ISSN 1440-771X

MONASH University

Australia

Department of Econometrics and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

Nonparametric Estimation and Parametric Calibration of Time-Varying Coefficient Realized Volatility Models

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

September 2013

Working Paper 21/13

Nonparametric Estimation and Parametric Calibration of Time–Varying Coefficient Realized Volatility Models

Xiangjin B. Chen, Jiti Gao*, Degui Li and Param Silvapulle

Department of Econometrics and Business Statistics, Monash University

September 17, 2013

Abstract

This paper introduces a new specification for the heterogeneous autoregressive (HAR) model for the realized volatility of S&P500 index returns. In this new model, the coefficients of the HAR are allowed to be time-varying with unknown functional forms. We propose a local linear method for estimating this TVC-HAR model as well as a bootstrap method for constructing confidence intervals for the time varying coefficient functions. In addition, the estimated nonparametric TVC-HAR was calibrated by fitting parametric polynomial functions by minimising the L_2-type criterion. The calibrated TVC-HAR and the simple HAR models were tested separately against the nonparametric TVC-HAR model. The test statistics constructed based on the generalised likelihood ratio method augmented with bootstrap method provide evidence in favour of calibrated TVC-HAR model. More importantly, the results of conditional predictive ability test developed by Giacomini and White (2006) indicate that the nonparametric TVC-HAR model consistently outperforms its calibrated counterpart as well as the simple HAR and the HAR-GARCH models in out-of-sample forecasting.

Keywords: Bootstrap Method, Heterogeneous Autoregressive Model, Locally Stationary Process, Nonparametric Method.

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 has been one of the most researched topics in both theoretical developments and practical applications in finance. Following the work of Engle (1982) and Bolleslev (1986) on ARCH and GARCH models for deterministic volatility modelling of financial market, modelling of the volatility has been very popular for nearly three decades. In addition, the work of Heston (1993), Ruiz (1994), and Jacquier et al. (1994) on stochastic volatility models raised their prominence 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.

Seminal papers by Andersen and Bollerslev (1998), Andersen et al. (2001) and Bollerslev et al. (2009), among others, introduced nonparametric realized volatility (RV) measures and advocated the use of these measures, which retain most of the pertinent information in the intraday data for measuring, modelling, and forecasting volatilities over daily and long horizons. The other reason for the popularity of the nonparametric measure of this RV is that it 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 realized volatility has created a new innovative research direction on the modelling of time series models for volatility in the recent literature in financial econometrics. The construction of (observable) realized volatility series from intra-day transaction data and the use of standard time-series techniques have led to promising approaches for modelling and predicting (daily) volatility.

The RV is known to possess long memory property, and as a result several studies modelled the RV as 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 AFIMA, Corsi (2009) proposed a heterogeneous autoregressive (HAR) model, which is relatively a simple autoregressive specification and shown to capture the crucial long memory feature of the RV series. In spite of its simplicity, Corsi (2009) showed the HAR-RV model to successfully encapsulate the main empirical features of financial returns distribution's fat tails and has a remarkable good forecasting performance relative to its competitors (McAleer and Medeiros 2008). Furthermore, $\log RV$ series is found to be normally distributed, despite RV being close to normal. McAleer and Medeiros (2008) provides an excellent review of the literature on the development and the relative performance of RV model, especially HAR model for RV relative to latent GARCH-type and stochastic volatility (SV) models.

Corsi et al. (2008) showed that the residuals of commonly used time-series models for realized volatility and logarithmic realized variance exhibit volatility clustering. They proposed extensions to explicitly account for these properties and assessed their relevance for modelling and forecasting realized volatility. In an empirical application to compare the one-step-ahead out-of-sample forecasting performance of HAR-GARCH against the simple HAR model, the authors found overall improved accuracy in volatility point forecasting by allowing for time-varying volatility of RV/$\log RV$.

The principal contributions of this paper are the following: (i) it introduces an alternative specification for HAR-RV model of S&P500 index returns in that the parameters of the HAR model are allowed to be time-varying with unknown functional forms; (ii) it proposes a local linear method for estimating this new specification of TVC-HAR model as well as a bootstrap method for constructing confidence intervals for the time varying coefficient functions; (iii) it calibrates the nonparametric TVC-HAR by fitting parametric polynomial functions; and

(iv) it evaluates the in-sample fitting and out-of-performances of proposed models against the simple HAR and HAR-ARCH models.

In our preliminary analysis of HAR-RV model of daily S&P500 stock index returns series over the period 1999-2010, we found that the coefficients of this model across the sub-sample periods were significantly different. Without making any assumptions as to the way in which these coefficients vary over time, we allow these HAR model parameters to be time-varying with unknown functional forms. If indeed these parameters are time varying and they are assumed to be constants, then this model misspecification can give rise to time-varying volatility for HAR model, which was introduced and extensively studied by Corsi et al. (2008). Thus, our proposed nonparametric TVC-HAR model is a natural competitor to the HAR-GARCH model. To assess the out-of-sample performance of this proposed model relative to several parametric counterparts, we use the conditional predictive ability (CPA) test developed recently by Giocomoni and White (2006). We will also 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 rapidly growing literature on realized volatility.

Modelling TVCs is common in finance. The TVC models have been extensively applied to several areas in finance, including the popular capital asset pricing model and the term structure of interest rates model; see, for example, Cochrane (2001), Tsay (2002) and Cai (2007). For more applications of the TVC time series model in econometrics and finance, as well as in other fields, see Jagannathan and Wang (1996), Cui et al. (2002), Akdeniz et al. (2003), Chang and Martinez-Chombo (2003) and Li et al. (2011), among others. Despite many studies imposing various parametric structures to TVC functions, in practice, the underlying true feature of time-varying coefficient 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, despite its simplicity to implement, the parametric structure imposed for TVC functions is often too restricted and can be unrealistic, leading to inaccurate forecasting. By contrast, without imposing any structure, nonparametric estimation of TVC models would be an attractive alternative, because it allows the data to "speak for themselves".

The specific aims of this paper are to: (i) introduce nonparametric methods based on Nadaraya-Watson and the local linear kernels for estimating the proposed TVC-HAR model for realized volatility, and a bootstrap based method to construct the point-wise confidence intervals for the nonparametric estimates; (ii) calibrate the nonparametric TVC-HAR by its parametric counterpart using higher-order polynomial functions; and (iii) ascertain the merit of the proposed models relative to the HAR-GARCH and the simple HAR models in terms of in-sample fitting as well as out-of-sample forecasting. The results will be established in the context of empirical applications to daily S&P500 index returns for the period 1999–2010.

There exists a vast literature on nonparametric estimation of regression models. Robinson (1989, 1991) contributed to the early development of the nonparametric estimation of TVC regression models with exogenous explanatory variables, while Chang and Martinez-Chombo (2003) and Cai (2007), among others, contributed to the recent literature on this topic. However, despite the popularity and widespread use of autoregressive (AR) models in econometrics and financial econometrics, the literature on nonparametric regression has not been extended to such time series models (except for Kim 2001). Since the HAR model is effectively a simple autoregressive (AR) model, we will extend the nonparametric method of Kim (2001) for estimating the TVC-HAR model that we propose in our paper. Kim (2001) developed this method for a locally stationary autoregressive process, and then extended to a general non-stationary process by allowing the autoregressive coefficients to change smoothly over time. Under some regularity conditions, Kim showed that the nonparametric estimators of the TVC-AR model can be derived in a similar fashion as for the conventional regression model with exogenous variables. Furthermore, we introduce an algorithm for nonlinear parametric calibration of the nonparametric TVC-HAR. The optimal calibrated specification is achieved via an automated algorithm, in which the L_2–type minimum distance function criterion is utilized.

The rest of the paper is organized as follows. Section 2 introduces the model, estimation methods and the construction of the confidence bands for the coefficient functions of TVC-HAR, as well as the specification testing of the nonparametric TVC-HAR model. Section 3 provides a description of the parametric and nonparametric out-of- sample forecasting and evaluation. Section 4 describes the S&P 500 data series and presents some stylized facts about $\log RV$ measures and a preliminary analysis, followed by the results of estimation and testing of the proposed nonparametric TVC–HAR model, its parametric calibration and the forecasting evaluation results. Section 5 concludes this paper. The assumptions and the necessary regularity conditions of the proposed nonparametric TVC-HAR model are given in the Appendix.

2 METHODOLOGY

2.1 TVC–HAR Model Specification

We consider the simple HAR model for RV proposed by Corsi (2008), which is defined by

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

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 constant over time. We relax the assumption that the HAR model parameters are constant and assume that they are time varying coefficients (TVC) with unknown functional forms, we denote such model as TVC–HAR model.

Before introducing the nonparametric TVC–HAR model specification, we introduce some notations through the following general TVC autoregressive model:

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

where $X_t = (X_{t,1}, X_{t,2}, \cdots, X_{t,d})^\top$ are lagged values of Y_t with $X_{t,1} = 1$, $\beta_t = (\beta_{t,1},\ \beta_{t,2},\ \cdots,\ \beta_{t,d})^\top$, β_t is a vector of unknown functions of time t, and $\{\varepsilon_t\}$ is a sequence of stationary errors. Using the same notations and specifications introduced by Robinson (1989) and Cai (2007), we assume that

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

where n is the sample size. Such a specification of β_t ensures the consistent estimation of the nonparametric regression estimator of $\beta(\cdot)$, via increasingly intense sampling of data points at each point in the interval $[0, 1]$; see Robinson (1989, 1991) for more details.

To specify the nonparametric TVC-HAR of model (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 these coefficients replaced with those defined by (3). Thus, the nonparametric TVC–HAR model for log RV can be specified as follows:

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

where $\tau_t = \frac{t}{n}$ for $1 \le t \le n$.

2.2 Nonparametric Estimation of TVC-HAR Model

We propose to use the popular local linear method for the estimation of $\beta_j(\tau_t)$, and $j = 0, d, w, m$, as defined in the model (4) for $\{\log RV_t,\ t = 1, \cdots, n\}$. This 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. See Fan and Gijbels (1996) and Cai (2007), among others, for details.

Assuming each $\beta_j(\cdot)$ has a continuous second derivative at any fixed point $\tau \in [0, 1]$, $\beta_j(\tau_t)$ can be approximated by a linear function, using the first–order Taylor series expansion as follows:

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

where $a_j = \beta_j(\tau)$, and $b_j = \beta_j'(\tau)$, the first derivative of $\beta_j(\tau)$. Thus, the model (4) can be approximated by

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

where $Z_t = \left(X_t^\top, \quad X_t^\top(\tau_t - \tau)\right)^\top$, and $\theta(\tau) = \left(\beta(\tau), \quad \beta'(\tau)\right)^\top$.

We then define the locally weighted sum of squares as,

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

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 estimation, and $K(\cdot)$ is the Epanechnikov kernel, defined as $K(u) = 0.75(1 - u^2)I(|u| \leq 1)$.

Then $\widehat{\theta}(\tau)$ can be obtained by minimizing (6) with respect to θ. Thus, we obtain the local linear kernel estimate of $\beta(\tau)$, denoted by $\widehat{\beta}(\tau)$, which consists of the first d elements of $\widehat{\theta}(\tau)$. The local linear estimator of derivative $\beta'(\tau)$ is $\widehat{\beta}'(\tau)$, and it consists of the last d elements of $\widehat{\theta}(\tau)$. By some elementary calculation, the expression of $\widehat{\theta}(\tau)$, which minimizes (6), is given by

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

where, for $k = 0, 1, 2$,

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

and

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

It can be shown that $\widehat{\beta}(\tau)$ is asymptotically normally distributed. For more detail, see Propositions 1 and some regularity conditions in the Appendix.

To obtain the local constant (Nadaraya–Watson) estimator of $\beta(\tau)$, we can replace (5) by

$$\log RV_t \simeq X_t^T \beta(\tau) + \varepsilon_t,$$

and the locally weighted sum of squares becomes

$$\sum_{t=1}^{n}[\log RV_t - X_t^\top \beta(\tau)]^2 K_h(\tau_t - \tau), \tag{8}$$

By minimizing (8) with respect to β, we can obtain the Nadaraya–Watson estimator $\widetilde{\beta}(\tau)$ by

$$\widetilde{\beta}(\tau) = S_{n0}^{-1}(\tau) M_{n0}(\tau). \tag{9}$$

Similarly to Propositions 1 in the Appendix, one may show that $\widetilde{\beta}(\tau)$ is also normally distributed.[1]

Note that many other nonparametric smoothing methods such as the series approximation method and cubic splines, can also be used to estimate the TVC–HAR model. There are also many studies which compared different nonparametric estimation methods (see, for example, Fan and Yao, 2003; Gao, 2007; Li and Racine, 2007). It is well documented that the local

[1] In our empirical application of both Nadaraja-Watson estimator and the local linear kernel estimators of the TVC-HAR model for log(RV) of S&P500 return series were quantitatively very similar. Therefore, we will neither discuss the former method nor report the results. But they are available upon request from the authors.

linear method is superior in theory and applications among many nonparametric smoothing methods.

As is well-known, the local linear estimator is very sensitive to the choice of the bandwidth h, and thus it is critical to choose the optimum h in empirical applications. This paper employs the popular cross–validation (CV) method for the choice of a suitable bandwidth. Let $\widehat{\theta}_{-t}(\tau_t)$ be the local linear estimator of $\theta(\tau_t)$ by using the sample $\{\log RV_s : s \neq t\}$. The optimal bandwidth h_{opt} is chosen such that

$$CV(h) = \frac{1}{n} \sum_{t=1, s\neq t}^{n} \left(\log RV_t - X_t^{\top} \widehat{\theta}_{-t}(\tau_t) \right)^2$$

is minimized.

2.3 Bootstrap method Construction of Point-wise Confidence Intervals for Time-Varying Coefficients

For given $0 < \alpha < 1$, the $1 - \alpha$ confidence interval of $\beta_j(\tau)$ can be defined by

$$\widehat{\beta}_j(\tau) \pm c_{\alpha/2} \times SD(\widehat{\beta}_j(\tau)) \tag{10}$$

for $j = 0, d, w, m$, where $c_{\alpha/2}$ is the upper $\alpha/2$ percentile of $Q_j = \frac{\widehat{\beta}_j(\tau) - \beta_j(\tau)}{SD(\widehat{\beta}_j(\tau))}$ and $SD(\widehat{\beta}_j(\tau))$ is the standard deviation of $\widehat{\beta}_j(\tau)$. However, (10) cannot be directly used to construct the confidence interval of $\beta_j(\tau)$ as both $c_{\alpha/2}$ and $SD(\widehat{\beta}_j(\tau))$ are unknown.

We next estimate $c_{\alpha/2}$ and $SD(\widehat{\beta}_j(\tau))$ by using the bootstrap procedure (see, for example, Zhang and Peng, 2010). Then, by using the bootstrap estimates of $c_{\alpha/2}$ and $SD(\widehat{\beta}_j(\tau))$, denoted by $\widehat{c}^*_{\alpha/2}$ and $SD^*(\widehat{\beta}_j(\tau))$, the $1 - \alpha$ confidence interval of $\beta_j(\tau)$ can be obtained by

$$\widehat{\beta}_j(\tau) \pm \widehat{c}^*_{\alpha/2} \times SD^*(\widehat{\beta}_j(\tau))$$

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

1. Estimate $\beta(\tau)$ by the estimation method introduced above, and get the resulting estimator $\widehat{\beta}(\tau)$.

2. For each $t = 1, \cdots, n$, generate

$$\log RV_t^* = X_t^\top \widehat{\beta}(\tau) + \varepsilon_t^*,$$

where $\{\varepsilon_t^*\}_{t=1}^n$ is sampled from the centred nonparametric residuals $\{\widetilde{\varepsilon}_t\}_{t=1}^n$, with $\widetilde{\varepsilon}_t = \widehat{\varepsilon}_t - \overline{\widehat{\varepsilon}}_t$ and $\widehat{\varepsilon}_t = \log RV_t - X_t^\top \widehat{\beta}(\tau)$, $\overline{\widehat{\varepsilon}}_t = \frac{1}{n}\sum_{t=1}^n \widehat{\varepsilon}_t$. 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 $\widehat{\beta}(\tau)$. Denote the resulting estimator as $\widehat{\beta}^*(\tau)$.

3. Repeat Step 2 B times to obtain B bootstrap samples, $\widehat{\beta}^*(\tau, i), i = 1, \cdots, B$, of $\widehat{\beta}(\tau)$. The estimator of $SD(\widehat{\beta}_j(\tau))$ is the sample standard deviation of $\{\widehat{\beta}_j^*(\tau, i) : i = 1, \cdots, B\}$. Denote the resulting estimator as $SD^*(\widehat{\beta}_j(\tau))$.

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

2.4 Specification Testing of the Simple HAR against the Nonparametric TVC-HAR models

In this section, we introduce a test of the null hypothesis that the coefficients of HAR model are constant against the nonparametric TVC–HAR alternative. The testing procedure is motivated by the generalized maximum likelihood ratio test method introduced by Fan et al. (2001) and augmented with a bootstrap method. To find statistical evidence in favour of the proposed TVC-HAR model of $\log RV_t$ for the given S&P500 stock returns series, the null hypothesis H_0 that the simple HAR defined in (4) is the valid model specification for the returns series, is tested against the alternative hypothesis H_1 that the proposed nonparametric TVC-HAR defined in (5) is valid. We use the following test statistic (TS):

$$TS = \frac{RSS_0 - RSS_1}{RSS_1}, \tag{11}$$

where RSS_0 is the residual sum of square (RSS) under the null hypothesis, and RSS_1 is the RSS under the alternative hypothesis. Let $\widehat{\gamma}$ be an estimator of γ (for example, MLE or OLS), then $RSS_0 = \frac{1}{n}\sum_{t=1}^{n}\widehat{\varepsilon}_{t0}^2$ with $\widehat{\varepsilon}_{t0} = \log RV_t - X_t^\top \delta(\tau_t, \widehat{\gamma})$, and $RSS_1 = \frac{1}{n}\sum_{t=1}^{n}\widehat{\varepsilon}_{t1}^2$, where $\widehat{\varepsilon}_{t1} = \log RV_t - X_t^\top \widehat{\beta}(\tau_t)$. The null hypothesis is rejected for a p–value that is smaller than the nominal level. The p–value is computed using the wild bootstrap procedure, which is outlined as follows:

1. For each $t = 1, \cdots, n$, generate $\log RV_t^* = X_t^\top \delta(\tau_t, \widehat{\gamma}) + \varepsilon_t^*$, where the definition of $\{\varepsilon_t^*\}$ is the same as that in Step 1 of the bootstrap procedure for the construction of the confidence interval.

2. Use the data set $\{(\log RV_t^*, X_t) : t = 1, \cdots, n\}$ to estimate γ and $\beta_j(\cdot)$, the resulting estimates are $\widehat{\gamma}^*$ and $\widehat{\beta}_j^*(\cdot)$. Calculate the corresponding RSS_0^* and RSS_1^*, and then compute the TS^* by (11).

3. Repeat Steps 1 and 2 B times to obtain the empirical distribution for TS^*. Then, the p–value of the test is computed by $\frac{1}{B}\sum_{i=1}^{B} I(TS_i^* \geq TS)$, where $I(\cdot)$ is an indicator function.

2.5 Calibration of Nonparametric TVC–HAR Model

As said the proposed nonparametric TVC-HAR model will capture the underlying true nature of these time-varying parameters. However, in general, one would expect the multi-step outer-sample forecasting by this nonparametric model to be more time consuming than its parametric counterpart. If a parametric TVC-HAR model of $\log RV_t$ for the data can be estimated by calibrating nonparametric TVC-HAR model, and it outperforms the simple HAR model consistently in outer-sample forecasting, then the calibrated TVC-HAR model would be the winner in empirical applications.

Consider the following specification for the parametric TVC–HAR model,

$$\log RV_t = \beta_0^*(\tau_t) + \beta_d^*(\tau_t)\log RV_{t-1} + \beta_w^*(\tau_t)\log RV_{t-5,t-1} + \beta_m^*(\tau_t)\log RV_{t-22,t-1} + \varepsilon_t, \quad (12)$$

where the coefficient functions $\beta_j^*(\cdot)$ are unknown parametric functions of τ_t. To achieve the optimal parametric TVC specification for the HAR model which is as close as possible to the nonparametric counterpart, we propose an automated model selection algorithm for calibrating the nonparametric TVCs by minimising the L_2- type distance function proposed by Härdle and Mammen (1993), which is defined as follows:

$$\min\{\frac{1}{n}\sum_{t=1}^{n}[\widehat{m}_h(x_t) - \widehat{m}_{\widehat{\theta}}(x_t)]^2\}$$

where $\widehat{m}_h(x_t)$ is the nonparametric estimator, and $\widehat{m}_{\widehat{\theta}}(\cdot)$ is the estimator of the corresponding parametric specification, which can be a higher order polynomial as in our empirical application.

The automated estimation of parametric TVCs is set up as outlined follows:

- Estimate the TVC-HAR model (4) by the local linear method discussed in Section 2.2. Let this estimator be $\widehat{m}_h(x_t)$;
- Define $D(\theta) = \frac{1}{n}\sum_{t=1}^{n}[\widehat{m}_h(x_t) - \widehat{m}_{\widehat{\theta}}(x_t)]^2$;
- Find the optimal $\widehat{\theta}$ by minimizing $D(\theta)$ for various higher order polynomials for each of the coefficients.

The empirical results presented below show that the optimal calibration resulted from the above algorithm provides a good parametric approximation of the nonparametric estimators of the time-varying coefficients of HAR model. In what follows, the relative merits of nonparametric TVC-HAR, calibrated TVC-HAR and the simple HAR models will be assessed in-sample specification testing and outer-of-sample forecasting.

3 FORECASTING VOLATILITY

In this section we assess the out-of-sample forecasting performance of the proposed nonparametric TVC-HAR model of $\log RV$ relative to the simple HAR and other competing forecasting models, such as HAR-GARCH. To do this, parametric and nonparametric forecasting methods will be employed.

3.1 Parametric Multi-step-ahead Forecasting

We partition the total data sample, of size T, into an estimation period ($t = 1, 2, ..., K$) and an evaluation period ($t = K + 1, K + 2, .., T$), as follows,

$$t = \underbrace{1, 2, ..., K,}_{\text{estimation period}} \quad \underbrace{K + 1, K + 2, .., T}_{\text{evaluation period}}$$

The HAR and TVC-HAR are nested models, and they are effectively AR(22) in specification. Traditionally, multi-step-ahead forecasting with parametric autoregressive models is made via iterative methods. This is also the approach that is adopted in this paper.

We generate the volatility point forecasts using a rolling window, of fixed length K, estimation scheme. At time K, the parameters, $\beta_{j,K}$, of model (1) are estimated using the first K observations, the δ-step-ahead out-of-sample forecasts are formulated and compared to the realization $\log RV_{K+\delta}$. At time $K + 1$, $\widehat{\beta}_{j,K+1}$ are estimated using the previous K observations, the second set of δ-step-ahead forecasts are formulated and compared to the realization $\log RV_{K+\delta+1}$. Iterating this procedure to generate $m \equiv T-\delta-K+1$ out-of-sample forecasts and relative forecast errors[2].

In summary, setting $\delta = 1, 5$ and 22 as the daily, weekly and monthly forecasting horizons, the conditional forecasts of $\log RV_{t+\delta}$ at any time point $t \in (K, K+1, .., T-\delta+1)$, are computed recursively as follows:

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

where $\mathcal{F}_t$ represents the information set available up to time t.

3.2 Nonparametric multi-step-ahead Forecasting

Nonparametric multi-step-ahead forecasting for non-linear AR(d) models is made by estimating the conditional mean $E(Y_{t+\delta}|Y_t, \ldots, Y_{t-d+1})$ via nonparametric smoothing of $Y_{t+\delta}$ on $(Y_t, \ldots, Y_{t-d+1})$ directly; see Robison (1983), Härdle and Vieu (1992), among others, for detailed discussion on the direct smoothing techniques. These direct nonparametric estimators, however, ignore the substantial information contained in the intermediate variables $Y_{t+1}, \ldots, Y_{t+\delta-1}$ about the conditional mean. To improve this estimator, Chen, Yang and Hafner (2004) introduced a multi-stage nonparametric predictor, which utilises information in pseudo observations $Y^*_{t+1}, \ldots, Y^*_{t+\delta-1}$ to generate an estimate for $Y_{t+\delta}$. Their paper showed that this multi-stage smoother improves the estimation of the conditional mean, and demonstrated that this new predictor is more efficient than the direct smoother.

In our empirical application, we employ the multistage nonparametric predictor introduced by Chen et al. (2004) for the conditional forecasts of $\log RV_{t+\delta}$ in the nonparametric

[2]Following Patton (2011), we measure the relative forecast error using the simple mean squared error (MSE) loss function. The paper derives necesary and sufficient conditions for the loss function to yield rankings of volatility forecasts that are robust to noise in the volatility proxy, such as realized volatility measure. It also proposes a new family of robust and homogeneous loss functions, which nest both squared-error and the "QLIKE" loss functions. For more detailed discussions, refer to the paper.

TVC-HAR specification (4), with adjustments made to the predictor to account for the time-varying nature of coefficient functions. The approach to nonparametric multi-step-ahead forecasting is described as follows:

1. Use the first K observations of $\{Y_t, X_t\}$ sample values and estimate the nonparametric TVC-HAR (4) by the method introduced in Section 2.2. Let the estimated optimum bandwidth be h_{opt}.

2. For one-step-ahead forecast at time t (and $\tau = \frac{t}{T}$), apply the bandwidth h_{opt} from step 1 and obtain the local linear least squares estimates $\widehat{\beta}_{j,t}$ of $\beta_{j,t}$, using the past K observations up to time t. Use these $\widehat{\beta}_{j,t}$ 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 t (and $\tau = \frac{t+1}{T}$), update the vector of regressor values, of fixed length K, by including the pseudo observation $\widehat{Y}^*_{t+1}$. Then, apply the same bandwidth h_{opt}, and estimate the parameters of this new model. Let this estimator be $\widehat{\beta}_{j,t+1}$. Use this estimator to generate the two-step-ahead forecast, say $\widehat{Y}^*_{t+2}$, and compare it to the realization $\log RV_{t+2}$.

4. Similarly, for δ-step-ahead forecast at t (and $\tau = \frac{t+\delta-1}{T}$), update the set of regressors by adding the pseudo observations $\widehat{Y}^*_{t+1}, ..., \widehat{Y}^*_{t+\delta-1}$. Apply the same bandwidth h_{opt} and estimate the updated model. Let this estimator be $\widehat{\beta}_{j,t+\delta-1}$. Use this estimator to generate $\widehat{Y}^*_{t+\delta}$, which is compared to the realization $\log RV_{t+\delta}$.

For simplicity, we assume the same optimum bandwidth, h_{opt} from Step 1, across different rolling samples of fixed length K. In practice however, one could re-estimate the optimal bandwidth h_{opt}, as it may differ as the data sample changes. The obvious advantage of bandwidth re-estimation is the gain (which may not be significant) in the forecast accuracy, but it comes at a cost of the significant increase in the computation time for nonparametric forecasting.

3.3 Evaluation of Volatility Forecasts: Conditional Predictive Ability (CPA) Testing

It is of practical importance to assess and compare the out-of-sample forecasting performance of competing models. To evaluate the forecasting ability of the TVC–HAR models against the simple HAR as well as other competing realized volatility forecasting models, we conduct the test of conditional predictive ability (CPA) of Giacomini and White (2006).

Relative to the early studies in the literature of out-of-sample predictive ability evaluation, represented by Diebold and Mariano (1995) and West (1996), henceforth referenced as DMW, the CPA framework provides a forecast evaluation criterion when the forecasting model(s) may be unwittingly misspecified[3], which is common in most practical applications. In this paper, we do not claim the TVC-HAR model to be the true underlying data generating process (DGP) for $\log RV_t$. It is a well known fact that the realized volatility exhibits volatility clustering pattern, which is adequately captured by adding a GARCH componet to the HAR model (Corsi et al., 2008). In the empirical forecasting exercise, the proposed TVC-HAR models are compared with HAR as well as HAR-GARCH model. To this end, the CPA test has an advantage over the DMW approach in that it is well suited for comparing forecasting methods based on both nested and nonnested models. In addition, the CPA test can be applied to multi-step point, interval, probability or density forecast valuation for a general loss function. Due to the fact that $\log RV_t$ being a long memory process, for a thorough forecasting evaluation, we follow Andersen et al. (2007) for formulating and evaluating the $1, 5$ and $22-$step–ahead forecasts of the TVC-HAR and the HAR models. In addition,

[3]Following Giacomini and White (2006), forecasting model is possibly misspecified, due to unmodeled dynamics, unmodeled heterogeneity, incorrect functional form, or any combination of these.

the CPA approach accommodates conditional forecast evaluation objectives (i.e. which forecast will be more accurate at a specific future date), and it nests the unconditional objectives (i.e. which forecast was more accurate on average) of the DMW approach. Although both unconditional and conditional approaches are informative, looking at the global (or average) relative forecasting performance may hide important information about the relative forecasting performance over time. The foregoing discussion justifies the use of CPA test for assessing the merit of the TVC-HAR models against the simple HAR and HAR-GARCH models in terms of forecast accuracy. The CPA test is conducted via the following steps:

1. Based on the rolling samples of fixed length K, conditional forecasts $\widehat{f}_{1,t+\delta}|\Omega_t$ and $\widehat{f}_{2,t+\delta}|\Omega_t$, of the TVC-HAR and HAR (-GARCH), respectively, for a conditioning set $\Omega_t = \mathcal{F}_t$, are generated for the target date $t+\delta$ of the evaluation period at $t = K, K+1, \ldots, T-\delta$.

2. For both forecasting models, generate a sequences of losses, $L_{p,t+\delta}|\mathcal{F}_t = L(Y_{t+\delta}, \widehat{f}_{j,t+\delta})|\mathcal{F}_t$, with $p=1$ denoting the benchmark simple HAR or HAR-GARCH model, and $p=2$ the alternative TVC-HAR models. The two forecasts are compared via the time series of loss differentials, $\Delta L_{t+\delta}|\mathcal{F}_t = L_{1,t+\delta}|\mathcal{F}_t - L_{2,t+\delta}|\mathcal{F}_t$.

3. A test of whether or not the benchmark model is outperformed by the alternative model is conducted by testing $H_0 : E[\Delta L_{t+\delta}|\mathcal{F}_t] = 0$ against $H_1 : E[\Delta L_{t+\delta}|\mathcal{F}_t] > 0$, using the test statistic $T_{K,m} = m\overline{Z}_{K,m}^{\top}\widehat{\Theta}_m^{-1}\overline{Z}_{K,m}$, where $\overline{Z}_{K,m} = \frac{1}{m}\sum_{t=K}^{T-\delta} Z_{t+\delta}$, $Z_{t+\delta} = \pi_t \Delta L_{t+\delta}$, and π_t is a chosen test function[4], $\widehat{\Theta}_{K,m} = \frac{1}{m}\sum_{t=K}^{T-\delta} Z_{t+\delta}Z_{t+\delta}^{\top} + \frac{1}{m}\sum_{l=1}^{\delta-1} w_{m,l} \times \sum_{t=K+l}^{T-\delta}[Z_{t+\delta}Z_{t+\delta-l}^{\top} + Z_{t+\delta-l}Z_{t+\delta}^{\top}]$, and $w_{m,l}$ is a weight function such that $w_{m,l} \to 1$ as

[4]In practice, as discussed in Giacomini and White (2006), π_t is chosen by the researcher 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) or Stinchcombe and White (1998) for ways of choosing test function.

$m \to \infty$ for each $l = 1, ...\delta - 1$.

A level α test can be conducted by rejecting the null hypothesis of equal conditional predictive ability of the models when $T_{K,m} > \chi^2_{q,1-\alpha}$, where $\chi^2_{q,1-\alpha}$ is the $(1-\alpha)$ quantile of a χ^2_q distribution. As noted by the authors, rejection occurs when the test functions $\{\pi_t\}$ can predict the out of sample loss differences $\{\Delta L_{t+\delta}\}$. Furthermore, in cases where the null hypothesis of equal conditional predictive ability is rejected, Giacomini and White (2006) also proposed an approach to make forecast model selection decisions. To assess the relative performance of the forecasting model under the null, their approach can be applied in the following three steps:

1. Regress $\Delta L_{t+\delta}|\mathcal{F}_t$ on π_t over the out-of-sample period for $t = K, K+1, \ldots, T-\delta$, and let $\widehat{\varphi}_m$ denote the regression coefficient matrix. Apply the above χ^2_q test and if the null is rejected, then proceed to step 2.

2. Approximate $E[\Delta L_{t+\delta}|\mathcal{F}_t]$ using $\widehat{\varphi}_m^\top \pi_t$, and model 1, with the lower loss, is considered superior if $\widehat{\varphi}_m^\top \pi_t < c$, and model 2 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 $\frac{\sum_{t=K}^{T-\delta} I\{\widehat{\varphi}_m^\top \pi_t > 0\}}{m}$, the relative out-of-sample performance of models 1 and 2, where $I\{\cdot\}$ is an indicator function. Thus, model 1 is a better forecasting model at $t + \delta$ if the ratio < 0.5, and model 2 otherwise.

4 ANALYSIS OF THE US MARKET DATA

4.1 Data and Preliminary Analysis

Our primary data set consists of tick–by–tick transaction prices for the $S\&P500$ Index for the period from 10/May/1999 to 26/October/2010. All Index data has been supplied by the Securities Industries Research Centre of Asia Pacific (SIRCA) on behalf of Reuters, with the raw index data having been filtered[5] prior to the construction of realized volatility 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 nonoverlapping 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 minimises 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.

(←--Insert Figure 1 here--→)

The realized volatility series exhibits the volatility clustering effect that is well documented in the literature. The daily $\log RV_t$ is generally much closer to being normally distributed than the raw realized volatility series. From a time series modelling perspective, the $\log RV_t$ is more conformable to standard time series analysis techniques. In line with the literature, we find that the $\log RV_t$ series is approximately normally distributed. Further, following the semiparametric local whittle estimator[6] of Robinson (1995), the estimated

[5] The key aspects of the data cleansing of the raw S&P500 index data are as follows: (1) Only data recorded between 10am and 4pm are used, with a view to eliminate the effect of "off-market" transactions; (2) Omitting days on which the trading was less than 3 hours; (3) Median index value is used in cases of simultaneous recording; (4) Abnormally large jumps in the index value that were reversed within a short time interval, are omitted.

[6] There are other semiparametric estimator for the memory parameter d in the literature, such as log periodogram and averaged periodogram estimators. However, local whittle estimator is the most preferred

memory parameter d of the $\log RV_t$ series is approximately 0.5376, i.e. $\widehat{d} > 0.5$, by definition, the volatility series is nonstationary[7]. The time series plot in Figure 1 does not appear to be stationary, it has a significant structural change after the crisis. In addition, the $\log RV_t$ series exhibits highly significant serial correlations, as evidenced by the Ljung–Box test statistic of 15222 for up to tenth order autocorrelation. The plot of sample autocorrelation functions (ACF) of the $\log RV_t$ series in Figure 1 exhibits the characteristic hyperbolic decay pattern, with autocorrelation coefficients being significant up to the 260^{th} order, according to the 95% Bartlett confidence intervals. These results indicate the presence of long memory in $\log RV_t$.

To approximate the long memory process for the $\log RV_t$, we first fit the HAR model (1) to the data set, and the least–square estimates of the coefficients are presented in Table 1.

(←--Insert Table 1 here--→)

All coefficients are significantly different from zero at the 1% significance level. Furthermore, to examine if the coefficients, β_d, β_w and β_m indeed change significantly over time, we depict the moving window plots (with window size $n = 200$) of bivariate correlations between $\log RV_t$ and $\log RV_{t-1}$, $\log RV_{t-5,t-1}$, $\log RV_{t-22,t-1}$, respectively. These plots are given in Figure 2.

(←--Insert Figure 2 here--→)

It appears that the all correlation measures are time–varying. All three moving window plots of correlations exhibit nonlinear patterns over the sample period. When the three

method due to its satisfying statistical properties. In addition, it is well known the above mentioned estimators are sensitive to the choice of bandwidth. Henry (2001) provides data dependent method for the choice of bandwidth, and this is also the choice in our study.

[7]However, if the period of global financial crisis (after 30/Jun/2007) is removed from the sample, the $\log RV_t$ series becomes stationary, with the estimated $\widehat{d} \approx 0.4727$.

curves are plotted together in a single graph, we see that these correlation plots move together in unison over time. In particular, all three correlation measures exhibit increasing trends during crises (the 2001 dotcom crisis and the 2008 global financial crisis), but move downwards and stay low during the tranquil period between 2003 and early 2007.

We subdivided the full sample period into four different subsample periods, an (unrestricted) HAR model[8] is fitted for the subsamples, and the results are presented in Table 1. Subsample 1 covers the periods before the dotcom crisis, subsample 2 represents the dotcom crisis, subsample 3 covers the period before the global financial crisis and subsample 4 represents the recent global financial crisis period. The least-squares estimates and standard errors of the coefficients of the HAR model for each subsample are also listed in Table 1. A wald test of the equality[9] of the regression coefficients across the four subsamples indicates inequality of these coefficients over time, with the test statistic 61.1675 and associated $p-$value 0. We propose a time-varying coefficient HAR model to capture the changes in the model coefficients over time. Moreover, the ACF plots of $\log RV_t$ series for all subsamples are presented in Figure 3, which exhibit different patterns across the subsamples, indicating the dynamic changes of the serial dependencies in the $\log RV_t$ series.

(←--Insert Figure 3 here--→)

The above preliminary analysis indicates that the TVC–HAR model (4) may be a better alternative to HAR specification for $\log RV_t$ series.

[8] The un-restricted HAR model is of the form: $\log RV_t = D_1 \times (\beta_{0,1} + \beta_{d,1} \log RV_{t-1} + \beta_{w,1} \log RV_{t-5,t-1} + \beta_{m,1} \log RV_{t-22,t-1}) + D_2 \times (\beta_{0,2} + \beta_{d,2} \log RV_{t-1} + \beta_{w,2} \log RV_{t-5,t-1} + \beta_{m,2} \log RV_{t-22,t-1}) + D_3 \times (\beta_{0,3} + \beta_{d,3} \log RV_{t-1} + \beta_{w,3} \log RV_{t-5,t-1} + \beta_{m,3} \log RV_{t-22,t-1}) + D_4 \times (\beta_{0,4} + \beta_{d,4} \log RV_{t-1} + \beta_{w,4} \log RV_{t-5,t-1} + \beta_{m,4} \log RV_{t-22,t-1}) + \varepsilon_t$, where D_1, D_2, D_3 and D_4 are binary dummy variables and equal to 1 when $t \in$ subsample $1, 2, 3$ and 4, respectively, 0 otherwise.

[9] The null of the wald test is: $\beta_{0,1} = \beta_{0,2} = \beta_{0,3} = \beta_{0,4}, \beta_{d,1} = \beta_{d,2} = \beta_{d,3} = \beta_{d,4}, \beta_{w,1} = \beta_{w,2} = \beta_{w,3} = \beta_{w,4}, \beta_{m,1} = \beta_{m,2} = \beta_{m,3} = \beta_{m,4}$.

4.2 Empirical Results of Nonparametric Estimation of TVC–HAR model

We use both the local constant and local linear methods to estimate the TVC–HAR model. The bandwidths for these two methods are selected by using the CV method discussed in Section 2.2, and they are found to be 0.1444 and 0.1862 respectively. The patterns exhibited by estimates from both estimators are similar, and the time–varying coefficients from local linear method are plotted in Figure 5.

(←--Insert Figure 5 here--→)

The intercept function $\beta_0(\cdot)$ is high during the crisis time due to the higher perceived risk level from the investors, and low during the less volatile period, reflecting the dynamic change of the $\log RV_t$ measure. The other three plots are related to coefficients that measure the respective effect of $\log RV_{t-1}$, $\log RV_{t-5,t-1}$, $\log RV_{t-22,t-1}$ on $\log RV_t$. Note that all three coefficient plots lie above the zero horizontal line, and a rise in the plot corresponds to an increase in the effect of respective regressor on $\log RV_t$. $\beta_d(\cdot)$ measures the effect of $\log RV_{t-1}$ on $\log RV_t$, and the corresponding plot indicate that $\log RV_{t-1}$ becomes a more important factor in explaining $\log RV_t$ during the crisis periods, but less so during the tranquil period. However, the opposite patterns emerge for the coefficient $\beta_w(\cdot)$, which measures the relevance of the $\log RV_{t-5,t-1}$: it has a greater importance when the financial market is stable, than when the market is volatile. The plot of the estimate of $\beta_m(\cdot)$ shows a downwards trend until 2004, indicating declining effect of $\log RV_{t-22,t-1}$ on $\log RV_t$ during this period. $\beta_m(\cdot)$ reached its trough during the period 2004–2006, but increased on the rise of the global financial crisis, and peaked around mid–2008 when the crisis was deepening. In short, all coefficients of the HAR model are time varying.

The regression residuals from both nonparametric estimators, as shown in Figure 4, are close to normal. Overall, it appears that the local linear estimator generates smoother patterns in the estimated time varying coefficients, largely due to a larger estimated optimum bandwidth. The residuals output from the local linear estimator seems to be closer to normal, but the difference is minimal. It is well documented that the local linear method is superior in theory and applications among many nonparametric smoothing methods. In what follows, we report on the findings based on the local linear estimator.

(←--Insert Figure 4 here--→)

To evaluate the model fitting performance, the estimated mean squared error (MSE) measures for the simple HAR model and the local linear TVC–HAR model are 0.2692 and 0.2606, respectively, see Table 3. The local linear TVC–HAR model offers 3.19% improvement of overall fit, compared to the simple HAR model.

`Model Misspecification Test for the Simple HAR model`

The 90% pointwise confidence intervals of the local linear estimator, are computed using the bootstrap method discussed in Section 2.2, and the results are plotted in Figure 5. There are four panels in this plot, each represents, respectively, a coefficient function of the TVC-HAR model. The least-square estimates of the HAR coefficients are also plotted as the horizontal lines in each of the panels. We find that, the estimates of constant coefficients, β_0, β_d and β_w of the simple HAR model (1) do not lie inside the confidence intervals at all times, with all three coefficients are clearly outside the intervals most of the time during the period between 2002 and 2006, and those of β_d and β_w lie partially outside the intervals during the GFC period. The coefficient function β_m, which lies inside the confidence intervals almost at all times for the selected sample period, was very close to and at times touching the upper bound of confidence intervals after 2004. These results indicate that, if tested jointly, the coefficients of the simple HAR model are likely to be time varying[10].

To examine, statistically, whether or not the popular simple HAR model is adequate for the data, we test the null hypothesis of constant coeffients that

$$H_0: \ \beta_0(\tau) = \gamma_0, \beta_d(\tau) = \gamma_1, \beta_w(\tau) = \gamma_2, \beta_m(\tau) = \gamma_3.$$

against the nonparametric time–varying alternative. The testing procedure is described in Section 2.3. With $B = 200$, the wild bootstrap method produces a p–value of approximately 0. Therefore, when testing the coefficients jointly, we do not have enough evidence to support the simple HAR model against the nonparametric TVC–HAR specification.

4.3 Results of the Conditional Predictive Ability (CPA) Test

In the comparative forecasting evaluation of the local linear TVC-HAR (TVC-HAR-LL) model against the simple HAR or HAR-GARCH model, β_j of the volatility models are estimated using the rolling samples of K observations, the conditional point forecasts of $\log RV_t$ are formulated and compared them with the realizations to generate the measures of loss function $L_{t+\delta}$ for the three different forecasting horizons: $\delta = 1, 5$ and 22. Following Giacomini and White (2006) described in Section 3.3, we choose the test function $\pi_t = (1, \Delta L_{t+\delta-1})$, where $\Delta L_{t+\delta-1}$ is the lagged value of the loss difference $\Delta L_{t+\delta}$ for the CPA testing method.

For nonparametric forecasting, the optimum bandwidth h_{opt} is estimated using the first K observations and, for simplicity, the same bandwidth is used throughout the forecasting exercise. To cater for the practice of single optimum bandwidth in the nonparametric forecasting, we perform stability check on the bandwidth and the below data sampling is

[10] We also observe the unusual wide confidence intervals for $\beta_0(\cdot)$ between early 2004 and mid–2007. To explore it further, a sub–sampling method was also considered, the findings were quite similar. Such issues might be left for future research.

chosen, such that the bandwidth is within 10% variance across the estimation and forecating evaluation periods.

	Estimation (K)	Evaluation (m) pre-GFC	Evaluation (m) 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

4.3.1 HAR-GARCH versus Simple HAR

At the time of introducing the HAR-GARCH model, Corsi et al. (2008) also compares its one-step-ahead out-of-sample forecasting performance against the simple HAR model. By computing and comparing the selected loss measures, the authors found overall improved accuracy in volatility point forecasting by allowing for time-varying volatility of $\log RV_t$. To validate this result for our data series, we perform the δ-step-ahead out-of-sample forecasting, and evaluate the performance of the volatility model using the CPA testing procedure. The simple HAR model is treated as the benchmark model, hence Model 1 in the CPA test. Tables 4 lists the test results.

(←--Insert Table 4 here--→)

For the full forecasting evaluation period, the $p-$values for the test statistic are 0.0025, 0.0034 and 0.0739 for $\delta = 1, 5$ and 22, respectively. For $\alpha = 0.05$, the null hypothesis of equal conditional predictive ability is not rejected for $\delta = 22$, but rejected for $\delta = 1$ and 5. The HAR-GARCH model is as good as the simple HAR model when performing point forecasting for $\log RV_t$ for $\delta = 22$, but inferior to the simple HAR model when forecasting for $\delta = 1$ and 5, with the simple HAR model being a better forecasting model 97.74% and 64.91% of times, repectively.

Since the evaluation period covers both pre- and post-GFC time-frames, it is of interest to know how the relative forecasting performance of two models would change. We found that, for the pre-GFC evaluation period, the $p-$values for the test statistic are $0.0137, 0.0835$ and 0.0064 for $\delta = 1, 5$ and 22, respectively. The HAR-GARCH model is as good as the simple HAR model when performing point forecast for $\log RV_t$ for $\delta = 5$, but inferior to the simple HAR model when forecasting for $\delta = 1$ and 22, with the simple HAR model being a better forecasting model 93.97% and 64.32% of times, respectively. For the post-GFC evaluation period, the $p-$values for the test statistic are 0.0679, 0.0016 and 0.0276 for $\delta = 1, 5$ and 22, respectively. The HAR-GARCH model is as good as the simple HAR model when performing point forecast of $\log RV_t$ for $\delta = 1$, but inferior to the simple HAR model when forecasting further out for $\delta = 5$ and 22, with the simple HAR model being a better forecasting model 59.30% and 67.34% of times.

4.3.2 TVC-HAR-LL versus Simple HAR

The simple HAR model is treated as the benchmark model. The results of the CPA test are reported in Table 5a).

(←--Insert Table 5 here--→)

For the full forecasting evaluation period, the $p-$values for the test statistic are 0.4974, 0.0000 and 0.0016 for $\delta = 1, 5$ and 22, respectively. Our TVC-HAR-LL model is as good as the simple HAR model when performing point forecast for $\log RV_t$ at near future date for $\delta = 1$, but superior to the simple HAR model when forecasting further out for $\delta = 5$ and 22, with TVC-HAR-LL model being a better forecasting model 62.91% and 57.14% of times, repectively.

For the pre-GFC evaluation period, the $p-$values for the test statistic are 0.5575, 0.0000 and 0.0907 for $\delta = 1, 5$ and 22, respectively. The TVC-HAR-LL model is as good as the simple HAR model when performing point forecast of $\log RV_t$ for $\delta = 1$ and 22, but superior to the simple HAR model when forecasting for $\delta = 5$, with TVC-HAR-LL model being a better forecasting model 57.79% of times. For the post-GFC evaluation period, the $p-$values for the test statistic are 0.7656, 0.0000 and 0.0008 for $\delta = 1, 5$ and 22, respectively. The TVC-HAR-LL model is as good as the simple HAR model when performing point forecast of $\log RV_t$ for $\delta = 1$, but superior to the simple HAR model when forecasting further out for $\delta = 5$ and 22, with the TVC-HAR-LL model being a better forecasting model 65.83% and 73.37% of times.

4.3.3 TVC-HAR-LL versus HAR-GARCH

The HAR-GARCH model is treated as the benchmark model. The results of the CPA test are reported in Table 5b).

For the full forecasting evaluation period, the $p-$values for the test statistic are 0.2562, 0.0000 and 0.0041 for $\delta = 1, 5$ and 22, respectively. Our TVC-HAR-LL model is as good as the HAR-GARCH model when performing point forecast for $\log RV_t$ at near future date for $\delta = 1$, but superior to the HAR-GARCH model when forecasting further out for $\delta = 5$ and 22, with TVC-HAR-LL model being a better forecasting model 60.40% and 55.64% of times, repectively.

For the pre-GFC evaluation period, the $p-$values for the test statistic are 0.4134, 0.0001 and 0.0972 for $\delta = 1, 5$ and 22, respectively. The TVC-HAR-LL model is as good as the HAR-GARCH model when performing point forecast of $\log RV_t$ for $\delta = 1$ and 22, but superior to the HAR-GARCH model when forecasting for $\delta = 5$, with TVC-HAR-LL model being a better forecasting model 58.79% of times. For the post-GFC evaluation period, the $p-$values for the test statistic are 0.5026, 0.0001 and 0.0035 for $\delta = 1, 5$ and 22, respectively. The TVC-HAR-LL model is as good as the HAR-GARCH model when performing point forecast of $\log RV_t$ for $\delta = 1$, but superior to the HAR-GARCH model when forecasting further out for $\delta = 5$ and 22, with the TVC-HAR-LL model being a better forecasting model 60.80%

and 66.33% of times.

In summary, the CPA test indicates that when compared with the simple HAR and HAR-GARCH models, the TVC-HAR-LL model is a better point forecasting model for $\log RV_t$ for forecating horizons that are further out. With the varying nature of the coeffient functions, the TVC-HAR-LL model is more capable, than the HAR and HAR-GARCH, of capturing the dynamic changes in the data, this is demonstrated by the improved forecasting performance of the TVC-HAR-LL in the post-GFC period, where the volatility series fluctuates in a more dramatic fashion. The resutls of the CPA test also show that the simple HAR is a superior volatility point forecasting model to the HAR-GARCH model, a result that is different to the findings in Corsi et al. (2008).

4.4 Calibration of TVC–HAR Model

Based on the patterns observed in the nonparametric estimator of the TVC-HAR model, one can apply the polynomial and other nonlinear parametric families to approximate the estimated nonparametric coefficient functions. In this paper, we consider higher-order polynomial functions of time t for the coefficient functions of the TVC–HAR model.

To calibrate the local linear estimators, we applied the automated algorithm discussed in Section 2.5 for fitting the 7^{th} order polynomial function for $\beta_0^*(\tau_t)$, and 2^{nd}, 8^{th} and 8^{th} order polynomials for $\beta_d^*(\tau_t)$, $\beta_w^*(\tau_t)$ and $\beta_m^*(\tau_t)$ respectively. The least square estimates of the parameters of the β_j^*'s are listed in Table 2, and are all, except the intercept coefficients of $\beta_w^*(\tau_t)$ and $\beta_m^*(\tau_t)$[11], significant at the 10% level. The plots of the calibrated coefficient functions are illustrated in Figure 6. However, not all the polynomial functions appear to adequately fit the patterns of local linear estimators.

[11] The local linear estimator of the TVC-HAR model, indicates non-zero intercept for all the coefficient functions. For this reason, as part of the calibration algorithm, the intercept coefficients of $\beta_{j,1}^*$ are not to be removed even if the p-value is large.

(←--Insert Table 2 here--→)

(←--Insert Figure 6 here--→)

The validity of the calibrated model is also tested against the nonparametric alternative, following the similar specification testing procedure in Section 2.4. We consider testing the null hypothesis

$$H_0: \ \beta_j(\tau) \text{ follows the parametric fittings as in model (12).}$$

against the nonparametric alternative. The wild bootstrap method, with $B = 200$, produces a p–value that is approximately 0.4950. Therefore, we do not have sufficient statistical evidence to reject the calibrated parametric TVC–HAR model against the nonparametric TVC–HAR specification.

To further assess the merit of the above calibrated TVC–HAR models as volatility forecasting models, we perform out-of-sample forecasting exercise. The forecasting performance is also evaluated using the CPA test. The results of CPA test are summarized in Table 6.

(←--Insert Table 6 here--→)

Clearly, all the predictive performance measures associated with the calibrated TVC-HAR are notably lower than those of its nonparametric counterpart, reported in Table 5. In particular, for the full evaluation period considered, the calibrated TVC-HAR model is the worst performer in out-of-sample forecasting overall, in comparison to both the simple HAR and HAR-GARCH models. Moreover, it has the same performance as both of these models at $\delta = 1$ and 22, while under performs at $\delta = 5$. We also find the calibrated TVC-HAR model to have the same performance as the both simple HAR and HAR-GARCH in the pre-GFC period at $\delta = 1$ and 22, but under-performs at $\delta = 5$. For post–GFC period, on the other hand, the calibrated TVC–HAR model is as good as the simple HAR and HAR-GARCH model at all forecasting horizons considered.

Due to lack of out-of-sample forecasting performance of the above calibrated TVC-HAR model, our intention is to improve its parametric calibrations that would adequately approximate the nonparametric coefficient functions of TVC-HAR. To do this, we calibrate the polynomial functions of time with an art of "trial and error" rather than using the automated algorithm. Based on the observed patterns of local linear estimator, we calibrated a 2^{nd} order polynomial function for $\beta_0^*(\tau_t)$, and 4^{th}, 3^{rd} and 2^{nd} order polynomials for $\beta_d^*(\tau_t)$, $\beta_w^*(\tau_t)$ and $\beta_m^*(\tau_t)$ respectively. The least square estimates of the parameters of β_j^*'s are listed in Table 3, and are all significant at the 10% level. The calibrated coefficient functions are plotted in Figure 7. Although not perfect, these polynomial functions appear to adequately approximate the patterns of local linear estimators for each of the coefficients. When plotted against the bootstrapped 90% confidence intervals for the local linear estimator, we find that all the calibrated beta functions fully lie inside the confidence intervals. This provide statistical evidence that the calibrated polynomial functions of τ_t are indeed valid specifications for the $\beta_j(\tau)$'s.

(←--Insert Table 3 here--→)

(←--Insert Figure 7 here--→)

The validity of the calibrated model is also tested against the nonparametric alternative. The wild bootstrap method, with $B = 200$, produces a p–value that is approximately 0.495. Therefore, we do not have sufficient statistical evidence to reject the calibrated parametric TVC–HAR model against the nonparametric TVC–HAR specification, indicating that the former is the better in-sample fit than the latter. In the out-of-sample forecasting exercise, the relative performances are also evaluated by the CPA test and the results are summarized in Table 7.

(←--Insert Table 7 here--→)

For all the evaluation periods considered, the calibrated TVC-HAR model is overall the worst performer relative to the simple HAR or HAR-GARCH model in the out-of-sample forecasting. It has the same performance as HAR and HAR-GARCH at $\delta = 1$, but it under-performs at $\delta = 5$ and $\delta = 22$.

In summary, calibration of nonparametric TVC model appear to be more of an art than a skill. Despite our two attempts, we have not still come up with parametric specifications for the coefficient functions of the TVC-HAR to outperform the local linear counterpart in out-of-sample forecasting. However, the calibrated TVC-HAR models for $\log RV_t$ outperforms the other parametric models as well as the the local linear counterpart in in-sample fitting. There is a need to come up with an improved method for this calibration so that its out-of-sample forecasting performance would be superior to all other models studied in this paper, which is a topic for the future research.

5 CONCLUSION

In this paper, we introduce a nonparametric TVC–HAR specification for the realized volatility, and propose both the local linear and local constant (Nadaraya–Watson) methods for estimating this model. Additionally, we construct the confidence interval estimates for time–varying coefficient functions by adapting a bootstrap method. Then, the simple parametric HAR model was tested against the nonparametric TVC-HAR model, and the test statistic was constructed based on the generalised likelihood ratio method augmented with bootstrap method. The results provided statistical evidence against the simple linear HAR. We also assessed the out-of sample forecasting performance of the local linear TVC–HAR model against the simple HAR model as well as the HAR-GARCH model. The results of the conditional predictive ability (CPA) test indicate that the nonparametric TVC-HAR model consistently outperforms the simple HAR and the HAR-GARCH models.

Furthermore, we used higher order polynomials of time t to calibrate the nonparametric estimates of TVC-HAR model. We find overwhelming statistical evidence to say that the calibrated polynomial specification is the best in-sample fit in comparison to the local linear TVC-HAR model as well as the other two parametric models, the simple HAR and HAR-GARCH. On the other hand, the calibrated TVC-HAR models has the worst out-of-sample volatility forecasting performance. The results of the overall methodological developments and empirical analysis suggest that there is a need to come up with an improved method for the calibration of local linear TVC-HAR model, so that the calibrated model would outperform its nonparametric counterpart in out-of-sample-forecasting. This is a topic for the future research.

ACKNOWLEDGEMENTS

The authors are grateful to Sirimon Treepongkaruna of University of Western Australia, for providing the Reuters tick–by–tick transaction prices data. Also, the authors thank Gael Martin of Monash University, and her associated research fellows for kindly sharing the programming codes for the construction of five-minute logarithmic returns. This research is supported by the ARC Discovery project under Grant number: DP1096374.

APPENDIX

`Local Stationary Autoregressive Process`

Kim (2001) adopts a general class of evolutionary processes that can accommodate a variety of forms of nonstationary behavior, and extends the application of autoregressive models to a general nonstationary process, allowing the autoregressive coeffcients to change smoothly over time. An evolutionary $AR(p)$ process, $\{V_t\}_{t=1}^n$ is defined in the following form,

$$V_t = \beta_0(\frac{t}{n}) + \sum_{k=1}^{p} \beta_k(\frac{t}{n}) V_{t-k} + \varepsilon_t, \tag{13}$$

where $\frac{t}{n} = \tau_t$ as in (3), and $\beta_0(\frac{t}{n})$ is the intercept coefficient function. The evolutionary AR process is locally stationary, if the coefficients are smooth functions of time. Following Dahlhaus (1996), a process is locally stationary when it has a stationary structure on the process at each time point. This is the similar idea to that underlies the nonparametric technique of fitting a line locally to a nonlinear curve. Thus, a locally stationary process behaves like a stationary process in the neighbourhood of each instant in time, but has global nonstationary behaviour. Kim (2001) estimates $\beta(\frac{t}{n})$ nonparametrically, and derives the asymptotical properties for local linear estimators.

`Autoregressive TVC-HAR Model`

In order to apply the results from Kim (2001), we need to convert the TVC–HAR model into a time–varying $AR(p)$ process. Recall our TVC–HAR model (4),

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

where $\log RV_{t-5,t-1} = \frac{1}{5} \sum_{i=1}^{5} \log RV_{t-i}$, $\log RV_{t-22,t-1} = \frac{1}{22} \sum_{i=1}^{22} \log RV_{t-i}$, and $\tau_t = \frac{t}{n}$, hence

$\beta_0(\tau_t) = \beta_0(\frac{t}{n}), \beta_d(\tau_t) = \beta_1(\frac{t}{n}), \beta_w(\tau_t) = \beta_2(\frac{t}{n})$ and $\beta_m(\tau_t) = \beta_3(\frac{t}{n})$. Then

$$\log RV_t = \beta_0(\tau_t) + \beta_d(\tau_t)\log RV_{t-1} + \beta_w(\tau_t)[\frac{1}{5}\sum_{i=1}^{5}\log RV_{t-i}] + \beta_m(\tau_t)[\frac{1}{22}\sum_{i=1}^{22}\log RV_{t-i}] + \varepsilon_t,$$

$$= \beta_0(\tau_t) + \{\beta_d(\tau_t) + \frac{\beta_w(\tau_t)}{5} + \frac{\beta_m(\tau_t)}{22}\}\log RV_{t-1} + \{\frac{\beta_w(\tau_t)}{5} + \frac{\beta_m(\tau_t)}{22}\}\sum_{i=2}^{5}\log RV_{t-i}$$

$$+ \{\frac{\beta_m(\tau_t)}{22}\}\sum_{i=6}^{22}\log RV_{t-i} + \varepsilon_t,$$

As illustrated, our TVC–HAR model is an $AR(22)$ process. Using the backward shift lag operator L, where $L^i X_t = X_{t-i}$, $i = 0, \pm 1, \pm 2, \ldots$, we have,

$$\log RV_t = \beta_0(\tau_t) + [\{\beta_d(\tau_t) + \frac{\beta_w(\tau_t)}{5} + \frac{\beta_m(\tau_t)}{22}\}L + \{\frac{\beta_w(\tau_t)}{5} + \frac{\beta_m(\tau_t)}{22}\}\sum_{i=2}^{5}L^i$$

$$+ \{\frac{\beta_m(\tau_t)}{22}\}\sum_{i=6}^{22}L^i]\log RV_t + \varepsilon_t,$$

$$\equiv \varphi_0(\tau_t) + [\varphi_1(\tau_t)L^1 + \varphi_2(\tau_t)L^2 + \ldots + \varphi_{22}(\tau_t)L^{22}]\log RV_t + \varepsilon_t,$$

$$\equiv \sum_{i=0}^{22}[\varphi_i(\tau_t)L^i\log RV_t] + \varepsilon_t,$$

where $\varphi_0(\tau_t) = \beta_0(\tau_t), \varphi_1(\tau_t) = \beta_d(\tau_t) + \frac{\beta_w(\tau_t)}{5} + \frac{\beta_m(\tau_t)}{22}, \varphi_2(\tau_t) = \varphi_3(\tau_t) = \ldots = \varphi_5(\tau_t) = \frac{\beta_w(\tau_t)}{5} + \frac{\beta_m(\tau_t)}{22}$, and $\varphi_6(\tau_t) = \varphi_7(\tau_t) = \ldots = \varphi_{22}(\tau_t) = \frac{\beta_m(\tau_t)}{22}$. Re-arranging the equation, $\log RV_t$ has the form

$$[1 - \sum_{i=0}^{22}\varphi_i(\tau_t)L^i]\log RV_t = \varepsilon_t, \tag{14}$$

$$\log RV_t \equiv \textstyle\sum_{l=0}^{\infty}\Psi_{i,l}(\tau_t)\varepsilon_{t-l}$$

`Asymptotic Theory of Nonparametric Estimator`

Before presenting the asymptotic properties of the local linear estimator for model (13), the following assumptions are necessary.

Assumption 1. The function $\{\beta_k(\tau_t)\}_{k=0}^p$ is twice continuously differentiable in t with uniformly bounded second order derivatives, and the roots of $\sum_{k=0}^p \beta_k(\tau_t)z^j$ are outside the unit circle.

Assumption 2. The kernel function $K(\cdot)$ is a continuous symmertric nonnegative function on a compact support, satisfying $\sup_r |K(r)|^p = \|K\|_\infty^p < \infty$.

Assumption 3. $\int K(r)dr = 1, \mu_k^2 = \int K(r)r^2dr < \infty$, $\int K^2(r)dr = \|K\|_2^2 < \infty$, and $\int K^2(r)r^2dr < \infty$.

Assumption 4. $\{\varepsilon_t\}$ is $i.i.d.(0, \sigma^2, \kappa_4)$, where κ_4 is finite fourth cumulant.

Assumption 5. (a)$\sum_{i=0}^{22} \Psi_{i,l-i}(\frac{t}{n})\varphi_i(\frac{t+i-l}{n}) = \begin{Bmatrix} 1, & \text{if } l = 0, \\ 0. & \text{if } l \neq 0. \end{Bmatrix}$, and $\Psi_{i,l-i}(\frac{t}{n}) = 0$ if $l < i$; (b) $\sup_{t\leq n} \sum_{i=0}^\infty i^{1/2}\varphi_i^2(\frac{t}{n}) < \infty$, (c) $\sup_{t\leq n} \sum_{i=0}^\infty i^{1/2}[\varphi_i'(\frac{t}{n})]^2 = o(n^2)$, recall that $\frac{t}{n} = \tau_t$;

Remark Assumption 1 is commonly imposed in literature and can be interpreted in terms of smoothness constraints; see Robinson (1989,1991). The commonly used kernel functions, such as the Epanechnikov kernel that we used here satisfies Assumption 2 and 3. Following the works by Dahlhaus (1996) and Kim (2001), Assumption 4 and 5 are required for a locally stationary time series.

Proposition 1.

Let Assumption 1–5 hold, if $h \to 0$ and $nh \to \infty$ as $n \to \infty$, then

$$\sqrt{nh}[\widehat{\beta}(\tau) - \beta(\tau) - B_n] \xrightarrow{D} N(0, \textstyle\sum_\beta(\tau)),$$

where $\sum_\beta(\tau) = \|K\|_2^2 \Gamma^{-1}(\tau)$, and the bias term $B_n \xrightarrow{p} \frac{h^2}{2}\mu_k^2\beta''(\tau)$. Let $\widehat{\varepsilon}_t = Y_t - Y_{t-k}^\top \widehat{\beta}_k(\tau_t)$ and $\widehat{\sigma}_\varepsilon^2 = \sum_{t=p+1}^n \widehat{\varepsilon}_t^2/(n-p)$, then $\Gamma(\tau)$ is consistently estimated by

$$\widehat{\Gamma}(\tau) \equiv S_{n0}(\tau)/\widehat{\sigma}_\varepsilon^2 = \widehat{\sigma}_\varepsilon^{-2}\frac{1}{n}\sum_{t=1}^n Y_{t-1}Y_{t-1}^\top K_h(\tau_t - \tau),$$

and $\sum_\beta(\tau)$ by

$$\widehat{\sum}_\beta(\tau) \equiv \|K\|_2^2 \widehat{\Gamma}^{-1}(\tau) = \|K\|_2^2 S_{n0}^{-1}(\tau)\widehat{\sigma}_\varepsilon^2.$$

For detailed proofs for the above asymptotic result, please refer to Dahlhaus (1996) and Kim (2001).

References

[1] Akdeniz, L., Altay-Salih, A., and Caner, M., 2003. Time-Varying Betas Help in Asset Pricing: The Threshold CAPM. Studies in Nonlinear Dynamics and Econometrics, Vol. 6, 2003.

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

[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 Diebold, F.X., 2007. Roughing It Up: Including Jump Components in the Measurement, Modeling and Forecasting of Return Volatility. Review of Economics and Statistics, Vol. 89 (4), 701 – 720.

[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&P500 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] Chang, Y., and Martinez-Chombo, E., 2003. Electricity Demand Analysis using Cointegration and Error-Correction Models with Time Varying Parameters: The Mexican Case. Working paper, Department of Economics, Rice University.

[10] Chen, R., Yang, L., and Hafner C., 2004. Nonparametric Multistep-Ahead Prediction in Time Series Analysis. Journal of the Royal Statistical Society, Series B (Statistical Methodology) Vol. 66, No. 3, $669 - 686$.

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

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

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

[14] Cui, H., He, X., and Zhu, L., 2002. On Regression Estimators with De-noised Variables. Statistica Sinica 12, $1191 - 1205$.

[15] Dahlhaus, R., 1996. On the Kullback-Leibler Information Divergence of Locally Stationary Processes. Stochastic Process. Appl. 62, $139 - 168$.

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

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

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

[19] Fan, J., and Yao, Q., 2003. Nonlinear Time Series: Nonparametric and Parametric Methods. New York, Springer.

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

[21] Gao, J., 2007. Nonlinear Time Series: Semiparametric and Nonparametric Methods. Chapman & Hall/CRC, London.

[22] Giacomini, R., and White, H., 2006. Tests of Conditional Predicative Ability. Econometrica, Vol. 74, No. 6, 1545 – 1578.

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

[24] Henry, M., 2001. Robust Automatic Bandwidth for Long Memory. Journal of Time Series Analysis, 22, 292 – 316.

[25] 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.

[26] 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.

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

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

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

[30] Li, Q., and Racine, J., 2007. Nonparametric Econometrics: Theory and Practice. Princeton University Press, New Jersey.

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

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

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

[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] Robinson, P. M., 1991. Time-Varying Nonlinear Regression. In: Hackl, P., Westland, A.H. (Eds.), Economic Structure Change Analysis and Forecasting. Springer, Berlin, $179 - 190$.

[37] Robinson, P. M., 1995. Gaussian Semiparametric Estimation of Long Range Dependence. Annals of Statistics, 23 (5), $1630 - 1661$.

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

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

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

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

[42] Zhang, W. and Peng, H., 2010. Simultaneous Confidence Band and Hypothesis Test in Generalised Varying-Coefficient Models. Journal of Multivariate Analysis, 101, $1656 - 1680$.

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) OLS estimates of HAR-RV 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 1: Subsamples and the simple HAR–RV model parameter estimates for the subsample periods. A wald test of the equality of the regression coefficients across the four subsamples warrants that a time-varying coefficient HAR model is necessary, the reported test statistic is 61.1675 with associated $p-$value equals 0. $*$ indicates insignificance of parameter at 10% level.

a) Parametric Specifications of TVC-HAR Model

$$
\begin{aligned}
\beta_0(\tau_t) &= \gamma_{0,1}+\gamma_{0,2}\tau_t+\gamma_{0,3}\tau_t^4+\gamma_{0,4}\tau_t^6+\gamma_{0,5}\tau_t^7 \\
\beta_d(\tau_t) &= \gamma_{d,1}+\gamma_{d,2}\tau_t^2 \\
\beta_w(\tau_t) &= \gamma_{w,1}+\gamma_{w,2}\tau_t+\gamma_{w,3}\tau_t^2+\gamma_{w,4}\tau_t^3+\gamma_{w,5}\tau_t^4+\gamma_{w,6}\tau_t^5+\gamma_{w,7}\tau_t^6+\gamma_{w,8}\tau_t^7+\gamma_{w,9}\tau_t^8 \\
\beta_m(\tau_t) &= \gamma_{m,1}+\gamma_{m,2}\tau_t+\gamma_{m,3}\tau_t^2+\gamma_{m,4}\tau_t^3+\gamma_{m,5}\tau_t^4+\gamma_{m,6}\tau_t^5+\gamma_{m,7}\tau_t^6+\gamma_{m,8}\tau_t^7+\gamma_{m,9}\tau_t^8
\end{aligned}
$$

b) Estimates of Calibrated TVC-HAR Model

	Coefficient	P-value
$\gamma_{0,1}$	0.3202	0.0008
$\gamma_{0,2}$	-4.1228	0.0000
$\gamma_{0,3}$	61.6397	0.0203
$\gamma_{0,4}$	-297.8402	0.0478
$\gamma_{0,5}$	272.0742	0.0583
$\gamma_{d,1}$	0.1249	0.0188
$\gamma_{d,2}$	0.6012	0.0005
$\gamma_{w,1}$	0.2673	0.4328*
$\gamma_{w,2}$	40.3538	0.0785
$\gamma_{w,3}$	-994.2419	0.0318
$\gamma_{w,4}$	9,491.8247	0.0210
$\gamma_{w,5}$	-45,393.9405	0.0176
$\gamma_{w,6}$	119,912.0461	0.0166
$\gamma_{w,7}$	-177,881.3972	0.0167
$\gamma_{w,8}$	138,842.8646	0.0174
$\gamma_{w,9}$	-44,372.1746	0.0186
$\gamma_{m,1}$	0.5213	0.2580*
$\gamma_{m,2}$	-58.6669	0.0421
$\gamma_{m,3}$	1,358.1538	0.0137
$\gamma_{m,4}$	-12,640.9258	0.0076
$\gamma_{m,5}$	59,555.3428	0.0058
$\gamma_{m,6}$	-155,821.3586	0.0051
$\gamma_{m,7}$	229,790.9028	0.0049
$\gamma_{m,8}$	-178,814.9305	0.0050
$\gamma_{m,9}$	57,104.9293	0.0051

Table 2: Least-squared estimates of coefficients of the calibrated TVC-HAR model, using the automated algorithm as in Section 2.5. $*$indicates insignificance of parameter at 10% level.

a) Parametric Specifications of TVC-HAR Model

$$\begin{aligned}\beta_0(\tau_t) &= \gamma_{0,1} + \gamma_{0,2}\tau_t + \gamma_{0,3}\tau_t^2 \\ \beta_d(\tau_t) &= \gamma_{d,1} + \gamma_{d,2}\tau_t^4 \\ \beta_w(\tau_t) &= \gamma_{w,1} + \gamma_{w,2}\tau_t^3 \\ \beta_m(\tau_t) &= \gamma_{m,1}\tau_t + \gamma_{m,2}\tau_t^2\end{aligned}$$

b) Estimates of Calibrated TVC-HAR Model

Coefficient	Least Squared estimates	P-value
$\gamma_{0,1}$	0.2428	0.0000
$\gamma_{0,2}$	-3.9508	0.0000
$\gamma_{0,3}$	4.7864	0.0000
$\gamma_{d,1}$	0.1681	0.0001
$\gamma_{d,2}$	1.1913	0.0001
$\gamma_{w,1}$	0.5867	0.0000
$\gamma_{w,2}$	-0.7645	0.0157
$\gamma_{m,1}$	-0.7083	0.0615
$\gamma_{m,2}$	1.1564	0.0513

Table 3: Least-squared estimates of coefficients of the (manually) calibrated TVC-HAR model. Note that all parameters are significant at 10% level.

	$\delta = 1$		$\delta = 5$		$\delta = 22$	
Subsamples	pval	Alternative better	pval	Alternative better	pval	Alternative better
Full sample	0.0025	97.74%	0.0034	64.91%	0.0739	Equally good
pre-GFC	0.0137	93.97%	0.0835	Equally good	0.0064	64.32%
post-GFC	0.0679	Equally good	0.0016	59.30%	0.0276	67.34%

Table 4: Results of Conditional Predictive Ability (CPA) test for out-of-sample forecasting performance between the alternative model (the simple HAR) and the bench model (HAR-GARCH), with selected test function $\pi_t = (1, \Delta L_{t+\delta-1})$. Note that large $p-$values ($>$ nominated 5% significancelevel) suggest that the forecasts generated from the two models are equally accurate, small $p-$values of the CPA test lead to the rejection of the null, and it further indicates the proportion of time (in %) that the alternative model offers a more accurate forecast over the bench.

a) TVC-HAR-LL *versus* Simple HAR

Subsamples	$\delta = 1$		$\delta = 5$		$\delta = 22$	
	pval	Alternative better	pval	Alternative better	pval	Alternative better
Full sample	0.4974	Equally good	0.0000	62.91%	0.0016	57.14%
pre-GFC	0.5575	Equally good	0.0000	57.79%	0.0907	Equally good
post-GFC	0.7656	Equally good	0.0000	65.83%	0.0008	73.37%

b) TVC-HAR-LL versus HAR-GARCH

Subsamples	$\delta = 1$		$\delta = 5$		$\delta = 22$	
	pval	Alternative better	pval	Alternative better	pval	Alternative better
Full sample	0.2562	Equally good	0.0000	60.40%	0.0041	55.64%
pre-GFC	0.4134	Equally good	0.0001	58.79%	0.0972	Equally good
post-GFC	0.5026	Equally good	0.0001	60.80%	0.0035	66.33%

Table 5: Results of Conditional Predictive Ability (CPA) test for out-of-sample forecasting performance between the alternative model (TVC-HAR-LL) and the bench model (the simple HAR or HAR-GARCH), with selected test function $\pi_t = (1, \Delta L_{t+\delta-1})$. Note that large $p-$values ($>$ nominated 5% significancelevel) suggest that the forecasts generated from the two models are equally accurate, small $p-$values of the CPA test lead to the rejection of the null, and it further indicates the proportion of time (in %) that the alternative model offers a more accurate forecast over the bench.

a) TVC-HAR-Calibrated *versus* Simple HAR

	$\delta = 1$		$\delta = 5$		$\delta = 22$	
Subsamples	pval	Alternative better	pval	Alternative better	pval	Alternative better
Full sample	0.2247	Equally good	0.0039	5.01%	0.2073	Equally good
pre-GFC	0.3333	Equally good	0.0385	11.06%	0.4260	Equally good
post-GFC	0.4304	Equally good	0.0570	Equally good	0.2963	Equally good

b) TVC-HAR-Calibrated *versus* HAR-GARCH

	$\delta = 1$		$\delta = 5$		$\delta = 22$	
Subsamples	pval	Alternative better	pval	Alternative better	pval	Alternative better
Full sample	0.3693	Equally good	0.0045	5.01%	0.2123	Equally good
pre-GFC	0.5002	Equally good	0.0358	10.55%	0.4261	Equally good
post-GFC	0.5404	Equally good	0.0667	Equally good	0.3158	Equally good

Table 6: Results of Conditional Predictive Ability (CPA) test for out-of-sample forecasting performance between the alternative model (TVC-HAR-Calibrated) and the bench model (the simple HAR or HAR-GARCH), with selected test function $\pi_t = (1, \Delta L_{t+\delta-1})$. Note that large $p-$values ($>$ nominated 5% significancelevel) suggest that the forecasts generated from the two models are equally accurate, small $p-$values of the CPA test lead to the rejection of the null, and it further indicates the proportion of time (in %) that the alternative model offers a more accurate forecast over the bench.

a) TVC-HAR-Calibrated *versus* Simple HAR

	$\delta = 1$		$\delta = 5$		$\delta = 22$	
Subsamples	pval	Alternative better	pval	Alternative better	pval	Alternative better
Full sample	0.1945	Equally good	0.0000	35.34%	0.0008	30.33%
pre-GFC	0.9287	Equally good	0.0000	35.68%	0.0101	44.22%
post-GFC	0.1656	Equally good	0.0000	34.67%	0.0001	20.60%

b) TVC-HAR-Calibrated *versus* HAR-GARCH

	$\delta = 1$		$\delta = 5$		$\delta = 22$	
Subsamples	pval	Alternative better	pval	Alternative better	pval	Alternative better
Full sample	0.1403	Equally good	0.0000	35.84%	0.0009	31.08%
pre-GFC	0.8153	Equally good	0.0000	35.68%	0.0110	45.23%
post-GFC	0.1258	Equally good	0.0000	35.18%	0.0001	24.12%

Table 7: Results of Conditional Predictive Ability (CPA) test for out-of-sample forecasting performance between the alternative model (TVC-HAR-Calibrated) and the bench model (the simple HAR or HAR-GARCH), with selected test function $\pi_t = (1, \Delta L_{t+\delta-1})$. Note that large $p-$values ($>$ nominated 5% significancelevel) suggest that the forecasts generated from the two models are equally accurate, small $p-$values of the CPA test lead to the rejection of the null, and it further indicates the proportion of time (in %) that the alternative model offers a more accurate forecast over the bench.

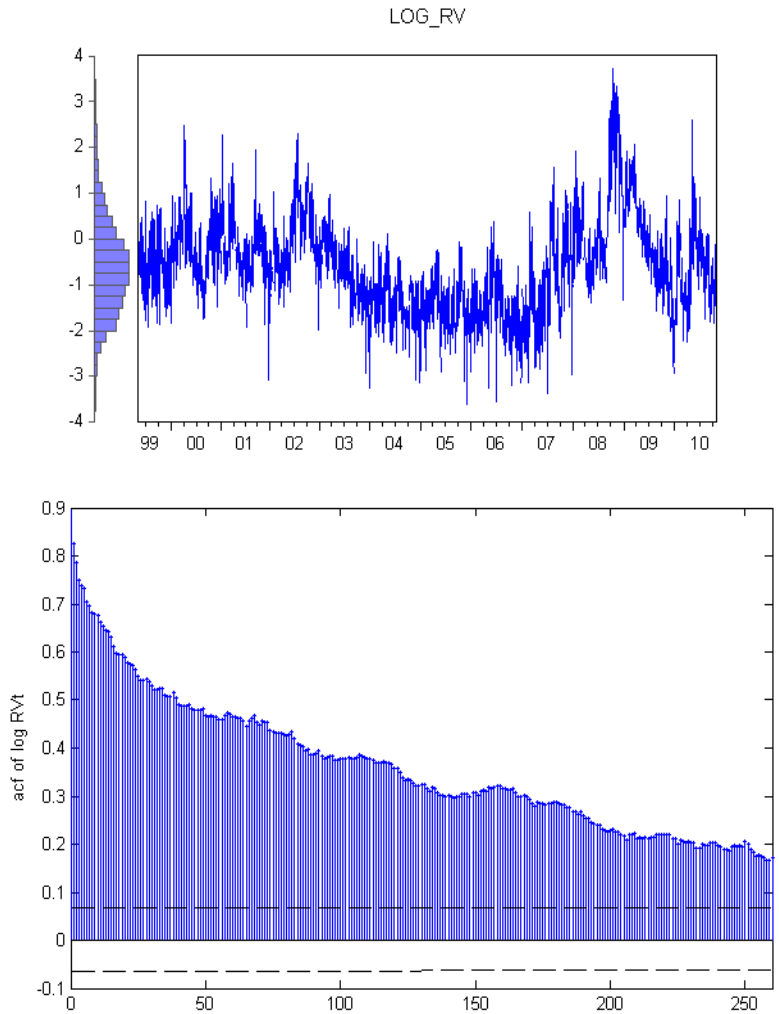

Figure 1: Time series of logarithmic realized volatility, and its sample autocorrelations up to 260th order, or one year. The dashed lines gives the lower and upper bounds of the Bartlett 95% confidence interval.

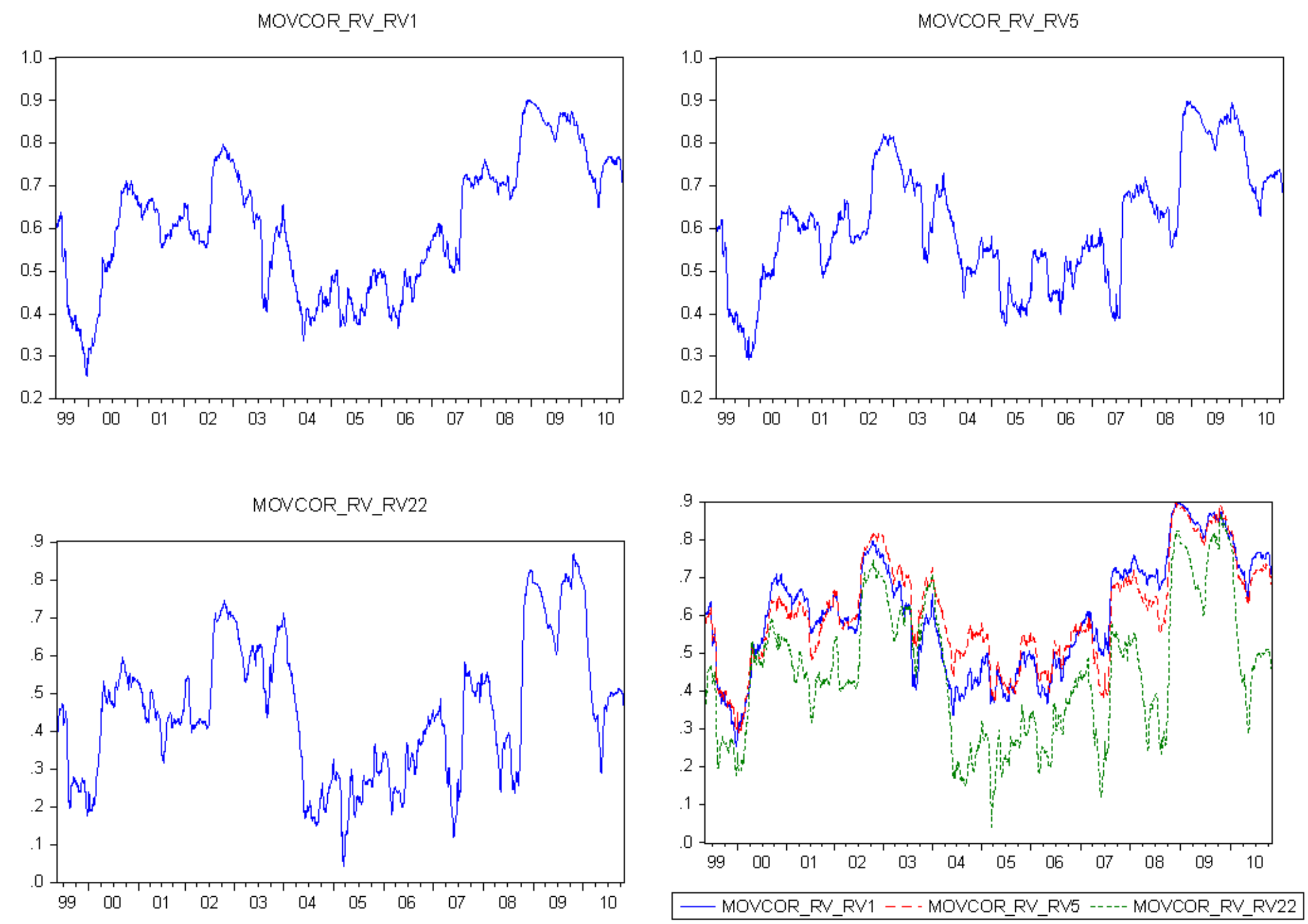

Figure 2: The moving window (n=200) correlation plot between $\log RV_t$ and $\log RV_{t-1}$, $\log RV_{t-5}$, $\log RV_{t-22}$, respectively. All plots show nonlinear patterns over time, and the correlation measures are time varying.

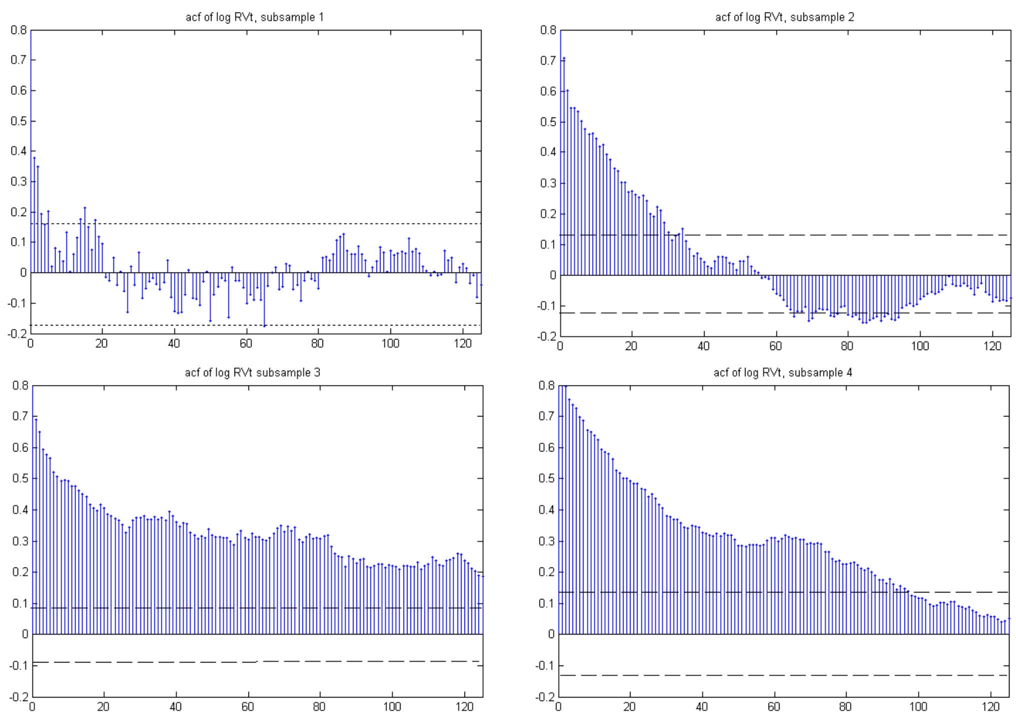

Figure 3: acf plots for $\log RV_t$ up to 125th order, or half a year. The plots exhibit different patterns across the four subsamples, reflecting the dynamic changes of the serial dependency of the $\log RV_t$ series over the sample period.

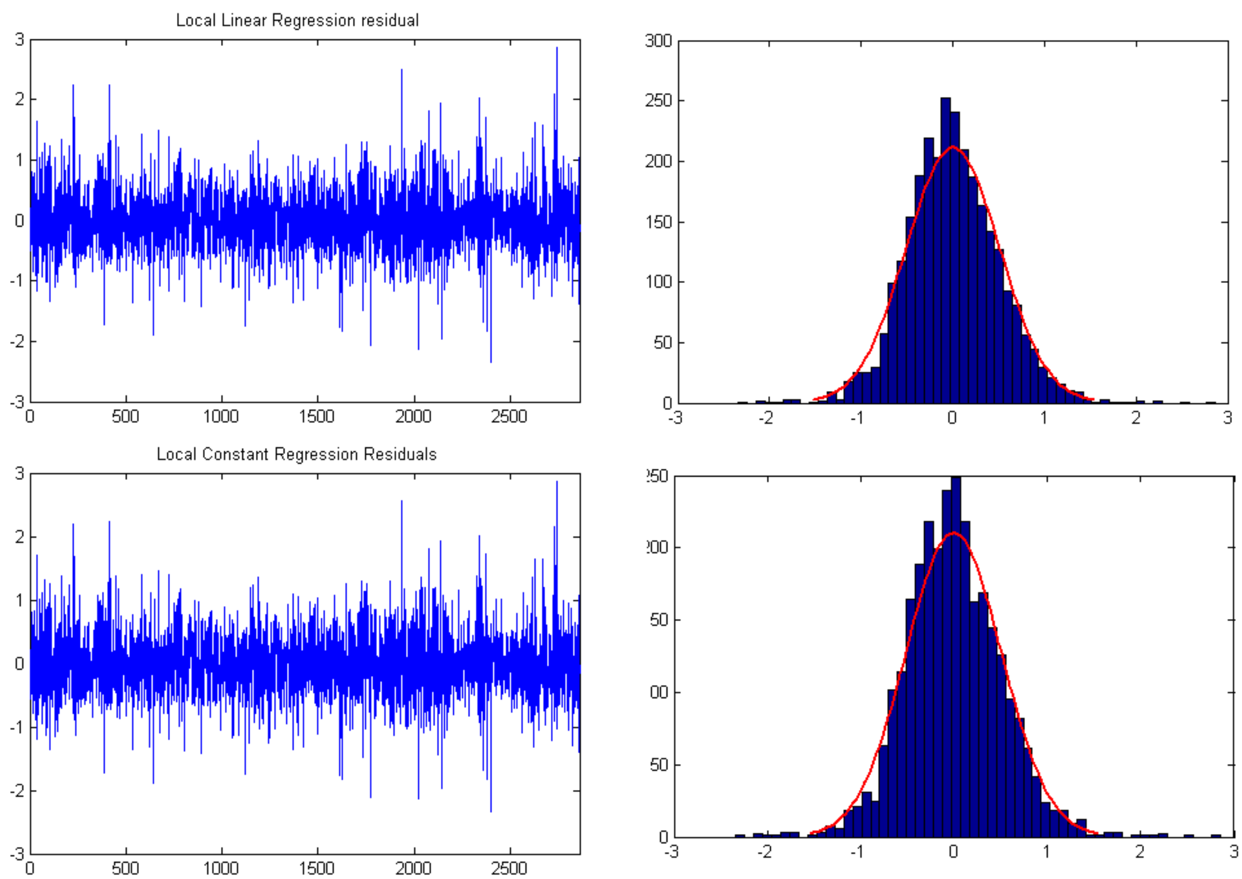

Figure 4: Residuals plots from local linear and local constant regressions. The residuals are close to normal, and the local least-square estimates are expected to be efficient.

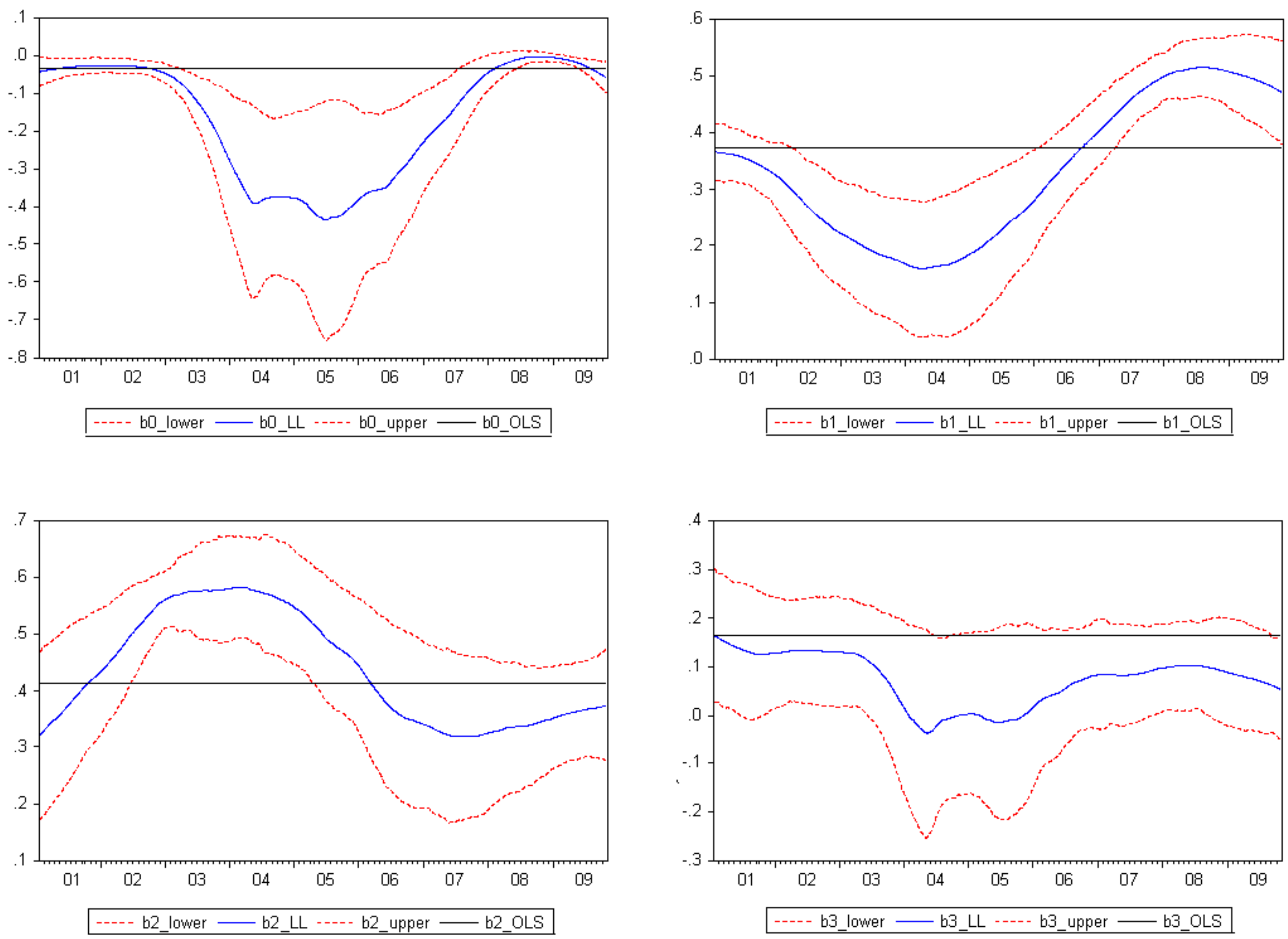

Figure 5: The local linear estimator (solid nonlinear line) and least square estimate (solid horizontal line) of the coefficient functions $\beta(\tau_t)$ with 90% pointwise confidence intervals (dotted lines).

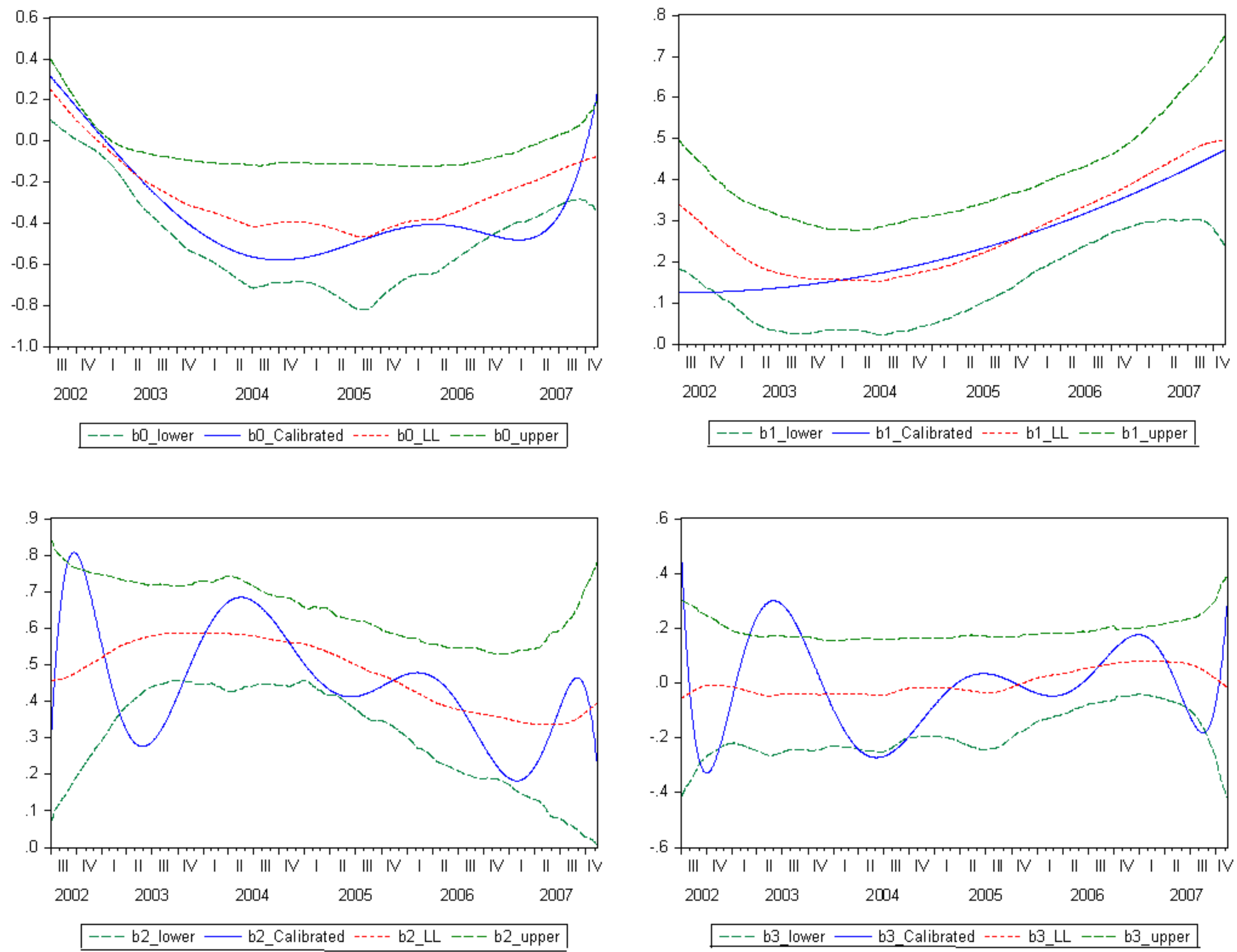

Figure 6: Plots of the (automatically) calibrated polynomial functions (solid line) against the local linear estimator (dotted line) of the coefficient functions of the TVC-HAR model, with the 90% confidence intervals (dashed lines). The estimation period is between 01/07/2002 and 13/11/2007.

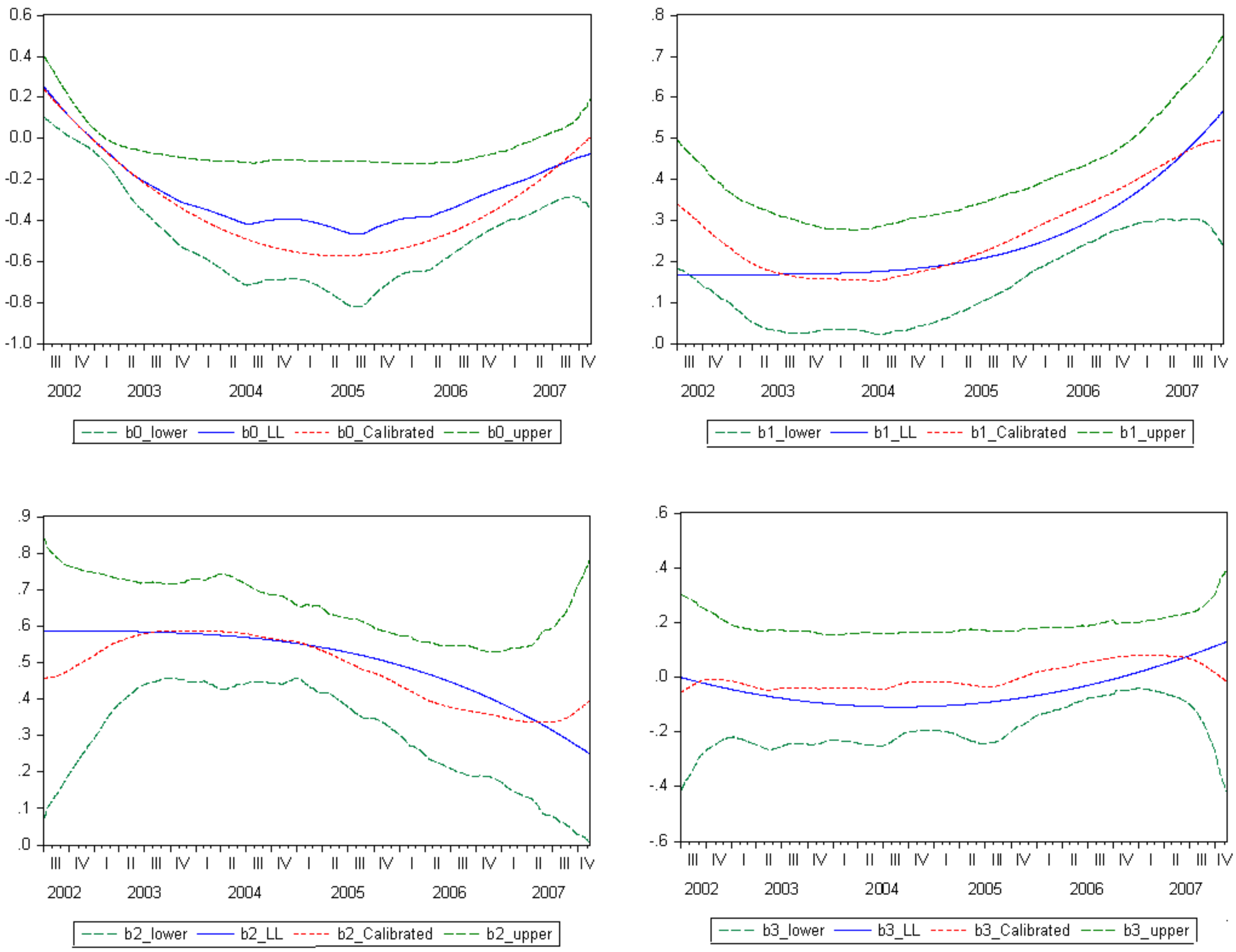

Figure 7: Plots of the (manually) calibrated polynomial functions (solid line) against the local linear estimator (dotted line) of the coefficient functions of the TVC-HAR model, with the 90% confidence intervals (dashed lines). The estimation period is between 01/07/2002 and 13/11/2007.