ARTICLE IN PRESS

Computational Statistics and Data Analysis ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

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

Extended stochastic volatility models incorporating realised measures

J.H. Venter *, P.J. de Jongh

Centre for Business Mathematics and Informatics, North-West University, Private Bag X6001, Potchefstroom 2520, South Africa

ARTICLE INFO

Article history:
Received 14 May 2010
Received in revised form 1 November 2012
Accepted 5 November 2012
Available online xxxx

Keywords:
Stochastic volatility
Realised volatility
Latent variables
Intraday price data
Combined volatility estimator

ABSTRACT

Extended stochastic volatility models are studied which use the daily returns as well as the volatility information in intraday price data summarised in terms of a number of realised measures. These extended models treat the logarithm of daily volatility as a latent process with autoregressive structure, relate to daily returns via their variance models and relate to the logarithms of the realised measures via linear models. Fitting such an extended stochastic volatility model automatically combines the realised measures and daily returns into an overall daily volatility estimator. This process is technically rather demanding: Kalman filter and efficient importance sampling approaches are used here. The extended models are illustrated empirically using both high and low trading rate data. Simulation studies are reported which confirm that the model delivers volatility estimates that have better mean squared error and bias performance than individual realised measures.

1. Introduction

The volatility of asset returns plays an important role in many areas of financial decision making including portfolio management, hedging and option pricing. This importance and widespread use makes volatility modelling a very active research area in financial time series analysis. To be specific this paper shall focus on volatility in the context of the daily returns of an asset price. Two approaches to modelling the daily volatility have been studied widely, namely GARCH type models and stochastic volatility (SV) models. In both these approaches the return volatility on a given day is taken to be expressed by the conditional variance of the return of that day. Thus the starting point of volatility modelling in both approaches is the specification of this conditional variance. Once this is done the model is fitted to the observed daily return series. Clearly an individual return observed on a given day can provide only limited information about the variance of that return. If volatility changes only slowly in time then the information provided by the returns around a given day may help to estimate the volatility on that day more accurately. However, this will not be the case during periods of high instability since then the information contained in the neighbouring returns will be less relevant to the volatility of a given day.

In the context of GARCH models Hansen et al. (2012) conclude that "A single return only offers a weak signal about the current level of volatility. The implication is that GARCH models are poorly suited for situations where volatility changes rapidly to a new level". They then go on to develop so-called "realized GARCH" models that incorporate additional information provided by high-frequency intraday price data, assuming this is available. Traditional SV models using only daily returns suffer from the same weakness as pointed out above. This motivates the need to extend them also to take advantage of additional volatility information provided by high-frequency price data. An important step in this direction was made in the paper of Takahashi et al. (2009) which originally motivated this paper and to which we return below.

* Corresponding author. Tel.: +27 18 2992551.
E-mail address: hennie.venter@nwu.ac.za (J.H. Venter).

0167-9473/$ – see front matter
doi:10.1016/j.csda.2012.11.005

In the context of high frequency intraday price data the first volatility estimator studied was the realised volatility or realised variance (RV), defined as the sum of the squared returns over short calendar time intervals during the trading day (see e.g. Andersen et al., 2003). Subsequently many different types of estimators within this context were studied. Examples include using trade time intervals rather than calendar time intervals (see e.g. Oomen, 2006), replacing squared returns over intervals by squared ranges (see e.g. Martens and van Dijk, 2007), replacing squared returns with power and bi-power variation and the so-called AC1 estimator (Barndorff-Nielsen and Shephard, 2004). These developments were stimulated by the need to reduce bias introduced by so-called market micro structure noise effects and to achieve more efficient estimation (see e.g. Barndorff-Nielsen et al., 2004, 2008; Hansen and Lunde, 2006; Zhang, 2006; Ait-Sahalia et al., 2011; Carrasco and Kotchoni, 2010). The class of daily volatility estimators based on intraday data will be referred to as realised measures (RMs). In the context of this paper they serve as summaries of the volatility information in intraday price data.

Returning to the paper of Takahashi et al. (2009), they introduce a bivariate model for daily returns and a given daily RM jointly in time. The observed daily returns are assumed to follow a standard SV model (Taylor, 1986) in which the true daily log-volatility is an AR(1) latent process and the logarithms of the observed daily RM are assumed to be linearly related to this latent log-volatility. This model is also treated in the more recent paper of Dobrev and Szerzen (2010). Their approach will be extended in several directions. In the first place the AR(1) process assumption is relaxed, in the second place multiple RMs are considered simultaneously and thirdly it is shown how the leverage effect can be handled using the leverage function recently introduced by Hansen et al. (2012) in the context of GARCH models. Among the results of fitting the extended model are that it allows one to estimate the biases of the RMs used and to calculate an overall daily volatility estimator that automatically takes the biases in the RMs into account. Combining a number of given RMs was studied by Patton (2009) and Patton and Sheppard (2009) who reported that this leads to better volatility estimation.

The remainder of the paper is organised as follows. The extended stochastic volatility model is formulated in Section 2 and some implications following from the model are discussed as well as our approach towards fitting the model to data via Kalman filtering and efficient importance sampling. This is different from the Markov Chain Monte Carlo method used in Takahashi et al. (2009) and Dobrev and Szerzen (2010) and required further development of the efficient importance sampling (EIS) method of Liesenfeld and Richard (2003). Section 3 describes the results of simulation based studies, aimed at confirming that the fitting of the extended SV model do indeed produce a daily combined volatility estimator (CVE) that is better than the individual RMs used in the process. Section 4 illustrates the methodology extensively using two empirical data sets, one with a high trading rate and the other with a low trading rate. The high rate data set is the intraday price data of the world's largest resources company BHP-Billiton (code BLT) listed on the London stock exchange. The low trading rate data set is the intraday price data of an exchange traded fund Satrix-40 (code STX) tracking the so-called TOP40 index on the Johannesburg stock exchange. Section 4 also contains a further contribution to the realised volatility literature in terms of the use of the sample variance of the logarithm of an RM to aid in the selection of the parameters to use for the RM. Section 5 closes with a summary and an indication of further research issues. Finally, Appendix A sets out some results on minimum relative mean square estimators and Appendix B gives the technical details on the model fitting procedures used here and the issues that had to be dealt with to make it possible.

2. Stochastic volatility models for daily returns incorporating realised measures

In this section, an extended form of the stochastic volatility model of Takahashi et al. (2009) is proposed and it is indicated how the biases and mean squared errors of the relevant RMs can be quantified using the model. The approach to fitting the model and several ways of deriving estimators and predictors of the daily volatilities from the model are also discussed.

2.1. An extended stochastic volatility model

An extended form of the stochastic volatility (SV) model for the daily return Y_t of a stock on trading day t may be written in the form

$$Y_t = \exp\left(\frac{1}{2}U_t\right)Z_t \quad \text{with } Z_t \text{ i.i.d. } N(0,1)\text{-distributed where}$$
$$U_t = \mu + \sum_{m=1}^{M}\phi_m(U_{t-m}-\mu) + \nu(Z_{t-1}) + \tau\eta_t$$
$$\text{with } \eta_t \text{ i.i.d. } N(0,1)\text{-distributed for } t = 1,\ldots,T \text{ independently of the } Z_t\text{'s.} \tag{2.1}$$

Here $\mathrm{Var}(Y_t|U_t) = \exp(U_t) = H_t$ is by definition the return volatility and $U_t = \log(H_t)$ will be referred to as the log-volatility on day t. H_t and U_t are not observed directly, i.e. they are latent variables. In case the leverage function $\nu(Z_{t-1}) = 0$ the U_t's are assumed to be an AR(M) process; in the existing literature on stochastic volatility models it is usually assumed that $M = 1$ but greater generality is allowed here, the need for which will be shown by the illustrations below. Two sets of innovations are present in the model, namely the Z_t's and the η_t's and it is often assumed that the innovations Z_{t-1} and η_t are correlated rather than independent in order to cater for leverage or news impact effects between daily return and