

FORECASTING REALIZED VOLATILITY USING A NONNEGATIVE SEMIPARAMETRIC MODEL

DANIEL PREVE, ANDERS ERIKSSON, AND JUN YU

ABSTRACT. This paper introduces a parsimonious and yet flexible nonnegative semiparametric model to forecast financial volatility. The new model extends the linear nonnegative autoregressive model of Barndorff-Nielsen & Shephard (2001) and Nielsen & Shephard (2003) by way of a power transformation. It is semiparametric in the sense that the distributional form of its error component is left unspecified. The statistical properties of the model are discussed and a novel estimation method is proposed. Asymptotic properties are established for the new estimation method. Simulation studies validate the new estimation method. The out-of-sample performance of the proposed model is evaluated against a number of standard methods, using data on S&P 500 monthly realized volatilities. The competing models include the exponential smoothing method, a linear AR(1) model, a log-linear AR(1) model, and two long-memory ARFIMA models. Various loss functions are utilized to evaluate the predictive accuracy of the alternative methods. It is found that the new model generally produces highly competitive forecasts.

Key words and phrases. Autoregression, nonlinear/non-Gaussian time series, realized volatility, semiparametric model, volatility forecast.

1. INTRODUCTION

Financial market volatility is an important input for asset allocation, investment, derivative pricing, and financial market regulation. Not surprisingly, how to model and forecast financial volatility has been a subject of extensive research. Numerous survey papers are now available on the subject, with hundreds of reviewed research articles. Excellent survey articles on the subject include Bollerslev, Chou & Kroner (1992), Bollerslev, Engle & Nelson (1994), Ghysels, Harvey & Renault (1996), Poon & Granger (2003), and Shephard (2005).

In this extremely extensive literature, ARCH and stochastic volatility (SV) models are arguably the most popular parametric tools. These two classes of models are motivated by the fact that volatilities are time-varying. Moreover, they offer ways to estimate past volatility and forecast future volatility from return data. In recent years, however, many researchers have argued that one could measure latent volatility by realized volatility (RV), see e.g. Andersen, Bollerslev, Diebold & Labys (2001, ABDL hereafter) and Barndorff-Nielsen & Shephard (2002), and then build a time series model for volatility forecasting using observed RV, see e.g. Andersen, Bollerslev, Diebold & Labys (2003). An advantage of this approach is that “models built for the realized volatility produce forecasts superior to those obtained from less direct methods” (ABDL 2003). In an important study, ABDL (2003) introduced a new Gaussian time series model for logarithmic RV (log-RV) and established its superiority for RV forecasting over some standard methods based on squared returns. Their choice of modeling log-RV rather than raw RV is motivated by the fact that the logarithm of RV, in contrast to RV itself, is approximately normally distributed. Moreover, conditional heteroskedasticity is greatly reduced in log-RV.

Following this line of thought, in this paper we introduce a new time series model for RV. For S&P 500 monthly RV, we show that although the distribution of log-RV is closer to a normal distribution than that of raw RV, normality is still rejected at all standard significance levels. Moreover, although conditional heteroskedasticity is reduced in log-RV, there is still evidence of remaining conditional heteroskedasticity. These two limitations associated with the logarithmic transformation motivate us to consider a more flexible transformation which is closely related to the well known Box-Cox transformation – the power transformation. In contrast to the logarithmic transformation, the power transformation is generally not compatible with a normal error distribution as the support for the normal distribution covers the entire real line. This well known truncation problem inherent to the power and Box-Cox transformations further motivates us to use nonnegative

error distributions. The new model is flexible, parsimonious, and has a simple forecast expression. Moreover, the numerical estimation of the model is very fast and can easily be implemented using standard computational software.

The new model is closely related to the linear nonnegative models described in Barndorff-Nielsen & Shephard (2001) and Nielsen & Shephard (2003). In particular, it generalizes the discrete time version of the nonnegative Ornstein-Uhlenbenck process of Barndorff-Nielsen & Shephard (2001) by (1) applying a power transformation to volatility and (2) leaving the dependency structure and the distribution of the nonnegative error term unspecified. Our work is also related to Yu, Yang & Zhang (2006) and Goncalves & Meddahi (2009) where the Box-Cox transformation is applied to stochastic volatility and RV, respectively. The main difference between our model specification and theirs is that a unspecified (marginal) distribution with nonnegative support, instead of the normal distribution, is induced by the transformation. Moreover, our model is loosely related to Higgins & Bera (1992), Hentschel (1995) and Duan (1997) where the Box-Cox transformation is applied to ARCH volatility, and to Fernandes & Grammig (2006) and Chen & Deo (2004). Finally, our model is related to a recent study by Cipollino, Engle & Gallo (2006) where an alternative model with nonnegative errors is used for RV. The main difference here is that the dynamic structure for the transformed RV is linear in our model, whereas the dynamic structure for the RV is nonlinear in theirs. In the terminology of Fan, Fan & Jiang (2007), our approach is in the time domain, but it can easily be integrated with methods in the state domain.

The proposed model is estimated using a two-stage estimation method. In the first stage, a nonlinear least squares procedure is applied to a nonstandard objective function. In the second stage an extreme value estimator is applied. The asymptotic properties are studied. The finite sample performance of the proposed estimation method is examined via simulations.

The new specification is used to model and forecast S&P 500 monthly RV. Its forecasting performance is compared to a number of standard time series methods previously used in the literature, including the exponential smoothing method and two logarithmic long-memory ARFIMA models. Under various evaluation criteria, we find that our simple nonnegative model generally produces highly competitive forecasts.

While our model is related to several models in the literature, to the best of our knowledge, our specification is new in two ways. First, it is based on the power transformation. Second, the distribution of its error component is unspecified. Also, it appears that our

paper is the first to establish the empirical usefulness of a nonnegative process for financial time series. Moreover, the estimation method that we introduce is new.

The paper is organized as follows. Section 2 motivates and presents the new model. In Section 3 a novel estimation method is proposed to estimate the parameters of the new model. Also, its asymptotic properties are studied and its finite sample performance is examined via simulations. Section 4 outlines the competing models for volatility forecasting and Section 5 presents the loss functions used to assess their forecast performances. Section 6 describes the S&P 500 realized volatility data and the empirical results. Finally, Section 7 concludes the paper and the Appendix collects the proofs of the propositions.

2. A NONNEGATIVE SEMIPARAMETRIC MODEL

Before introducing the new model, we first review two related time series models previously used in the volatility literature, namely, a simple nonnegative autoregressive (AR) model and the Box-Cox AR model.

2.1. Some Existing Time Series Models for Volatility. Barndorff-Nielsen & Shephard (2001) introduced the following continuous time model for volatility, $\sigma^2(t)$,

$$(2.1) \quad d\sigma^2(t) = -\lambda\sigma^2(t)dt + dz(\lambda t), \quad \lambda > 0$$

where $\{z\}$ is a Lévy process with independent nonnegative increments which ensures the positivity of $\sigma^2(t)$ (see Equation (2) in Barndorff-Nielsen & Shephard 2001). Applying the Euler approximation to the continuous time model in (2.1) yields the following discrete time model

$$(2.2) \quad \sigma_{t+1}^2 = \phi\sigma_t^2 + V_{t+1},$$

where $\phi = 1 - \lambda$ and $V_{t+1} = z[\lambda(t+1)] - z(\lambda t)$ is a sequence of independent identically distributed (iid) random variables whose distribution has a nonnegative support. A well known nonnegative random variable is the generalized inverse Gaussian, whose tails can be quite fat. Barndorff-Nielsen & Shephard (2001) discuss the analytical tractability of this model. In the case when V_{t+1} is exponentially distributed, Nielsen & Shephard (2003) derive the exact finite sample distribution of an extreme value estimator of ϕ for the stationary, unit root and explosive cases.¹ Simulated paths from Model (2.2) typically match real realized volatility data very well. See, for example, Figure 1(c) in Barndorff-Nielsen & Shephard (2001). Unfortunately, so far no empirical evidence that establishes the usefulness of this model has been reported.

¹See Section 3 for a detailed discussion on the extreme value estimator.

Two restrictions seem to apply to Model (2.2). First, since $\{V_{t+1}\}$ is an iid sequence, conditional heteroskedasticity is not allowed in the specification. However, conditional heteroskedasticity in volatility is well documented empirically. Important volatility models with conditional heteroskedasticity include the ones by Heston (1993) and Meddahi & Renault (2004). The second restriction in the specification concerns the ratio of two successive volatilities. More specifically, from (2.2) it can be seen that $\sigma_{t+1}^2/\sigma_t^2$ is bounded from below by ϕ , almost surely, implying that σ_{t+1}^2 cannot decrease by more than $100(1 - \phi)\%$ compared to σ_t^2 . Since the parameter ϕ in the model is typically estimated by an extreme value method, in practice, this restriction is automatically satisfied by construction. For example, the estimate of ϕ in our empirical study is 0.2615, implying that σ_t^2 cannot decrease by more than 73.85% from one time period to the next. Indeed, 73.85% is the maximum percentage drop in successive monthly volatilities in the sample, which took place on November 1987.

In a discrete time framework, a popular parametric time series model for volatility is the lognormal SV model of Taylor (1986) given by

$$(2.3) \quad X_t = \sigma_t \varepsilon_t,$$

$$(2.4) \quad \ln \sigma_t^2 = \mu + \phi(\ln \sigma_{t-1}^2 - \mu) + \eta_t,$$

where X_t is the return, σ_t^2 is the latent volatility, and $\{\varepsilon_t\}$ and $\{\eta_t\}$ are two uncorrelated sequences of independent zero-mean Gaussian random variables. In this specification volatility clustering is modeled as a first-order autoregression for the log-volatility. The logarithmic transformation in (2.4) serves three important purposes. First, it ensures the positivity of σ_t^2 . Second, it reduces heteroskedasticity. Third, it induces normality.

Yu et al. (2006) introduce a closely related SV model by replacing the logarithmic transformation in Taylor's volatility equation with the more general Box-Cox transformation Box & Cox (1964),

$$(2.5) \quad h(\sigma_t^2, \lambda) = \mu + \phi[h(\sigma_{t-1}^2, \lambda) - \mu] + \eta_t,$$

where

$$(2.6) \quad h(x, \lambda) = \begin{cases} (x^\lambda - 1)/\lambda, & \lambda \neq 0 \\ \ln x, & \lambda = 0. \end{cases}$$

Compared to the logarithmic transformation, the Box-Cox power transformation provides a more flexible way to induce normality and reduce heteroskedasticity. A nice feature of the Box-Cox model is that it includes several standard specifications as special cases, including the logarithmic transformation ($\lambda = 0$) and a linear specification ($\lambda = 1$). In the context of SV, Yu et al. (2006) document empirical evidence against the logarithmic transformation.

Chen & Deo (2004) and Goncalves & Meddahi (2009) are interested in the optimal power transformation. In the context of RV, Goncalves & Meddahi (2009) find evidence of non-optimality for the logarithmic transformation. They further report evidence of negative values of λ as the optimal choice under various data generating processes. Our empirical results reinforce this important conclusion, although our approach is vastly different.

While both the nonnegative model and the Box-Cox model have proven useful for modeling volatility, nothing is documented on their usefulness for forecasting volatility. Moreover, it is well known that the Box-Cox transformation is incompatible with a normal error distribution. This is the well known truncation problem associated with the Box-Cox transformation in the context of Gaussianity.

2.2. Realized Volatility. In the ARCH or SV models, volatilities are estimated parametrically from returns observed at the same frequency. In recent years, it has been argued that one can measure volatility in a model-free framework using an empirical measure of the quadratic variation of the underlying efficient price process, that is, RV. RV has several advantages over ARCH and SV models. First, by treating volatility as directly observable, RV overcomes the well known curse-of-dimensionality problem in the multivariate ARCH or SV models. Second, compared with the squared return, RV provides a more reliable estimate of integrated volatility. This improvement in estimation naturally leads to gains in volatility forecasting.

Let RV_t denote the RV at a lower frequency (say daily or monthly) and $p(t, k)$ denote the log-price at a higher frequency (say intra-day or daily). Then RV_t is defined by

$$(2.7) \quad RV_t = \sqrt{\sum_{k=2}^N [p(t, k) - p(t, k-1)]^2},$$

where N is the number of higher frequency observations in a lower frequency period.²

The theoretical justification of RV as a measure of volatility comes directly from standard stochastic process theory, according to which the empirical quadratic variation converges to integrated volatility as the infill sampling frequency goes to zero (ABDL 2001, Barndorff-Nielsen & Shephard 2002, Jacod 1994). The empirical method inspired by this consistency has recently become more popular with the availability of ultra high frequency data.

²In ABDL (2003) RV is referred to as the realized variance, $\sum_{k=2}^N [p(t, k) - p(t, k-1)]^2$. Although the authors build time series models for the realized variance, they forecast the realized volatility. In contrast, the present paper builds time series models for, and forecasts, the realized volatility, which seems more appropriate. Consequently, the bias correction, as described in ABDL (2003), is not needed.

TABLE 1. Summary statistics for S&P 500 monthly RV, log-RV and power-RV. JB is the p -value of the Jarque-Bera test under the null hypothesis that the data are from a normal distribution.

	Mean	Med	Max	Skew	Kurt	JB
RV	0.0037	0.0033	0.0256	3.3073	28.7912	0.000
log-RV	-5.6873	-5.7257	-3.6661	0.3887	3.6574	0.000
power-RV	4.8939	4.9123	6.9076	0.0320	3.2879	0.277

In a recent important contribution, ABDL (2003) show that a Gaussian long-memory model for the logarithmic daily realized variance provides more accurate forecasts than the GARCH(1,1) model and the RiskMetrics method of J.P. Morgan (1996). The logarithmic transformation is used since it is found that the distribution of logarithmic realized variance, but not of realized variance, is approximately normal. In Table 1, we report some summary statistics for monthly RV, log-RV and power-RV for the S&P 500 data over the period Jan 1946-Dec 2004, including the skewness, kurtosis and p -values of the Jarque-Bera test statistic for normality.³ For RV, the departure from normality is overwhelming. While the distribution of log-RV is much closer to a normal distribution than that of RV, there is still strong evidence against normality.

In order to compare the conditional heteroskedasticities, in Figure 1 we plot the squared residuals ($\hat{\varepsilon}_{it}^2$), obtained from the AR(1) regressions for RV, log-RV and power-RV, respectively, against each corresponding explanatory variable (lagged RV, log-RV and power-RV). For ease of comparison, superimposed are smooth curves fitted using the LOESS method. It is clear that while the logarithmic transformation reduces the conditional heteroskedasticity there is still evidence of conditional heteroskedasticity in the residuals. The power transformation further reduces the conditional heteroskedasticity of RV. While the logarithmic transformation reduces the impact of large observations (extreme deviations from the mean), the second plot of Figure 1 suggests that it is not as effective as anticipated. In contrast, the power transformation with a negative power parameter is able to reduce the impact of large observations further. Thus, the results indicate that there is room for further improvements over the logarithmic transformation. A more detailed analysis of the S&P 500 data is provided in Section 6.

2.3. The Nonnegative Semiparametric Model–NonNeg. In this paper, we focus on modeling and forecasting RV. Let us first consider the RV version of Model (2.5),

$$(2.8) \quad h(RV_t, \lambda) = \alpha + \beta h(RV_{t-1}, \lambda) + \varepsilon_t, \quad t = 1, \dots, T$$

³The power parameter is -0.2780 which is the estimate of λ in the proposed model obtained using the entire S&P 500 monthly RV sample; see sections 3 and 6 for further details.

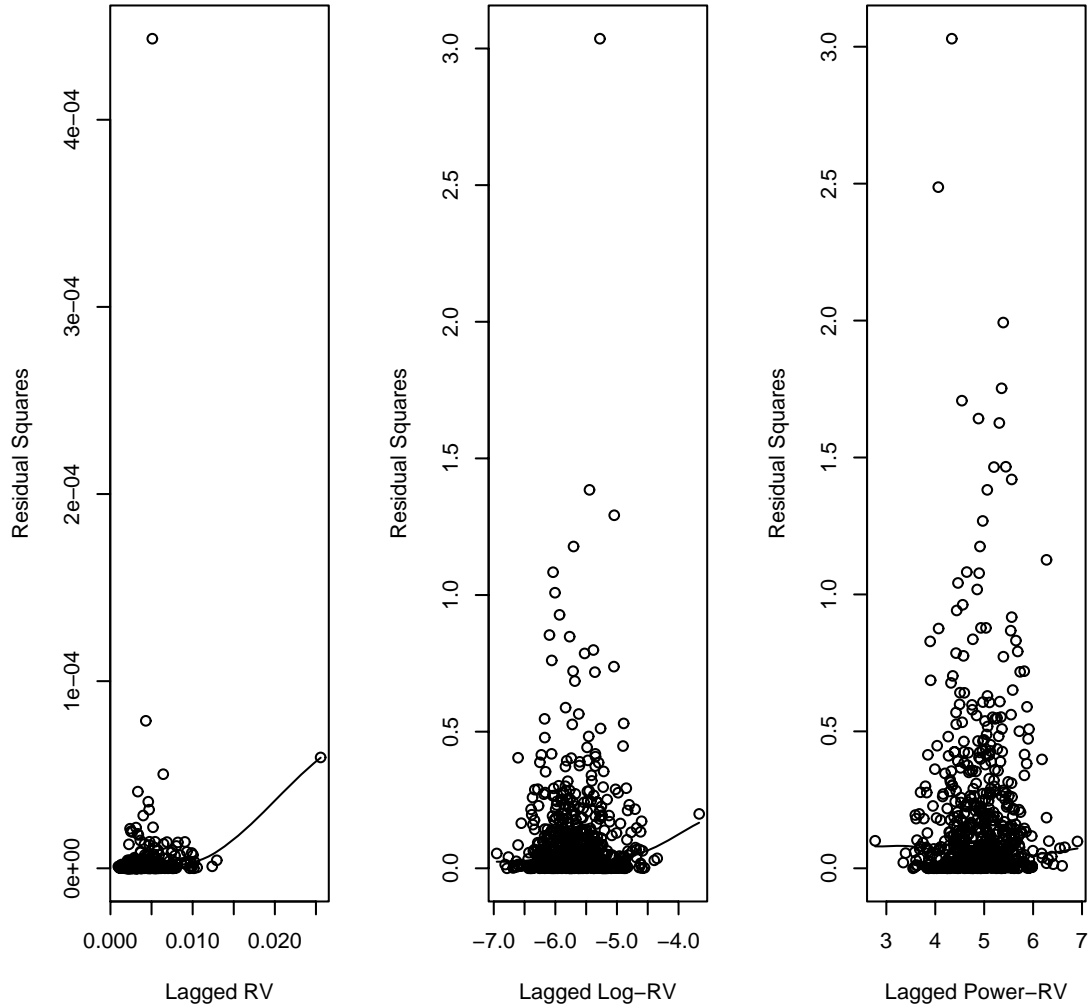


FIGURE 1. Plots of the squared residuals, obtained from the AR(1) regressions for RV, log-RV and power-RV, respectively, against each corresponding lagged explanatory variable. Superimposed are smooth curves fitted using the LOESS method.

where $\{\varepsilon_t\}$ is a sequence of independent $N(0, \sigma_\varepsilon^2)$ distributed random variables and $h(x, \lambda)$ is given by (2.6).

If $\lambda \neq 0$, we may rewrite (2.8) as

$$(2.9) \quad RV_t^\lambda = (\lambda\alpha + 1) + \beta(RV_{t-1}^\lambda - 1) + \lambda\varepsilon_t,$$

where RV_t^λ is a simple power transformation. A special case of (2.9) is the linear Gaussian AR(1) model when $\lambda = 1$:

$$(2.10) \quad RV_t = (\alpha + 1) + \beta(RV_{t-1} - 1) + \varepsilon_t.$$

If $\lambda = 0$ in (2.8), we have the log-linear Gaussian AR(1) model previously used in the literature:

$$(2.11) \quad \ln RV_t = \alpha + \beta \ln RV_{t-1} + \varepsilon_t.$$

While the specification in (2.8) is more general than the log-linear Gaussian AR(1) model (2.11), it has a serious drawback. In general, the right hand side of (2.9) has to be non-negative with probability 1. This requirement is violated since a normal error distribution has a support covering the entire real line.

This drawback motivates us to explore an alternative model specification for RV. Our nonnegative (NonNeg) model is of the form

$$(2.12) \quad RV_t^{\lambda_0} = \phi_0 RV_{t-1}^{\lambda_0} + V_t, \quad t = 1, \dots, T$$

where $\lambda_0 \neq 0$, $\phi_0 > 0$ and the initial value RV_0 is positive with probability 1. $\{V_t\}$ is assumed to be a sequence of independent and identically distributed, continuous random variables with nonnegative support $[\gamma, \infty)$, for some unknown $\gamma \geq 0$. It is assumed that $m \in \mathbf{N}$ is finite and potentially unknown. Hence, the distribution of V_t is left unspecified.

The nonnegative restriction on the support of the error distribution ensures the positivity of RV_t^λ . Hence, our model does not suffer from the truncation problem in the classical Box-Cox model (2.8). As the distribution of V_t is left unspecified, some very flexible tail behavior is allowed for. Consequently, the drawback in the classical Box-Cox model (2.8) is addressed in the NonNeg model.

One role that the transformation parameter, λ , plays in our model is to stabilize the variance, i.e. to induce homoskedasticity (cf. Figure 1). An intercept in the model is superfluous because γ can be strictly positive. Our model echoes Equation (2.8) where the normal distribution in (2.8) is replaced with a nonnegative error distribution. If, in addition, $\lambda = 1$, our model becomes the discrete time version of Equation (2) in Barndorff-Nielsen & Shephard (2001). In general, the distributional form is not assumed to be known for the error component. Hence, the NonNeg model combines a parametric component for the persistence with a nonparametric component for the error. On the one hand, the new model is highly parsimonious. In particular, there are only two parameters that have to be estimated for the purpose of volatility forecasting, namely ϕ and λ . On the other hand, the specification is sufficiently flexible for modeling the error.

As mentioned earlier, there exists a lower bound for the percentage change in volatility in Model (2.2). A similar bound applies to our model. It is easy to show that $RV_t/RV_{t-1} \leq \phi^{1/\lambda}$ if $\lambda < 0$ (upper bound), and $RV_t/RV_{t-1} \geq \phi^{1/\lambda}$ if $\lambda > 0$ (lower bound). Typical estimated values of ϕ and λ in (2.12) for our empirical study are 0.639 and -0.278, respectively, implying that RV_t cannot increase by more than 500% from one time period to the next. As we will see later, our proposed estimator for λ depends on the ratios of successive RV's and hence the bound is endogenously determined.

3. ROBUST ESTIMATION & FORECASTING

3.1. Estimation of ϕ . For the exceptional situation when λ_0 is known, we propose to estimate ϕ using the extreme value estimator (EVE) defined by

$$(3.1) \quad \hat{\phi}_T = \min \left\{ \frac{RV_t^{\lambda_0}}{RV_{t-1}^{\lambda_0}} \right\}_{t=1}^T .$$

This estimator is the maximum likelihood estimator (MLE) of ϕ when the errors in (2.12) are iid exponentially distributed random variables; see Nielsen & Shephard (2003). Interestingly, the EVE is strongly consistent for more general error specifications, including heteroskedasticity and m -dependence. It is robust in the sense that the conditions allow for certain model misspecifications in V_t . For example, the order of m -dependence in the errors and the conditional distribution of RV_t may be incorrectly specified. Moreover, the EVE is strongly consistent even under quite general forms of heteroskedasticity. For a more detailed account of the EVE, see Preve (2008).

Like the well known ordinary least squares (OLS) estimator, the EVE is distribution-free in the sense that its consistency does not rely on a particular distributional assumption for the error component. However, the EVE is in many ways superior to the OLS estimator. For example, its rate of convergence can be faster than $O_p(T^{-1/2})$ even when $\phi_0 < 1$, whereas the OLS estimator converges faster than $O_p(T^{-1/2})$ only when $\phi_0 \geq 1$ Phillips (1987). Furthermore, unlike the OLS estimator the consistency conditions of the EVE do not involve the existence of any higher order moments.

Under additional technical conditions, Davis & McCormick (1989) and Feigin & Resnick (1992) obtain the limiting distribution of a linear programming estimator (LPE) for which the EVE in (3.1) appear as a special case when $\lambda = 1$ and the errors are iid. The authors show that the accuracy of the LPE depends on the index of regular variation at zero (or infinity) of the error distribution function. For example, for standard exponential errors, the index of regular variation at zero is 1 and the LPE/EVE converges to ϕ at the rate of $O_p(T^{-1})$. In general, a difficulty in the application of the limiting distribution is that the

index of regular variation at zero appears both in a normalizing constant and in the limit. Datta & McCormick (1995) avoid this difficulty by establishing the asymptotic validity of a bootstrap scheme based on the LPE.

It is readily verified that the EVE is positively biased and nonincreasing in T , that is, $\phi_0 < \hat{\phi}_{T_2} \leq \hat{\phi}_{T_1}$ with probability 1 for any $T_1 < T_2$.⁴ Hence, the accuracy of the EVE either remains the same or improves as the sample size increases (cf. Figure 2).

To understand the robustness of the EVE, consider the covariance stationary AR(1) model

$$RV_t = \phi RV_{t-1} + V_t, \quad t = 0, \pm 1, \pm 2, \dots$$

under the possible model misspecification

$$V_t = Z_t + \sum_{i=1}^m \theta_i Z_{t-i},$$

where $\{Z_t\}$ is a sequence of non-zero mean iid random variables. For $m > 0$ the identically distributed errors V_t are serially correlated. In this setting the OLS estimator of ϕ is inconsistent while the EVE remains consistent. In the first panel of Figure 2 we plot 100 observations simulated from the nonnegative ARMA(1,1) model, $RV_t = \phi RV_{t-1} + Z_t + \theta Z_{t-1}$ with $\phi = 0.5$, $\theta = 0.75$ and standard exponential noise. In the second panel of Figure 2 we plot the paths of the recursive EVEs and the recursive OLS estimates for ϕ obtained from the simulated data. In each pass, a new observation is added to the sample. It can be seen that the EVEs quickly approach the true value of ϕ , whereas the OLS estimates do not. Additionally, the OLS estimates fluctuate much more than the EVEs when the sample size is small, suggesting that in small samples the EVE is less sensitive to extreme deviations from the mean than the OLS estimator.

We now list the assumptions under which the consistency of the EVE holds. The proof of the proposition is found in Preve (2008).

Assumption 1a: In Model (2.12), $\lambda_0 \neq 0$, $\phi_0 > 0$ and the initial value RV_0 is almost surely positive; $\{V_t\}$ is a sequence of m -dependent, identically distributed, nonnegative continuous random variables.

Assumption 2a: The error component in Model (2.12) satisfies

$$P(c_1 < V_t < c_2) < 1,$$

for all $0 < c_1 < c_2 < \infty$.

Remark 3.1. It is important to point out that Assumption 2a is satisfied for any error distribution with unbounded nonnegative support.

⁴Whenever necessary we use the subscript T to emphasize on the sample size.

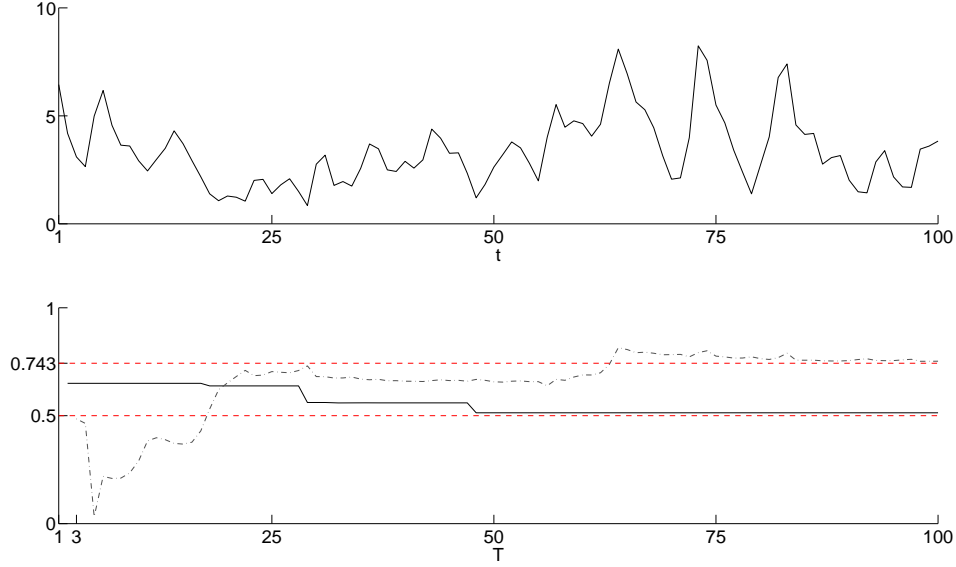


FIGURE 2. The top panel displays a time series plot of data simulated from the nonnegative ARMA(1, 1) process $RV_t = \phi RV_{t-1} + Z_t + \theta Z_{t-1}$ with $\phi = 0.5$, $\theta = 0.75$ and iid standard exponential noise. The bottom panel displays the paths of recursive EVEs and OLS estimates for ϕ in the misspecified AR(1) model $RV_t = \phi RV_{t-1} + Z_t$, obtained from $\{RV_1, \dots, RV_T\}$, where $T \in \{3, \dots, 100\}$. The solid line represents the EVEs and the dash-dotted line represents the OLS estimates.

Proposition 3.1. *For model (2.12) with a known λ_0 , let the EVE of ϕ (denoted by $\hat{\phi}_T$) be given by (3.1). Under Assumptions 1a and 2a, $\hat{\phi}_T \xrightarrow{a.s.} \phi_0$ as $T \rightarrow \infty$.*

3.2. Estimation of ϕ and λ . In practice, we do not know the true value of λ_0 . In this section we propose an EVE based two-stage estimation method for λ and ϕ in Model (2.12). The estimators are easily computable using standard computational software such as Matlab. In doing so, we also establish an expression for its one-step-ahead forecast. We then examine the asymptotic properties and the finite sample performance of the proposed estimation method.

Estimation of λ and ϕ is non-trivial, even under certain parametric and simplifying assumptions about V_t . For example, if $\{V_t\}$ is a sequence of independent exponentially distributed random variables it can be shown that the MLEs of λ and ϕ are inconsistent.

In our EVE based two-stage estimation method, we first choose $\hat{\lambda}_T$ to minimize the sum of squared one-step-ahead prediction errors

$$(3.2) \quad \hat{\lambda}_T = \arg \min_{\lambda \in \Lambda} \frac{1}{T} \sum_{t=1}^T [RV_t - \widehat{RV}_t(\lambda)]^2,$$

where

$$\widehat{RV}_t(\lambda) = \frac{1}{T} \sum_{i=1}^T \left[\hat{\phi}_T(\lambda) RV_{t-1}^\lambda + \widehat{V}_i(\lambda) \right]^{1/\lambda},$$

$$\hat{\phi}_T(\lambda) = \min \left\{ \left(\frac{RV_t}{RV_{t-1}} \right)^\lambda \right\}_{t=1}^T \quad \text{and} \quad \widehat{V}_i(\lambda) = RV_i^\lambda - \hat{\phi}_T(\lambda) RV_{i-1}^\lambda,$$

respectively. Although our estimator of λ looks like the standard nonlinear least squares (NLS) estimator of Jennrich (1969), the two approaches are quite different because in our model an explicit expression for $E(RV_t | RV_{t-1})$ is not available. In fact, the NLS estimators of λ and ϕ , that minimizes $\sum_t (RV_t^\lambda - \phi RV_{t-1}^\lambda)^2$, always take values of 0 and 1, respectively, and hence are inconsistent.

The intuition behind the proposed estimation method is that we expect $\widehat{RV}_t(\hat{\lambda}_T)$ to be close $E(RV_t | RV_{t-1})$ as T is large. This is not surprising since Model (2.12) implies that

$$RV_t = (\phi_0 RV_{t-1}^{\lambda_0} + V_t)^{1/\lambda_0},$$

and that

$$E(RV_t | RV_{t-1}) = E[(\phi_0 RV_{t-1}^{\lambda_0} + V_t)^{1/\lambda_0} | RV_{t-1}].$$

In the second stage, we use the EVE to estimate ϕ , i.e.,

$$(3.3) \quad \hat{\phi}_T = \hat{\phi}_T(\hat{\lambda}_T) = \min \left\{ \frac{RV_t^{\hat{\lambda}_T}}{RV_{t-1}^{\hat{\lambda}_T}} \right\}_{t=1}^T.$$

While we minimize the sum of squared one-step-ahead prediction errors when estimating λ , other criteria, such as minimizing the sum of absolute one-step-ahead prediction errors, can be used. We have experimented with absolute prediction errors using the S&P500 data and found that our out-of-sample forecasting results for the NonNeg model are quite insensitive to the choice of the objective function in the estimation stage. However, the objective function with squared prediction errors performs better in simulations.

With an estimated λ and ϕ , the proposed one-step-ahead semiparametric forecast expression for the NonNeg model is

$$\widehat{RV}_{T+1} = \frac{1}{T} \sum_{i=1}^T (\hat{\phi}_T RV_T^{\hat{\lambda}_T} + \widehat{V}_i)^{1/\hat{\lambda}_T},$$

where $\widehat{V}_i = RV_i^{\hat{\lambda}_T} - \hat{\phi}_T RV_{i-1}^{\hat{\lambda}_T}$ is the residual at time period i . Of course, in line with Granger & Newbold (1976), several forecasts of RV_{T+1} may be considered. For example, one could base a forecast on the well known approximation $E[h(Y)] \approx h[E(Y)]$ using $h(y) = y^{1/\lambda}$. However, this approximation does not take into account the nonlinearity of $h(y)$. For instance, if $Y \sim N(0, \sigma^2)$ and $h(y) = y^2$ then $E[h(Y)] = \sigma^2 \neq h[E(Y)] = 0$.

We now list the assumptions under which we can establish the asymptotic properties of the proposed estimators. The proof of the propositions is found in the Appendix.

Assumption 1b: In Model (2.12), $\lambda_0 \neq 0$ and $\phi_0 \in (0, 1)$; V_t is a sequence of iid nonnegative continuous random variables; for any $0 < c_1 < c_2 < \infty$, $P(c_1 < V_t < c_2) < 1$; RV_0 follows from the marginal distribution.

Assumption 2b: Λ in (3.2) is compact. Without loss of generality, assume that $\Lambda = [\lambda_{\min}, \lambda_{\max}]$ where $-\infty < \lambda_{\min} < \lambda_0 < \lambda_{\max} < 0$.⁵

Assumption 3b: For any $\lambda \neq \lambda_0$ and any i , we assume

$$E\left(\left(\phi_0^{\lambda/\lambda_0} RV_{t-1}^\lambda + RV_i^\lambda - \phi_0^{\lambda/\lambda_0} RV_{i-1}^\lambda\right)^{1/\lambda} \middle| RV_{t-1}\right) \neq E(RV_t | RV_{t-1})$$

Assumption 4b: Assume that $E(V_t) < \infty$.

Remark 3.2. All the conditions listed here are primitive and very easy to check. Assumption 1b ensures that V_t is stationary and ergodic, implying that RV_t is also stationary and ergodic. Assumption 2b is the standard compactness condition. Assumption 3b is closely related to the identification condition for NLS. Assumption 4b imposes a simple moment condition.

Proposition 3.2. For model (2.12) with a unknown λ , denote the estimators of λ and ϕ by $\hat{\lambda}_T$ and $\hat{\phi}_T$, given by (3.2) and (3.3), respectively. Under Assumptions 1b-4b, $\hat{\lambda}_T \xrightarrow{a.s.} \lambda_0$ and $\hat{\phi}_T \xrightarrow{a.s.} \phi_0$ as $T \rightarrow \infty$.

Assumption 5b: Assume

$$\frac{1}{T} \sum_{t=1}^T \left\{ (RV_t - E(RV_t | RV_{t-1}))^2 \left(\frac{\sum_{i=1}^T f_i(\lambda_0)^{1/\lambda_0} \ln f_i(\lambda_0) \frac{\partial f_i(\lambda_0)}{\partial \lambda}}{T} \right)^2 \right\} \xrightarrow{p} \Omega,$$

where

$$f_i(\lambda) = \hat{\phi}_T(\lambda) RV_{t-1}^\lambda + RV_i^\lambda - \hat{\phi}_T(\lambda) RV_{i-1}^\lambda.$$

Assumption 6b: For some $r > 2$,

$$E \left| (RV_t - E(RV_t | RV_{t-1})) \frac{1}{T} \sum_{i=1}^T f_i(\lambda_0)^{1/\lambda_0} \ln f_i(\lambda_0) \frac{\partial f_i(\lambda_0)}{\partial \lambda} \right|^r < \infty.$$

Assumption 7b: Assume

$$\frac{1}{T} \sum_{i=1}^T f_i(\lambda_0)^{1/\lambda_0} \ln f_i(\lambda_0) \frac{\partial f_i(\lambda_0)}{\partial \lambda} = O_p(1).$$

⁵If λ_0 is positive, a similar condition can be imposed and the proof can be easily modified.

Assumption 8b: Assume $g_T(\lambda) \xrightarrow{P} g(\lambda)$ with $g(\lambda_0) > 0$ (denote $g(\lambda_0)$ by g), where

$$(3.4) \quad g_T(\lambda) = \frac{1}{T} \sum_{t=1}^T \left\{ - \left(\frac{1}{T} \sum_{i=1}^T f_i(\lambda)^{1/\lambda} \ln f_i(\lambda) \frac{\partial f_i(\lambda)}{\partial \lambda} \right)^2 \right. \\ \left. + \left(RV_t - \frac{1}{T} \sum_{i=1}^T f_i(\lambda)^{1/\lambda} \right) \times \left[\frac{1}{T} \sum_{i=1}^T \left(f_i(\lambda)^{1/\lambda} \ln f_i(\lambda) \frac{\partial f_i(\lambda)}{\partial \lambda} \right)^2 \right] \right. \\ \left. + \frac{1}{T} \sum_{i=1}^T f_i(\lambda)^{1/\lambda-1} \left(\frac{\partial f_i(\lambda)}{\partial \lambda} \right)^2 + \frac{1}{T} \sum_{i=1}^T f_i(\lambda)^{1/\lambda} \ln f_i(\lambda) \frac{\partial^2 f_i(\lambda)}{\partial \lambda^2} \right\}$$

Remark 3.3. All the new conditions are of high level and less primitive than Assumptions 1b-4b. Given the fact that the estimators are quite nonstandard, some trade-offs may be needed.

Proposition 3.3. For model (2.12) with a unknown λ_0 , denote the estimators of λ and ϕ by $\hat{\lambda}_T$ and $\hat{\phi}_T$, given by (3.2) and (3.3), respectively. Under Assumptions 1b-8b,

$$\sqrt{T}(\hat{\lambda}_T - \lambda_0) \xrightarrow{d} N(0, g^{-1}\Omega g^{-1}).$$

Remark 3.4. Although we only establish the asymptotic distribution for $\hat{\lambda}_T$, the result is remarkable. This is because in the special case of a nonnegative AR(1) model with iid exponential errors, the distribution of the EVE for ϕ is nonstandard and asymptotically exponential; see Nielsen & Shephard (2003). As a result, it is anticipated that the asymptotic distribution for $\hat{\phi}_T$ is non-Gaussian in our model. It appears that the asymptotic distribution for $\hat{\lambda}_T$ does not seem to be dependent on that of $\hat{\phi}_T$.

3.3. Monte Carlo Evidence. We now examine the performance of our estimation method via simulations. We designed two experiments in which data are generated by the non-negative model

$$RV_t^\lambda = \phi RV_{t-1}^\lambda + Z_t + \theta Z_{t-1}, \quad t = 1, \dots, T$$

with iid standard exponential driving noise Z_t .

In the first Monte Carlo experiment λ is assumed to be known and we only estimate ϕ using the EVE given in (3.1). In this case the consistency is robust to the specification of the dependence structure in the error term. Hence, we simulate data from the model with the true value of θ being different from zero. In particular, the true values were set to $\lambda = -0.28$ and $\theta = 0.75$. The true value of ϕ is set at 0.4 and 0.7, respectively.

In the second experiment λ is assumed to be unknown and is estimated together with ϕ using the proposed two-step method. The true parameter values are $\lambda = -0.6$, $\phi = 0.8$ and $\theta = 0$.

TABLE 2. Simulation results for the proposed two-step estimation method. Summary statistics for $\hat{\lambda}_T$ and $\hat{\phi}_T$ based on data generated by the nonnegative process $RV_t^{-0.28} = \phi RV_{t-1}^{-0.28} + Z_t + 0.75Z_{t-1}$ with iid standard exponential noise. True values for ϕ are 0.4 and 0.7, respectively. Mean and SD denotes the sample mean and standard deviation, respectively. Results based on 5000 Monte Carlo replications.

True	Est	T = 200		T = 400		T = 800	
		Mean	SD	Mean	SD	Mean	SD
0.4	$\hat{\phi}_T$.4350	.0185	.4245	.0130	.4169	.0090
0.7	$\hat{\phi}_T$.7156	.0085	.7109	.0058	.7075	.0040

TABLE 3. Simulation results for the proposed two-step estimation method. Summary statistics for $\hat{\lambda}_T$ and $\hat{\phi}_T$ based on data generated by the nonnegative process $RV_t^\lambda = \phi RV_{t-1}^\lambda + Z_t$ with iid standard exponential noise. True values for λ and ϕ are -0.6 and 0.8. Mean and SD denotes the sample mean and standard deviation, respectively. Results based on 5000 Monte Carlo replications.

True	Est	T = 200		T = 400		T = 800	
		Mean	SD	Mean	SD	Mean	SD
$\lambda=-0.6$	$\hat{\lambda}_T$	-0.8035	0.2647	-0.7146	0.1825	-0.6638	0.1358
$\phi=0.8$	$\hat{\phi}_T$	0.7464	0.0708	0.7689	0.0514	0.7825	0.0394

The values chosen for λ and ϕ in the two experiments are empirically realistic (cf. the results of Section 6). We consider sample sizes of $T = 200, 400$, and 800 in the both experiments. The sample size of 400 is close to the smallest sample size used for estimation in our empirical study, while the sample size of 800 is close to the largest sample size in the study.

Simulation results based on 5000 Monte Carlo replications are reported in Tables 2 and 3. Several interesting results emerge from the tables. First, the smaller the value of T , the greater the sample bias in $\hat{\phi}_T$ in the first experiment and in $\hat{\lambda}_T$ and $\hat{\phi}_T$ in the second experiment. Second, as T increases, the standard error of $\hat{\phi}_T$ in the first experiment and those of $\hat{\lambda}_T$, and $\hat{\phi}_T$ in the second experiment, decreases. It may be surprising to see that the bias of $\hat{\phi}_T$ can be negative in the second experiment. The negative bias arises because λ is estimated. Figure 3 plots histograms of $\hat{\lambda}_T$ and $\hat{\phi}_T$ obtained from 5000 Monte Carlo replications with common sample size $T = 800$ in the second experiment. In sum, it seems that the proposed estimation method works well, especially when T is reasonably large.

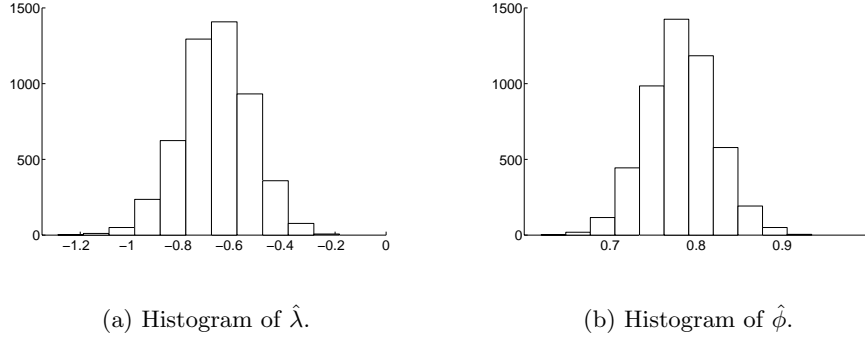


FIGURE 3. Histograms of $\hat{\lambda}_T$ and $\hat{\phi}_T$ based on data generated by the non-negative process $RV_t^\lambda = \phi RV_{t-1}^\lambda + Z_t$ with iid standard exponential noise. Results based on 5000 Monte Carlo replications with common sample size $T = 800$. True values for λ and ϕ are -0.6 and 0.8, respectively.

4. COMPETING MODELS

Numerous models and methods have been applied to forecast stock market volatility. For example, ARCH-type models are popular in academic publications and RiskMetrics is widely used in practice. Both methods use returns to forecast volatility at the same frequency. However, since the squared return is a noisy estimator of volatility ABDL (2003) instead consider RV and present strong evidence to support time series models based directly on RV in terms of forecast accuracy. Motivated by their empirical findings, we compare the forecast accuracy of the proposed model against three time series methods, all based on RV: (1) the linear Gaussian AR model (LinGau); (2) the log-linear Gaussian AR model (LogGau) and (3) the logarithmic autoregressive fractionally integrated moving average (ARFIMA) model. We also compare the performance of our model against the exponential smoothing method, a RV version of RiskMetrics. The LinGau and LogGau models are defined by equations (2.10) and (2.11), respectively. We now review the exponential smoothing method and the ARFIMA model.

Exponential Smoothing-ES. Exponential smoothing is a simple method of forecasting, where the one-step-ahead forecast of RV is given by

$$(4.1) \quad \widehat{RV}_{T+1} = (1 - \alpha)RV_T + \alpha\widehat{RV}_T = (1 - \alpha) \sum_{i=0}^{T-1} \alpha^i RV_{T-i},$$

with $0 < \alpha < 1$.

The exponential smoothing formula can be understood as the RV version of RiskMetrics, where the squared return, X_T^2 , is replaced by RV_T . Under the assumption of conditional

normality of the return distribution, X_T^2 is an unbiased estimator of σ_t^2 . RiskMetrics recommends $\alpha = 0.94$ for daily data and $\alpha = 0.97$ for monthly data.

To see why the squared return is a noisy estimator of volatility even under the assumption of conditional normality of the return distribution, suppose that X_t follows Equation (2.3). Conditional on σ_t , it is easy to show that Lopez (2001)

$$(4.2) \quad P\left(X_t^2 \in \left[\frac{1}{2}\sigma_t^2, \frac{3}{2}\sigma_t^2\right]\right) = 0.2588.$$

This implies that with a probability close to 0.74 the squared return is at least 50% greater, or at most 50% smaller, than the true volatility. Not surprisingly, Andersen & Bollerslev (1998) find that RiskMetrics is dominated by models based directly on RV. For this reason, we do not use RiskMetrics directly. Instead, we use (4.1) with $\alpha = 0.97$, which assigns a weight of 3% to the most recently observed RV. We remark that the forecasting results of Section 6 were qualitatively left unchanged when other values for α were used.

ARFIMA(p, d, q). Long range dependence is a well documented stylized fact for volatility of many financial time series. Fractional integration has previously been used to model the long range dependence in volatility and log-volatility. The autoregressive fractionally integrated moving average (ARFIMA) was considered as a model for logarithmic RV in ABDL (2003) and Deo, Hurvich & Lu (2006), among others. In this paper, we consider two parsimonious ARFIMA models for log-RV, namely, an ARFIMA(0,d,0) and an ARFIMA(1,d,0).

The ARFIMA($p, d, 0$) model for the log-RV is defined by

$$(1 - \beta_1 B - \dots - \beta_p B^p)(1 - B)^d(\ln RV_t - \mu) = \varepsilon_t,$$

where the parameters $\mu, \beta_1, \dots, \beta_p$ and the memory parameter d are real valued, and $\{\varepsilon_t\}$ is a sequence of independent $N(0, \sigma_\varepsilon^2)$ distributed random variables.

Following a suggestion of a referee, we estimate all the parameters of the ARFIMA model using an approximate ML method by minimizing the sum of squared one-step-ahead prediction errors. See Beran (1995), Chung & Baillie (1993) and Doornik & Ooms (2004) for detailed discussions about the method and Monte Carlo evidence supporting the method. Compared to the exact ML of Sowell (1992), there are two advantages in the approximate ML method. First, it does not require d to be less than 0.5. Second, it has smaller finite sample bias. Compared to the semi-parametric methods, it is also more efficient.⁶ The one-step-ahead forecast of an ARFIMA($p, d, 0$) with $p = 0$ at time period

⁶We also applied the exact ML method of Sowell (1992) and the exact local Whittle estimator of Shimotsu & Phillips (2005) in our empirical study and found that the forecasts remained essentially unchanged.

T is given by

$$\widehat{RV}_{T+1} = \exp \left\{ \hat{\mu} - \sum_{j=0}^{T-1} \hat{\pi}_j (\ln RV_{T-j} - \hat{\mu}) + \frac{\hat{\sigma}_\varepsilon^2}{2} \right\},$$

and for $p = 1$ by

$$\begin{aligned} \widehat{RV}_{T+1} = \exp \left\{ \hat{\mu} + \hat{\beta} (\ln RV_T - \hat{\mu}) \right. \\ \left. + \sum_{j=1}^{T-1} \hat{\pi}_j [\hat{\beta} (\ln RV_{T-j} - \hat{\mu}) - (\ln RV_{T-j+1} - \hat{\mu})] + \frac{\hat{\sigma}_\varepsilon^2}{2} \right\}, \end{aligned}$$

where

$$\hat{\pi}_j = \frac{\Gamma(j - \hat{d})}{\Gamma(j + 1)\Gamma(-\hat{d})},$$

and $\Gamma(\cdot)$ denotes the gamma function.

5. FORECAST ACCURACY MEASURES

It is not obvious which accuracy measure is more appropriate for the evaluation of the out-of-sample performance of alternative time series methods. Rather than making a single choice, we use four measures to evaluate forecast accuracy, namely, the mean absolute error (MAE), the mean absolute percentage error (MAPE), the mean square error (MSE), and the mean square percentage error (MSPE). Let \widehat{RV}_{it} denote the forecast of RV_t from model i at time period t and define the accompanying forecast error by $e_{it} = RV_t - \widehat{RV}_{it}$. The four accuracy measures are defined, respectively, by

$$\begin{aligned} \text{MSE} &= \frac{1}{P} \sum_{t=1}^P e_{it}^2, & \text{MSPE} &= \frac{100}{P} \sum_{t=1}^P \left(\frac{e_{it}}{RV_t} \right)^2, \\ \text{MAE} &= \frac{1}{P} \sum_{t=1}^P |e_{it}|, & \text{MAPE} &= \frac{100}{P} \sum_{t=1}^P \left| \frac{e_{it}}{RV_t} \right|. \end{aligned}$$

where P is the length of the forecast evaluation period.

An advantage of using MAE instead of MSE is that it has the same scale as the data. The MAPE and the MSPE are scale independent measures. For a comprehensive survey on these and other forecast accuracy measures see Hyndman & Koehler (2006).

When calculating the forecast error, it is implicitly assumed that RV_t is the true volatility at time t . However, in reality the volatility proxy RV_t is different from the true latent volatility. Several recent papers discuss the implications of using noisy volatility proxies when comparing volatility forecasts under certain loss functions. See, for example, Andersen & Bollerslev (1998), Hansen (2006) and Patton (2009). The impact is found to be particularly large when the squared return is used as a proxy for the true volatility, but diminishes with the approximation error. In this paper, the true (monthly) volatility is

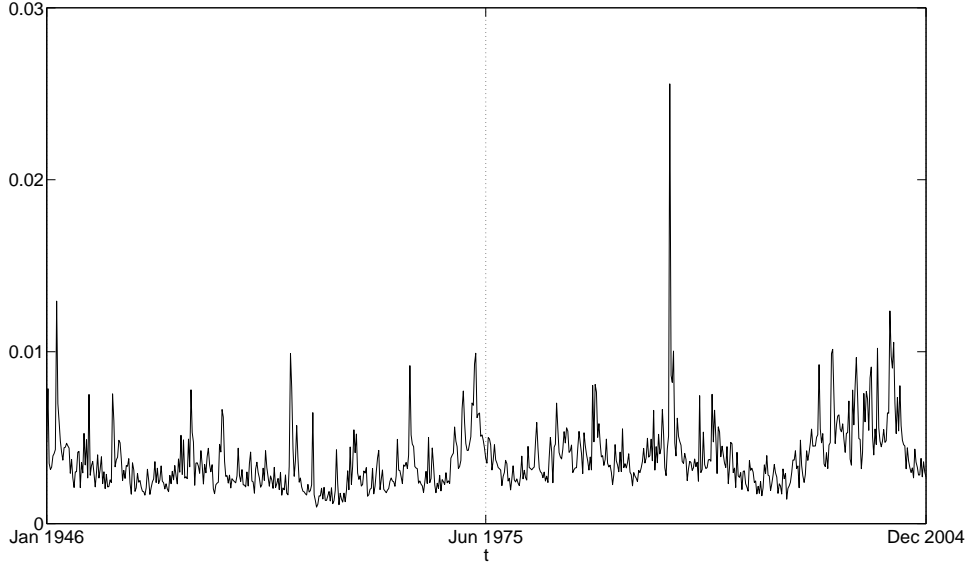


FIGURE 4. S&P 500 monthly realized volatilities, Jan 1946-Dec 2004. The vertical dashed line indicates the end of the initial sample period used for estimation in our first out-of-sample forecasting exercise.

approximated by the RV using 22 (daily) squared returns. As a result, the approximation error is expected to be much smaller than in the case of using a single squared return.

6. EMPIRICAL STUDY

Forecasting S&P 500 Monthly Realized Volatility. The data used in this paper consist of daily closing prices for the S&P 500 index over the period January 2, 1946-December 31, 2004, covering 708 months and 15,054 trading days. We measure the monthly volatility using realized volatility calculated from daily data. Denote the log-closing price on the k 'th trading day in month t by $p(t, k)$. Assuming there are T_t trading days in month t , we define the monthly RV as

$$RV_t = \sqrt{\frac{1}{T_t} \sum_{k=2}^{T_t} [p(t, k) - p(t, k-1)]^2}, \quad t = 1, \dots, 708$$

where $1/T_t$ serves the purpose of standardization.

In order to compare the out-of-sample predictive accuracy of the competing methods, we split the time series of monthly RV into two subsamples. The first time period is used for the initial estimation. The second period is the hold-back sample used for forecast evaluation. When calculating the forecasts we use a recursive scheme, where the size of the sample used for estimation successively increases as new forecasts are made. The time

TABLE 4. Summary statistics for the S&P 500 RV data. JB is the p -value of the Jarque-Bera test under the null hypothesis that the data are from a normal distribution.

	Mean	Max	Skew	Kurt	JB	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$
RV	0.0037	0.0256	3.307	28.791	0.000	0.576	0.477	0.408
log-RV	-5.6873	-3.6661	0.389	3.657	0.000	0.683	0.595	0.511

series plot of monthly RV for the entire sample is shown in Figure 4, where the vertical dashed line indicates the end of the initial sample period used for estimation in our first forecasting exercise.

Table 4 shows the mean, maximum, skewness, kurtosis, the p -value of the JB test statistic for normality, and the first three sample autocorrelations of the entire sample for RV and log-RV. For RV, the sample maximum is 0.0256 which occurred in October 1987. The sample kurtosis is 28.791 indicating that the distribution of RV is non-Gaussian. In contrast, log-RV has a much smaller kurtosis (3.657) and is less skewed (0.389). It is for this reason that we include Gaussian time series models for log-RV in the competition. However, a formal test for normality via the JB statistic strongly rejects the null hypothesis of normality of log-RV, suggesting that further improvements over log-linear Gaussian approaches are possible.

Higher order sample autocorrelations are in general slowly decreasing and not statistically negligible, indicating that RV and log-RV are predictable. To test for possible unit roots, augmented Dickey-Fuller (ADF) test statistics were calculated. The ADF statistic for the sample from 1946 to 2004 is -5.69 for RV and -5.43 for log-RV, which is smaller than -2.57, the critical value at the 10% significance level. Hence, we reject the null hypothesis that RV or log-RV has a unit root.

Forecasting Results. Each competing model was fitted to the in-sample RV data and used to obtain one-period-ahead out-of-sample forecasts. Since a forecast frequency of one month is sufficiently important in practical applications, we focus on one-period-ahead forecasts in this paper. However, multi-period-ahead forecasts can be obtained in a similar manner.

We perform two out-of-sample forecasting exercises. In both exercises, we use the recursive scheme, where the size of the sample used to estimate the competing models grows as we make forecasts for successive observations. More precisely, in the first exercise, we first estimate all the competing models with data from the period January 1946-June 1975 and use the estimated models to forecast the RV of July 1975. We then estimate all

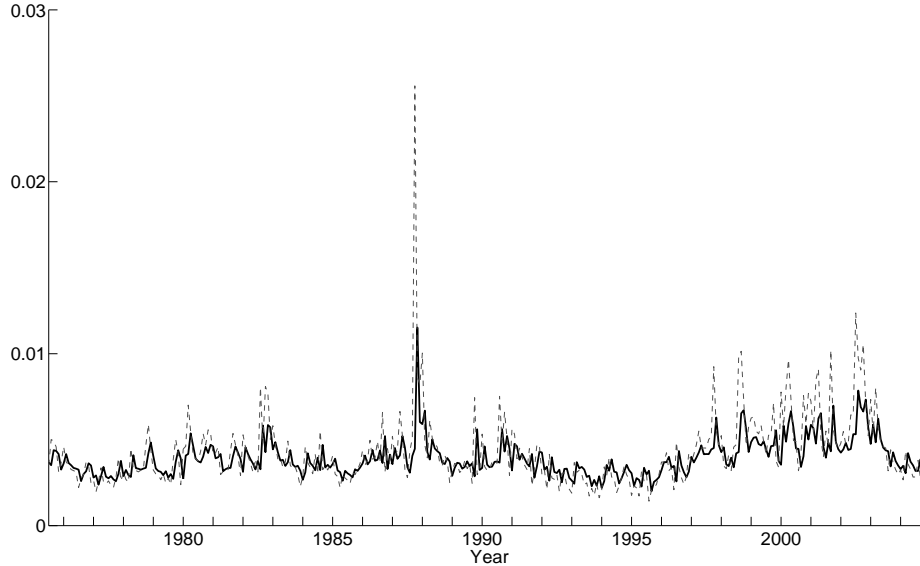


FIGURE 5. Realized volatility and out-of-sample NonNeg forecasts for the period Jul 1975-Dec 2004. *Dashed line*: S&P 500 monthly realized volatility. *Solid line*: one-step-ahead NonNeg forecasts.

models with data from January 1946-July 1975 and use the model estimates to forecast the RV of August 1975. This process is repeated until, finally, we estimate the models with data from January 1946-November 2004. The final model estimates are used to forecast the RV of December 2004, the last observation in the sample.

Sample including the 1987 Crash. In the first exercise, the first month for which an out-of-sample volatility forecast is obtained is July 1975. In total 354 monthly volatilities are forecasted, including the volatility of October 1987 when the stock market crashed and the RV is 0.0256.

In Figure 5, we plot the monthly RV and the corresponding one-month-ahead NonNeg forecasts for the out-of-sample period, July 1975 to December 2004. It seems that the NonNeg model captures the overall movements in RV reasonably well. The numerical computation of the 354 forecasts is fast and takes less than ten minutes on a standard desktop computer.

In Figure 6, we plot the recursive estimates, $\hat{\phi}_T$ and $\hat{\lambda}_T$. While $\hat{\lambda}_T$ takes values from -0.45 to -0.28, $\hat{\phi}_T$ ranges between 0.58 and 0.64. It may be surprising to see that the path of $\hat{\phi}_T$ is nonmonotonic. This is because the estimates of the transformation parameter, λ , are varying over time. Our empirical estimates of λ seem to corroborate well with the optimal value of λ obtained by Goncalves & Meddahi (2009) using simulations in the

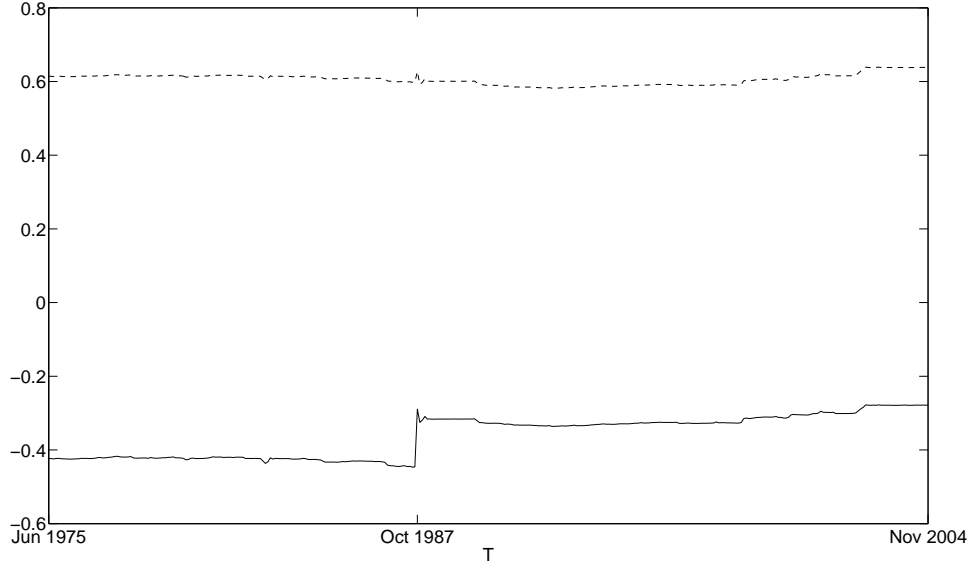


FIGURE 6. NonNeg recursive parameter estimates for the first out-of-sample forecasting experiment. *Solid line*: path of $\hat{\lambda}_T$. *Dashed line*: path of $\hat{\phi}_T$.

context of a GARCH diffusion and a two factor SV model. While $\hat{\phi}_T$ is quite stable, $\hat{\lambda}_T$ jumps in October 1987.

In Figure 7, we plot the NonNeg residuals, \hat{V}_t , for the entire sample. Also depicted is a histogram of the residuals. From the histogram, it can be seen that there is no evidence of outliers in \hat{V}_t and that there is an acute “peak” about the mean. These two features suggest that the error term in the NonNeg model tends to take a value around the mean and may explain why the NonNeg model can under-predict (cf. Figure 5).

Table 5 reports the forecasting performance of the competing models under the four forecast accuracy measures of Section 5. Several results emerge from the table. First, the relative performances of the competing models are sensitive to the measures. Under the MSE measure, the two ARFIMA models ranks as the best, followed by the NonNeg model and the log-Gaussian model. ABDL (2003) found that their ARFIMA models perform well in terms of R^2 in the Mincer-Zarnomitz regression. Since the MSE is closely related to the R^2 in the Mincer-Zarnomitz regression, our results reinforce their findings. However, the rankings obtained under MSE are very different from those obtained under the other three accuracy measures. The MAE and the MSPE ranks the NonNeg model the first while the MAPE and the MSPE ranks the log-Gaussian model the first. Second, the performances of the two ARFIMA models are very similar under all measures. To understand why,

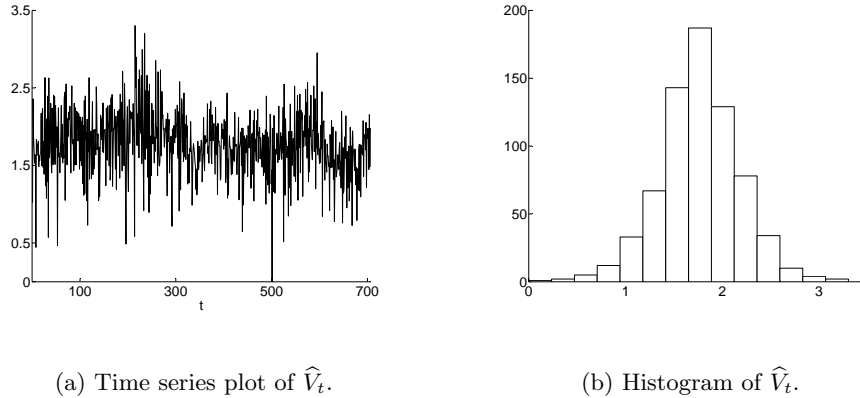


FIGURE 7. Time series plot and histogram of the residuals \widehat{V}_t , obtained when estimating the NonNeg model using the entire sample.

TABLE 5. Forecasting performance of the competing methods under four different accuracy measures. Results based on 354 one-step-ahead forecasts for the period *Jul 1975-Dec 2004*.

	$\text{MAE} \times 10^3$		MAPE		$\text{MSE} \times 10^6$		MSPE	
	value	rank	value	rank	value	rank	value	rank
ES	1.2681	6	31.04	6	3.8622	6	15.30	6
LinGau	0.9746	5	20.93	3	3.3117	5	7.80	3
LogGau	0.9544	2	20.74	1	3.0759	4	7.56	1
ARFIMA(0,d,0)	0.9615	3	22.09	4	2.8474	1	8.04	4
ARFIMA(1,d,0)	0.9615	3	22.08	5	2.8510	2	8.04	4
NonNeg	0.9536	1	20.78	2	3.0748	3	7.56	1

we plotted the sample autocorrelation functions of the ARFIMA(0,d,0) residuals for the entire sample and found that fractional differencing alone successfully removes the serial dependence in log-RV. Third, the improvement of ARFIMA(0,d,0) over NonNeg is 7.40% in terms of MSE. On the other hand, the improvement of NonNeg over ARFIMA(0,d,0) is 0.83%, 6.30% and 6.35% in terms of MAE, MAPE, and MSPE, respectively. These improvements are striking as we expect ARFIMA models to be hard to beat. Fourth, ES performs the worst in all cases.

Sample Post the 1987 Crash. To examine the sensitivity of our results with respect to the 1987 crash and the 1997 crash due to the Asian financial crisis, we redo the forecasting exercise so that the first month for which an out-of-sample volatility forecast is obtained is January 1988 and the last month is September 1997.

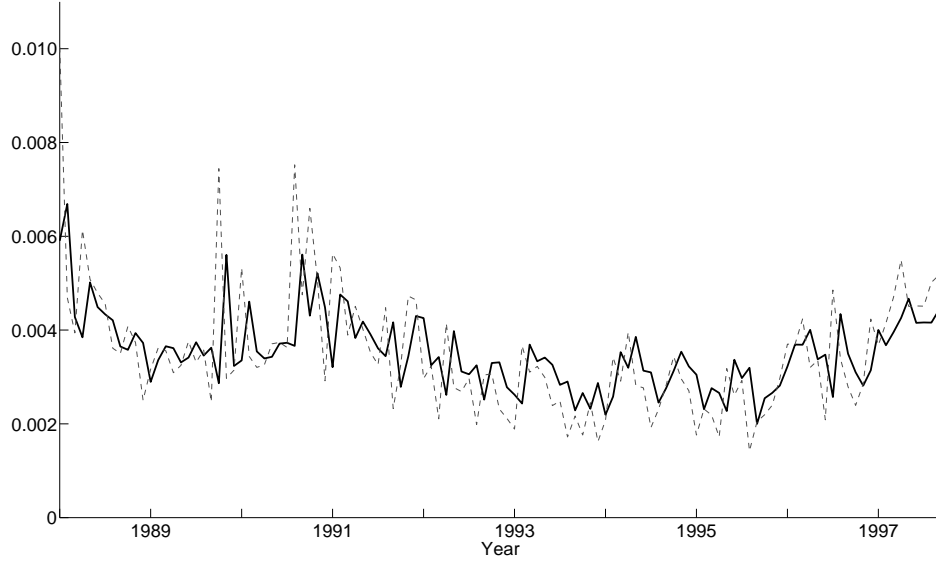


FIGURE 8. Realized volatility and out-of-sample NonNeg forecasts for the period Jan 1988-Sep 1997. *Dashed line*: S&P 500 monthly realized volatility. *Solid line*: one-month-ahead NonNeg forecasts.

In Figure 8, we plot the monthly RV and the corresponding one-month-ahead NonNeg forecasts for the out-of-sample period, January 1988-September 1997. As before, forecasts from the NonNeg model captures the overall movements in RV reasonably well. Table 6 reports the forecasting performance of the competing models under the four forecast accuracy measures. Since the RVs are smaller in this subsample, as expected, the MAE and the MSE are smaller than before. However, the relative performances of the competing models obtained from the subsample are very similar to those obtained from the entire sample. For example, as before the NonNeg model performs the best according to the MAE. Moreover, the two ARFIMA models perform better than the other competing models in terms of MSE.

7. CONCLUDING REMARKS & FUTURE RESEARCH

In this paper, a simple time series model is introduced to model and forecast RV. The new model combines a nonnegative valued process for the error term with the flexibility of a Box-Cox like power transformation. The transformation is used to induce homoskedasticity while the nonnegative support of the error distribution overcomes the truncation problem in the classical Box-Cox setup. The model is semiparametric as only the support

TABLE 6. Forecasting performance of the competing methods under four different accuracy measures. Results based on 117 one-step-ahead forecasts for the period *Jan 1988-Sep 1997*.

	MAE $\times 10^3$		MAPE		MSE $\times 10^6$		MSPE	
	value	rank	value	rank	value	rank	value	rank
ES	1.0770	6	35.38	6	1.7070	6	20.18	6
LinGau	0.7828	5	23.88	5	1.2579	5	10.73	5
LogGau	0.7792	4	23.38	1	1.2724	4	10.53	3
ARFIMA(0,d,0)	0.7789	3	23.65	4	1.1603	1	10.24	2
ARFIMA(1,d,0)	0.7782	2	23.61	3	1.1621	2	10.22	1
NonNeg	0.7769	1	23.45	2	1.2599	3	10.58	4

and not the functional form of the error distribution is assumed to be known. Consequently, the proposed model is highly parsimonious, containing only two parameters that have to be estimated for the purpose of forecasting. A two-stage estimation method is proposed to estimate the parameters of the new model. Asymptotic properties are established for the new estimation method. Simulation studies validate the new estimation method and suggest that it works reasonably well in finite samples.

We empirically examine the forecast performance of the proposed model relative to a number of existing methods, using monthly S&P 500 RV data. The out-of-sample performances were evaluated under four different forecast accuracy measures (MAE, MAPE, MSE and MSPE). We found strong empirical evidence that our nonnegative model produce highly competitive forecasts.

Why does the simple nonnegative model produce such competitive forecasts? Firstly, as shown in Section 2.2, the logarithmic transformation may not induce homoskedasticity and normality as well as anticipated. A more general transformation may be needed. Secondly, the nonnegative model is highly parsimonious. This new approach is in sharp contrast to the traditional approach which aims to find a model that removes all the dynamics in the original data. When the dynamics are complex, a model with a rich parametrization is called for. This approach may come with the cost of over-fitting and hence may not necessarily lead to superior forecasts. By combining a parametric component for the persistence and a nonparametric error component, our approach presents an effective utilization of more recent information.

Although we only examine the performance of the proposed model for predicting S&P 500 realized volatility one month ahead, the technique itself is quite general and can be applied in many other contexts. First, the method requires no modification when

applied to intra-day data to forecast daily RV. In this context, it would be interesting to compare our method with the preferred method in ABDL (2003). Second, our model can easily be extended into a multivariate context by constructing a nonnegative vector autoregressive model. Third, while we focus on stock market volatility in this paper, other financial assets and financial volatility from other financial markets can be treated in the same fashion. Fourth, as two alternative nonnegative models, it would be interesting to compare the performance of our model with that of Cipollino et al. (2006). Finally, it would be interesting to examine the usefulness of the proposed model for multi-step-ahead forecasting. These extensions will be considered in later work.

8. APPENDIX

8.1. Proof of Proposition 3.2. Following Theorem 2.1 of Newey and McFadden (1994), we treat $\widehat{\lambda}_T$ as an extremum estimator and check the four conditions that are sufficient for consistency. The consistency of $\widehat{\phi}_T$ follows from the consistency of $\widehat{\lambda}_T$ and a result developed in Preve (2008), as we will show later.

The first condition is about the identification, namely, λ_0 must be the unique minimizer of $Q_0(\lambda)$, where $Q_0(\lambda)$ is the probability limit of $Q_T(\lambda)$, the objective function which defines $\widehat{\lambda}_T$, i.e.,

$$Q_T(\lambda) = \frac{1}{T} \sum_{t=1}^T \left\{ RV_t - \frac{1}{T} \sum_{i=1}^T \left(\widehat{\phi}_T(\lambda) RV_{t-1}^\lambda + RV_i^\lambda - \widehat{\phi}_T(\lambda) RV_{i-1}^\lambda \right)^{1/\lambda} \right\}^2.$$

To obtain an expression for $Q_0(\lambda)$, first denote $RV_t^{\lambda_0}$ by X_t . Thus, $X_t^{\lambda/\lambda_0} = RV_t^\lambda$, and $\forall \lambda \in \Theta$,

$$\widehat{\phi}_T(\lambda) = \min_t \left\{ \frac{RV_t^\lambda}{RV_{t-1}^\lambda} \right\} = \min_t \left\{ \frac{X_t^{\lambda/\lambda_0}}{X_{t-1}^{\lambda/\lambda_0}} \right\} = \left(\min_t \left\{ \frac{X_t}{X_{t-1}} \right\} \right)^{\lambda/\lambda_0}.$$

As $T \rightarrow \infty$, Preve (2008) showed that under Assumption 1b, $\min_t \left\{ \frac{X_t}{X_{t-1}} \right\} \xrightarrow{a.s.} \phi_0$ and hence $\left(\min_t \left\{ \frac{X_t}{X_{t-1}} \right\} \right)^{\lambda/\lambda_0} \xrightarrow{a.s.} \phi_0^{\lambda/\lambda_0}$. Let

$$Q_0(\lambda) = E \left\{ RV_t - E \left[\left(\phi_0^{\lambda/\lambda_0} RV_{t-1}^\lambda + RV_i^\lambda - \phi_0^{\lambda/\lambda_0} RV_{i-1}^\lambda \right)^{1/\lambda} \mid RV_{t-1} \right] \right\}^2.$$

The pointwise strong consistency of $Q_T(\lambda)$ to $Q_0(\lambda)$ follows from Assumption 1b because RV_t is stationary and ergodic and the strong law of large numbers (LLN) applies.

Furthermore,

$$\begin{aligned}
Q_0(\lambda_0) &= E \left\{ RV_t - E \left[\left(\phi_0 RV_{t-1}^{\lambda_0} + RV_i^{\lambda_0} - \phi_0 RV_{i-1}^{\lambda_0} \right)^{1/\lambda_0} \mid RV_{t-1} \right] \right\}^2 \\
&= E \left\{ RV_t - E \left[\left(\phi_0 RV_{t-1}^{\lambda_0} + V_i \right)^{1/\lambda_0} \mid RV_{t-1} \right] \right\}^2 \\
&= E \left\{ RV_t - E \left[\sum_{k=0}^{\infty} \binom{1/\lambda_0}{k} \left(\phi_0 RV_{t-1}^{\lambda_0} \right)^{1/\lambda_0-k} V_i^k \mid RV_{t-1} \right] \right\}^2 \\
&= E \left\{ RV_t - \left[\sum_{k=0}^{\infty} \binom{1/\lambda_0}{k} \left(\phi_0 RV_{t-1}^{\lambda_0} \right)^{1/\lambda_0-k} E(V_i^k) \mid RV_{t-1} \right] \right\}^2 \\
&= E \left\{ RV_t - \left[\sum_{k=0}^{\infty} \binom{1/\lambda_0}{k} \left(\phi_0 RV_{t-1}^{\lambda_0} \right)^{1/\lambda_0-k} E(V_t^k) \mid RV_{t-1} \right] \right\}^2 \\
&= E \left\{ RV_t - E \left[\left(\phi_0 RV_{t-1}^{\lambda_0} + V_t \right)^{1/\lambda_0} \mid RV_{t-1} \right] \right\}^2 \\
&= E \{ RV_t - E [RV_t \mid RV_{t-1}] \}^2,
\end{aligned}$$

where the third equality follows the generalized binomial theorem, the fifth equality follows Assumption 1b as V_t is identically distributed, and the sixth equality follows Assumption 1b as V_t is independent of RV_{t-1} (if V_t is m -dependent with $m > 0$, this equality does not hold.).

By Assumption 3b, we have, for all $\lambda \neq \lambda_0$,

$$\begin{aligned}
Q_0(\lambda) &= E (RV_t - E(RV_t \mid RV_{t-1}))^2 \\
&\quad + \left\{ E(RV_t \mid RV_{t-1}) - E \left[\left(\phi_0^{\lambda/\lambda_0} RV_{t-1}^{\lambda} + RV_i^{\lambda} - \phi_0^{\lambda/\lambda_0} RV_{i-1}^{\lambda} \right)^{1/\lambda} \mid RV_{t-1} \right] \right\}^2 \\
&> Q_0(\lambda_0).
\end{aligned}$$

The second condition is the continuity of $Q_0(\lambda)$ in λ . This is easily satisfied.

The third condition is the compactness of the parameter space. It is guaranteed by Assumption 2b.

The fourth condition is the uniform convergence of $Q_T(\lambda)$ to $Q_0(\lambda)$. It is sufficient to show that

$$\frac{1}{T} \sum_{t=1}^T \left\{ RV_t - \frac{1}{T} \sum_{i=1}^T \left(\widehat{\phi}_T(\lambda) RV_{t-1}^{\lambda} + RV_i^{\lambda} - \widehat{\phi}_T(\lambda) RV_{i-1}^{\lambda} \right)^{1/\lambda} \right\}$$

converges uniformly. To do it, following Lemma 2.4 of Newey and McFadden (1994), we will find a dominance condition so that for all i, t ,

$$\left| RV_t - \left(\widehat{\phi}_T(\lambda) RV_{t-1}^{\lambda} + RV_i^{\lambda} - \widehat{\phi}_T(\lambda) RV_{i-1}^{\lambda} \right)^{1/\lambda} \right| < d$$

with $E(d) < \infty$. Note that

$$\begin{aligned}
& \left| RV_t - \left(\widehat{\phi}_T(\lambda) RV_{t-1}^\lambda + RV_i^\lambda - \widehat{\phi}_T(\lambda) RV_{i-1}^\lambda \right)^{1/\lambda} \right| \\
& \leq |RV_t| + \left| \left(\widehat{\phi}_T(\lambda) RV_{t-1}^\lambda + RV_i^\lambda - \widehat{\phi}_T(\lambda) RV_{i-1}^\lambda \right)^{1/\lambda} \right| \\
& \leq RV_t + \left| \left(\widehat{\phi}_T(\lambda) RV_{t-1}^\lambda \right)^{1/\lambda} \right| \\
& = RV_t + \left| \widehat{\phi}_T^{1/\lambda_0}(\lambda_0) \right| RV_{t-1} \\
& \leq RV_t + \phi_0^{1/\lambda_0} RV_{t-1} := d.
\end{aligned}$$

The first inequality follows by the triangle inequality. The second inequality follows because $RV_i^\lambda \geq \widehat{\phi}_T(\lambda) RV_{i-1}^\lambda$ by construction and the fact that $\lambda < 0$. The last inequality follows as $\widehat{\phi}_T(\lambda_0)$ is positively biased. By Assumption 4b, $E(RV_t) = E(RV_{t-1}) < \infty$, which implies that $E(d) < \infty$. Therefore, the uniform convergence is established for $Q_T(\lambda)$.

By Theorem 2.1 of Newey and McFadden (1994), $\widehat{\lambda}_T \xrightarrow{a.s.} \lambda_0$.

To prove the consistency of $\widehat{\phi}_T$, note that

$$\widehat{\phi}_T(\lambda) = \left(\min_t \left\{ \frac{X_t}{X_{t-1}} \right\} \right)^{\lambda/\lambda_0} \xrightarrow{a.s.} \phi_0^{\lambda/\lambda_0}.$$

Hence, $\widehat{\phi}_T = \widehat{\phi}_T(\widehat{\lambda}_T) \xrightarrow{a.s.} \phi_0$ because $\widehat{\lambda}_T \xrightarrow{a.s.} \lambda_0$ and $\widehat{\phi}_T(\lambda)$ is continuous in λ .

8.2. Proof of Proposition 3.3. By Assumption 2b, the derivative of $Q_T(\widehat{\lambda}_T)$ must be zero, i.e.,

$$\begin{aligned}
\frac{\partial Q_T(\widehat{\lambda}_T)}{\partial \lambda} &= -\frac{2}{T} \sum_{t=1}^T \left\{ \left[RV_t - \frac{\sum_{i=1}^T f_i^{1/\widehat{\lambda}_T}(\widehat{\lambda}_T)}{T} \right] \frac{\sum_{i=1}^T f_i^{1/\widehat{\lambda}_T}(\widehat{\lambda}_T) \ln f_i(\widehat{\lambda}_T) \frac{\partial f_i(\widehat{\lambda}_T)}{\partial \lambda}}{T} \right\} \\
&= 0
\end{aligned}$$

where $f_i(\lambda) = \widehat{\phi}_T(\lambda) RV_{t-1}^\lambda + RV_i^\lambda - \widehat{\phi}_T(\lambda) RV_{i-1}^\lambda$. Although $f_i(\lambda)$ is also dependent on t , this dependency is omitted for brevity.

The mean value theorem is applied to $\frac{\partial Q_T(\widehat{\lambda}_T)}{\partial \lambda}$ and gives

$$\begin{aligned}
0 &= \frac{\partial Q_T(\widehat{\lambda}_T)}{\partial \lambda} \\
&= -\frac{2}{T} \sum_{t=1}^T \left\{ \left[RV_t - \frac{\sum_{i=1}^T f_i^{1/\lambda_0}(\lambda_0)}{T} \right] \frac{\sum_{i=1}^T f_i^{1/\lambda_0}(\lambda_0) \ln f_i(\lambda_0) \frac{\partial f_i(\lambda_0)}{\partial \lambda}}{T} \right\} \\
(8.1) \quad &-2(\widehat{\lambda}_T - \lambda_0) g_T(\widetilde{\lambda}_T)
\end{aligned}$$

where $\widetilde{\lambda}_T$ is between λ_0 and $\widehat{\lambda}_T$ and $g_T(\lambda)$ is given in (3.4).

Note that $(RV_t - E(RV_t | RV_{t-1})) \frac{1}{T} \sum_{i=1}^T f_i^{1/\lambda_0}(\lambda_0) \ln f_i(\lambda_0) \frac{\partial f_i(\lambda_0)}{\partial \lambda}$ is a martingale difference sequence (MDS) for which the central limited theorem (CLT) is applied (Hall and

Heyde, 1980). Thus,

$$\begin{aligned}
& \sqrt{T} \frac{1}{T} \sum_{t=1}^T \left\{ \left[RV_t - \frac{\sum_{i=1}^T f_i^{1/\lambda_0}(\lambda_0)}{T} \right] \frac{\sum_{i=1}^T f_i^{1/\lambda_0}(\lambda_0) \ln f_i(\lambda_0) \frac{\partial f_i(\lambda_0)}{\partial \lambda}}{T} \right\} \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^T \left\{ \left[RV_t - E(RV_t | RV_{t-1}) + E(RV_t | RV_{t-1}) - \frac{\sum_{i=1}^T f_i^{1/\lambda_0}(\lambda_0)}{T} \right] \right. \\
&\quad \left. \times \frac{\sum_{i=1}^T f_i^{1/\lambda_0}(\lambda_0) \ln f_i(\lambda_0) \frac{\partial f_i(\lambda_0)}{\partial \lambda}}{T} \right\} \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^T \underbrace{\left\{ [RV_t - E(RV_t | RV_{t-1})] \frac{\sum_{i=1}^T f_i^{1/\lambda_0}(\lambda_0) \ln f_i(\lambda_0) \frac{\partial f_i(\lambda_0)}{\partial \lambda}}{T} \right\}}_A + \\
&\quad \underbrace{\left\{ \left[E(RV_t | RV_{t-1}) - \frac{\sum_{i=1}^T f_i^{1/\lambda_0}(\lambda_0)}{T} \right] \frac{\sum_{i=1}^T f_i^{1/\lambda_0}(\lambda_0) \ln f_i(\lambda_0) \frac{\partial f_i(\lambda_0)}{\partial \lambda}}{T} \right\}}_B
\end{aligned}$$

$$\xrightarrow{d} N(0, \Omega) + o_p(1) = N(0, \Omega)$$

where $A \xrightarrow{d} N(0, \Omega)$ follows from the MDS central limited theorem under Assumptions 5b-6b; $B \xrightarrow{p} 0$ because $\frac{1}{T} \sum_{i=1}^T f_i^{1/\lambda_0}(\lambda_0) \xrightarrow{p} E(RV_t | RV_{t-1})$ and Assumption 7b.

Since $\hat{\lambda}_T \xrightarrow{p} \lambda_0$, we have $\tilde{\lambda}_T \xrightarrow{p} \lambda_0$. By Assumption 8b, $g_T(\tilde{\lambda}_T) \xrightarrow{p} g(\lambda_0) = g$. From Equation (8.1) and the Slutsky theorem, we have,

$$\sqrt{T}(\hat{\lambda}_T - \lambda_0) \xrightarrow{d} N(0, g^{-1}\Omega g^{-1}).$$

ACKNOWLEDGMENTS

The authors gratefully acknowledge research support from the Jan Wallander and Tom Hedelius Research Foundation under Grant No. P2006-0166:1 and from the Singapore Ministry of Education AcRF Tier 2 Fund under Grant No. T206B4301-RS. We wish to thank the editor, an associate editor, three referees, Torben Andersen, Federico Bandi, Frank Diebold, Marcelo Medeiros, Bent Nielsen, and Neil Shephard for very helpful comments. The OX language of J. A. Doornik and ARFIMA package of Doornik & Ooms (2004) were used to estimate the two ARFIMA models. Matlab code and data used in this paper is available from <http://www.mysmu.edu/faculty/yujun/research.html>.

REFERENCES

- Andersen, T. & Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts, *International Economic Review* **39**: 885–905.
- Andersen, T., Bollerslev, T., Diebold, F. & Labys, P. (2001). The distribution of realized exchange rate volatility, *Journal of the American Statistical Association* **96**: 42–55.
- Andersen, T., Bollerslev, T., Diebold, F. & Labys, P. (2003). Modeling and forecasting realized volatility, *Econometrica* **71**: 579–625.

- Barndorff-Nielsen, O. E. & Shephard, N. (2001). Non-gaussian ornstein-uhlenbeck-based models and some of their uses in financial economics, *Journal of the Royal Statistical Society, Series B* **63**: 167–241.
- Barndorff-Nielsen, O. E. & Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models, *Journal of the Royal Statistical Society, Series B* **64**: 253–280.
- Beran, J. (1995). Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models, *Journal of the Royal Statistical Society, Series B* **57**: 659–72.
- Bollerslev, T., Chou, R. Y. & Kroner, K. F. (1992). Arch modeling in finance: A review of the theory and empirical evidence, *Journal of Econometrics* **52**: 5–59.
- Bollerslev, T., Engle, R. F. & Nelson, D. (1994). Arch models, in R. F. Engle & D. McFadden (eds), *Handbook of Econometrics, Vol IV*, Vol. 14, Elsevier Science B.V., Amsterdam, pp. 2959–3038.
- Box, G. E. P. & Cox, D. R. (1964). An analysis of transformations, *Journal of the Royal Statistical Society, Series B* **26**(2): 211–252.
- Chen, W. & Deo, R. (2004). Power transformations to induce normality and their applications, *Journal of Royal Statistical Society, Series B* **66**: 117–130.
- Chernov, M., Gallant, A. R., Ghysels, E. & Tauchen, G. (2003). Alternative models for stock price dynamics, *Journal of Econometrics* **116**: 225–257.
- Chung, C. & Baillie, R. (1993). Small sample bias in conditional sum-of-squares estimators of fractionally integrated arma models, *Empirical Economics* **18**(4): 791–806.
- Cipollino, F., Engle, R. F. & Gallo, G. (2006). Vector multiplicative error models: Representation and inference, *Technical report*, NBER.
- Datta, S. & McCormick, W. P. (1995). Bootstrap inference for a first-order autoregression with positive innovations, *Journal of the American Statistical Association* **90**: 1289–1300.
- Davis, R. A. & McCormick, W. P. (1989). Estimation for first-order autoregressive processes with positive or bounded innovations, *Stochastic Processes and their Applications* **31**: 237–250.
- Deo, R., Hurvich, C. & Lu, Y. (2006). Forecasting realized volatility using a long-memory stochastic volatility model, *Journal of Econometrics* **131**: 29–58.
- Doornik, J. & Ooms, M. (2004). Inference and forecasting for arfima models with an application to us and uk inflation, *Studies in Nonlinear Dynamics and Econometrics* **8**: 1218–1218.
- Duan, J.-C. (1997). Augmented garch(p,q) process and its diffusion limit, *Journal of Econometrics* **79**(1): 97–127.
- Fan, J., Fan, Y. & Jiang, J. (2007). Dynamic integration of time- and state-domain methods for volatility estimation, *Journal of the American Statistical Association* **102**: 618–631.
- Feigin, P. D. & Resnick, S. I. (1992). Estimation for autoregressive processes with positive innovations, *Communications in Statistics - Stochastic Models* **8**(3): 479–498.
- Fernandes, M. & Grammig, J. (2006). A family of autoregressive conditional duration models, *Journal of Econometrics* **127**(1): 1–23.
- Ghysels, E., Harvey, A. C. & Renault, E. (1996). Stochastic volatility, in C. R. RAO, & G. S. Maddala (eds), *Handbook of Statistics*, Vol. 14, North-Holland, Amsterdam, pp. 119–191.
- Goncalves, S. & Meddahi, N. (2009). Box-cox transforms for realized volatility, *Journal of Econometrics*, forthcoming.
- Granger, C. W. J. & Newbold, P. (1976). Forecasting transformed series, *Journal of the Royal Statistical Society, Series B* **38**: 189–203.
- Hall, P. & C. C. Heyde (1980). *Martingale Limit Theory and Its Application*, Academic Press.
- Hansen, P. R. (2006). Consistent ranking of volatility models, *Journal of Econometrics* **131**: 97–12.
- Hentschel, L. (1995). All in the family: Nesting symmetric and asymmetric garch models, *Journal of Financial Economics* **39**: 71–104.
- Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options, *Review of Financial Studies* **6**: 327–343.
- Higgins, M. L. & Bera, A. (1992). A class of nonlinear arch models, *International Economic Review* **33**: 137–158.
- Hyndman, R. J. & Koehler, A. B. (2006). Another look at measures of forecast accuracy, *International Journal of Forecasting* **22**: 679–688.
- Jacod, J. (1994). Limit of random measures associated with the increments of a brownian semimartingale, *Technical report*, Laboratoire de Probabilités, Paris.

- Jennrich, R.I. (1969). Asymptotic Properties of Non-Linear Least Squares Estimators, *Annals of Math. Statist* **2**: 633-643.
- Lopez, J. A. (2001). Evaluating the predictive accuracy of volatility models, *Journal of Forecasting* **20**: 87–109.
- Meddahi, N. & Renault, E. (2004). Temporal aggregation of volatility models, *Journal of Econometrics* **119**: 355–379.
- Morgan, J. P. (1996). Riskmetrics technical document, *New York*. 4th Ed.
- Newey, W. K. & D. McFadden (1994). Large sample estimation and hypothesis testing. In Engle, R.F. and D. McFadden, editors, *Handbook of Econometrics, Vol 4*. North-Holland, Amsterdam.
- Nielsen, B. & Shephard, N. (2003). Likelihood analysis of a first-order autoregressive model with exponential innovations, *Journal of Time Series Analysis* **24**(3): 337–344.
- Patton, A. (2009). Volatility forecast comparison using imperfect volatility proxies, *Journal of Econometrics*, forthcoming.
- Phillips, P. C. B. (1987). Time series regression with a unit root, *Econometrica* **55**: 277–301.
- Poon, S.-H. & Granger, C. W. J. (2003). Forecasting volatility in financial markets: A review, *Journal of Economic Literature* **41**(2): 478–539.
- Preve, D. (2008). *Essays on Time Series Analysis - With Applications to Financial Econometrics*, PhD thesis, Uppsala University. Paper II.
- Shephard, N. (2005). *Stochastic Volatility: Selected Readings*, Oxford University Press.
- Shimotsu, K. & Phillips, P. C. B. (2005). Exact local whittle estimation of fractional integration, *The Annals of Statistics* **33**(4): 1890–1933.
- Sowell, F. (1992). Maximum likelihood estimation of stationary univariate fractionally integrated time series models, *Journal of Econometrics* **53**: 165–188.
- Taylor, S. J. (1986). *Modeling Financial Time Series*, John Wiley and Sons, New York.
- Yu, J., Yang, Z. & Zhang, X. (2006). A class of nonlinear stochastic volatility models and its implications for pricing currency options, *Computational Statistics and Data Analysis* **51**: 2218–2231.
- Current address*: School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903.
E-mail address: `danielpreve@smu.edu.sg`
- Current address*: Department of Information Science/Statistics, University of Uppsala, Box 513 SE-751 20, Sweden.
E-mail address: `anders.eriksson@dis.uu.se`
- Current address*: School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903.
E-mail address: `yujun@smu.edu.sg`