR model is not formally a long-memory model, long-memory behavior results from the sum of volatility components constructed for different time horizons. Some more long-memory aspects of realized volatility have been discussed by Raggi and Bordignon (2012) in terms of the Markov switching approach.

The HAR model has been successfully used in forecasting realized volatility. Ghysels et al. (2006) and Forsberg and Ghysels (2007) compared HAR with the MIDAS model. Andersen et al. (2007) used the HAR model for prediction of the volatility of stock prices, foreign exchange rates, and bond prices. Corsi et al. (2008) showed that non-Gaussianity and time-varying volatility of reduced-form RV models, such as ARFIMA and HAR, might be partly attributable to time variations in

* Corresponding author at: Department of Statistics, Ewha University, Seoul, Republic of Korea. Tel.: +82 2 3277 2614; fax: +82 2 3277 3607.
E-mail addresses: shindw@mm.ewha.ac.kr, shindw@ewha.ac.kr (D.W. Shin).

0167-9473/$ – see front matter © 2013 Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.csda.2013.08.009

the volatility of the RV estimator. McAleer and Medeiros (2008) proposed a model called heterogeneous autoregression with multiple-regime smooth transition as an extension of the HAR model. This model contains long memory and nonlinearity, and incorporates sign and size asymmetries. Motivated by the use of HAR models in practice, Craioveanu and Hillebrand (2009) provided a critical review of the advantages of HAR models for daily RV. Hillebrand and Medeiros (2010) considered log–linear and neural network HAR models of realized volatility. They also applied bagging, a data mining technique, to realized volatility. Tang and Chi (2010) addressed test methods and models for long memory and found that the HAR model showed better predictive ability than the ARFIMA-RV model. Other uses of HAR models include risk management with VaR measures (Clements et al., 2008), risk-return tradeoff (Bollerslev et al., 2009), serial correlation (Bianco et al., 2009), implied volatility (Buscha et al., 2011), and realized volatility errors (Asai et al., 2012).

In the HAR model of Corsi (2009), a hierarchical model is considered with three volatility components corresponding to time horizons of 1 day ($1d$), 1 week ($1w$), and 1 month ($1m$). The HAR(3)-RV time series representation of the proposed cascade model can be written in the form

$$RV_{t+1}^{(d)} = c + \beta^{(d)} RV_t^{(d)} + \beta^{(w)} RV_t^{(w)} + \beta^{(m)} RV_t^{(m)} + \omega_{t+1},$$

where $RV_t^{(d)}$ is the realized variance on day t, and $RV_t^{(w)}$ and $RV_t^{(m)}$ are moving averages given by

$$RV_t^{(w)} = \frac{1}{5}\left(RV_t^{(d)} + RV_{t-1}^{(d)} + \cdots + RV_{t-4}^{(d)}\right),$$
$$RV_t^{(m)} = \frac{1}{22}\left(RV_t^{(d)} + RV_{t-1}^{(d)} + \cdots + RV_{t-21}^{(d)}\right).$$

Therefore, the HAR(3) model can be expressed as an AR(22) model. By adding lags of $RV_t^{(d)}$ up to lag 21, the model captures the long-memory properties of RV in a parsimonious way, but is theoretically a short-memory model.

It is important to study the prediction performance of the HAR(3) model of Corsi (2009) when the data-generating process has long memory. For this purpose, we extend the HAR(3) model to long memory and develop an asymptotic estimation theory. As noted by Corsi (2009), the HAR(3) model has short memory because it is an AR(22) model. To obtain the long-memory property, an infinite-order model, HAR(∞), is proposed. We present stationarity conditions for the HAR(∞) model and give necessary and sufficient conditions for the long-memory property. The essential part of these conditions is that the HAR(∞) coefficients decay exponentially.

Poskitt (2007) and Baillie and Kapetanios (2009) suggested the use of high-order autoregressions to approximate long-memory processes. Poskitt (2007) considered long AR approximations for general fractionally integrated processes. Poskitt (2007) established convergence rates for AR estimates and gave a CLT for the coefficient estimates. Baillie and Kapetanios (2009) dealt with practical investigation of a time series with long-memory characteristics using a semi-parametric estimation of the long-memory parameter.

One approach for predictions of the HAR(∞) process is to estimate a HAR(p) model of order p that increases with the sample size. The finite-order approach was used by Ing and Wei (2003) for an infinite-order autoregressive process and by Kuersteiner (2005) for infinite-order vector autoregressions, but their analyses were not for long-memory processes. We establish consistency and limiting normality of the OLSE of the fitted HAR(p) model.

The remainder of the paper is organized as follows. In Section 2, we describe the long-memory HAR(∞) model and discuss its properties. In Section 3, asymptotic estimation theory is developed for HAR(p) fitting as an approximation for the HAR(∞) model. In Section 4, the large-sample mean-squared error for HAR(p) prediction is derived. In Section 5, a Monte Carlo experiment is conducted. Realized volatilities for log returns of the Korean stock price index (KOSPI) and the Korean Won–US Dollar exchange rate are analyzed in terms of HAR estimation and HAR prediction in Section 6. Section 7 contains conclusions. Proofs are given in Section 8.

2. HAR(∞) model and long-memory properties

We extend the HAR(3) model of Corsi (2009) to an infinite-order model given by

$$Y_t = \beta_0 + \beta_1 Y_{t,h_1} + \beta_2 Y_{t,h_2} + \cdots + \epsilon_t, \tag{1}$$

where $\{\beta_j : j = 0, 1, 2, \ldots\}$ is a sequence of real numbers tending to 0, $\{h_j : j = 1, 2, \ldots\}$ is a given sequence of positive integers increasing to ∞,

$$Y_{t,h_j} = \frac{1}{h_j}(Y_{t-1} + Y_{t-2} + \cdots + Y_{t-h_j}),$$

and $\{\epsilon_t\}$ is a sequence of i.i.d. random variables with mean zero and variance $E[\epsilon_t^2] = \sigma^2$. We denote this as the HAR(∞) model.

We discuss basic probabilistic properties of Y_t in (1). Note that Y_t is an AR(∞) process:

$$Y_t = \beta_0 + \sum_{j=1}^{\infty} \frac{\beta_j}{h_j}(Y_{t-1} + Y_{t-2} + \cdots + Y_{t-h_j}) + \epsilon_t = \beta_0 + \sum_{j=1}^{\infty} \alpha_j \left(\sum_{k=h_{j-1}+1}^{h_j} Y_{t-k} \right) + \epsilon_t = \beta_0 + \sum_{i=1}^{\infty} \phi_i Y_{t-i} + \epsilon_t$$