UNIVERSITY *of York*

This is a repository copy of *Nonparametric Estimation and Forecasting for Time-Varying Coefficient Realized Volatility Models*.

Nonparametric Estimation and Forecasting for Time-Varying Coefficient Realized Volatility Models

Xiangjin B. Chen, Jiti Gao, Degui Li & Param Silvapulle

Nonparametric Estimation and Forecasting for Time-Varying Coefficient Realized Volatility Models

Xiangjin B. Chen[1], Jiti Gao[1]*, Degui Li[2] and Param Silvapulle[1]

[1]*Monash University, Australia and* [2]*University of York, United Kingdom*

Abstract

This paper introduces a new specification for the heterogeneous autoregressive (HAR) model for the realized volatility of S&P 500 index returns. In this modelling framework, the coefficients of the HAR are allowed to be time-varying with unspecified functional forms. The local linear method with the cross-validation (CV) bandwidth selection is applied to estimate the time-varying coefficient HAR (TVC-HAR) model, and a bootstrap method is used to construct the point-wise confidence bands for the coefficient functions. Furthermore, the asymptotic distribution of the proposed local linear estimators of the TVC-HAR model is established under some mild conditions. The results of the simulation study show that the local linear estimator with CV bandwidth selection has favorable finite sample properties. The outcomes of the conditional predictive ability test indicate that the proposed nonparametric TVC-HAR model outperforms the parametric HAR and its extension to HAR with jumps and/or GARCH in terms of multi-step out-of-sample forecasting, in particular in the post-2003 crisis and 2007 GFC periods, during which financial market volatilities were unduly high.

Keywords: Bootstrap Method, Heterogeneous Autoregressive Model, Local Linear Estimation, Locally Stationary Process, Nonparametric Method, Time-Varying Coefficient Models.

JEL Classifications: C14, C22, C52, C58, G32

Correspondence to: Jiti Gao, Department of Econometrics and Business Statistics, Monash University, 26 Sir John Monash Drive, Caulfield East, Victoria 3145, Australia.. Email: Jiti.Gao@monash.edu.

1   Introduction

Financial return volatility is fundamental to portfolio diversification, pricing financial assets and derivatives, and risk management, among others. As a consequence, volatility modelling and forecasting have been two of the most researched topics in both theoretical developments and practical applications in financial econometrics. Engle (1982) and Bolleslev (1986) developed ARCH/GARCH models for deterministic volatility modelling of financial market, which led to various extensions and empirical applications in the last three decades. Recently, several authors studied the GARCH model with time-varying coefficients (TVC). For example, Frijns et al. (2011) introduced a TVC-GARCH model based on multinormial switching mechanism, while Polzehl and Spokoiny (2006) proposed an adaptive procedure to estimate the GARCH coefficients as a function of time. In addition, Heston (1993), Ruiz (1994), and Jacquier et al. (1994) developed stochastic volatility models in both the continuous and discrete time series frameworks. In these models, however, the volatility is assumed to be a latent factor and the daily volatility series are estimated largely from daily return series.

The seminal papers by Andersen and Bollerslev (1998) and Andersen et al. (2001), among others, introduced the nonparametric realized volatility (RV) measures constructed from intra-day transaction data, which retain most of the pertinent information for measuring, modelling, and forecasting RV over the daily and long horizons. Additionally, RV was shown to be an efficient and consistent estimator of the latent volatility of asset returns series. Evidently, the availability of nonparametric measures of latent volatility constructed from high frequency intraday data such as the RV has created a new innovative research direction on the modelling and forecasting of volatility in the recent literature on financial econometrics. The construction of observable RV series and the use of standard time-series techniques have led to promising approaches for modelling and forecasting daily return volatility.

The RV is known to possess the long memory property, and as a result several studies modelled

the RV as the ARFIMA process, which is known to be difficult for estimating and forecasting. Based on the *Heterogeneous Market Hypothesis* introduced by Müller et al. (1993), as an alternative to ARFIMA model, Corsi (2009) proposed the heterogeneous autoregressive (HAR) model, which is a relatively simple autoregressive specification and shown to capture the crucial long memory feature of the RV series. In spite of its simplicity, Corsi (2009) showed that the HAR model can successfully encapsulate the main empirical features of financial returns distribution's fat tails and thus has a remarkably better forecasting performance than its competitors. McAleer and Medeiros (2008) provided an excellent literature review and compared the performance of the HAR with that of the latent GARCH-type and stochastic volatility models. The authors further proposed a HAR with multiple-regime switching smooth transition model for RV, combining the HAR and the smooth transition autoregressive models. Liu and Maheu (2008) introduced a jump component as well as asymmetric components to HAR model. See also Raggi and Bordignon (2012) and Nonejad (2014) who proposed various extensions to HAR model for the log-transformed RV (logRV) series.

Corsi et al. (2008) showed that the residual series of the HAR model exhibits volatility clustering and thus proposed HAR-GARCH model to explicitly account for this property. They found that the HAR-GARCH model outperforms the HAR model in terms of out-of-sample point forecasting. Subsequently, the HAR with various jump components in modelling and forecasting the logRV series has received notable attention in the recent literature. Anderson et al. (2007) introduced an HAR-J model, in which the price process includes daily jumps associated with short-lived busts in realised volatility, and proposed an HAR-CJ model, in which logRV is modelled in terms of multi-period rough jumps and smooth continuous components. Some further advances can be found in Bollerslev et al. (2009) and Andersen et al. (2011). These studies raise the HAR model with jumps to the prominence by showing that the HAR models with jump components produce better RV forecasts than those without jump components. In this paper, we propose a flexible data-driven nonparametric approach and apply this to estimate the TVC-HAR model, and compare the results

with those of the aforementioned HAR-type models in terms of out-of-sample forecasting of daily, weekly and monthly logRV.

The principal contributions of the paper are the following: (i) it introduces an alternative specification for HAR model of S&P 500 index returns by allowing the parameters of the HAR model to be time-varying with unknown functional forms; (ii) it applies a local linear smoothing method to estimate the TVC-HAR model and uses a bootstrap method to construct the point-wise confidence bands for the coefficient functions; and (iii) it evaluates the multi-period out-of-sample forecasting performance of the proposed TVC-HAR model against the HAR, HAR-GARCH models with and without jumps. To our knowledge, this paper is among the first to generate nonparametric multi-step ahead forecasts for logRV.

In a preliminary analysis of the daily S&P 500 stock index returns over the period 1999-2010, we found that the coefficients of the simple HAR model across the sub-sample periods were significantly different. Without making any parametric assumption as to the way in which these coefficients vary over time, in this paper we allow the coefficients of the HAR models to vary over time with unknown functional forms. If these coefficients are assumed to be constants but indeed time-varying, then the model misspecification can give rise to time-varying volatility for HAR model, which is introduced and extensively studied by Corsi et al. (2008). Thus, the nonparametric TVC-HAR model that we propose in this paper is a natural competitor to the well-known HAR-GARCH model. To assess the out-of-sample performance of the model proposed in this paper relative to several parametric counterparts, we use the conditional predictive ability (CPA) test developed by Giocomoni and White (2006). We further employ a hypothesis testing method to find supporting evidence for its relative in-sample performance. In this respect, this paper makes methodological as well as empirical contributions to the growing literature on realized volatility.

Modelling the TVC is common in finance. The TVC models have been applied to several areas in finance, including the popular capital asset pricing model and the term structure of interest rates model (c.f., Cochrane, 2001; Tsay, 2002; and Cai, 2007). Despite many studies imposing

various parametric structures to TVC functions, in practice, the underlying true feature of the TVC function is largely unknown. However, the traditional approach to establish a TVC model relies on the assumption made on the parametric functional forms for the time-varying coefficients. Moreover, although it is simple to implement, the parametric structure imposed for the TVC functions is often too restricted and can be unrealistic, leading to model misspecification, and thus to inaccurate forecasting. By contrast, without imposing any specific structure, nonparametric estimation of the TVC models would be an attractive alternative, because it allows the data to "speak for themselves". There exists a vast literature on nonparametric estimation of the mean regression models as well as their econometric applications. For instance, Robinson (1989) contributed to the early development of the nonparametric mean estimation of TVC regression models with exogenous explanatory variables, while Cai (2007) and Li et al. (2011), among others, contributed to the recent literature on this topic. Meanwhile, there has also been increasing interest on nonparametric estimation of the dynamic time series models under the assumption of local stationarity (c.f., Kim, 2001; Giraitis et al., 2012; and Zhang and Wu, 2012). However, so far as we know, there is little literature on nonparametric estimation of the TVC realized volatility models with local stationarity assumption. This paper fills this gap by further deriving the asymptotic distribution theory of the proposed nonparametric methodology which is given in the appendix.

The rest of the paper is organized as follows. Section 2 introduces the TVC-HAR model specification for logRV, local linear estimation method with the cross-validation (CV) bandwidth selection and the construction of the point-wise confidence bands for the coefficient functions, and conducts a simulation study for assessing the finite sample performance of the proposed model and estimation approach. Section 3 provides a description of the parametric and nonparametric out-of-sample forecasting and evaluation. Section 4 describes the S&P 500 series and presents a preliminary analysis of the series, followed by the results of local linear estimation and CPA testing of the proposed nonparametric TVC-HAR model. Section 5 concludes this paper. An asymptotic theory for the proposed methodology with some regularity conditions is given in Appendix A.1

and the generalized likelihood ratio test for the simple HAR model against the TVC-HAR model is outlined in Appendix A.2.

2 Model and Methodology

In this section, we first specify the TVC-HAR modelling framework, and then introduce the methodologies including the local linear estimator and the bootstrap method to construct the point-wise confidence bands for the coefficient functions. Furthermore, we also give a simulation study and report the results.

2.1 Specification of the TVC-HAR Model

We start with Corsi (2009)'s simple HAR model, which is defined by

$$\log RV_t = \beta_0 + \beta_d \log RV_{t-1} + \beta_w \log RV_{t-5,t-1} + \beta_m \log RV_{t-22,t-1} + \varepsilon_t, \tag{1}$$

where $\log RV_{t+1-k,t} = k^{-1}(\log RV_{t-1} + \log RV_{t-2} + \cdots + \log RV_{t-k}) = \frac{1}{k}\sum_{j=1}^{k}(\log RV_{t-j})$ is the k-period normalized realized variation, β_0, β_d, β_w and β_m are the four unknown parameters which are assumed to be invariant over time. We next relax the restriction that the coefficients of the HAR model (1) are constant which might be unrealistic in practical application as mentioned in Section 1. Instead, we allow that the coefficients of the model can vary over the time with certain unknown functional forms, which leads to a flexible TVC-HAR model for logRV.

Before specifying the nonparametric TVC-HAR model, we introduce some notation through the following general TVC autoregressive time series model:

$$Y_t = X_t^\top \beta_t + \varepsilon_t, \quad t = 1, \cdots, n, \tag{2}$$

where $X_t = (1, X_{t,1}, X_{t,2}, \cdots, X_{t,d})^\top$ are the lagged values of the response variable Y_t with $X_{t,k} = Y_{t-k}$ for $k = 1, \cdots, d$, $\beta_t = (\beta_{t,0}, \beta_{t,1}, \cdots, \beta_{t,d})^\top$ is a vector of unknown functions of time t, and $\{\varepsilon_t\}$ is a

sequence of stationary errors. As in Robinson (1989), Cai (2007) and Li et al. (2011), we assume that

$$\beta_{t,j} = \beta_j(\tau_t),\, \tau_t = \frac{t}{n},\quad t = 1,\cdots,n,\quad j = 0,\cdots,d, \tag{3}$$

where n is the sample size. Such a specification of β_t makes it possible to construct a consistent nonparametric kernel regression estimation of the coefficient functions, via increasingly intense sampling of data points at each point in the interval $[0, 1]$.

Let $Y_t = \log RV_t$, $X_t = (1, \log RV_{t-1}, \log RV_{t-5,t-1}, \log RV_{t-22,t-1})^\top$ and $\beta_t = (\beta_{t,0}, \beta_{t,d}, \beta_{t,w}, \beta_{t,m})^\top$ in the model (2), with the coefficient functions replaced with those defined by (3). We then obtain the nonparametric TVC-HAR model for the logRV series through:

$$\log RV_t = \beta_0(\tau_t) + \beta_d(\tau_t) \log RV_{t-1} + \beta_w(\tau_t) \log RV_{t-5,t-1} + \beta_m(\tau_t) \log RV_{t-22,t-1} + \varepsilon_t. \tag{4}$$

The simple HAR model (1) can be seen as a special case of the TVC-HAR model by restricting the coefficient functions to be constants over the time. Furthermore, note that model (4) is different from those in Cai (2007) and Li et al. (2011), as there is a dynamic structure involved in model (4) and the commonly-assumed α-mixing condition may not be satisfied in our setting.

2.2 Local Linear Estimation Method

We next introduce the local linear method to estimate the coefficient functions $\beta_{t,j} = \beta_j(\tau_t)$ for $j = 0, d, w, m$. The local linear method has been employed in the nonparametric regression estimation in the literature, due to its attractive properties such as efficiency, bias reduction, and adaptation of boundary effects (c.f., Fan and Gijbels, 1996). Assuming each $\beta_j(\cdot)$ has a continuous second-order derivative in the interval $[0, 1]$, $\beta_j(\tau_t)$ can be approximated by a linear function through the first-order Taylor expansion:

$$\beta_j(\tau_t) \simeq a_j + b_j(\tau_t - \tau),$$

where $a_j = \beta_j(\tau)$, and $b_j = \beta'_j(\tau)$ which is the first-order derivative of $\beta_j(\tau)$. Thus, the model (4) can be locally approximated by

$$\log RV_t \simeq \widetilde{X}_t^\top \theta(\tau) + \varepsilon_t, \tag{5}$$

where $\widetilde{X}_t^\top = [X_t^\top, X_t^\top(\tau_t - \tau)]^\top$ and $\theta(\tau) = \{\beta^\top(\tau), [\beta'(\tau)]^\top\}^\top$. Note that the dimension for $\widetilde{X}_t$ and $\theta(\tau)$ in the above model is 8.

Based on the local approximation (5), we define the locally weighted sum of squares:

$$\sum_{t=1}^{n} [\log RV_t - \widetilde{X}_t^\top \theta]^2 K_h(\tau_t - \tau), \tag{6}$$

where $K_h(u) = \frac{1}{h} K(\frac{u}{h})$, $K(\cdot)$ is a kernel function, and $h = h_n > 0$ is the bandwidth satisfying the conditions that $h \to 0$ and $nh \to \infty$ as $n \to \infty$. Note that h controls the amount of smoothing used in the local linear estimation. Then, $\widehat{\theta}(\tau)$ can be obtained by minimizing the weighted loss function (6) with respect to θ. By some elementary calculation, we may show that the expression of $\widehat{\theta}(\tau)$ is given by

$$\widehat{\theta}(\tau) = \begin{bmatrix} S_{n0}(\tau) & S_{n1}^\top(\tau) \\ S_{n1}(\tau) & S_{n2}(\tau) \end{bmatrix}^{-1} \begin{bmatrix} M_{n0}(\tau) \\ M_{n1}(\tau) \end{bmatrix}, \tag{7}$$

where

$$S_{nk}(\tau) = \frac{1}{n} \sum_{t=1}^{n} X_t X_t^\top (\tau_t - \tau)^k K_h(\tau_t - \tau), \ k = 0, 1, 2,$$

and

$$M_{nk}(\tau) = \frac{1}{n} \sum_{t=1}^{n} X_t (\log RV_t)(\tau_t - \tau)^k K_h(\tau_t - \tau), \ k = 0, 1.$$

The local linear estimate of $\beta(\tau) = [\beta_0(\tau), \beta_d(\tau), \beta_w(\tau), \beta_m(\tau)]^\top$ is $\widehat{\beta}(\tau)$ which consists of the first 4 elements of $\widehat{\theta}(\tau)$, and the local linear estimate of the derivative $\beta'(\tau)$ is $\widehat{\beta'}(\tau)$ which consists of the last 4 elements of $\widehat{\theta}(\tau)$. Proposition 1 in the appendix shows that the local linear estimation $\widehat{\beta}(\tau)$ is asymptotically normally distributed under some regularity conditions.

As is well-known, the local linear estimator is sensitive to the choice of the bandwidth h, and thus it is critical to choose an appropriate bandwidth in empirical applications. In this paper, we apply the commonly-used CV bandwidth choice for the time series case (c.f., Härdle and Vieu, 1992). Let $\widehat{\beta}_{-t}(\tau_t)$ be the local linear estimated value of $\beta(\tau_t)$ by using the sample $\{\log RV_s : s \neq t\}$. The optimal bandwidth h_{opt} is chosen such that

$$\mathsf{CV}(h) = \frac{1}{n} \sum_{t=1}^{n} \left[\log RV_t - X_t^{\top} \widehat{\beta}_{-t}(\tau_t) \right]^2$$

is minimized. In Section 2.4 below, we conduct a simulation study to evaluate the finite sample properties of the local linear estimator of the TVC-HAR model with the CV bandwidth selection.

2.3 Construction of the Confidence Bands

For given $0 < \alpha < 1$, the $(1 - \alpha)$ point-wise confidence bands of $\beta_j(\cdot)$ at point $0 \leq \tau \leq 1$ can be defined by

$$\left[\widehat{\beta}_j(\tau) - c_{\alpha/2}(\tau) \times \mathsf{sd}(\widehat{\beta}_j(\tau)), \; \widehat{\beta}_j(\tau) + c_{\alpha/2}(\tau) \times \mathsf{sd}(\widehat{\beta}_j(\tau)) \right] \tag{8}$$

for $j = 0, d, w, m$, where $\widehat{\beta}_j(\tau)$ is the local linear estimate of $\beta_j(\tau)$, $c_{\alpha/2}(\tau)$ is the upper $\alpha/2$ percentile of $Q_j(\tau) = (\widehat{\beta}_j(\tau) - \beta_j(\tau))/\mathsf{sd}(\widehat{\beta}_j(\tau))$ and $\mathsf{sd}(\widehat{\beta}_j(\tau))$ is the standard deviation of $\widehat{\beta}_j(\tau)$. However, (8) cannot be directly used to construct the confidence bands of $\beta_j(\tau)$ as neither $c_{\alpha/2}(\tau)$ nor $\mathsf{sd}(\widehat{\beta}_j(\tau))$ is known. We next estimate $c_{\alpha/2}(\tau)$ and $\mathsf{sd}(\widehat{\beta}_j(\tau))$ by using the bootstrap procedure. Then, by using the estimates of $c_{\alpha/2}(\tau)$ and $\mathsf{sd}(\widehat{\beta}_j(\tau))$ which are denoted by $\widehat{c}^*_{\alpha/2}(\tau)$ and $\mathsf{sd}^*(\widehat{\beta}_j(\tau))$, respectively, the $(1 - \alpha)$ point-wise confidence bands of $\beta_j(\tau)$ can be obtained by

$$\left[\widehat{\beta}_j(\tau) - \widehat{c}^*_{\alpha/2}(\tau) \times \mathsf{sd}^*(\widehat{\beta}_j(\tau)), \; \widehat{\beta}_j(\tau) + \widehat{c}^*_{\alpha/2}(\tau) \times \mathsf{sd}^*(\widehat{\beta}_j(\tau)) \right].$$

The bootstrap procedure to estimate $c_{\alpha/2}(\tau)$ and $\mathsf{sd}(\widehat{\beta}_j(\tau))$ is described as follows.

1. Estimate β_t by the local linear smoothing method introduced in Section 2.2, and denote the resulting estimates by $\widehat{\beta}_t$ for $t = 1, \cdots, n$.

2. For each $t = 1, \cdots, n$, generate $\log RV_t^* = X_t^\top \widehat{\beta_t} + \varepsilon_t^*$, where $\{\varepsilon_t^*\}_{t=1}^n$ is sampled from the centred residuals $\{\widetilde{\varepsilon_t}\}_{t=1}^n$, with $\widetilde{\varepsilon_t} = \widehat{\varepsilon_t} - \overline{\widehat{\varepsilon_t}}$, $\widehat{\varepsilon_t} = \log RV_t - X_t^\top \widehat{\beta_t}$ and $\overline{\widehat{\varepsilon_t}} = \frac{1}{n} \sum_{t=1}^n \widehat{\varepsilon_t}$. In practice, $\varepsilon_t^* = \widetilde{\varepsilon_t} \cdot \eta_t$, where $\{\eta_t\}$ is a sequence of independent and identically distributed random variables drawn from a prespecified distribution with mean zero and unit variance, such as $N(0, 1)$. Use the data set $\{(\log RV_t^*, X_t) : t = 1, \ldots, n\}$ to estimate the coefficient functions and denote the resulting estimator as $\widehat{\beta}^*(\cdot)$.

3. Repeat Step 2 for B times and obtain B bootstrap local linear estimated values, $\widehat{\beta}^*(\tau, i)$, $i = 1, \cdots, B$. The estimate of $\mathsf{sd}(\widehat{\beta}_j(\tau))$ is the sample standard deviation of $\{\widehat{\beta}_j^*(\tau, i) : i = 1, \cdots, B\}$ and is denoted as $\mathsf{sd}^*(\widehat{\beta}_j(\tau))$.

4. For each $i = 1, \cdots, B$, use the sequence $\{\widehat{\beta}_j^*(\tau, i)\}$ and $\mathsf{sd}^*(\widehat{\beta}_j(\tau))$ to compute $Q_{j,i}^*(\tau) = (\widehat{\beta}_j^*(\tau, i) - \widehat{\beta}_j(\tau))/\mathsf{sd}^*(\widehat{\beta}_j(\tau))$, and then obtain the estimate of $c_{\alpha/2}(\tau)$, denoted by $\widehat{c}_{\alpha/2}^*(\tau)$, by calculating the upper $\alpha/2$ percentile of $\{Q_{j,i}^*(\tau) : i = 1, \cdots, B\}$.

As in Teräsvirta et al. (2010), we let the covariates Y_{t-k}, for $k = 1, 5, 22$ in the dynamic autoregression be invariant when generating bootstrap samples, although other bootstrap methods may also be applicable in our setting. In Section 4, the results of our empirical application show that the above bootstrap procedure works reasonably well. Additionally, we find that the point-wise confidence bands for the coefficient functions of TVC-HAR are sensitive to minor variations to the optimum bandwidth which produces the narrowest confidence bands. The simultaneous confidence bands for the coefficient functions proposed in Zhang and Peng (2010) for the case where the observations are independent does not appear to work well in our setting, where the data is weakly dependent and locally stationary, and they are generally much wider than the point-wise confidence bands. Therefore, in the empirical study presented in Section 4, we report only the point-wise confidence bands.

2.4 Simulation Study

In this section, we conduct a simulation study to evaluate the finite sample behavior of the local linear estimator with the CV bandwidth, which is introduced in Section 2.2, to estimate the coefficient functions of the TVC-HAR model. In this simulation study, we specify the TVC-HAR model as

$$Y_t = \beta_0(\tau_t) + \beta_d(\tau_t)Y_{t-1} + \beta_w(\tau_t)Y_{t-5,t-1} + \beta_m(\tau_t)Y_{t-22,t-1} + \varepsilon_t \tag{9}$$

where $\tau_t = \frac{t}{n}$, and $Y_{t-k,t-1} = k^{-1}(Y_{t-1} + Y_{t-2} + \cdots + Y_{t-k}) = \frac{1}{k}\sum_{j=1}^{k} Y_{t-j}$, $k = 1, 5, 22$ and $t = 1, \cdots, n$. The functional forms for $\beta_j(\cdot)$, $j = 0, d, w, m$, are generated as: $\beta_0(\tau) = 0.1 \times \sin(2\pi\tau)$, $\beta_d(\tau) = 0.1 \times \cos(2.5\pi\tau)$, $\beta_w(\tau) = 0.1 \times \cos(2\pi\tau)$, and $\beta_m(\tau) = -0.1 \times \sin(\pi\tau)$. We consider two data generating processes: DGP 1 and DGP 2. In DGP 1, ε_t in (9) is $N(0,1)$; and in DGP 2, ε_t is Student-t with 5 degrees of freedom. We set $n = 250, 500$ and 1000 sample sizes, and $M = 500$ replications. Moreover, to ensure that the initial values for the DGP do not affect the results, we generate $n + n^*$ observations, with $n^* = 1000$ being the burn-in sample size.

The simulation study is conducted in the following steps:

1. Set $n = 250$ and generate $\{\varepsilon_t\}$ from $N(0,1)$.

2. Generate the initial values of the elements of $\{Y_0, Y_{-1}, Y_{-2}, ..., Y_{-21}\}$.

3. Generate a $T \times 1$ series of Y_t recursively from DGP1: TVC-HAR with the known coefficient functions and the standard normal ε_t.

4. Then, construct the RHS dynamic variables $(1, Y_{t-1}, Y_{t-5,t-1}, Y_{t-22,t-1})$ as defined above.

5. Discard the first n^* observations of Y_t and X_t, where $X_t = (1, Y_{t-1}, Y_{t-5,t-1}, Y_{t-22,t-1})$. Use the remaining n observations to estimate the four unknown coefficient functions $\beta(.)_j$ by the local linear method with the CV bandwidth, and denote the resulting estimates by $\widehat{\beta}_j(.)$ for $j = 0, d, w, m$.

6. Repeat the above Steps 3 to 5 M times.

7. Compute the bias and the standard deviation (std) for $\phi = [\beta_0, \beta_d, \beta_w, \beta_m]^\top$, as follows:

$$\text{bias}_j = \frac{1}{n} \sum_{t=1}^{n} (\overline{\widehat{\beta}}_j(\tau_t) - \beta_j(\tau_t)) \ \text{ and } \ \text{std}_j = \sqrt{\frac{1}{M}\frac{1}{n} \sum_{i=1}^{M} \sum_{t=1}^{n} (\widehat{\beta}_{j,i}(\tau_t) - \overline{\widehat{\beta}}_j(\tau_t))^2},$$

where $\beta_j(\cdot)$ for $j = 0, d, w$ and m are the known coefficient functions, $\overline{\widehat{\beta}}_j(\cdot) = \frac{1}{M} \sum_{i=1}^{M} \widehat{\beta}_{j,i}(\cdot)$, and $\widehat{\beta}_{j,i}(\cdot)$ denotes the value of $\widehat{\beta}_j(\cdot)$ at the i-th replication.

8. Repeat the steps 3–7 for $n = 500$ and 1000.

9. Repeat the steps 1–7 with the DGP2 for $n = 250, 500$ and 1000.

The results of the simulation study are summarized in Table 1. The overall results indicate that the local linear estimator with CV bandwidth is not very sensitive to various error distributions, and this estimator has favorable finite sample properties in terms of bias and standard deviation (std). Furthermore, an attractive property is that these two quantities rapidly decrease as the sample size increases from 250 to 1000.

Table 1 near here

3 Forecasting Volatility

In this section we discuss how to assess the out-of-sample forecasting performance of the proposed nonparametric TVC-HAR model relative to other parametric HAR models including the simple HAR and the HAR-GARCH models, with or without the jump components. The parametric and nonparametric forecasting methods are introduced in Sections 3.1 and 3.2, respectively, and the so-called CPA test is discussed in Section 3.3 to evaluate the volatility forecasts.

3.1 Parametric Multi-Step-Ahead Forecasting

In this section we first provide a summary of some selected parametric forecasting models of the logRV series, and then describe how to generate the multi-step-ahead forecasts of logRV using these models.

Andersen et al. (2007) extended Corsi (2009)'s daily HAR model to longer horizons, i.e.,

$$\log RV_{t,t+\delta} = \beta_{\delta,0} + \beta_{\delta,d} \log RV_{t-1} + \beta_{\delta,w} \log RV_{t-5,t-1} + \beta_{\delta,m} \log RV_{t-22,t-1} + \varepsilon_{t,t+\delta}, \tag{10}$$

The δ-step-ahead forecasts of logRV can be generated as the one-step-ahead forecast from model (10) for a given δ. The estimates of the parameters in model (10) are expected to vary for different δ. In this paper, for the purpose of forecasting comparison, we include the parametric HAR-family including simple HAR, HAR-GARCH(1,1) with normal errors, and HAR-J and HAR-CJ models (Equations 13 and 28 in Andersen et al. (2007), respectively), and HAR-J with GARCH(1,1) errors, where the letter C refers to the log continuous sample path component variation, while the letter J refers to the log jump component of logRV.

To generate multi-step-ahead out-of-sample forecasts, we partition the total data sample of size n into two periods: an estimation period and an evaluation period:

$$t = \underbrace{1, 2, \cdots, k,}_{\text{estimation period}} \underbrace{k+1, k+2, \cdots, n}_{\text{evaluation period}}.$$

Next, we describe the out-of-sample forecasting methodology only for the HAR model (10), and the methodology is similar for other parametric HAR-type models included in this forecasting exercise; see also Andersen et al. (2007) for HAR models with jumps. We generate the volatility point forecasts using a rolling window of fixed length k with the estimation scheme. At the time point k, let the estimators of parameters in model (10) be $\widehat{\beta}_{\delta,j}^{k}$ estimated using the first k observations, and then the δ-step-ahead out-of-sample forecasts are generated and compared to the realization $\log RV_{k+\delta}$. Similarly, at the time point $k+1$, $\widehat{\beta}_{\delta,j}^{k+1}$ are obtained using the k observations ending at $k+1$, the second set of δ-step-ahead forecasts are generated and compared to the realization

$\log RV_{k+\delta+1}$, and so on. This iterative procedure generates $m \equiv n - \delta - k + 1$ number of out-of-sample forecasts and relative forecast errors. Following Patton (2011), we measure the relative forecast error using the simple mean squared error loss function. In summary, setting $\delta = 1, 5$ and 22 as the daily, weekly and monthly forecasting horizons, respectively, the conditional forecasts of $\log RV_{t,t+\delta}$ at any time point $t \in \{k, k+1, \cdots, n-\delta+1\}$, are computed directly as follows:

$$\mathsf{E}(\log RV_{t,t+\delta}|\mathcal{F}_t) = \widehat{\beta}^t_{\delta,0} + \widehat{\beta}^t_{\delta,d} \log RV_{t-1} + \widehat{\beta}^t_{\delta,w} \log RV_{t-5,t-1} + \widehat{\beta}^t_{\delta,m} \log RV_{t-22,t-1}$$

for $\delta = 1, 5$ and 22, where $\mathcal{F}_t$ represents the information set available up to time t.

3.2 Nonparametric Multi-Step-Ahead Forecasting

The nonparametric multi-step-ahead forecasting for non-linear autoregression models with order d can be made by estimating the conditional mean $\mathsf{E}(Y_{t+\delta}|Y_t, \cdots, Y_{t-d+1})$ via nonparametric smoothing of $Y_{t+\delta}$ on $(Y_t, \cdots, Y_{t-d+1})$ directly (c.f., Robinson, 1983; Härdle and Vieu, 1992; Teräsvirta et al., 2010). The direct nonparametric approach, however, ignores the substantial information contained in the intermediate variables $Y_{t+1}, \cdots, Y_{t+\delta-1}$ about the conditional mean. Hence, to improve this estimator, Chen et al. (2004) introduced a multi-stage nonparametric predictor, which utilises information in pseudo observations $Y^*_{t+1}, \cdots, Y^*_{t+\delta-1}$ (which will be defined later) to generate an estimate for $Y_{t+\delta}$. They showed that such multi-stage smoother improves the estimation of the conditional mean, and demonstrated that this new predictor is more efficient than the direct nonparametric smoother. In the empirical application, we thus employ the multi-stage nonparametric predictor introduced by Chen et al. (2004) for the conditional forecasts of $\log RV_{t+\delta}$ under the nonparametric TVC-HAR modelling framework (4), with adjustments made to the predictor to account for the time-varying nature of the coefficient functions. The approach of the nonparametric multi-step-ahead forecasting is described as follows.

1. Choose the optimal bandwidth h_{opt} by the CV selection introduced in Section 2.2, where

only the first k observations of $\{Y_t, X_t\}$ are used.

2. For one-step-ahead forecast at time t, use the optimal bandwidth h_{opt} from Step 1 and obtain the local linear estimates $\widehat{\beta}_{t,j}$ of $\beta_{t,j}$, using the past observations up to time t. Use these $\widehat{\beta}_{t,j}$ and model (4) to predict Y_{t+1}, denoted as $\widehat{Y}^*_{t+1}$, and compare it to the realization $\log RV_{t+1}$.

3. For two-step-ahead forecast at time t, update the vector of regressor values by including the pseudo observation $\widehat{Y}^*_{t+1}$. Then, estimate the coefficient functions using the same bandwidth h_{opt}, and let the resulting estimator be $\widehat{\beta}_{t+1,j}$. Use this estimator to generate the two-step-ahead forecast, $\widehat{Y}^*_{t+2}$, and compare it to the realization $\log RV_{t+2}$.

4. Similarly, for δ-step-ahead forecast at t, update the set of regressors by adding the pseudo observations $\widehat{Y}^*_{t+1}, \cdots, \widehat{Y}^*_{t+\delta-1}$. Estimate the coefficient functions in the updated model using the same bandwidth h_{opt} and let this estimator be $\widehat{\beta}_{t+\delta-1,j}$. Use this estimator to generate $\widehat{Y}^*_{t+\delta}$ and compare it to the realization $\log RV_{t+\delta}$.

For simplicity, in Steps 2–4, we apply the same optimal bandwidth, h_{opt}, which is constructed from Step 1. In practice, however, one may re-estimate the optimal bandwidth h_{opt}, as it may differ as the data sample changes. The advantage of bandwidth re-estimation is the gain in the forecast accuracy (which may not be significant), but it comes at a cost of a significant increase in the computation time for nonparametric forecasting.

3.3 Evaluation of Volatility Forecasts: CPA Testing

It is of practical importance to evaluate and compare the out-of-sample forecasting performance between the proposed TVC-HAR model and its competitors including the simple HAR and the HAR-GARCH models. To achieve this, we employ the CPA test introduced by Giacomini and White (2006).

Diebold and Mariano (1995) and West (1996) (henceforth referenced as DM-W), contributed to the early literature on the unconditional out-of-sample predictive ability evaluation of forecasting models. However, the recent CPA test provides a framework for an unconditional forecast evaluation criterion, which is robust to misspecified forecasting models. In our forecast evaluation which involves three models, the proposed TVC-HAR and the simple HAR are nested models, whereas the proposed TVC-HAR and HAR-GARCH are non-nested models. The CPA test has an advantage over the DM-W approach in that the former is well suited for evaluating CPA of nested as well as non-nested forecasting models. In addition, the CPA test can be applied to multi-step point, interval, probability or density forecast valuation for a general loss function. Furthermore, the CPA approach accommodates conditional forecast evaluation objectives (which is more accurate at a specific future date), as well as nesting the unconditional objectives (which is more accurate on average) of the DM-W approach. Although both the unconditional and conditional approaches are informative, the global (or average) relative forecasting performance may conceal important information about the relative forecasting performance over time. Hence, the use of the CPA test appears promising for assessing the merit of the TVC-HAR model against the parametric HAR-type models in terms of out-of-sample forecasting accuracy. The CPA test is implemented via the following three steps.

1. Based on the rolling samples of fixed length k, the conditional forecasts of $Y_{t+\delta}$ by using the TVC-HAR and the benchmark models (say, the simple HAR model), respectively, for a given set $\mathcal{F}_t$ (defined in Section 3.1), are generated for the target date $t + \delta$ of the evaluation period at $t = k, k + 1, \cdots, n - \delta$. Let $\widehat{Y}_{1,t+\delta}$ and $\widehat{Y}_{2,t+\delta}$ be the conditional forecasts of $Y_{t+\delta}$ by using the simple HAR and TVC-HAR models, respectively.

2. For each of the two forecasting models, generate a sequences of losses, $L_{j,t+\delta} = L(Y_{t+\delta}, \widehat{Y}_{j,t+\delta})$, with $j = 1$ denoting the benchmark model, and $j = 2$ denoting the TVC-HAR model, where $L(\cdot, \cdot)$ is the mean squared error loss function. The two conditional forecasts are compared

through calculating $\Delta L_{t+\delta} = L_{1,t+\delta} - L_{2,t+\delta}$.

3. To test whether the alternative TVC-HAR model outperforms the benchmark model, we consider testing

$$H_{0*} : \ \mathsf{E}[\Delta L_{t+\delta}|\mathcal{F}_t] = 0 \ a.s., \ \text{versus } H_{1*} : \ \mathsf{E}[\Delta L_{t+\delta}|\mathcal{F}_t] > 0 \ a.s.$$

We use the test statistic $T_{k,m} = m\overline{Z}_{k,m}^{\top}\widehat{\Theta}_m^{-1}\overline{Z}_{k,m}$, where $\overline{Z}_{k,m} = \frac{1}{m}\sum_{t=k}^{n-\delta}Z_{t+\delta}$, $Z_{t+\delta} = \pi_t\Delta L_{t+\delta}$, $m = n - \delta - k + 1$, π_t is a chosen test function,

$$\widehat{\Theta}_{k,m} = \frac{1}{m}\sum_{t=k}^{n-\delta}Z_{t+\delta}Z_{t+\delta}^{\top} + \frac{1}{m}\sum_{l=1}^{\delta-1}w_{m,l} \times \sum_{t=k+l}^{n-\delta}[Z_{t+\delta}Z_{t+\delta-l}^{\top} + Z_{t+\delta-l}Z_{t+\delta}^{\top}],$$

and $w_{m,l}$ is a weight function such that $w_{m,l} \to 1$ as $m \to \infty$ for each $l = 1, \cdots, \delta - 1$.

With respect to the choice of π_t, as discussed in Giacomini and White (2006), π_t is chosen by researchers to include variables that are considered helpful to distinguish the relative forecast performance of the two competing models. It can be an indicator of past relative performance (such as lagged loss differences or moving averages of past loss differences) or business cycle indicators, see Bierens (1990) and Stinchcombe and White (1998) for various ways of choosing the test function. Section 4.3 below will specify the choices of π_t and $w_{m,l}$ in the analysis of the logRV of the S&P 500 index returns.

The null hypothesis H_{0*} of equal conditional predictive ability of the models is rejected when $T_{k,m} > \chi_{q,1-\alpha}^2$, where $\chi_{q,1-\alpha}^2$ is the $(1 - \alpha)$ quantile of the χ_q^2 distribution and q is the dimension of π_t. This rejection occurs when the out-of-sample loss difference $\{\Delta L_{t+\delta}\}$ is significantly different from zero. Furthermore, in case where the null hypothesis of equal conditional predictive ability is rejected, Giacomini and White (2006) also proposed an approach to make forecasting model selection decisions. Their approach consists of the following three steps.

1. Regress $\Delta L_{t+\delta}$ on π_t over the out-of-sample period for $t = k, k + 1, \cdots, n - \delta$, and let $\widehat{\varphi}_m$

denote the regression coefficient matrix. Apply the above χ_q^2 test and if the null hypothesis is rejected, then we proceed to step 2.

2. Approximate $\mathsf{E}[\Delta L_{t+\delta}|\mathcal{F}_t]$ using $\widehat{\varphi}_m^\top \pi_t$, and model **A** (the benchmark model), with the lower loss, is considered superior if $\widehat{\varphi}_m^\top \pi_t < c$, and model **B** (the alternative TVC-HAR model) is superior otherwise. In this paper, we specify $c = 0$, as we desire to choose a model that yields lower loss at $t + \delta$.

3. Compute the ratio $\frac{1}{m}\sum_{t=k}^{n-\delta} I\{\widehat{\varphi}_m^\top \pi_t > 0\}$, the relative out-of-sample performance of models **A** and **B**, where $I\{\cdot\}$ is an indicator function. Thus, model **A** is a better forecasting model at $t + \delta$ if the ratio < 0.5, and model **B** otherwise.

4 Analysis of the US Market Data

4.1 Preliminary Analysis of the Data

The primary data set consists of tick-by-tick transaction prices for the S&P 500 index for the period from 10/May/1999 to 26/October/2010. All the index data have been supplied by the Securities Industries Research Centre of Asia Pacific (SIRCA) on behalf of Reuters, with the raw index data filtered prior to the construction of the RV data. Construction of RV does not impose any particular requirement on the way in which prices are sampled as long as the corresponding returns are non-overlapping and span the time period of interest. There are variety of different sampling schemes used in the literature. However, it is well established that ultra high-frequency returns would lead to bias in the volatility measures, due to market microstructure effects such as the bid-ask bounce, stale prices and price discreteness. These effects cause the observed asset prices to behave differently to the assumptions underlying the construction of RV. In the literature, there is a general consensus that the five-minute interval minimizes the influence of such microstructure effects. This is also the approach adopted in our empirical study. Following Andersen et al. (2007)

and Bollerslev et al. (2009), we compute the daily realized variance from five-minute logarithmic returns constructed using the nearest price to each five-minute mark. The resulting daily $\log RV_t$ time series is displayed in Figure 1.

We first fit the simple HAR model (1) to the data set, and the ordinary least squares (OLS) estimates of the coefficients are presented in Table 2. All of the coefficients are significantly different from zero at the 10% significance level. Furthermore, to examine if the four coefficients, β_0, β_d, β_w and β_m change significantly over time, we also estimate the moving window (with size 250) bivariate correlations between $\log RV_t$ and $\log RV_{t-1}$, $\log RV_{t-5,t-1}$, and $\log RV_{t-22,t-1}$, respectively. All the three sets of correlations are plotted in Figure 2. All three moving window correlations appear to be nonlinear over the sample period and they move in unison over time. In particular, all of the three correlation measures move upwards during the crises such as the 2002-03 stock market downturn and the 2007 global financial crisis (GFC), and then downwards and stay low during the tranquil period between 2003 and early 2007.

Table 2 near here

To learn more about the time-varying nature of the coefficients of the HAR models, we subdivide the full sample period into four different subsample periods. A simple HAR model is fitted to each of the four subsamples, and the results are also presented in Table 2. Subsample 1 covers the periods before the dotcom crisis, subsample 2 covers the US stock market downturn during the dotcom crisis, subsample 3 covers the period before the GFC, and subsample 4 represents the recent GFC period. The OLS estimates and their standard errors (in the parentheses) of the coefficients in the simple HAR model for each subsample are listed in Table 2. The result of the Wald test for the equality of the regression coefficients indicates that the coefficients are significantly different across the four sample periods, as the value of the test statistic is 61.168. Moreover, the ACF plots of $\log RV_t$ series for all the four subsamples are presented in Figure 3, which exhibit different patterns across the subsamples, indicating the dynamic changes of the serial dependence

in the $\log RV_t$ series. Therefore, from the above our preliminary analysis, the TVC-HAR model (4) may be a better alternative to the simple HAR specification for the $\log RV_t$ series.

4.2   Estimation of the TVC-HAR model

In this section, we implement the methodologies introduced in Sections 2.2 and 2.3 to the logRV of the S&P 500 series. We first compare the local linear estimation method with the Nadaraya-Watson local constant estimation method. The optimal bandwidths chosen by the CV method for the local constant and local linear methods are 0.137 and 0.174, respectively. Overall, the local linear estimates of the coefficient functions are somewhat smoother than those by using the local constant method, mainly due to a slightly larger optimal bandwidth for the local linear method. In addition, as shown in Figure 4, the plot of the residuals from the latter estimation are closer to a normal distribution than those from the former. To evaluate the in-sample estimation performance of the simple HAR model with OLS estimation, the TVC-HAR model with local constant smoothing and the TVC-HAR model with local linear smoothing, the mean squared errors are calculated as 0.269, 0.266, and 0.260, respectively, which indicate that the local linear smoothing method of the TVC-HAR model provides the best estimation overall. Following the preceding observations, we will not further discuss the results of local constant estimation on TVC-HAR.

The local linear estimates of the time-varying coefficient functions in the TVC-HAR models are plotted in Figure 5 along with their 90% point-wise confidence bands. The horizontal lines in all four panels in Figure 5 indicate the OLS estimates: beta0-OLS, beta1-OLS, beta2-OLS and beta3-OLS, respectively of the constant coefficients, $\beta_0, \beta_d, \beta_w$ and β_m of the simple HAR model (1). The plot of the intercept function $\hat{\beta}_0(\cdot)$ in Figure 5 is significantly smaller than the corresponding constant coefficient during the tranquil period. The $\widehat{\beta}_d(\cdot), \widehat{\beta}_w(\cdot)$ and $\widehat{\beta}_m(\cdot)$ in Figure 5 are respectively the estimated coefficient functions of $\log RV_{t-1}, \log RV_{t-5,t-1}$, and $\log RV_{t-22,t-1}$ on $\log RV_t$, Reported along with these estimates are their 90% point-wise confidence bands which

lie either above or below the horizontal lines. The local linear estimates of coefficient functions of TVC-HAR in Figure 5 are denoted by beta0-LL, beta1-LL, beta2-LL and beta3-LL, respectively.

The plot of $\widehat{\beta}_d(\cdot)$ which measures the daily effect of $\log RV_{t-1}$ on $\log RV_t$ is smaller during the tranquil period, while it is larger during the GFC period. However, the opposite effects emerge from the plot of the $\widehat{\beta}_w(\cdot)$ function, which measures the weekly effect of $\log RV_{t-5,t-1}$ on $\log RV_t$. It is larger during the non-crisis period while smaller during the GFC period. The plot of the estimate of monthly effect $\widehat{\beta}_m(\cdot)$ shows a downwards trend until 2004 indicating declining effect of $\log RV_{t-22,t-1}$ on $\log RV_t$ during this period, reaches its trough during the period 2004–2006, but then increases on the occurrence of the GFC and peaks around mid-2008 when this crisis was deepening. It is evident that all the coefficients of the HAR model are time-varying.

To find further statistical evidence of whether or not the widely used simple HAR model is adequate for the data, we apply the generalized likelihood ratio (GLR) test proposed by Fan et al. (2001), with the wild bootstrap method to compute the p-value, is outlined in Appendix A.2 below. We use the GLR to test whether H_0: simple HAR against H_1: TVC-HAR is rejected or not. By setting $B = 200$ in the bootstrap procedure, we obtain the p-value of 0, which indicates that the simple HAR model is rejected in favour of the TVC-HAR specification for the logRV of the S&P 500 index returns. Despite the GLR test originally proposed for the independent data, the result of our analysis indicates that this test works well numerically for the weakly dependent and locally stationary logRV series.

4.3   Forecasting and CPA Test

In this section we evaluate the out-of-sample forecasting ability of the TVC-HAR model with the local linear method relative to the simple HAR, HAR-CJ, HAR-J, HAR-GARCH and HAR-J-GARCH models. To do this, we first define the in-sample estimation and the out-of-sample prediction periods as follows:

		Evaluation (m)	
	Estimation (k)	pre-GFC	post-GFC
Dates	01/07/2002–13/11/2007	14/11/2007–31/8/2008	02/09/2008–16/07/2009
Data size	1345	200	220

The reason for splitting the out-of-sample forecasting period into pre- and post-GFC periods is that during the post-crisis period the stock market volatility has been unduly high. Finding a model that generates superior multi-step-ahead out-of-sample forecasts relative to its competitors during high volatility is very useful to investors and practitioners. To apply the nonparametric forecasting method discussed in Section 3.2, the optimal bandwidth h_{opt} estimated for the full sample period is used for simplicity and time efficiency. All the models included in this forecasting are estimated using the rolling samples of k observations and the conditional forecasts of $\log RV_{t,t+\delta}$ are generated and compared them with the realizations. Following Giacomini and White (2006)'s method described in Section 3.3, we choose the weight function $w_{m,l} = 1 - \frac{l}{\delta}$ and the test function $\pi_t = (1, \Delta L_{t+\delta-1})$, where $\Delta L_{t+\delta-1}$ is the lagged value of the difference of the loss $\Delta L_{t+\delta}$, from which the test statistic for the CPA test is constructed. In what follows, we report the results of the CPA test.

4.3.1 CPA Test of the HAR-type model against the TVC-HAR model

The CPA test applied for testing the H_{0*} that the parametric HAR-type model (benchmark model) and the TVC-HAR model have the same out-of-sample forecasting performance against the H_{1*} that these two models do not have the same forecasting performance. The p-values of the test statistic are reported in Table 3. Furthermore, the proportion of the times (in percentages) that the alternative model TVC-HAR offers a more accurate forecast over the benchmark model are also reported in this table. When the simple HAR model is considered as the benchmark model, the

results are given in Table 3(a). For the entire forecasting evaluation period, the *p*-values indicate that the TVC-HAR model is as good as the HAR model in forecasting of $\log RV_{t,t+\delta}$ at $\delta = 1$. On the other hand, at $\delta = 5$ and 22, the performance of TVC-HAR is superior to HAR in forecasting. The results in Table 3(c) and (e) show that these observations are very similar when the HAR-J and HAR-CJ model are considered as the benchmark model. More importantly, there is an overwhelming evidence showing that the TVC-HAR model consistently outperforms the three HAR-type models in the post-2007 GFC period.

The out-of-sample forecasting evaluations of these models were also made during the 2003 crisis, for which we used the estimation period 10/05/1999–20/09/2001, and the out-of-sample forecasting evaluation periods: pre-2003 crisis period 21/09/2001–16/07/2002 and the post-2003 crisis period 17/07/2002–06/06/2003. The forecasting performance of the TVC-HAR relative to the three HAR-type models are very similar to what was observed during the 2007 crisis. To save space the results for during the 2003 crisis period are not tabulated in this paper.

<div style="text-align:center; border:1px solid black; display:inline-block;">Table 3 near here</div>

4.3.2 CPA Test for HAR-GARCH and HAR-J-GARCH against TVC-HAR

In this section, the HAR and HAR-J models with GARCH(1,1) errors are treated as benchmark models against the TVC-HAR model. The results of the CPA test are reported in Table 3(b) and (d), respectively, and they show that as a consequence of adding the GARCH component, the out-of-sample forecasting performances of HAR-GARCH and HAR-J-GARECH models have improved for the full forecasting evaluation period, which includes the 2007-crisis. However, it is noticeable that the relative performance of the TVR-HAR model consistently better than the parametric counterparts during the post-GFC period when $\delta = 22$ and during the pre-GFC period when $\delta = 5$.

5   Conclusions

In this paper, we introduce a flexible TVC-HAR modelling framework for the logRV of the S&P 500 index returns. In this flexible model, the coefficients of the HAR model are allowed to be time-varying with unspecified functional forms, which are estimated by a local linear method. In a simulation study, we find that the local linear estimator for the coefficient functions of a TVC-HAR time series model has good finite sample properties in terms of bias and standard deviation. We construct the point-wise confidence bands for the coefficient functions using the bootstrap procedure, which provide a statistical evidence to show that the HAR coefficients are indeed time-varying. Additionally, the GLR test statistic augmented with the wild bootstrap method is conducted to provide further statistical evidence against the simple linear HAR model for logRV of the S&P 500 series in favour of the TVC-HAR model proposed in this paper. We evaluate the multi-step-ahead out-of-sample forecasting performance of the TVC-HAR model relative to the simple HAR model and its extensions to HAR with jumps and/or GARCH. The results of the CPA test developed by Giacomini and White (2006) indicate that the one-step-ahead (daily) forecasting performance of these models are equal. On the other hand, the TVC-HAR model consistently outperforms the other parametric HAR-type models in 5-step-ahead (weekly) and 22-step-ahead (monthly) out-of-sample forecasting, particularly in the post-2003 and 2007 crises periods during which the financial market volatilities were extremely high.

6   Acknowledgements

Appendix A

A.1. Assumptions and an Asymptotic Distribution

In this part of the appendix, we derive the asymptotic distribution theory of the local linear estimation $\widehat{\beta}(\tau)$ defined in Section 2.2, which is of independent interest. Existing literature, such as Robinson (1989), Cai (2007) and Li et al. (2011), mainly considers the limiting distribution of the nonparametric kernel or local linear estimation of the TVC models where the covariates are exogenous and do not include the lagged term of the response. Hence, their results cannot be directly extended to the setting here.

Before giving the asymptotic results, we introduce a general class of evolutionary processes which can accommodate a variety of forms of nonstationary (or locally stationary) behavior. Let $\tau_t = \frac{t}{n}$ and

$$X_t = \beta_{t,0} + \sum_{j=1}^{p} \beta_{t,j} X_{t-j} + \varepsilon_t = \beta_0(\tau_t) + \sum_{j=1}^{p} \beta_j(\tau_t) X_{t-j} + \varepsilon_t, \tag{11}$$

which allows the autoregressive coefficients to change smoothly over time, and thus provides a much more general framework than the traditional autoregression models. Following Dahlhaus (1996), the process (11) is locally stationary under some mild conditions on the coefficient functions. The locally stationary process behaves like a stationary process in a small neighbourhood of each instant in time, but has a nonstationary behavior globally. We can refer to Kim (2001), and Zhang and Wu (2012) for recent development on nonparametric estimation of the coefficient functions $\{\beta_j(\cdot) : 1 \le j \le p\}$.

We next convert our TVC-HAR model into the framework of the locally stationary process (11). It is easy to see that model (4) can be re-written as

$$
\begin{aligned}
\log RV_t &= \beta_0(\tau_t) + \beta_d(\tau_t) \log RV_{t-1} + \beta_w(\tau_t)\Big[\frac{1}{5}\sum_{i=1}^{5} \log RV_{t-i}\Big] + \beta_m(\tau_t)\Big[\frac{1}{22}\sum_{i=1}^{22} \log RV_{t-i}\Big] + \varepsilon_t \\
&= \beta_0(\tau_t) + \Big[\beta_d(\tau_t) + \frac{\beta_w(\tau_t)}{5} + \frac{\beta_m(\tau_t)}{22}\Big] \log RV_{t-1} + \Big[\frac{\beta_w(\tau_t)}{5} + \frac{\beta_m(\tau_t)}{22}\Big] \sum_{i=2}^{5} \log RV_{t-i} \\
&\quad + \Big[\frac{\beta_m(\tau_t)}{22}\Big] \sum_{i=6}^{22} \log RV_{t-i} + \varepsilon_t \\
&\equiv \varphi_0(\tau_t) + \sum_{j=1}^{22} \varphi_j(\tau_t) \log RV_{t-j} + \varepsilon_t, \tag{12}
\end{aligned}
$$

where $\tau_t = \frac{t}{n}$. Hence, the TVC-HAR model falls into the locally stationary modelling framework defined in (11). Before presenting the asymptotic properties of the local linear estimator for the TVC-HAR models, we introduce the following regularity conditions.

Assumption 1. The coefficient functions $\beta_j(\cdot)$, $j = 0, d, w, m$, are twicely continuous differentiable in $[0, 1]$. Meanwhile, $\sum_{j=0}^{22} \varphi_j(\tau) z^j \neq 1$ for all $|z| \leq 1 + c$ with $c > 0$ uniformly in $\tau \in [0, 1]$, where $\varphi_j(\cdot)$ is defined on the right hand side of (12).

Assumption 2. The sequence $\{\varepsilon_t\}$ is independent and identically distributed (*i.i.d.*) with $\mathsf{E}[\varepsilon_t] = 0$, $0 < \sigma^2 \equiv \mathsf{E}[\varepsilon_t^2] < \infty$ and $\mathsf{E}[|\varepsilon_t|^{2+\kappa}] < \infty$ for $\kappa > 0$.

Assumption 3. The kernel function $K(\cdot)$ is a continuous, symmetric and nonnegative function with a compact support.

Assumption 4. Let the bandwidth h satisfy $h \to 0$ and $nh \to \infty$.

Assumption 1 imposes certain smoothness condition on the coefficient functions which ensures that the local linear approach is applicable and it also entails local stationarity and short-range dependence, respectively (c.f., Dahlhaus, 1996; Kim, 2001; and Zhang and Wu, 2012). The *i.i.d.* condition on the error term in Assumption 2 can be relaxed to some stationary and weakly dependent conditions such as the stationary martingale differences. Assumptions 3 and 4 are two commonly-used conditions on the kernel function and bandwidth, respectively.

Let $\mathcal{F}_j^* = (\cdots, \varepsilon_{j-1}, \varepsilon_j)$ be a shift process of *i.i.d.* random variables ε_i. We next give the asymptotic distribution for the local linear estimation $\widehat{\beta}(\tau)$ defined in Section 2.2, where $\tau \in [0, 1]$.

Proposition 1. *Suppose that Assumptions 1–4 are satisfied. Furthermore, there exists a measurable and stochastically Lipschitz continuous function $X(\cdot, \cdot)$ such that*

$$\max_{1 \leq t \leq n} \left\| X_t^* - X(\tau_{t-1}, \mathcal{F}_{t-1}^*) \right\| = O_P\left(\frac{1}{n}\right), \tag{13}$$

where $X_t^* = (\log RV_{t-1}, \log RV_{t-5,t-1}, \log RV_{t-22,t-1})^\top$. Then, we have

$$\sqrt{nh} \left[\widehat{\beta}(\tau) - \beta(\tau) - b_\tau(h) \right] \xrightarrow{d} \mathsf{N}(0, v_0 \sigma^2 \Sigma_\tau^{-1}), \tag{14}$$

where $b_\tau(h) = \frac{1}{2} h^2 \mu_2 \beta''(\tau) + o_P(h^2)$ with $\mu_2 = \int u^2 K(u) du$, $v_0 = \int K^2(u) du$ and $\Sigma_\tau = \mathsf{E}\left[\widetilde{X}_\tau \widetilde{X}_\tau^\top \right]$, $\widetilde{X}_\tau = [1, X^\top(\tau, \mathcal{F}_0^*)]^\top$.

Proof. By (13), Assumption 1 and Proposition 4.2 in Zhang and Wu (2012), we can prove that $X_t = (1, \log RV_{t-1}, \log RV_{t-5,t-1}, \log RV_{t-22,t-1})^\top$ in the local linear estimation can be replaced by $\widetilde{X}_t = [1, X^\top(\tau_{t-1}, \mathcal{F}^*_{t-1})]^\top$, which would not affect the asymptotic distribution. Noting that $\{\varepsilon_t\}$ is *i.i.d.* and ε_t is independent of $X(\tau_{t-1}, \mathcal{F}^*_{t-1})$, we can complete the proof of (14) by using the central limit theorem and the argument in Zhang and Wu (2012).

A.2. Specification Testing

In this part of the appendix, we give an outline for the generalized likelihood ratio test developed by Fan et al. (2001) for finding a statistical evidence to check whether the proposed TVC-HAR model fits the data better than the simple HAR model. To outline this test, we consider the null hypothesis:

$$H_0 : \beta_0(\tau_t) = \beta_0, \ \beta_d(\tau_t) = \beta_d, \ \beta_w(\tau_t) = \beta_w, \ \beta_m(\tau_t) = \beta_m,$$

against the alternative hypothesis:

$$H_1 : \beta_j(\tau_t) \neq \beta_j \text{ for at least one of } j = 0, d, w, m.$$

If H_0 is rejected, then we would infer that the nonparametric TVC-HAR model fits the logRV series better than HAR. The test statistic is constructed based on the generalized maximum likelihood ratio test proposed by Fan et al. (2001) and defined as follows:

$$TS_n = \frac{RSS_0 - RSS_1}{RSS_1}, \tag{15}$$

where RSS_0 is the residual sum of squares under the null hypothesis H_0, and RSS_1 is the residual sum of squares under the alternative hypothesis H_1. Letting $\widehat{\beta}$ be the OLS estimate of $\beta = (\beta_0, \beta_d, \beta_w, \beta_m)^\top$, then $RSS_0 = \frac{1}{n} \sum_{t=1}^{n} \widehat{\varepsilon}_{t0}^2$ with $\widehat{\varepsilon}_{t0} = \log RV_t - X_t^\top \widehat{\beta}$ and $RSS_1 = \frac{1}{n} \sum_{t=1}^{n} \widehat{\varepsilon}_{t1}^2$ with $\widehat{\varepsilon}_{t1} = \log RV_t - X_t^\top \widehat{\beta}(\tau_t) = \log RV_t - X_t^\top \widehat{\beta}_t$.

The null hypothesis is rejected if the *p*-value of the test is smaller than a nominal level (say 5%). The *p*-value is computed using the following wild bootstrap procedure introduced by Stinchcombe and White (1998).

1. Under the null hypothesis H_0, for each $t = 1, \cdots, n$, generate $\log RV_t^* = X_t^\top \widehat{\beta} + \varepsilon_t^*$, where the $\{\varepsilon_t^*\}$ is defined as in Step 2 of the bootstrap procedure in Section 2.3.

2. Use the data set $\{(\log RV_t^*, X_t) : t = 1, \cdots, n\}$ to estimate the models under both H_0 and H_1 and then calculate the corresponding RSS_0^* and RSS_1^*, and TS_n^* defined in (15).

3. Repeat Steps 1 and 2 for B times and obtain the empirical distribution of TS_n^*. Then, the p-value of the test statistics is computed by $\frac{1}{B} \sum_{i=1}^{B} I(TS_n^*(i) \geq TS_n)$, where $I(\cdot)$ is an indicator function and $TS_n^*(i)$ is calculated as TS_n^* by using the i-th bootstrap sample.

References

[1] Andersen, T.G., and Bollerslev, T., 1998. Answering The Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts. International Economic Review, 39, 885 − 905.

[2] Andersen, T.G., Bollerslev, T., and Diebold, F.X., 2007. Roughing it Up: Including Jump Components in the Measurement, Modeling and Forecasting of Return Volatility. Review of Economics and Statistics, 89, 701−720.

[3] Andersen, T.G., Bollerslev, T., Diebold, F.X., and Labys, P., 2001. The Distribution of Realized Exchange Rate Volatility. Journal of the American Statistical Association, 96, 42 − 55.

[4] Andersen, T. G., Bollerslev, T., and Huang, X., 2011. A Reduced form Framework for Modeling Volatility of Speculative Prices based on Realized Variation Measures. Journal of Econometrics, 160(1), 176 − 189.

[5] Bierens, H. B. 1990. A Consistent Conditional Moment Test of Functional Form. Econometrica, 58, 1443−1458.

[6] Bollerslev, T. 1986. Generalized Autoregressive Conditional Heteroskedasticity. Journal of Econometrics, 31, 307 − 327.

[7] Bollerslev, T., Kretschmer, U., Pigorsch, C., and Tauchen, G., 2009. A Discrete-Time Model for Daily S&P 500 Returns and Realized Variations: Jumps and Leverage Effects. Journal of Econometrics, 150, 151 − 166.

[8] Cai, Z. 2007. Trending Time-Varying Coeffcient Time Series Models with Serially Correlated Errors. Journal of Econometrics, 136, 163 − 188.

[9] Chen, R., Yang, L., and Hafner C., 2004. Nonparametric Multistep-Ahead Prediction in Time Series Analysis. Journal of the Royal Statistical Society, Series B, 66, 669 − 686.

[10] Cochrane, J.H., 2001. *Asset Pricing*. Princeton University Press, Englewood Cliffs, NJ.

[11] Corsi, F., 2009. A Simple Approximate Long-Memory Model of Realized Volatility. Journal of Financial Econometrics, 7, 174 − 196.

[12] Corsi, F., Mittnik, S., Pigorsch, C., and Pigorsch, U., 2008. The Volatility of Realized Volatility. Econometric Reviews, 27, 46 − 78.

[13] Dahlhaus, R., 1996. On the Kullback-Leibler Information Divergence of Locally Stationary Processes. Stochastic Processes and Their Applications, 62, 139 − 168.

[14] Diebold, F. X. and Mariano, R. S., 1995. Comparing Predictive Accuracy. Journal of Business & Economic Statistics, 13, 253 − 263.

[15] Engle R. F., 1982. Autoregressive Conditional Heteroscedasticity with Estimates of The Variance of United Kingdom Inflation. Econometrica, 50, 987 − 1007.

[16] Fan, J., and Gijbels, I., 1996. *Local Polynomial Modelling and its Applications*. Chapman and Hall, London.

[17] Fan, J., Zhang, C., and Zhang, J., 2001. Generalized Likelihood Ratio Statistics and Wilks Phenomenon. Annals of Statistics, 29, 153 − 193.

[18] Frijns, B., Lehnert, T., and Zwinkels, R. C., 2011. Modeling Structural Changes in the Volatility Process. Journal of Empirical Finance, 18(3), 522 − 532.

[19] Giacomini, R., and White, H., 2006. Tests of Conditional Predicative Ability. Econometrica, 74, 1545 − 1578.

[20] Giraitis, L., Kapetanios, G., and Yates, T., 2012. Inference on Stochastic Time-Varying Coefficient Models. Working Paper at Queen Mary, University of London.

[21] Härdle, W., and Vieu, P., 1992. Kernel Regression Smoothing for Time Series. Journal of Time Series Analysis, 13, 209 − 232.

[22] Heston, S. L., 1993. A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. The Review of Financial Studies, 6, 327 − 343.

[23] Jacquier, E., Polson, N. G., and Rossi, P. E. 1994. Bayesian Aanalysis of Stochastic Volatility Models (with Discussion). Journal of Business & Economic Statistics, 12, 371 − 417.

[24] Jagannathan, R., and Wang, Z., 1996. The Conditional CAPM and the Cross-Section of Expected Returns. Journal of Finance, 51, 3 − 53.

[25] Kim, W., 2001. Nonparametric Kernel Estimation of Evolutionary Autoregressive Processes. Discussion paper 103, Sonderforschungsbereich 373, Berlin.

[26] Li, D., Chen, J., and Gao, J., 2011. Nonparametric Time-Varying Coefficient Panel Data Models with Fixed Effects. The Econometrics Journal 14, 387 − 408.

[27] Liu, C., and Maheu, J. M., 2008. Are There Structural Breaks in Realized Volatility ? Journal of Financial Econometrics, 6(3), 326 − 360.

[28] McAleer, M., and Medeiros, M. C., 2008. A Multiple Regime Smooth Transition Heterogeneous Autoregressive Model for Long Memory and Asymmetries. Journal of Econometrics, 147, 104 − 119.

[29] Müller, U., Dacorogna, M., Dav, R., Pictet, O., Olsen, R. and Ward, J., 1993. Fractals and Intrinsic Time – A Challenge to Econometricians. 39th International AEA Conference on Real Time Econometrics, $14 - 15$ October 1993, Luxembourg.

[30] Nonejad, N., 2014. Long Memory and Structural Breaks in Realized Volatility: An Irreversible Markov Switching Approach. Available at SSRN 2427675.

[31] Patton, A. J., 2011. Volatility Forecast Comparison using Imperfect Volatility Proxies. Journal of Econometrics, 160, $246 - 256$.

[32] Polzehl, J., and Spokoiny, V., 2006. Varying Coefficient GARCH versus Local Constant Volatility Modeling: Comparison of the Predictive Power (No. 2006, 033). SFB 649 Discussion Paper.

[33] Raggi, D., and Bordignon, S., 2012. Long Memory and Nonlinearities in Realized Volatility: a Markov Switching Approach. Computational Statistics & Data Analysis, 56(11), $3730 - 3742$.

[34] Robinson, P. M., 1983. Nonparametric Estimation for Time Series Model. Journal of Time Series Analysis, 4, $185 - 208$.

[35] Robinson, P. M., 1989. Nonparametric Estimation of Time-Varying Parameters. Statistical Analysis and Forecasting of Economic Structural Change. Hackl, P. (Ed.). Springer, Berlin, $164 - 253$.

[36] Ruiz, E., 1994. Quasi-maximum Likelihood Estimation of Stochastic Volatility Models. Journal of Econometrics, 63, July 1994, $289 - 306$.

[37] Stinchcombe, M. B., and White, H., 1998. Consistent Specification Testing with Nuisance Parameters Present Only under the Alternative. Econometric Theory, 14, $295 - 325$.

[38] Teräsvirta, T., Tjøstheim, D., and Granger, C. W. J., 2010. *Modelling Nonlinear Economic Time Series*. Oxford University Press.

[39] Tsay, R. S., 2002. *Analysis of Financial Time Series*. Wiley, New York.

[40] West, K. D., 1996. Asymptotic Inference about Predictive Ability. Econometrica, 64, $1067 - 1084$.

[41] Zhang, T. and Wu, W., 2012. Inference of Time-Varying Regression Models. Annals of Statistics, 40, $1376 - 1402$.

[42] Zhang, W. and Peng, H., 2010. Simultaneous Confidence Band and Hypothesis Test in Generalised Varying-Coefficient Models. Journal of Multivariate Analysis, 101, $1656 - 1680$.
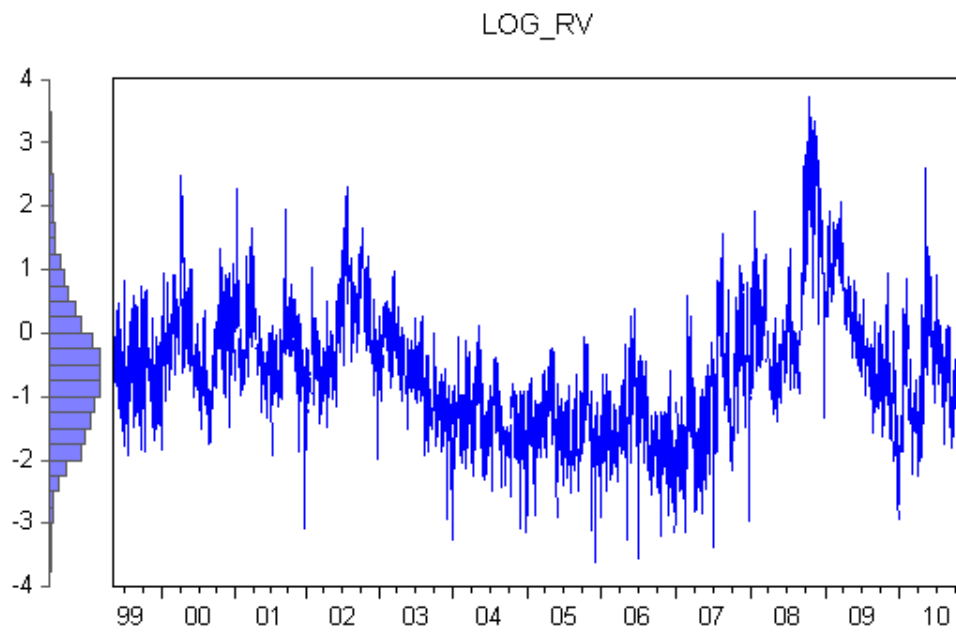
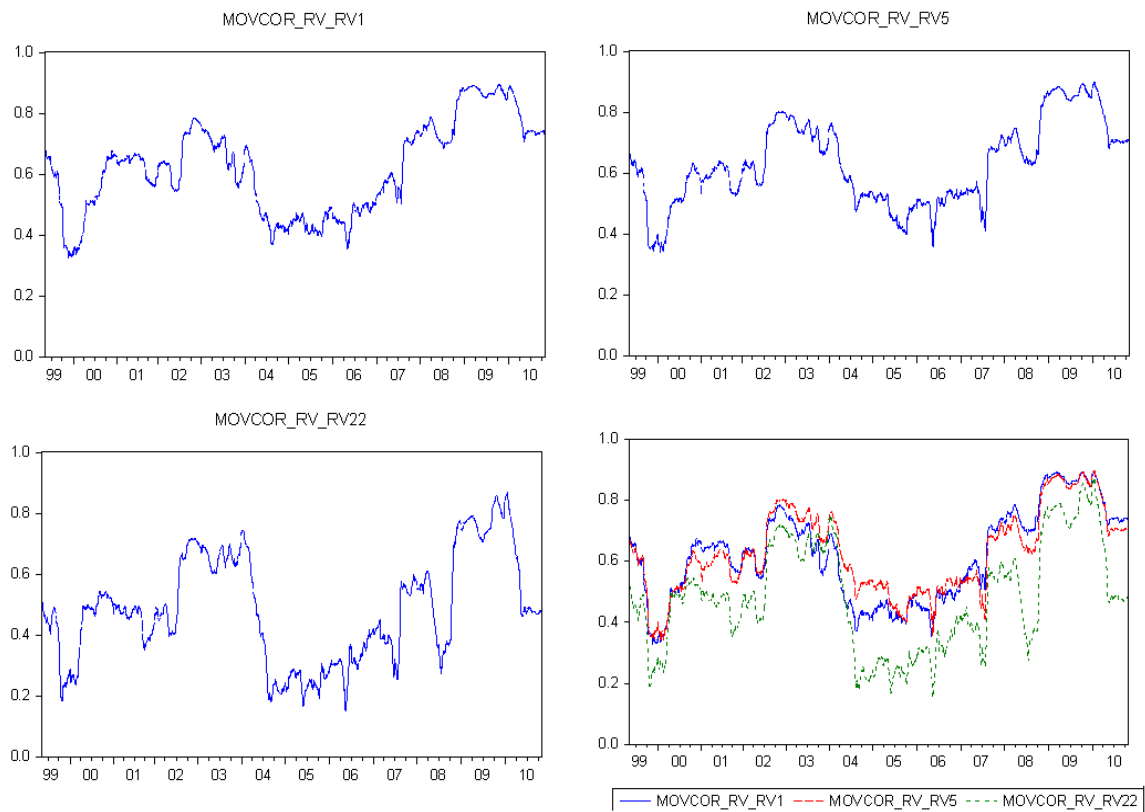*Figure 1. Time series of logarithmic realized volatility.*

Figure 2. *The moving window correlation plots between* $\log RV_t$ *and* $\log RV_{t-1}$, $\log RV_{t-5}$, $\log RV_{t-22}$, *respectively. The bottom-right plot combines the curves in the other three plots.*
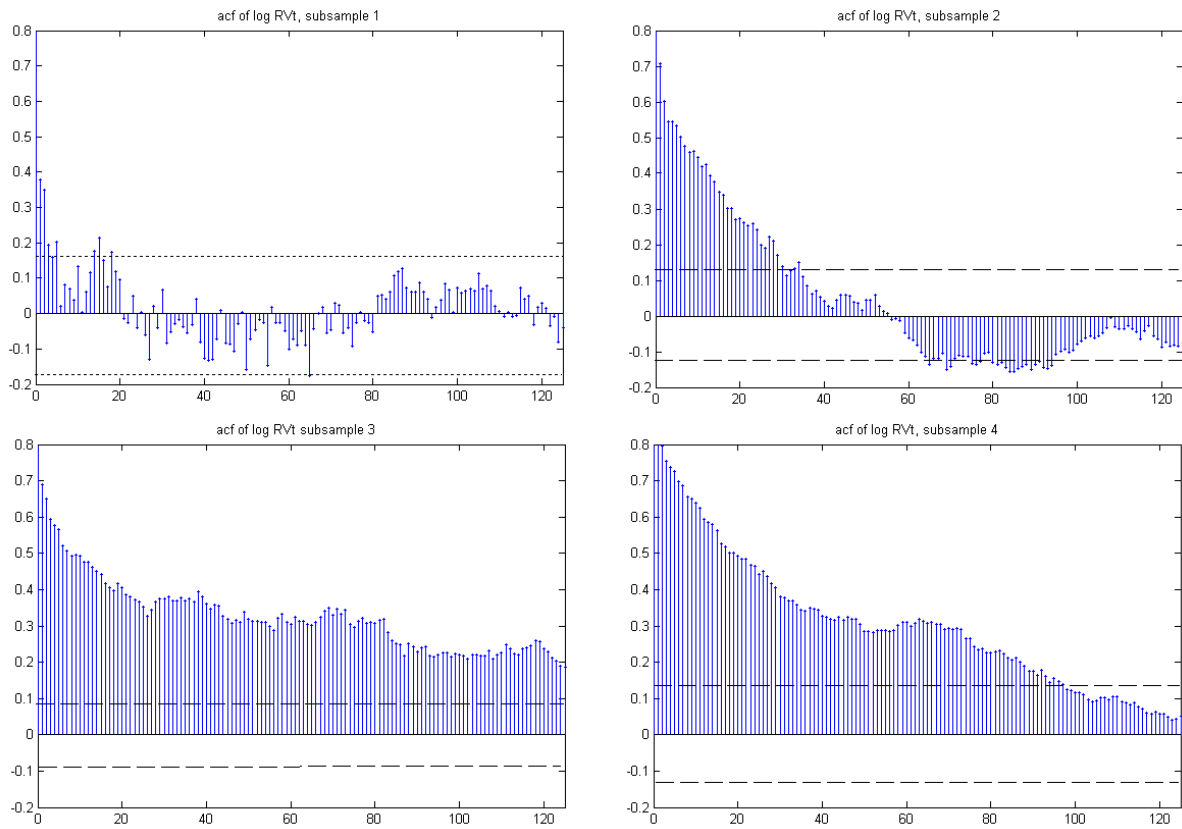
Figure 3. ACF plots for $\log RV_t$ up to 125th order. The plots exhibit different patterns across the four subsamples, reflecting the dynamic changes of the serial dependency of the $\log RV_t$ series over the sample period.
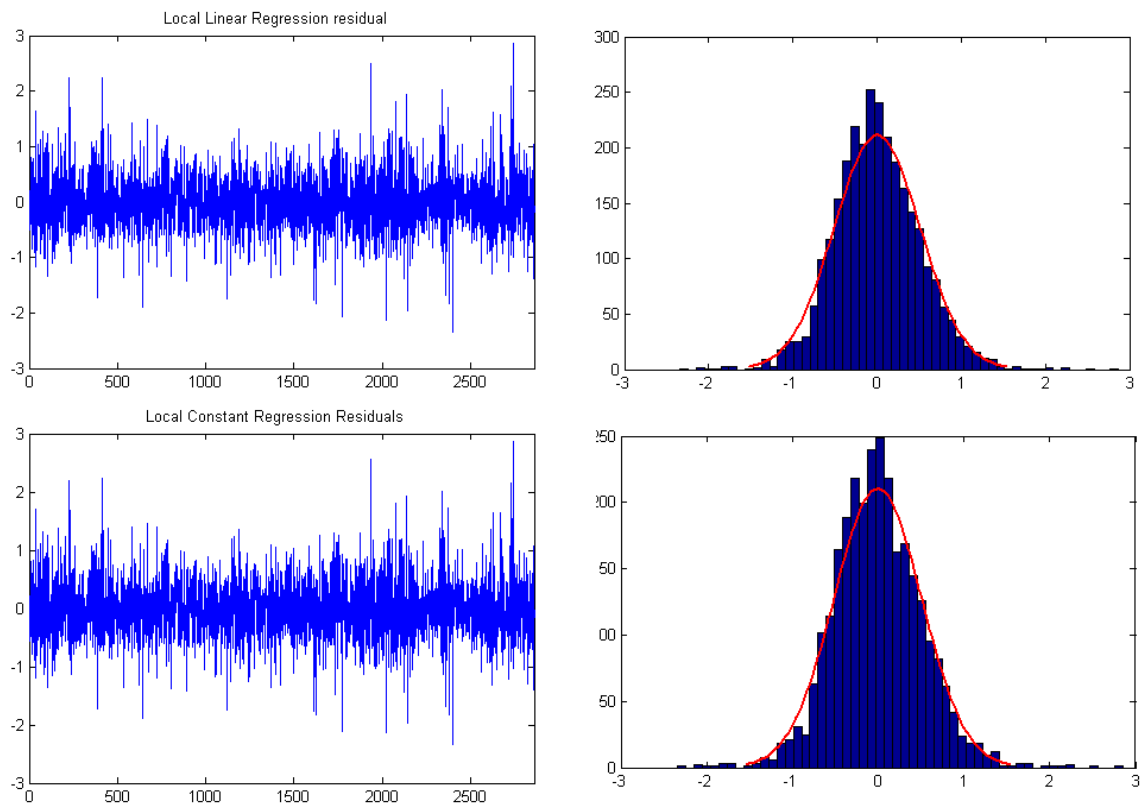
Figure 4. Plots of the residuals from local linear and local constant regressions as well as their distributions.
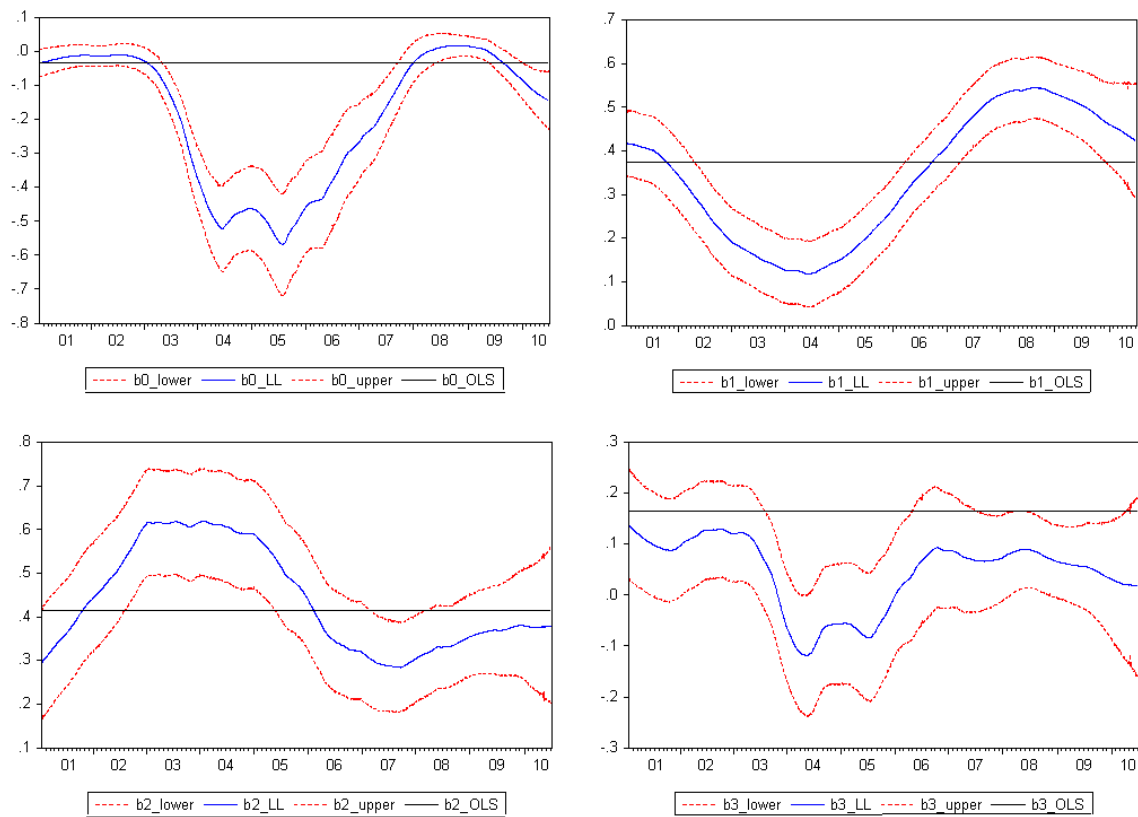
Figure 5. The local linear estimates of the coefficient functions β(·) (solid nonlinear line) with 90% point-wise confidence bands (dotted lines), and OLS estimates (solid horizontal line).

Table 1: Bias and Standard Deviation (Std) of the local linear estimates of the TVC-HAR model of based on 500 replications

	n	DGP 1				DGP 2			
		$\widehat{\beta_0}$	$\widehat{\beta_d}$	$\widehat{\beta_w}$	$\widehat{\beta_m}$	$\widehat{\beta_0}$	$\widehat{\beta_d}$	$\widehat{\beta_w}$	$\widehat{\beta_m}$
	250	0.0004	0.0039	−0.1099	−0.9151	0.0015	0.0018	−0.0937	−0.9062
Bias	500	−0.0045	0.0035	−0.0272	−0.3331	−0.0006	0.0044	−0.0033	−0.3149
	1000	−0.0018	0.0060	−0.0021	−0.1540	−0.0053	0.0047	0.0026	−0.1427
	250	0.1263	0.0327	0.0387	0.2087	0.1301	0.0281	0.0282	0.2179
Std	500	0.0771	0.0296	0.0250	0.0919	0.0718	0.0261	0.0230	0.0582
	1000	0.0634	0.0356	0.0409	0.0279	0.0632	0.0341	0.0372	0.0332

Table 2: Subsample periods and the associated simple HAR model parameter estimates. The standard errors are given in the parentheses and $*$ indicates insignificance of parameter at 10% level.

(a) Subsample Periods

	Periods	
Subsample 1	10/05/1999	10/03/2000
Subsample 2	11/03/2000	09/10/2002
Subsample 3	10/10/2002	13/07/2007
Subsample 4	16/07/2007	26/10/2010

(b) The OLS estimates of the simple HAR model

	β_0	β_d	β_w	β_m
Full sample	-0.0313 (0.0121)	0.3732 (0.0220)	0.4137 (0.0343)	0.1655 (0.0274)
Subsample 1	-0.1347 (0.0764)	0.2264 (0.0708)	0.2705 (0.1333)	0.2253^* (0.1820)
Subsample 2	-0.0100^* (0.0217)	0.4077 (0.0480)	0.3588 (0.0771)	0.1374 (0.0696)
Subsample 3	-0.1357 (0.0338)	0.2731 (0.0355)	0.4011 (0.0584)	0.2249 (0.0495)
Subsample 4	-0.0065^* (0.0181)	0.4930 (0.0388)	0.3833 (0.0555)	0.0629^* (0.0408)

Table 3: Results of the CPA test for out-of-sample forecasting performance between the TVC-HAR model and the benchmark model (HAR-RV, HAR-GARCH, HAR-RV-J, HAR-J-GARCH or HAR-RV-CJ model), with selected test function $\pi_t = (1, \Delta L_{t+\delta-1})$. The p-values denoted by pval and the proportion of time (in %) that the alternative model offers a more accurate forecast over the benchmark are given in the table.

	$\delta = 1$		$\delta = 5$		$\delta = 22$	
Subsamples	pval	Alternative better	pval	Alternative better	pval	Alternative better
(a) TVC-HAR *versus* HAR-RV						
Full sample	0.4813	Equally good	0.0000	93.48%	0.0000	81.70%
pre-GFC	0.5550	Equally good	0.0000	95.98%	0.0727	Equally good
post-GFC	0.7415	Equally good	0.0000	91.96%	0.0000	98.49%
(b) TVC-HAR *versus* HAR-GARCH						
Full sample	0.2454	Equally good	0.0000	47.12%	0.0002	52.38%
pre-GFC	0.4099	Equally good	0.0015	62.31%	0.0003	48.24%
post-GFC	0.4841	Equally good	0.0001	42.71%	0.0000	55.28%
(c) TVC-HAR *versus* HAR-J						
Full sample	0.5182	Equally good	0.0000	92.98%	0.0000	78.95%
pre-GFC	0.6574	Equally good	0.0000	95.98%	0.0693	Equally good
post-GFC	0.7152	Equally good	0.0000	91.96%	0.0000	92.46%
(d) TVC-HAR *versus* HAR-J-GARCH						
Full sample	0.3383	Equally good	0.0000	46.87%	0.0002	51.38%
pre-GFC	0.5033	Equally good	0.0017	63.32%	0.0004	49.25%
post-GFC	0.6034	Equally good	0.0001	42.71%	0.0000	53.77%
(e) TVC-HAR *versus* HAR-CJ						
Full sample	0.8182	Equally good	0.0000	56.14%	0.0000	64.16%
pre-GFC	0.9855	Equally good	0.0003	63.82%	0.0001	52.76%
post-GFC	0.6227	Equally good	0.0008	51.26%	0.0000	70.85%