Abstract

In this paper, we propose an alternative approach to estimation and forecasting seasonal long-memory stochastic volatility based on the generalized long-memory stochastic volatility (GLMSV) model. We investigate the predictive ability of the proposed method vis-à-vis other estimation and forecasting approaches on real data – Microsoft stock intraday volatility.

1. Introduction

Many economic time series exhibit some form of seasonality. In particular, seasonal long-range dependence or long-memory is a common feature of time-varying volatility of financial time series. The seasonal or periodic components are usually represented by pronounced peaks at certain frequencies in the periodogram of a series. This seasonal variation may account for the preponderance of the total variance, which when ignored may severely affect the results in modeling and forecasting volatility dynamics.

Long-memory of volatility in time series is widely recognized in literature. See e.g. Hurvich and Ray (2003), Deo and Hurvich (2001), Deo, Hurvich and Lu (2006), Jensen (2004), and Bollerslev and Mikkelsen (1996). In the context of stochastic volatility (SV) models, Breidt, Crato and de Lima (1998) and Harvey (1998) independently proposed the long-memory stochastic volatility (LMSV) model by replacing the autoregressive moving average (ARMA) in the log-volatility term of the SV model by the fractionally integrated (FI) process. The LMSV model accounts for long-range persistence in volatility; however, it does not capture the cyclical behavior in the volatility series, which could significantly affect modeling and forecasting of volatility dynamics. This cyclical behavior has been observed in the volatility of financial time series, and it is an important characteristic of intraday data (see e.g. Andersen and Bollerslev (1997), Bisaglia et al. (2003)). To account for this persistent cyclic dynamic Arteche (2004) proposed an extension of the LMSV model in the form of seasonal or cyclical asymmetric long memory (SCALM) as defined in Arteche and Robinson (2000). It is a local Whittle-LMSV model that allows a single pole with known location and different slopes on both sides. Local Whittle estimators are known to be sensitive to the choice of the bandwidth parameter (e.g. Taqqu and Teverovsky (1997)).

In this paper, we consider an extension of the LMSV model in Breidt et al. (1998) by representing the log-volatility by a \( k \)-factor Gegenbauer autoregressive moving average (\( k \)-GARMA) process that accounts for \( k \) persistent periodicities in the volatility series. We call this model a generalized long-memory stochastic volatility (GLMSV) model. We propose a Whittle estimator of the GLMSV model that is consistent for all its parameters. We derive the analytic expression of the prediction function of the GLMSV model, and we provide the \( h \)-step prediction error. We investigate the predictive ability of the proposed method on real data vis-à-vis the forecasting approach based on deterministic deseasonalization in Deo, Hurvich and Lu (2006).
This paper is organized as follows. In Section 2 we discuss long-memory stochastic volatility and deterministic seasonality. In Section 3 we present the generalized long-memory stochastic volatility (GLMSV) model, its Whittle likelihood function, and some simulation results. In Section 4 we present the proposed forecasting approach using stochastic seasonality. In Section 5 we provide an application of the proposed method to the Microsoft stock realized volatility, comparing the forecasts based on stochastic seasonality, deterministic seasonality and combinations of these. Finally, some concluding remarks are given in Section 6.

2. Long-memory Stochastic Volatility and Deterministic Seasonality

A long-memory process is a stationary stochastic process for which the autocorrelations decay to zero slowly at a hyperbolic rate. It is often modeled by the autoregressive fractionally integrated moving average (ARFIMA) process. An ARFIMA\((p,d,q)\) process \(\{X_t\}\) is a stationary process such that

\[
\Phi(B)(1-B)^d X_t = \Theta(B)\varepsilon_t
\]

where \(\varepsilon_t\) is white noise, where \(B\) denotes the backshift operator, \(\Phi(B)=1-\phi_1 B^{-1} - \ldots - \phi_p B^{-p}\) and \(\Theta(B)=1+\theta_1 B^{-1} + \ldots + \theta_q B^{-q}\) are polynomials of order \(p\) and \(q\), respectively, with all roots outside the unit circle, and \((1-B)^d\) is the fractional difference operator. Its spectral density is given by

\[
S_X(f) = \sigma^2 |1 - \exp(2\pi if)|^{-2d} \left| \frac{\Theta(e^{2\pi if})}{\Phi(e^{2\pi if})} \right|^2.
\]

We consider the long-memory stochastic volatility (LMSV) model for the mean corrected returns of financial assets \(\{r_t\}\) given by

\[
r_t = \sigma \exp\{X_t/2\} \varepsilon_t,
\]

with \(\sigma > 0\), \(\{\varepsilon_t\}\) are shocks with zero mean and unit variance, and \(\{X_t\}\) is a stationary Gaussian with possible long-memory and persistent cyclical or seasonal behavior, which is independent of \(\{\varepsilon_t\}\). Breidt, et al. (1998) and Deo et al. (2006) considered for \(\{X_t\}\) ARFIMA\((p,d,q)\) model, which is also assumed in this section.

Let \(Y_t = \log r_t^2\) be the logarithm of squared returns, hereinafter referred to as the log squared process. Hence, we have

\[
Y_t = \mu + X_t + \eta_t,
\]

where \(\mu = \log \sigma^2 + E(\log \varepsilon_t^2)\) and \(\eta_t = \log \varepsilon_t^2 - E(\log \varepsilon_t^2)\) is iid with mean zero and finite variance \(\sigma^2_{\eta}\) independent of \(\{X_t\}\). If \(\varepsilon_t\) is standard Gaussian, then \(E(\log \varepsilon_t^2) = -\gamma - \log(2)\), where \(\gamma \approx 0.5772\ldots\) is the Euler constant, is non-Gaussian, and \(\eta_t\) is distributed \(\log X^2_{(v)}\) with variance \(\sigma^2_{\eta} = \pi^2/2\). Heavy tails in the returns may also require more kurtosis in \(\varepsilon_t\). If \(\varepsilon_t\) is assumed to have the \(t\)-distribution with \(v\) degrees of freedom, normalized so that its variance is one, then \(E(\log \varepsilon_t^2) = \log(v-2) - \gamma - 2\log(2) + \psi(v/2)\) and \(\eta_t\) again non-Gaussian with
variance $\sigma^2_n = \psi'(v/2) + \pi^2/2$, where $\psi$ and $\psi'$ are the digamma and the derivative of the digamma function, respectively (Deo et al. (2006)).

In the presence of seasonality, Deo et al. (2006) proposed a method of forecasting that accounts for slowly varying seasonality in volatility. It involves the removal of slowly varying seasonality, represented by a deterministic linear combination of sines and cosines evaluated at the Fourier frequencies that show seasonal peaks. Forecasts are initially obtained from the deseasonalized series, which are then combined with the corresponding forecasts of the seasonality model to obtain forecast values of the volatility series.

The basic model is for the seasonal adjustment procedure is given by

$$R_t = \exp(S_t/2)\eta_t,$$

where $\eta_t$ is the demeaned return, $S_t$ is the seasonal component and $r_t$ is the deseasonalized return. The deterministic model for the seasonality $S_t$ is a linear combination of sines and cosines evaluated at the Fourier coefficients corresponding to seasonal peaks. This can be written as

$$S_t = \sum_{p=1}^{k} a_p \cos(f_{jp} t) + \sum_{p=1}^{k} b_p \sin(f_{jp} t),$$

where $B = \{f_{jp}\}_{p=1}^k$ is the collection of all frequencies corresponding to the poles and their neighboring frequencies that exhibit large values. The fitted values in the regression

$$Y_t = \sum_{p=1}^{k} a_p \cos(f_{jp} t) + \sum_{p=1}^{k} b_p \sin(f_{jp} t) + \xi_t,$$

yield the estimated seasonal component $\hat{S}_t = \sum_{p=1}^{k} \hat{a}_p \cos(f_{jp} t) + \sum_{p=1}^{k} \hat{b}_p \sin(f_{jp} t)$. The residuals represent the deseasonalized series $\hat{\epsilon}_t = Y_t^{DS} = Y_t - \hat{S}_t = \mu + X_t + \eta_t$ and $X_t$ is an ARFIMA($p,d,q$) process. The $h$-step ahead forecast is then given by

$$\hat{Y}_{t+h} = \hat{Y}_{t+h}^{DS} + \hat{S}_{t+h}$$

where $\hat{Y}_{t+h}^{DS}$ is the $h$-step ahead forecast of the deseasonalized series and $\hat{S}_{t+h}$ is the $h$-step ahead forecast of the deterministic seasonal component. The $h$-step ahead forecast of the squared return is simply

$$\hat{r}_{t+h}^2 = \exp(\hat{Y}_{t+h}^2)$$

We intend to improve the performance of such approach by assuming that the poles represent stochastic seasonality. This is possible by considering a stochastic model that accounts for several persistent periodicities represented by poles in its spectral density.

3. **Generalized Long-memory Stochastic Volatility**

A fairly general model of long-memory that accounts for persistent cyclic behavior at $k$ frequencies is the $k$-factor Gegenbauer autoregressive moving-average ($k$-GARMA) model (see Woodward et al. 1998). A $k$-GARMA($p,d,u,q$) process $\{X_t\}$ is given by

$$\Phi(B)\prod_{i=1}^{k}(1 - 2u_i B + B^2)^d X_t = \Theta(B)\epsilon_t,$$

where $\Phi(B)$ and $\Theta(B)$ are polynomials in the backshift operator $B$. The parameter $u_i$ represents the frequency of the $i$-th component, $d$ is the degree of the polynomial, and $\epsilon_t$ is a white noise process. This model allows for a flexible representation of long-memory and cyclic behavior, which can be useful in financial time series analysis.
where \( B \) denotes the backshift operator, \( \{ \varepsilon_i \} \sim \text{iid } \mathcal{N}(0, \sigma^2_\varepsilon) \), \( \Phi(B) = 1 - \phi_1 B^{-1} - \cdots - \phi_p B^{-p} \) and \( \Theta(B) = 1 + \theta_1 B^{-1} + \cdots + \theta_q B^{-q} \) are polynomials of order \( p \) and \( q \), respectively, with all roots outside the unit circle; \( d \) and \( u \) are vectors of length \( k \), with \( d_i \neq 0 \) and distinct \( u_i \), with \( |u_i| \leq 1 \), \( i=1, \ldots, k \). The vector \( v \), with components \( v_i = \frac{\cos^{-1}(u_i)}{2\pi} \in [0,0.5] \), \( i=1, \ldots, k \), gives the frequencies of the persistent periodic cycles. The process is causal and invertible (see Gray et al. (1989) and Giraitis and Leipus (1995)) if for \( i=1, \ldots, k \)
\[
|d_i| < \begin{cases} Y_1/2 \ , & 0 < v_i < Y_2/2 \ , \\ Y_4 \ , & v_i = 0 \text{ or } Y_2/2 \ . \end{cases} \tag{11}
\]

Its spectral density function (SDF) is then
\[
S_X(f) = \frac{\sigma^2_\varepsilon}{\Phi(e^{-i2\pi f})} \prod_{i=1}^k \left| 2(\cos(2\pi f) - u_i) \right|^{-2d_i}, f \in (-0.5,0.5] \tag{12}
\]
(Gray et al., 1998). We will collect all unknown parameters in the vector \( \theta_X = (\sigma^2_\varepsilon, d_1, \ldots, d_k, v_1, \ldots, v_k, \theta_1, \ldots, \theta_q, \phi_1, \ldots, \phi_p) \).

The \( k \)-GARMA process generalizes several known long-memory models. It generalizes the seasonal fractionally differenced process \((1-B^S)^d X_t = \varepsilon_t \) (see e.g. Porter-Hudak (1990)), which has \( S \) poles on \((-0.5,0.5)\) for \( d>0 \). The \( k \)-vector \( d \) of the \( k \)-GARMA model with \( k=[(S+1)/2] \) has then the form \( d=1, (d/2, \ldots, d) \), and fixed \( v_i = 2\pi/(S \cdot i), i=0, \ldots, k \). On the other hand, if there is only one pole at \( v=0 \) the model simplifies to the well-known autoregressive fractionally integrated moving average ARFIMA\((p,d,q)\) process.

We use this model to generalize the long-memory stochastic volatility in (3). Instead of an ARFIMA\((p,d,q)\) process for \( X_t \), we consider \( k \)-GARMA process. We now call (3) a generalized long-memory stochastic volatility (GLMSV) model. This allows for stochastic seasonal components, an alternative to the deterministic deseasonalization in Deo, Hurvich and Lu (2006).

Since \( X_t \) and \( \eta_t \) are independent, the autocovariance function of \( Y_t \) in (4) has a simple structure. It is the sum of the covariances of the long-memory signal and the noise series given by \( \gamma_Y(s) = \gamma_X(s) + \sigma^2_\eta I_{[s=0]} \). Hence, its spectral density is
\[
S_Y(f) = S_X(f) + \sigma^2_\eta, \tag{13}
\]
where \( S_X(f) \) is defined by (12), the constant \( \sigma^2_\eta \) is the spectral density of the iid process \( \{ \eta_t \} \). The associated parameter vector of \( S_Y(f) \) is \( \theta_X \) augmented by \( \sigma^2_\eta \) and denoted by \( \theta_Y = (\sigma^2_\varepsilon, \sigma^2_\eta, d_1, \ldots, d_k, v_1, \ldots, v_k, \theta_1, \ldots, \theta_q, \phi_1, \ldots, \phi_p) \).

We can easily extend the discrete Whittle estimator in Breidt, Crato and de Lima (1998) by replacing the spectral density of the FD process by that of the \( k \)-GARMA. Given a sample of \( N \) observations, the Whittle (negative log-)likelihood is
\[
\ell(\theta_Y) = N^{-1} \left\{ \sum_{j=1}^{\lceil (N-1)/2 \rceil} \left[ \frac{I_Y(\lambda_j)}{S_Y(\lambda_j)} + \log S_Y(\lambda_j) \right] \right\}, \tag{14}
\]
where \( \lceil \cdot \rceil \) is the integer part, \( \lambda_n = n/N \) is \( n \)th frequency, and
\[
I_N(\lambda_n) = \frac{1}{N} \left| \sum_{t=0}^{N-1} Y_t e^{-2\pi i \lambda_n t} \right|^2,
\]
(15)
is the \( n \)th normalized periodogram ordinate of the series \( Y_t, t=0, ..., N-1 \).

To show consistency, we consider the following conditions.
A1: The polynomials \( \Theta(z) \) and \( \Phi(z) \) have no common zeros, their zeros lie outside the unit circle, and they satisfy the normalization condition \( \Theta(0) = \Phi(0) = 1 \).
A2: The long-memory parameters satisfy (11).
A3: The GLMSV model is identifiable, that is, \( \sigma^2 \) is known or \( \{X_t\} \) is not white noise.

Conditions A1 and A2 ensure that the \( k \)-GARMA process is causal, invertible and normalized. Condition A3 ensures that the estimates are unique.

The theoretical properties of the Whittle estimators of \( k \)-GARMA are not yet completely known for \( k > 1 \). If \( k = 1 \), a relatively complete account of asymptotic properties is given in Giraitis, Hidalgo and Robinson (2001). On the other hand, Giraitis and Leipus (1995) have shown that the Whittle estimator of the \( k \)-GARMA process is consistent. In the following theorem we show that the Whittle estimators of the parameters of a GLMSV defined are consistent. It is an extension of the consistency property of the Whittle estimators for the LMSV parameters in Breidt, et al.(1998). The result holds when poles are known or unknown.

**Theorem.** Let assumptions A1- A3 be satisfied, and let the parameter vectors \( \theta^1_Y \) and \( \theta^2_Y \) be elements of the compact parameter space \( \Theta_Y \) such that \( S(f; \theta^1_Y) = S(f; \theta^2_Y) \) implies that \( \theta^1_Y = \theta^2_Y \). Then \( \hat{\theta}^Y \rightarrow \theta^0_Y \) almost surely as \( N \rightarrow \infty \), where \( \theta^0_Y \) denotes the vector of true parameters.

The proof of this theorem is given in the Appendix. It is just a modification of the proof for consistency of Whittle estimator of LMSV model in Breidt, et al.(1998). We also verify this property from the finite sample performance by Monte Carlo simulation. Table 1 presents the results of our simulation study of a GLMSV model based on 3-GARMA (\( v_1 = 0.05 \), \( d_1 = 0.3 \), \( v_2 = 0.3 \), \( d_2 = 0.2 \), \( v_3 = 0.45 \), \( d_3 = 0.3 \)) and \( \epsilon_t \sim iid N(0,1) \). Hence, \( \eta_t \sim iid \log \chi^2_{(1)} \) with variance \( \sigma^2 \eta = \pi^2/2 \). We consider pole frequencies such that one pole is close to frequency 0, one close to 0.25 and one close to 0.5. The sample sizes are 128, 512 and 1024 with 50 iterations each.

Table 1 shows the estimates, bias and mean squared error (MSE) of the long-memory parameters and the variances. The MSE’s of the long-memory parameters tend to zero monotonically rather quickly at all poles. Their absolute biases decrease monotonically at \( v = 0.05 \) and \( v = 0.45 \), but the decrease is non-monotonic at \( v = 0.3 \). The bias and MSE of \( \sigma^2 \) tend to zero; however, the decrease of its MSE seems slow. The decrease of the bias and MSE of \( \sigma^2 \) is non-monotonic and is rather slow.
4. Forecasting using Stochastic Seasonality

We propose an alternative approach to forecasting seasonal long-memory stochastic volatility based on the generalized long-memory stochastic volatility (GLMSV) model. In this case, we assume that some, if not all, seasonal components are not deterministic, but can be accounted for by the stochastic model.

We obtain the forecast of \( r_t^2 \) from the best linear prediction of \( y_t = \log r_t^2 \), the logarithm of squared returns. If Conditions A1 and A2 are satisfied, Giraitis and Leipus (1995) have shown that the \( k \)-GARMA process \( \{X_t\} \) has a one-sided moving average (MA) representation given by

\[
X_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} = \sum_{i=0}^{\infty} \zeta_i \nabla_{d_1,...,d_k} (B) \varepsilon_{t-i},
\]

where \( \nabla_{d_1,...,d_k} (B) = \prod_{i=1}^{k} (1 - 2u_i B^{-1} + B^{-2})^{-d_i} \), and \( \zeta_i \) is determined by the Laurent expansion \( \sum_{i=2}^{\infty} \zeta_i z^i = \frac{\theta(z)}{\varphi(z)} \) in some annulus of \( |z|=1 \). On the other hand, the autoregressive (AR) representation of \( \{X_t\} \) is given by

\[
\sum_{j=0}^{\infty} \pi_j X_{t-j} = \sum_{j=0}^{\infty} \varepsilon_{t-j} \nabla_{d_1,...,d_k} (B) X_{t-j} = \varepsilon_t.
\]

where \( \nabla_{d_1,...,d_k} (B) = \prod_{i=1}^{k} (1 - 2u_i B^{-1} + B^{-2})^{-d_i} \), and \( \zeta_i \) is determined by the Laurent expansion

\[
\sum_{i=2}^{\infty} \zeta_i z^i = \frac{\phi(z)}{\varphi(z)} \] in some annulus of \( |z|=1 \). The corresponding weights satisfy \( \sum_{i=0}^{\infty} \psi_i^2 < \infty \) and \( \sum_{j=0}^{\infty} \pi_j^2 < \infty \). For a \( k \)-GARMA(0,d,u,0) the autoregressive and moving average coefficients are defined by

\[
\pi_j (d,v) = \sum_{l_1 + ... + l_k = j} C_{l_1}(-d_1,v_1)...C_{l_k}(-d_k,v_k) \quad \text{and} \quad \psi_j (d,v) = \sum_{l_1 + ... + l_k = j} C_{l_1}(d_1,v_1)...C_{l_k}(d_k,v_k)
\]

where \( 0 \leq l_1,...,l_k \leq j \) (Giraitis and Leipus (1995)). \( C_l(d,u) \) is the Gegenbauer polynomial defined by

\[
|z| = 1, \quad \varphi(z) = \prod_{i=1}^{k} (1 - 2u_i z + z^2)^{-d_i}, \quad \theta(z) = \sum_{i=2}^{\infty} \zeta_i z^i = \frac{\phi(z)}{\varphi(z)}
\]

<table>
<thead>
<tr>
<th>N</th>
<th>Statistics</th>
<th>( d_1=0.3 ) (v=0.05)</th>
<th>( d_2=0.2 ) (v=0.3)</th>
<th>( d_3=0.3 ) (v=0.45)</th>
<th>( \sigma^2 = 1 )</th>
<th>( \sigma^2 = \pi^2/2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>128</td>
<td>Mean</td>
<td>0.24428</td>
<td>0.20555</td>
<td>0.24800</td>
<td>0.84391</td>
<td>4.62433</td>
</tr>
<tr>
<td></td>
<td>Bias</td>
<td>-0.05572</td>
<td>0.00555</td>
<td>-0.05200</td>
<td>-0.15609</td>
<td>-0.31047</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>0.01570</td>
<td>0.00984</td>
<td>0.01451</td>
<td>0.41782</td>
<td>0.59618</td>
</tr>
<tr>
<td>512</td>
<td>Mean</td>
<td>0.27961</td>
<td>0.20962</td>
<td>0.27158</td>
<td>0.96624</td>
<td>4.43736</td>
</tr>
<tr>
<td></td>
<td>Bias</td>
<td>-0.02034</td>
<td>0.00962</td>
<td>-0.02842</td>
<td>-0.03376</td>
<td>-0.49744</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>0.00342</td>
<td>0.00239</td>
<td>0.00367</td>
<td>0.31588</td>
<td>1.59182</td>
</tr>
<tr>
<td>1024</td>
<td>Mean</td>
<td>0.28289</td>
<td>0.20673</td>
<td>0.27838</td>
<td>0.98260</td>
<td>4.51022</td>
</tr>
<tr>
<td></td>
<td>Bias</td>
<td>-0.01711</td>
<td>0.00673</td>
<td>-0.02162</td>
<td>-0.01740</td>
<td>-0.42458</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>0.00255</td>
<td>0.00117</td>
<td>0.00150</td>
<td>0.31338</td>
<td>1.22993</td>
</tr>
</tbody>
</table>

Table 1. Monte Carlo simulations for GLMSV Model
\[ C_t(d,u) = \sum_{k=0}^{\lfloor n/2 \rfloor} (-1)^k (2u)^{n-2k} \Gamma((d)-k+n) \], \quad (19) \]

\([n/2]\) is the largest integer less than or equal to \(n/2\), and \(u = \cos(2\pi t)\) (Gray, et al. (1989)).

Now using the AR representation in (17), log-squared process may be written as

\[ Y_{t+1} = \mu - \sum_{j=1}^{\infty} \pi_j X_{t+1-j} + \varepsilon_t + \eta_t. \quad (20) \]

Since \( \hat{Y}_{t+1} = \mu - \sum_{j=1}^{\infty} \pi_j X_{t+1-j} \) is an element of \( M_X \), the linear subspace generated by \( \{X_s \leq t\} \), and \( Y_{t+1} - \hat{Y}_{t+1} = \varepsilon_t + \eta_t \) is orthogonal to all of the generators in \( \{X_s \leq t\} \) with respect to the inner product \( \langle U, V \rangle = E(UV) \) in \( L^2 \), then \( \hat{Y}_{t+1} \) is the projection of \( Y_{t+1} \) on \( M_X \); hence, \( \hat{Y}_{t+1} \) is the best one-step linear predictor of \( Y_{t+1} \). In general, the best \( h \)-step linear predictor is

\[ \hat{Y}_{t+h} = \mu - \sum_{j=1}^{h-1} \pi_j \hat{Y}_{h-j} - \sum_{j=h}^{\infty} \pi_j Y_{t+1-j}. \quad (21) \]

Therefore, the \( h \)-step predictor of \( \hat{r}_t^2 \) is given by

\[ \hat{r}_{t+h}^2 = \exp(\hat{Y}_{t+h}). \quad (22) \]

An alternative approach is to assume that the poles not included in the GLMSV model be treated as deterministic components as in Deo, Hurvich and Lu(2006) discussed in Section 2. For this purpose, we fit the deterministic function \( S_t \) with the residuals of the GLMSV model, instead of the original data set. This ensures mutual exclusivity of the stochastic and deterministic components. We compute the residuals as follows: If \( J_{Ry}(f) \) and \( J_{y^*}(f) \) are the discrete Fourier transforms of the residuals of the estimated GLMSV process and the observations \( Y_t^* \), respectively, then the residuals may be computed by taking the inverse Fourier transform of

\[ J_{Ry}(f) = \xi_t^* \prod_{v=1}^{d_1} \prod_{v=1}^{d_2} \{e^{-i2\pi v f} \} J_{y^*}(f), \quad f \in (-0.5,0.5). \quad (23) \]

We then fit the regression equation

\[ RY_t = S_t + \xi_t, \quad \xi_t \sim \text{iidN}(0,\sigma^2), \quad (24) \]

to obtain the estimated seasonal component

\[ \hat{S}_t = \sum_{p=1}^{k} \hat{a}_p \cos(f_p t) + \sum_{p=1}^{k} \hat{b}_p \sin(f_p t), \quad (25) \]

and the deseasonalized series \( Y_t^{DS} = Y_t - \hat{S}_t \). The \( h \)-step ahead forecast is then given by

\[ \hat{Y}_{t+h} = \hat{Y}_{t+h}^{DS} + \hat{S}_{t+h}, \quad (26) \]

where \( \hat{Y}_{t+h}^{DS} \) is the best \( h \)-step linear predictor in (21) and \( \hat{S}_{t+h} \) is the \( h \)-step ahead forecast of the deterministic seasonal component. The \( h \)-step ahead forecast of the squared return is simply

\[ \hat{r}_{t+h}^2 = \exp(\hat{Y}_{t+h}^{DS}). \quad (27) \]

The MA representation in (16) implies that \( M_\varepsilon \), the linear subspace generated by \( \{\varepsilon_s \leq t\} \), is a subspace of \( M_X \). The AR representation in (17) implies that \( M_X \) is a
subspace of $M_\varepsilon$ such that $M_X = M_\varepsilon$. Hence, the best $h$-step predictor can be computed as the projection of $Y_{t+h}$ on $M_\varepsilon$. Using the moving average representation in (16), we have

$$Y_{t+h} = \mu + \sum_{s=0}^{\infty} \psi_s e_{t+h-s} + \eta_{t+h} = \mu + \sum_{s=-h}^{\infty} \psi_{h+s} e_{t-s} + \eta_{t+h}. \quad (28)$$

The best $h$-step predictor is then given by

$$\hat{Y}_{t+h} = \mu + \sum_{s=0}^{\infty} \psi_{h+s} e_{t-s}. \quad (29)$$

The prediction error of the $h$-step ahead forecast of the squared return is given by

$$E[r_{t+v}^2 - \hat{r}_{t+v}^2] = E[\exp(Y_{t+h}^2 + S_{t+h}) - \exp(\hat{Y}_{t+h}^2 + \hat{S}_{t+h})]^2$$

$$= E\left[\exp\left(\mu + \sum_{s=0}^{\infty} \psi_{v+s} e_{t-s} + S_t\right) \exp\left(\sum_{s=0}^{\infty} \psi_{v+s} e_{t-s} + \eta_{t+v} + \xi_{t+v}\right)\right]^2. \quad (30)$$

Since $\{e_t\}$ is iid and independent of $\{\eta_t\}$, we have

$$E[r_{t+v}^2 - \hat{r}_{t+v}^2] = \exp(2\mu)E[\exp(2\eta_{t+v})]E[\exp(2\xi_{t+v})]\prod_{s=0}^{\infty} E[\exp(2\psi_{v+s} e_{t-s})]. \quad (31)$$

where by the moment generating function of the normal distribution we have

$$\exp(\mu) = \sigma^2 \exp\left(\mathbb{E}[\log e_t^2]\right). \quad (32)$$

$$E[\exp(\eta_{t+v})] = \exp\left(\mathbb{E}[\log e_t^2]\right), \quad (33)$$

$$E[\exp(\xi_{t+v})] = \exp\left(\frac{1}{2} \mathbb{E}[\xi^2]\right), \quad (34)$$

$$E[\exp(\psi_{v+s} e_{t-s})] = \exp\left(\sigma^2 e^2 / 2\right). \quad (35)$$

Thus, the prediction error is given by

$$E[r_{t+v}^2 - \hat{r}_{t+v}^2] = \sigma^4 \exp\left(\sigma^2 e^2 + \sigma^2 \sum_{s=0}^{\infty} \psi_s^2\right). \quad (36)$$

This prediction error is an increasing function of the variances of the innovations in the GLMSV model. Conditions A1 and A2 ensure that it is finite since $\sum_{s=0}^{\infty} \psi_s^2 < \infty$.

5. Application

In this section, we assess the performance of the proposed method by applying the results to modeling and forecasting Microsoft stock intraday volatility. We compare the predictive performance of the deterministic seasonality approach in Deo, Hurvich and Lu(2006) in Section 2 denoted by “DetSes+LMSV”, the stochastic seasonality model based on “GLMSV”, and the GLMSV plus deterministic seasonality denoted by “GLMSV+DetSes” in Section 3.

The data consists of Microsoft stock realized volatility, $r_v$, using tick data sampled every minute. It covers 140 trading days in the period August 5, 2008 to November 26, 2008 each 9:30 to 16:00, with 391 prices per day. Returns are defined as $r(t + m\delta) = p(t + m\delta) - p(t + (m-1)\delta)$, where $p = \log(p)$, $m=1,\ldots,390$, $\delta$ is 1/390, i.e. 1 minute, and $t=1,\ldots,140$. $P$ denotes the prices of the share. We consider realized volatility at
30-minutes intervals $rv(t; h, \delta) = \sum_{m=1}^{30} [r(t + 30(h-1) + \delta m)]^2$ with $h=1,\ldots,13$. (See Andersen et al.(2003)) For instance, $rv(2;3,\delta)$ is the realized volatility of third half hour period of the second day. So a total of $13 \times 140$ observations are available, using the first $N = 2^{10}$. Overnight returns are ignored as we are interested primarily in the intraday seasonal pattern.

Figure 1. Log-periodogram of $Y_t$

The log-periodogram of the $Y_t$ series in Figure 1 of the Microsoft stock intraday realized volatility seems to indicate 3 to 4 strong seasonal components with varying decay, out of the total 7 possible seasonal components on the frequency range $[0, 0.5]$ or 13 seasonal components on $(-0.5, 0.5]$. Using the “DetSes + LMSV” approach, the series is first deseasonalized by removing the deterministic seasonal component $\hat{S}_t$, which is estimated by fitting the multiple regression equation $Y_t = \hat{S}_t$. Figure 2 gives us the deseasonalized series after removing $\hat{S}_t$ from $Y_t$. It shows a log-periodogram with hyperbolic decay with one pole at the origin. Using (14) restricted to only one pole at $\nu=0$, the estimated model is a
fractionally integrated (FI) process with long-memory parameter $d = 0.1772893$. Forecasts are obtained by adding the forecast values of the estimated FI process and the deterministic seasonal component $\hat{S}_t$. The mean squared errors (MSE) of these forecasts are given in Table 2 denoted by “DetSes+LMSV”.

To obtain a GLMSV model for $Y_t$, we start estimating the stochastic volatility model with two poles, then we successively add more seasonal components into the model. For this purpose, we use the Akaike Information Criterion (AIC) for model selection. The AIC is a penalized log-likelihood defined by $-2(\log\text{-likelihood}) + 2p$, where $p$ is the number of parameters in the model. A smaller AIC for a competing model indicates the need for inclusion of more parameters in order to fit an additional singularity in the spectrum. Using this criterion, the final model is a GLMSV with 4 poles with long-memory parameter estimates $d_1 = 0.23178$, $d_2 = 0.40606$, $d_3 = 0.25687$, $d_4 = 0.34496$ with AIC = $-827.8083$. Any further increase in the number of poles or ARMA parameters either results in higher AIC or parameter estimates outside the bounds for stationarity, which may require a different approach. Similar results are obtained using the Bayesian information criterion (BIC) or Schwarz Criterion.

To test the efficiency of deterministic deseasonalization further, we also estimate the deterministic seasonal component $\hat{S}_t$ by fitting a multiple linear regression between $S_t$ and the residuals of the estimated GLMSV model. The MSE of the forecasts are given in Table 2 denoted by “$k$-GLMSV” and “$k$-GLMSV+DetSes”, respectively.

<table>
<thead>
<tr>
<th></th>
<th>DetSes+LMSV</th>
<th>2-GLMSV</th>
<th>4-GLMSV</th>
<th>2-GLMSV+DetSes</th>
<th>4-GLMSV+DetSes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-step</td>
<td>9.26636(10⁻⁸)</td>
<td>9.24919(10⁻⁹)</td>
<td>7.27475(10⁻⁶)</td>
<td>9.24883(10⁻⁸)</td>
<td>7.27117(10⁻⁵)</td>
</tr>
<tr>
<td>2-step</td>
<td>9.29977(10⁻⁸)</td>
<td>9.28265(10⁻⁹)</td>
<td>7.31077(10⁻⁶)</td>
<td>9.28249(10⁻⁸)</td>
<td>7.30667(10⁻⁵)</td>
</tr>
<tr>
<td>3-step</td>
<td>9.37528(10⁻⁸)</td>
<td>9.35818(10⁻⁹)</td>
<td>7.38144(10⁻⁶)</td>
<td>9.35808(10⁻⁸)</td>
<td>7.37841(10⁻⁵)</td>
</tr>
<tr>
<td>4-step</td>
<td>9.51507(10⁻⁸)</td>
<td>9.49785(10⁻⁹)</td>
<td>7.53409(10⁻⁶)</td>
<td>9.49766(10⁻⁸)</td>
<td>7.52771(10⁻⁵)</td>
</tr>
<tr>
<td>5-step</td>
<td>9.54360(10⁻⁸)</td>
<td>9.52640(10⁻⁹)</td>
<td>7.57358(10⁻⁶)</td>
<td>9.52619(10⁻⁸)</td>
<td>7.56990(10⁻⁵)</td>
</tr>
<tr>
<td>6-step</td>
<td>9.48543(10⁻⁸)</td>
<td>9.46850(10⁻⁹)</td>
<td>7.49649(10⁻⁶)</td>
<td>9.46814(10⁻⁸)</td>
<td>7.48775(10⁻⁵)</td>
</tr>
<tr>
<td>7-step</td>
<td>9.45412(10⁻⁸)</td>
<td>9.43722(10⁻⁹)</td>
<td>7.49436(10⁻⁶)</td>
<td>9.43685(10⁻⁸)</td>
<td>7.48954(10⁻⁵)</td>
</tr>
<tr>
<td>8-step</td>
<td>9.48849(10⁻⁸)</td>
<td>9.47135(10⁻⁹)</td>
<td>7.50030(10⁻⁶)</td>
<td>9.47111(10⁻⁸)</td>
<td>7.49254(10⁻⁵)</td>
</tr>
<tr>
<td>9-step</td>
<td>9.50715(10⁻⁸)</td>
<td>9.48999(10⁻⁹)</td>
<td>7.52429(10⁻⁶)</td>
<td>9.48973(10⁻⁸)</td>
<td>7.51765(10⁻⁵)</td>
</tr>
<tr>
<td>10-step</td>
<td>9.53656(10⁻⁸)</td>
<td>9.51929(10⁻⁹)</td>
<td>7.55565(10⁻⁶)</td>
<td>9.51906(10⁻⁸)</td>
<td>7.54837(10⁻⁵)</td>
</tr>
<tr>
<td>11-step</td>
<td>9.61075(10⁻⁸)</td>
<td>9.59336(10⁻⁹)</td>
<td>7.63974(10⁻⁶)</td>
<td>9.59327(10⁻⁸)</td>
<td>7.63389(10⁻⁵)</td>
</tr>
<tr>
<td>12-step</td>
<td>9.66456(10⁻⁸)</td>
<td>9.64692(10⁻⁹)</td>
<td>7.66925(10⁻⁶)</td>
<td>9.64683(10⁻⁸)</td>
<td>7.66100(10⁻⁵)</td>
</tr>
<tr>
<td>13-step</td>
<td>9.71037(10⁻⁸)</td>
<td>9.69928(10⁻⁹)</td>
<td>7.71418(10⁻⁶)</td>
<td>9.69249(10⁻⁸)</td>
<td>7.70907(10⁻⁵)</td>
</tr>
<tr>
<td>1-day</td>
<td>9.25562(10⁻⁸)</td>
<td>9.23520(10⁻⁹)</td>
<td>7.21842(10⁻⁶)</td>
<td>9.23550(10⁻⁸)</td>
<td>7.21236(10⁻⁵)</td>
</tr>
<tr>
<td>2-day</td>
<td>9.97850(10⁻⁸)</td>
<td>9.96126(10⁻⁹)</td>
<td>7.96624(10⁻⁶)</td>
<td>9.96086(10⁻⁸)</td>
<td>7.95396(10⁻⁵)</td>
</tr>
<tr>
<td>3-day</td>
<td>10.3762(10⁻⁸)</td>
<td>10.3590(10⁻⁹)</td>
<td>8.34031(10⁻⁶)</td>
<td>10.3585(10⁻⁸)</td>
<td>8.33630(10⁻⁵)</td>
</tr>
<tr>
<td>4-day</td>
<td>10.9985(10⁻⁸)</td>
<td>10.9818(10⁻⁹)</td>
<td>9.03206(10⁻⁶)</td>
<td>10.9812(10⁻⁸)</td>
<td>9.02664(10⁻⁵)</td>
</tr>
<tr>
<td>5-day</td>
<td>11.6785(10⁻⁸)</td>
<td>11.6619(10⁻⁹)</td>
<td>9.71785(10⁻⁶)</td>
<td>11.6612(10⁻⁸)</td>
<td>9.71296(10⁻⁵)</td>
</tr>
</tbody>
</table>

Table 2 Mean squared errors of forecast values

Table 2 shows that the MSE of the forecasts non-monotonically increase with time. A notable decrease is at the 6-step ahead forecast, which corresponds to forecast after 3 hours from the start of the trading day. The results also show that forecasts based on models involving stochastic seasonality (GLMSV or GLMSV+DetSes) perform better than the deterministic seasonality approach (DetSes+LMSV). The final model selected by the AIC (4-
GLMSV) performs best among the competing models without deterministic deseasonalization. On the other hand, the inclusion of deterministic seasonality not accounted for by the GLMSV model (GLMSV+DetSes) does not contribute significantly to the improvement of the prediction errors. It should be noted, however, that the computational complexity in forecasting a $k$-GLMSV significantly increases with $k$. Hence, model selection criteria such as AIC and BIC are imperative in providing a trade-off between computing time and prediction accuracy.

6. Concluding Remarks

In this paper, we proposed an alternative approach to estimation and forecasting seasonal long-memory stochastic volatility based on the generalized long-memory stochastic volatility (GLMSV) model. Forecasts results applied to Microsoft stock realized volatility have demonstrated the superiority of the stochastic seasonality approach based on GLMSV over the deterministic seasonality approach. Deterministic deseasonalization of other possible seasonal components not accounted for by the GLMSV model does not significantly improve the forecasts of the corresponding GLMSV. However, these do not preclude possible disparate results based on series of different dynamics.

Appendix (Proof of Theorem 1)

Let $\theta_y$ be an element of the compact parameter space $\Theta_y$. We apply the normalization $S^*_y(f;\theta_y) = (\sigma^2_x + \sigma^2_\eta)^{-1} S_y(f;\theta_y)$ such that $\int_\Theta \log S^*_y(f;\theta_y) df = 0$. The likelihood function in (14) reduces to

$$\ell(\theta_y) = N^{-1} \sum_{j=1}^{[N-1/2]} \left| \frac{I_N(\lambda_j)}{S^*_y(\lambda_j)} \right|$$

We also define

$$\ell^o(\theta_y) = \int_{1/2} \left[ \frac{S^*_y(f;\theta_y)}{S^*_y(f;\theta^o_y)} \right] df.$$  (38)

We then have

$$|\ell_N^*(\theta_y) - \ell^o(\theta_y)| \leq N^{-1} \sum_{r=1}^{[N-1/2]} \frac{I_N(f)}{S^*_y(f;\theta_y)} - \int_{1/2} \frac{S^*_y(f;\theta^o_y)}{S^*_y(f;\theta^o_y)} df = M_N.$$  (39)

Now, $M_N$ can be shown to converge to zero almost surely by modifying the proof of Lemma 1 of Hannan(1973) as in the proof of Theorem 1 in Breidt, et al. (1998). First, since $\left[ S^*_y(f;\theta_y) \right]^{-1}$ is continuous in $(f, \theta_y)$, the Cesaro sum of its Fourier series converges uniformly in $(f, \theta_y)$. Second, by A1 and A2 the process $\{Y_t\}$ is ergodic since $\{X_t\}$ linear process with iid innovations and square summable coefficients (see Hannan(1970) p. 204) and $\{\eta_t\}$ are iid independent of $\{X_t\}$. Thus, by Lemma 1 of Hannan(1973), $M_N$ converges to zero almost surely such that

$$\sup_{\theta} |\ell_N^*(\theta_y) - \ell^o(\theta_y)| \longrightarrow 0 \quad a.s.$$  (40)
Since $\int_{-\frac{1}{2}}^{\frac{1}{2}} \log S_Y(f; \theta_0) df = 0$ and $-\log x \geq 1 - x$, then

$$\ell^o(\theta_y) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \left\{ -\log \left[ \frac{S_Y(f; \theta_0)}{S_Y(f; \theta_y)} \right] + \frac{S_Y(f; \theta_0)}{S_Y(f; \theta_y)} \right\} df,$$ (41)

$$\geq \int_{-\frac{1}{2}}^{\frac{1}{2}} \left\{ 1 - \frac{S_Y(f; \theta_0)}{S_Y(f; \theta_y)} + \frac{S_Y(f; \theta_0)}{S_Y(f; \theta_y)} \right\} df,$$ (42)

$$= \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[ \frac{S_Y(f; \theta_0)}{S_Y(f; \theta_y)} \right] df = \ell^o(\theta_0).$$ (43)

And using the identifiability condition A3, $\theta_y^0$ uniquely minimizes $\ell^o(\theta_y)$. This implies that $\ell^o(\theta_y^0) \leq \ell^o(\hat{\theta}_{y_n})$. Moreover, $\ell^o(\hat{\theta}_{y_n}) = \inf_{\theta_y} \ell^o(\theta_y) \leq \ell^o(\theta_y^0)$. These imply that $\ell^o(\hat{\theta}_{y_n}) \rightarrow \ell^o(\theta_y^0)$ a.s. Therefore $\hat{\theta}_{y_n} \rightarrow \theta_y^0$ a.s. by compactness of $\Theta_y$.

References


Daubechies I (1992) Ten Lectures on Wavelets. SIAM, Philadelphia


