2016/04

A Simple Model for Now-Casting Volatility Series

Jörg Breitung and Christian M. Hafner

50 YEARS OF

CORE

DISCUSSION PAPERS

CORE
Voie du Roman Pays 34, L1.03.01
Tel (32 10) 47 43 04
Fax (32 10) 47 43 01
Email: immaq-library@uclouvain.be
http://www.uclouvain.be/en-44508.html

A simple model for now-casting volatility series

Jörg Breitung[1] and Christian M. Hafner*[2]

[1] *University of Cologne*

[2] *Université catholique de Louvain*

October 2015

Abstract

Popular volatility models focus on the conditional variance given past observations, whereas the (arguably most important) information in the current observation is ignored. This paper proposes a simple model for now-casting volatilities based on a specific ARMA representation of the log-transformed squared returns that allows us to estimate current volatility as a function of current and past returns. The model can be viewed as a stochastic volatility model with perfect correlation between the two error terms. It is shown that the volatility nowcasts are invariant to this correlation and therefore the estimated volatilities coincide. An extension of our now-casting model is proposed that takes into account the so-called leverage effect. The alternative models are applied to estimate daily return volatilities from the S&P 500 stock price index.

Some key words: EGARCH, stochastic volatility, ARMA, realized volatility, leverage

JEL Classification Number: C22, C58

*Corresponding author, Institute of statistics, biostatistics and actuarial sciences, and CORE, Université catholique de Louvain, Voie du Roman Pays 20, 1348 Louvain-la-Neuve, Belgium, email: `christian.hafner@uclouvain.be`.

The authors would like to thank seminar participants at Humboldt-University Berlin, University of Cologne, University of Salerno, University of St Andrews, University of Cambridge, Tinbergen Institute Amsterdam, and in particular Giampiero Gallo, Siem Jan Koopman and Roman Liesenfeld, for helpful comments and discussions.

1 Introduction

The literature on volatility models continues to grow steadily, driven mainly by the success that these models encounter in modelling financial time series, but also by the non-exhausted understanding of some of their properties and their estimators. The main benchmark remains the classical GARCH model, introduced by Engle (1982) and Bollerslev (1986), due to its simplicity in estimation and widespread availability in software packages. The GARCH model is essentially a model for *predicting* volatility for today, given past observations. It does so quite well, as demonstrated by Andersen and Bollerslev (1998) by using a realized volatility target instead of the commonly used daily squared returns. However, the GARCH model does not offer the possibility to update a prediction with today's observed data. In other words, *nowcasting* volatility in the GARCH model corresponds to using predicted volatility, ignoring today's observation. Following Anderson and Bollerslev (1998) we consider a continuous time process where the instantaneous returns are generated by the martingale

$$dp(t) = \sigma(t) \cdot dW_p(t) \tag{1}$$

where $W_p(t)$ is a Wiener process with $\mathbb{E}[W_p(t) - W_p(t-1)]^2 = 1$. In discrete time with $t = 1, 2, \ldots, T$ the variance results as

$$h_t \equiv \mathbb{E}[p(t) - p(t-1)]^2 \;=\; \int_{t-1}^{t} \sigma(s)^2 ds \;.$$

For concreteness let us consider the diffusion limit of the GARCH(1,1) process given by

$$d\sigma(t) = a_1[a_2 - \sigma(t)^2] \cdot dt + \sqrt{2a_3a_1}\,\sigma(t) \cdot W_\sigma(t) \tag{2}$$

where a_1, a_2, a_3 are positive parameters and the standard Wiener process $W_\sigma(t)$ is independent of $W_p(t)$ (see also Anderson and Bollerslev (1998)).

Let $y_t = p(t) - p(t-1)$ with $\mathbb{E}(y_t|y_{t-1}, y_{t-2}, \ldots) = 0$ and consider the GARCH(1,1) discrete time approximation of the variance process

$$\begin{aligned} y_t &= \sqrt{h_{t|t-1}}\xi_t \\ h_{t|t-1} &= \mathbb{E}(y_t^2|y_{t-1}^2, y_{t-2}^2, \ldots) \\ &= \mu + \alpha y_{t-1}^2 + \phi h_{t-1|t-2} \end{aligned}$$

where ξ_t is i.i.d. with $\mathbb{E}(\xi_t) = 0$ and $\mathbb{E}(\xi_t^2) = 1$. Letting $y_t^2 = h_{t|t-1} + v_t$ we can replace $h_{t|t-1}$ by $y_t^2 - v_t$ yielding the ARMA representation of y_t^2:

$$y_t^2 = \mu + (\alpha + \phi)y_{t-1}^2 + v_t - \phi v_{t-1}. \tag{3}$$

Accordingly, the conditional variance is equivalent to the linear forecast of y_t^2 conditional on $\{y_{t-1}^2, y_{t-2}^2, \ldots\}$ and the conditional variance process results from a filtration of the form

$$h_{t|t-1} = \frac{\mu}{1-\phi} + \alpha \sum_{i=1}^{\infty} \phi^{i-1} y_{t-i}^2. \tag{4}$$

An important drawback of the standard GARCH model is that the observation y_t^2 does not enter the conditional variance process $h_{t|t-1}$, which is arguably the most important information about current volatility h_t. This pitfall has been noted, for instance, by Politis (2007) among others.

To appreciate the importance of the current observation for estimating ("nowcasting") the volatility process we simulate the discrete analog to the continuous time processes (1) and (2) for $t = 1, \ldots, 5000$. For our simulation experiment we employ the same parameters as for the DM-\$ process of Anderson and Bollerslev (1998). Our parameter estimates $\widehat{\alpha} = 0.064$ and $\widehat{\phi} = 0.91$ correspond well to the estimates presented in Table 1 of Anderson and Bollerslev (1998). To investigate how well the estimated conditional variances predict the variance process h_t we run a regression of h_t on the estimated GARCH(1,1) variances $\widehat{h}_{t|t-1}$ yielding

$$\begin{array}{ccccc} h_t = & 0.089 & + & 0.853\, \widehat{h}_{t|t-1} & + \widehat{u}_t\ , \\ & (0.038) & & (0.074) & \end{array}$$

where HAC standard errors are presented in parentheses. The regression R^2 is 0.467, which is slightly below the value reported by Anderson and Bollerslev (1998). The restrictions that the constant is zero and the slope is equal to one cannot be rejected at the 5 percent significance level.

Next we repeat the regression by including y_t^2 as an additional regressor resulting in

$$\begin{array}{ccccc} h_t = & 0.088 & + & 0.790\, \widehat{h}_{t|t-1} & + \ 0.065 y_t^2 \ + \ \widehat{u}_t\ . \\ & (0.036) & & (0.062) & \qquad (0.005) \end{array}$$

According to these results, the square of the current observation is highly significant and increases the R^2 to 0.508. This first experiment suggests that including the contemporaneous observation in the information set may provide much more reliable estimates of current volatility.

In this paper we propose a simple variant of the (exponential) GARCH model that exploits the information in the current observation y_t. Assuming normality of $\log \xi_t^2$ the

model parameters can be estimated efficiently by fitting an ARMA(1,1) model to the transformed series $x_t = \log y_t^2$. In contrast to the GARCH(1,1) the log variance process in our model results from the filtration

$$E(\log h_t | x_t, x_{t-1}, \ldots) = c + \left(1 - \frac{\theta}{\beta}\right) \sum_{i=0}^{\infty} \theta^i x_{t-i} \ , \tag{5}$$

where θ and β are typically positive parameters, close to unity with $\theta < \beta$ and c is a constant.

The plan of the remainder of the paper is as follows. In Section 2 we introduce our forecasting model and in Section 3 the reduced form ARMA(1,1) representation is developed. The relationship to the stochastic volatility model is studied in Section 4 and the small sample properties are studied in Section 5. An asymmetric extension for accommodating the leverage effect is proposed in Section 6. Section 7 presents an application to the S&P 500 stock price index. Finally, Section 8 concludes.

2 The nowcasting model

To exploit the information in the current observation y_t we consider the following model for a series of financial returns y_t,

$$y_t = \exp(h_t/2)\xi_t, \quad \xi_t \sim i.i.d.(0,1) \tag{6}$$

$$h_t = \alpha + \beta h_{t-1} + \kappa \varepsilon_t \ , \tag{7}$$

where $\varepsilon_t = \log(\xi_t^2) - C \sim i.i.d.(0, \sigma_\varepsilon^2)$ and $C = \mathbb{E}[\log(\xi_t^2)]$. Stationarity of the variance process requires $|\beta| < 1$ and in empirical practice we typically encounter values slightly less than unity. Furthermore, we expect κ to be small and positive because a large absolute value of ξ_t tends to increase volatility but in principle this parameter might also be negative. The invertibility of the reduced form ARMA representation (see below) only requires $1 + \kappa > \beta$ which would also allow for small negative values of κ.

Log volatility h_t in (7) follows an AR(1) process, but unlike in stochastic volatility models where this process is independent of ξ_t, the error term ε_t in (7) is an explicit function of the innovation term ξ_t in (6). Some more comparisons with the stochastic volatility model will be presented in Section 4.

Let us first discuss some properties of model (6)–(7). If the distribution of ξ_t is known, then the parameter C is identified. For example, for Gaussian ξ_t, $C \approx -1.27$. In what

follows we assume that C is unknown. At the end of Section 3 we discuss how to estimate this constant. Note also that mean and variance of log volatility are given by, respectively, $\mathbb{E}[h_t] = \alpha/(1-\beta)$ and $\mathrm{Var}(h_t) = \kappa^2\sigma_\varepsilon^2/(1-\beta^2)$, and σ_ε^2 depends on the distribution of ξ_t. If ξ_t is Gaussian, then $\sigma_\varepsilon^2 = \pi^2/2$.

Under the assumption that ξ_t has a symmetric distribution it follows that y_t is a martingale difference series. To see this, let $I_{t-1} := \sigma(y_{t-1}, y_{t-2}, \ldots)$ be the information set generated by the observations, and note that ξ_t is independent of I_{t-1}, while h_{t-1} is measurable with respect to I_{t-1}. Then,

$$\mathbb{E}[y_t|I_{t-1}] = \exp\{(\alpha + \beta h_{t-1})/2\}\mathbb{E}[\exp(\kappa/2 \log \xi_t^2)\xi_t],$$

where the expectation on the right hand side is zero since it is the expectation of an odd function of ξ_t. Thus, as in classical ARCH or stochastic volatility models, the return series y_t has a conditional mean of zero, and all temporal dependence is captured via the log volatility process h_t.

We now transform model (6) – (7) to obtain a linear process for the transformed variable. Defining $x_t = \log y_t^2$, we have

$$x_t = C + h_t + \varepsilon_t \tag{8}$$

and, replacing h_{t-1} in (7) by $x_{t-1} - \varepsilon_{t-1} - C$,

$$\begin{aligned} h_t &= \alpha - \beta C + \beta x_{t-1} + \kappa\varepsilon_t - \beta\varepsilon_{t-1} && (9) \\ x_t &= \alpha^* + \beta x_{t-1} + (1+\kappa)\varepsilon_t - \beta\varepsilon_{t-1}\ . && (10) \end{aligned}$$

where $\alpha^* := \alpha + (1-\beta)C$. Indeed, the transformed returns x_t in (10) follow an ARMA(1,1) process.

It is interesting to compare this model specification with two popular GARCH alternatives: First, the (symmetric version of the) EGARCH model suggested by Nelson (1991) replaces (7) by the equation

$$h_t = \alpha + \beta h_{t-1} + \psi|\xi_{t-1}|. \tag{11}$$

Here, log-volatilities are driven by lagged values ξ_{t-1} instead of the current values ξ_t. Moreover, by rewriting (7) as $h_t = \alpha + \beta h_{t-1} + \kappa \log(\xi_t^2)$ it becomes obvious that large shocks ξ_t have a much stronger effect in model (11). The proposed model (7) is actually closer to the so-called log-GARCH model, introduced independently by Geweke (1986) and Pantula (1986), where x_t is as in (8) with h_t given by

$$h_t = \alpha + \beta h_{t-1} + \psi \log y_{t-1}^2 \tag{12}$$

which leads to the ARMA representation

$$x_t = \alpha + (\psi + \beta)x_{t-1} + \varepsilon_t - \beta\varepsilon_{t-1}. \tag{13}$$

Notice the difference with respect to the ARMA representation (10), in which the coefficient κ captures the impact of the current observation on volatility in the moving average part, which is shifted to a lagged effect ψx_{t-1} in the autoregressive part of (13).

3 The reduced form ARMA representation

An observationally equivalent ARMA(1,1) model for x_t is obtained from

$$\begin{aligned} x_t &= \alpha^* + \beta x_{t-1} + (1+\kappa)\varepsilon_t - \frac{\beta}{1+\kappa}(1+\kappa)\varepsilon_{t-1} \\ &= \alpha^* + \beta x_{t-1} + u_t - \theta u_{t-1} \ , \end{aligned} \tag{14}$$

where $u_t = (1+\kappa)\varepsilon_t$ is white noise with variance $\sigma_u^2 = (1+\kappa)^2\sigma_\varepsilon^2$ and $\theta = \beta/(1+\kappa)$. Note that stationarity and invertibility of the model requires $\kappa > \beta - 1$. The relationship between the reduced form parameters θ, $\sigma_u^2 = \mathbb{E}(u_t^2)$, and the structural parameters κ, σ_ε^2 is given by

$$\kappa = \beta/\theta - 1 \tag{15}$$

$$\sigma_\varepsilon^2 = \left(\frac{\theta}{\beta}\right)^2 \sigma_u^2. \tag{16}$$

Another possiblity is to motivate our structural model by a Beveridge-Nelson (1981) type of decomposition.[1] Decomposing the ARMA polynomial as

$$\frac{1-\theta L}{1-\beta L} = \frac{a + b(1-\beta L)}{1-\beta L} \tag{17}$$

such that

$$x_t = C + \underbrace{\frac{\alpha}{1-\beta} + \frac{a}{1-\beta L}u_t}_{h_t} + \underbrace{bu_t}_{\varepsilon_t}.$$

yields the structural form (8). Accordingly, our structural model corresponds to a simple decomposition of an ARMA(1,1) series into an AR(1) and a white noise component.

[1] We are grateful to an anonymous referee for suggesting this interpretation to us.

Comparing the coefficients of the polynomial in equation (17) yields $b = \theta/\beta$ and $a = 1 - \theta/\beta$ and, thus,

$$\begin{aligned} h_t &= \alpha + \beta h_{t-1} + (1 - \theta/\beta) u_t \\ \varepsilon_t &= \left(\frac{\theta}{\beta}\right) u_t. \end{aligned} \tag{18}$$

The usual Beveridge-Nelson decomposition is obtained by letting $\beta = 1$.

Since $\varepsilon_t = u_t/(1+\kappa) = (\theta/\beta)u_t$, the variance component h_t is obtained from the reduced form as

$$h_t = x_t - \frac{\theta}{\beta} u_t - C. \tag{19}$$

Note that (19) is measurable w.r.t. present and past values of x_t, because the reduced form is invertible and $u_t = -\alpha^*/(1-\theta) + \phi(L)x_t$ with $\phi(L) = (1-\theta L)^{-1}(1-\beta L)$. By comparing coefficients of the lag polynomials, one obtains $\phi(L) = 1 + (1-\beta/\theta)\sum_{j=1}^{\infty} \theta^j L^j$. Inserting this result into (19), we obtain

$$h_t = \frac{\theta\alpha^*}{\beta(1-\theta)} + \left(1 - \frac{\theta}{\beta}\right) \sum_{j=0}^{\infty} \theta^j x_{t-j}. \tag{20}$$

This shows that the filtered volatility is a linear combination of present and past values of x_t with exponentially declining weights.

(Pseudo) ML estimators of the structural parameters $(\beta, \kappa, \sigma_\varepsilon^2)$ are obtained by inserting the ML estimators of the de-meaned reduced form $(\beta, \theta, \sigma_u^2)$ into (15) and (16). An estimator of the constant α^* is obtained from the equality $\alpha^* = \mu(1-\beta)$, where μ is the mean of x_t. Based on the consistency and asymptotic normality of the reduced form ML estimators, we can find similar results for the estimators of the structural form using the delta method. This gives closed form expressions for the asymptotic variances of $\sqrt{n}(\hat{\beta} - \beta)$ and $\sqrt{n}(\hat{\kappa} - \kappa)$, see Appendix A.1. Note that in practice, θ is often close to unity. Therefore, an exact ML estimation method with stationary initial values should be employed rather than the popular nonlinear least-squares estimator setting the initial values y_0 and u_0 equal to zero. Whenever ε_t is normally distributed, then the ML estimator is asymptotically efficient.

Estimation of the reduced form ARMA(1,1) model in (14) delivers parameter estimates of α^*, β and θ, which could be used to obtain filtered volatilities in equation (20). The simpler expression in (19) cannot be used directly since C is not known. We next propose an approach to estimate this constant.

Estimation of the constant. Suppose one uses (19) but ignores the unknown constant C, i.e. setting it to zero. This delivers a filtered volatility process $h_t^* = h_t + C$. From (8) it follows that

$$y_t^2 = e^{h_t^* - C}\xi_t^2 \;=\; c\, e^{h_t^*}\xi_t^2 \;,$$

where $c = \exp(-C)$ and

$$\xi_t^2 = \frac{y_t^2}{c\, e^{h_t^*}} \;.$$

Since we assume that $\mathbb{E}(\xi_t^2) = 1$ we can estimate the constant from the estimated values of ξ_t^2 as

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T}\widehat{\xi}_t^2 \;&=\; 1 \\ \Leftrightarrow \quad \widehat{c} \;&=\; \frac{1}{T}\sum_{t=1}^{T}\frac{y_t^2}{e^{\widehat{h}_t^*}} \;, \end{aligned}$$

where $\widehat{h}_t^* = x_t - (\widehat{\theta}/\widehat{\beta})\widehat{u}_t$ denotes the ARMA estimator of the volatility series.

Maximum likelihood estimation. By making distributional assumptions on the error ε_t it is possible to estimate the parameters by maximum likelihood. Assume for example that $\xi_t \sim \mathcal{N}(0,1)$. Then $\mathrm{Var}(\varepsilon_t) = \pi^2/2$ (e.g. Taylor, 1986) which implies a restriction between the reduced form parameters and the residual variance given by

$$\sigma_u^2 = \left(\frac{\pi\beta}{\sqrt{2}\theta}\right)^2 . \tag{21}$$

This restriction can be imposed to a pseudo ML estimator that treats u_t as being normally distributed with variance (21). Another gain in efficiency would result from setting up the likelihood function based on the more appropriate assumption that u_t is the logarithmic transformation of a χ^2 distributed random variable. As found in many empirical work, however, the GARCH innovation is typically fat-tailed and is therefore often modeled by invoking the t-distribution. We do not advocate these more sophisticated estimation techniques for several reasons. First, the computational effort for these refinements increase dramatically and the estimator cannot be obtained by usual software packages. Second, in financial applications a large sample size is typically available so that the estimation error is negligible relative to the magnitude of h_t and efficiency is of minor importance.

Third, it is not clear what class of distribution is best suited for ξ_t or ε_t. Note that if y_t tends to zero, then x_t tends to $-\infty$. To avoid large negative outliers it is therefore recommendable to add some small number (say $0.001 \cdot \widehat{\sigma}_y^2$) to y_t^2 before applying the logarithmic transformation. In our experience such slight adjustments are much more important for the performance of the estimator than the distributional assumptions.

4 Relationship to the stochastic volatility model

It is interesting to compare our approach to the stochastic volatility (SV) model, where (7) is replaced by

$$h_t = \alpha + \beta h_{t-1} + \eta_t \tag{22}$$

assuming that ξ_t and η_t are independent. The ARMA representation results as

$$x_t = \alpha^* + \beta x_{t-1} + \eta_t + \varepsilon_t - \beta \varepsilon_{t-1} \ . \tag{23}$$

where $\varepsilon_t = \log(\xi_t^2) - C$. Again we can find a second-order equivalent reduced form ARMA model as in (14), i.e.

$$x_t = \alpha^* + \beta x_{t-1} + u_t - \theta u_{t-1} \ , \tag{24}$$

that is, the autocovariance functions of x_t in (23) and (24) are identical. Accordingly, the model parameters of the SV model can be seen as transformations of the reduced form parameters in (24). Specifically we have

$$\sigma_u^2(1+\theta^2) = \sigma_\eta^2 + \sigma_\varepsilon^2(1+\beta^2) \tag{25}$$

$$\theta \sigma_u^2 = \beta \sigma_\varepsilon^2. \tag{26}$$

It follows that

$$\sigma_\varepsilon^2 = \frac{\theta}{\beta} \sigma_u^2 \tag{27}$$

$$\sigma_\eta^2 = \left[1 - \frac{\theta}{\beta} - \theta(\beta - \theta) \right] \sigma_u^2 \ . \tag{28}$$

The Kalman filter applied to the state space representation of this model delivers the filtered volatility

$$h_{t|t} = (1 - \theta/\beta) x_t + (\theta/\beta) h_{t|t-1}$$

where the predicted volatility $h_{t|t-1}$ is given by

$$h_{t|t-1} = \alpha^* + (\beta - \theta)x_{t-1} + \theta h_{t-1|t-2}$$

Hence, we obtain

$$\begin{aligned} h_{t|t} &= \frac{\theta}{\beta}\left(\frac{\alpha^*}{1-\theta}\right) + \frac{\kappa}{1+\kappa}\sum_{j=0}^{\infty}\theta^j x_{t-j} \\ &= \frac{\theta\alpha^*}{\beta(1-\theta)} + \left(1 - \frac{\theta}{\beta}\right)\sum_{j=0}^{\infty}\theta^j x_{t-j} \end{aligned}$$

which shows that the SV filtered volatility is equivalent to the filtered volatility using the ARMA model given by (20).

In the next proposition we show that this result extends to the class of models with an arbitrary error correlation:

Proposition 1 *Let $x_t = C + h_t + \varepsilon_t$, where $h_t = \alpha + \beta h_{t-1} + \eta_t$, $\eta_t \sim i.i.d.(0, \sigma_\eta^2)$, $\varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2)$ and arbitrary covariance $\mathbb{E}(\eta_t \varepsilon_t) = \rho\sigma_\varepsilon\sigma_\eta$ with $\rho \in [-1, 1]$. It follows that*

$$h_{t|t} = \frac{\theta\alpha^*}{\beta(1-\theta)} + \left(1 - \frac{\theta}{\beta}\right)\sum_{j=0}^{\infty}\theta^j x_{t-j}$$

The proof is provided in Appendix A.2.

Our model in (7) corresponds to the case $\rho = 1$, while the classical SV model (22) results from setting $\rho = 0$. It follows from Proposition 1 that for the estimation of h_t based on the information set $x_t, x_{t-1}, \ldots$ the correlation between ε_t and η_t does not matter. Therefore, there is no need to invoke Kalman filter recursions to estimate the variance process.

It is interesting to note that related results were found by Morley, Nelson and Zivot (2003) and Proietti (2006) for trend-cycle decompositions which can be seen as a special case with $\beta = 1$. It should be noted however that alternative structural representations involve different parameter estimates and may have very different implications on the reduced form. For example, the orthogonal decomposition with $\rho = 0$ implies that the spectral density of x_t is bounded from below by σ_η^2 which is not the case for our structural model with $\rho = 1$.

Note also that a non-zero correlation between ε_t and η_t does not imply that ξ_t and η_t are correlated. The latter case attracted some interest to model the so-called leverage

effect in stochastic volatility, see e.g. Harvey and Shephard (1996). For example, consider our model (6)-(7), i.e. the degenerate case of Proposition 1 with $\rho = 1$, $\eta_t = \kappa\varepsilon_t$, and suppose that the distribution of ξ_t is symmetric. Then, the correlation between ξ_t and η_t is zero even though ε_t and η_t are perfectly correlated. To include a leverage effect, the model needs to be extended, which we will do in Section 6.

Finally, if the distribution of ξ_t is symmetric, then it can be shown that the white noise u_t of the reduced form ARMA representation (24) is serially uncorrelated. In general, however, it is not a martingale difference, as e.g. $\mathbb{E}[u_t x_{t-1}^2] \neq 0$, see Francq and Zakoian (2006). The fact that u_t in the ARMA representation of the SV model is neither i.i.d. nor a martingale difference also has implications for inference. The general sandwich type formula for the asymptotic covariance matrix of QMLE estimators remains valid, but it is not available in closed form and it is different from the asymptotic covariance matrix of our model, given in Appendix A.1. Thus, although for given parameters both models yield the same filtered volatility estimates, estimation and inference are different due to the different properties of the error term u_t.

5 Finite sample properties

In this section we compare the finite sample properties of alternative estimators for volatilities. The data are generated as

$$y_t = e^{h_t/2}\xi_t \quad t = 1, \ldots, T \ ,$$

where h_t is either

$$\text{ARMA: } h_t = \alpha + \beta h_{t-1} + \kappa\varepsilon_t, \tag{29}$$

$$\text{SV}: \quad h_t = \alpha + \beta h_{t-1} + \eta_t, \tag{30}$$

$$\text{or EGARCH: } h_t = \alpha + \beta h_{t-1} + \psi|\xi_{t-1}|. \tag{31}$$

The error process $\varepsilon_t = \log(\xi_t^2) + 1.27$ with $\xi_t \overset{iid}{\sim} N(0,1)$ is independent of $\eta_t \overset{iid}{\sim} N(0, \sigma_\eta^2)$. Accordingly, in the stochastic volatility model (SV) $x_t = \log y_t^2$ is composed of two independent processes, whereas in the ARMA model x_t is driven by a single stochastic process ε_t.

First, consider the case where the generated volatility is a classical stochastic volatility process. We follow Sandmann and Koopman (1998) in specifying the parameters of the SV model. Defining the coefficient of variation, $CV = \mathbb{V}\text{ar}[\exp(h_t)]/\mathbb{E}[\exp(h_t)]^2$, one

obtains the expression $CV = \exp(\sigma_\eta^2/(1-\beta^2)) - 1$. The coefficient of variation for this model is directly related to the kurtosis of y_t, which is given by $\kappa = 3(CV+1)$. Here, α is an irrelevant scaling parameter, but Sandmann and Koopman (1998) determine α such that $E[\exp(h_t)] = 0.0009$, which gives a realistic annualized standard deviation of 22% for generated weekly data. To distinguish between highly and moderately persistent volatility processes, we fix β alternatively at 0.98 and 0.90. Similarly, to evaluate the effects of high versus low coefficients of variation (or, equivalently, high versus low kurtosis), we fix CV alternatively at 10 and 1, with corresponding kurtosis coefficients 33 and 6, respectively. This gives four different parameterizations. The sample sizes are $T = 500$ and 2000. Each process is simulated $k = 1000$ times.

The volatilities of the process y_t are estimated by fitting a symmetric EGARCH model, a symmetric SV model, and the ARMA approach proposed in Section 3, where the constant is estimated as suggested in Section 3. The performance is measured by an R^2 type criterion computed as

$$\widetilde{R}_h^2 = 1 - \frac{\sum_{t=1}^{T}(h_t - \widehat{h}_t)^2}{\sum_{t=1}^{T}(h_t - \bar{h})^2}$$

where $\bar{h} = T^{-1}\sum_{t=1}^{T} h_t$. This variant of the usual R^2 imposes a zero constant and a unit scaling coefficient in order to measure the correspondence of the estimates with the original volatility process. Table 1 reports the results.

Not surprisingly, the R^2 measures of the EGARCH model are substantially smaller in all cases, due to the smaller information set that is used in estimation, but also because the model is mis-specified. Also not surprisingly, the R^2 of SV and ARMA are close to each other, since both models deliver the same filtered volatility estimate for given parameters. Hence, the differences between them are solely due to differences in parameter estimates. Note that the ARMA R^2 tends to be higher when the persistence is moderate ($\beta = 0.9$). Note also that for increasing sample size, the R^2 does not need to improve, because the sample size affects the estimation error but not the signal to noise ratio determined essentially by σ_η^2 and σ_ε^2.

In the second simulation setup we generate reduced form ARMA processes for h_t with parameters chosen analogously to the SV case. More precisely, the persistence parameter β and the intercept α are the same as in SV. The moving average parameter θ is chosen such that $CV \in \{1, 10\}$, as before, by expressing θ as a function of σ_η, β, and σ_ε. Results

are reported in Table 2. Overall, the R^2 tends to be higher than in the SV case, which is plausible as there is no second noise term in the volatility equation. Furthermore, volatility is a measurable function of today's and lagged information. Thus, for increasing sample size we expect the R^2 to converge to unity, which happens for both estimation methods based on ARMA and SV. Again we observe the same effect as for a true SV process where the ARMA R^2 is higher for moderate persistence.

Finally, we generate EGARCH processes as in (31) with $\xi_t \sim N(0,1)$, setting as before $\beta \in \{0.9, 0.98\}$, and $CV \in \{1, 10\}$. Then one can calculate ψ from the equation $CV = \exp\{\psi^2(1-2/\pi)/(1-\beta^2)\} - 1$, and α from the equation $\mathbb{E}[\exp(h_t)] = \exp\{\alpha/(1-\beta) + \psi^2/2(1-\beta^2)\} = 0.0009$ as above. In the case of a true EGARCH process, the ARMA and SV models should have no advantage of including the current observation in the volatility, since in the EGARCH case volatility is a function of past values only. The results of the correctly specified EGARCH model are now much better than before. However, for some constellations such as high persistence and high CV, the performance is substantially worse according to the R^2 criterion.

6 An asymmetric extension

In order to account for the leverage effect that is often encountered in empirical applications, we define the dummy variable $d_t = I(y_t > \tau)$, where $I(\cdot)$ is the indicator function, and τ is a predefined threshold (which is typically zero), the mean of y_t, or some other value of interest.

An asymmetric extension of the above model is given by

$$h_t = \alpha + \beta x_{t-1} + \kappa^+ d_t \varepsilon_t + \kappa^-(1-d_t)\varepsilon_t - \beta\varepsilon_{t-1}, \tag{32}$$

which we call ARMA model with leverage, or ARMA-L. Note that in contrast to Nelson's EGARCH model and other asymmetric GARCH models, the asymmetric effect in this model is contemporaneous and not lagged.

The structural form for x_t results as

$$x_t = \alpha + \beta x_{t-1} + (1+\kappa^+)d_t\varepsilon_t + (1+\kappa^-)(1-d_t)\varepsilon_t - \beta\varepsilon_{t-1}. \tag{33}$$

Denote again the MA part of this model by $v_t = (1+\kappa^+)d_t\varepsilon_t + (1+\kappa^-)(1-d_t)\varepsilon_t - \beta\varepsilon_{t-1}$. We have the following conditional second order moment structure.

$$\begin{aligned} \mathrm{Var}(v_t|d_t) &= ((1+\kappa^+)^2 d_t + (1+\kappa^-)^2(1-d_t) + \beta^2)\sigma^2_{\varepsilon t} && (34) \\ \mathbb{E}[v_t v_{t-1}|d_t] &= -\{(1+\kappa^+)d_t + (1+\kappa^-)(1-d_t)\}\beta\sigma^2_{\varepsilon t} && (35) \end{aligned}$$

We can find an observationally equivalent ARMA(1,1) process

$$x_t = \alpha + \beta x_{t-1} + u_t - \theta^+ d_t u_{t-1} - \theta^-(1-d_t)u_{t-1}. \tag{36}$$

This process has the same conditional second order moment structure provided that

$$\begin{aligned} \kappa^+ &= \beta/\theta^+ - 1 && (37)\\ \kappa^- &= \beta/\theta^- - 1 && (38)\\ \sigma^2_{\varepsilon t} &= \{(1+\kappa^+)^2 d_t + (1+\kappa^-)^2(1-d_t)\}^{-1}\sigma^2_u && (39) \end{aligned}$$

Note that the error term ε_t is conditionally heteroskedastic. If the estimated model (36) is invertible, then it is easy to check that the model (33) with parameters given by (37)-(39) will also be invertible.

We could have chosen the alternative solution

$$\begin{aligned} \kappa^+ &= \beta\theta^+ - 1 && (40)\\ \kappa^- &= \beta\theta^- - 1 && (41)\\ \sigma^2_\varepsilon &= \frac{\sigma^2_u}{\beta^2} && (42) \end{aligned}$$

which is conditionally homoskedastic. However, if the estimated model (36) is invertible, then the model (33) with parameters given by (40)-(42) will not be invertible, and is therefore excluded.

Note that the process (36) is similar to the asymmetric ARMA model proposed by Brännäs and De Gooijer (1994), the difference being that in their model, the indicator variable is specified as $d_t = I(u_{t-1} > 0)$. The model (36) can be estimated by quasi maximum likelihood. To obtain the information matrix, the Hessian can be approximated by the sum of the outer products of the gradient as in Brännäs and De Gooijer (1994).

7 An empirical application

We apply our model on a large dataset, the demeaned daily (close to close) return on the S&P 500 index from 1/1/1950 — 25/10/2012, a total of 16,058 observations. We first estimate the classical EGARCH(1,1) with $N(0,1)$ innovations, as proposed by Nelson (1991):

$$\begin{aligned} y_t &= \exp(h_t/2)\xi_t, \quad \xi_t \sim N(0,1)\\ h_t &= \alpha + \beta h_{t-1} - \theta\xi_{t-1} + \gamma|\xi_{t-1}| \end{aligned}$$

This model is estimated by maximum likelihood, and the results are shown in Table 4.

We estimate the SV model (22) by QMLE and the Kalman filter, assuming $\eta_t \sim N(0, \sigma_\eta^2)$ and $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$. Results are also presented in Table 4. The estimators of σ_η^2 and σ_ε^2 correspond to (28) and (27).

The ARMA(1,1) model in (14) is estimated using nonlinear least squares with numerical optimization. For the nonlinear ARMA model with leverage (ARMA-L), see equation (32), we choose a threshold $\tau = -0.01$, corresponding to one negative unconditional standard deviation of returns y_t. Estimation results for the three models are also reported in Table 4. All three models pass portmanteau specification tests applied to the squared residuals $\hat{\xi}_t^2$.

The persistence of shocks to volatility measured by β is even higher in the ARMA models than for EGARCH. The parameter estimate of κ implied by the estimates of θ and β is given by $\hat{\kappa} = \hat{\beta}/\hat{\theta} - 1 = 0.0391$ for the ARMA model, and $\hat{\kappa^+} = \hat{\beta}\hat{\theta^+} - 1 = 0.0353$ and $\hat{\kappa^-} = \hat{\beta}\hat{\theta^-} - 1 = 0.0605$ for the ARMA-L model. The estimated volatility process $\hat{h}_t^*$ is adjusted by the estimated constant $\hat{C} = -\log(\hat{c})$, where $\hat{c}$ is the sample mean of $y_t^2/\exp(\hat{h}_t^*)$, see Section 3.

Table 4 also presents a goodness-of-fit criterion analogous to the pseudo-R^2 measure used in the simulations, but replacing the unknown h_t by the observed x_t:

$$\widetilde{R}_x^2 = 1 - \frac{\sum_{t=1}^{T}(x_t - \widehat{h}_t^*)^2}{\sum_{t=1}^{T}(x_t - \hat{\mu})^2} \tag{43}$$

where $\hat{\mu}$ is the sample mean of x_t. Note that due to the additional noise in x_t compared to the true but unknown h_t, this R^2 is smaller than that using the true h_t as target, see Andersen and Bollerslev (1998). We see that the fit of the ARMA and SV models are roughly similar, while the EGARCH model fit is clearly worse according to this criterion.

Figure 1 shows the nowcast of log volatility using the ARMA-L model (32), and the predicted log-volatility of the EGARCH model. The sample correlation between both volatility series is 91%. The predicted EGARCH volatility was higher after the October 1987 crash than after the Lehman crisis 2008, while the updated ARMA volatility was higher for the Lehman crisis. An explanation might be that the 1987 crash was mainly driven by an exceptionally severe one-day drop of returns, while absolute returns were exceptionally high during a longer time period around the Lehman crisis.

8 Conclusions

The proposed ARMA representation of log squared returns provides a simple method for estimating current volatility given the past and current information on the underlying returns. Our results suggest that it outperforms predictions of GARCH-type models, and similarly to stochastic volatility models while being easier to estimate.

We have proposed an important extension of the model to incorporate the so-called leverage effect. Many other extensions are possible and indeed object of future work. For example, it is straightforward to include a "GARCH-in-mean"-type risk premium in the conditional mean of returns, where the risk premium would depend on the current volatility, not on the predicted one. Second, multivariate extensions are possible. For example, one could use a factorization as in the orthogonal GARCH model of Alexander (2001). We believe that these are important topics of future research.

Appendix

A.1 Asymptotic distribution of estimators

Under our conditions, the maximum likelihood estimator $\hat{\gamma} = (\hat{\beta}, \hat{\theta})$ of the reduced form ARMA model (14) is consistent and asymptotically normal with asymptotic distribution given by

$$\sqrt{n}(\hat{\gamma} - \gamma) \to_d N\left(0, \frac{1-\beta\theta}{(\beta-\theta)^2}\begin{bmatrix} (1-\beta^2)(1-\beta\theta) & -(1-\theta^2)(1-\beta^2) \\ -(1-\theta^2)(1-\beta^2) & (1-\theta^2)(1-\beta\theta), \end{bmatrix}\right)$$

see e.g. Brockwell and Davis (1991). This gives directly the asymptotic variance of $\sqrt{n}(\hat{\beta} - \beta)$, while that of $\sqrt{n}(\hat{\kappa} - \kappa)$ is obtained via the delta method. Straightforward calculations yield

$$n\mathrm{Var}(\hat{\kappa}) \to \frac{(1-\beta\theta)^2}{(\beta-\theta)^2}\frac{1-\theta^2}{\theta^2}(1-\beta^2)\left(\frac{1}{1-\theta^2} + \frac{2\beta}{\theta(1-\theta\beta)} + \frac{\beta^2}{\theta^2(1-\beta^2)}\right).$$

A.2 Proof of Proposition 1

Denote $X_t = \sigma(x_t, x_{t-1}, x_{t-2}, \ldots)$ and let $h_{t|t-1} = \mathbb{E}[h_t|X_{t-1}]$, $h_{t|t} = \mathbb{E}[h_t|X_t]$, and $\mathcal{V}_t = \mathrm{Var}(h_t|X_{t-1})$. First we note that

$$\begin{aligned} u_t &= x_t - \mathbb{E}(x_t|X_{t-1}) = h_t - h_{t|t-1} + \varepsilon_t \\ &= \alpha^* + \eta_t + \beta(h_{t-1} - h_{t-1|t-1}) + \varepsilon_t \ . \end{aligned}$$

The estimator of the log-variance process is

$$h_{t|t} = h_{t|t-1} + \frac{\mathcal{V}_t + \mathbb{E}[(h_t - h_{t|t-1})\varepsilon_t]}{\mathcal{V}_t + 2\mathbb{E}[(h_t - h_{t|t-1})\varepsilon_t] + \sigma_\varepsilon^2}\ u_t \tag{44}$$

where $\mathbb{E}[(h_t - h_{t|t-1})\varepsilon_t] = \mathbb{E}(\eta_t\varepsilon_t)$, which follows from the Kalman filter with correlated measurement and transition errors, see e.g. Section 3.2.4 of Harvey (1989). For the case of correlated errors, equation (26) generalizes to

$$\begin{aligned} \mathrm{Var}(u_t) &= \mathcal{V}_t + \sigma_\varepsilon^2 + 2\mathbb{E}(\eta_t\varepsilon_t) \\ &= \frac{\beta}{\theta}[\sigma_\varepsilon^2 + \mathbb{E}(\eta_t\varepsilon_t)] \end{aligned}$$

which is the denominator in the second term of the right hand side of (44). The numerator is obtained by subtracting from this expression $\sigma_\varepsilon^2 + \mathbb{E}(\eta_t\varepsilon_t)$, and we obtain

$$\begin{aligned} \mathcal{V}_t + \mathbb{E}(\eta_t\varepsilon_t) &= \left(\frac{\beta}{\theta} - 1\right)[\sigma_\varepsilon^2 + \mathbb{E}(\eta_t\varepsilon_t)] \\ &= \kappa[\sigma_\varepsilon^2 + \mathbb{E}(\eta_t\varepsilon_t)] \end{aligned}$$

We finally obtain

$$\begin{aligned} h_{t|t} &= h_{t|t-1} + \frac{\mathcal{V}_t + \mathbb{E}(\eta_t \varepsilon_t)}{[\mathcal{V}_t + \mathbb{E}(\eta_t \varepsilon_t)] + [\sigma_\varepsilon^2 + \mathbb{E}(\eta_t \varepsilon_t)]}\, u_t \\ &= h_{t|t-1} + \frac{\kappa}{1+\kappa}\, u_t \ . \end{aligned}$$

Since $h_{t|t-1}$ is identical to the forecast of x_t based on the $x_{t-1}, x_{t-2}, \ldots$, the estimator $h_{t|t}$ is invariant to the covariance $\mathbb{E}(\eta_t \varepsilon_t)$. Therefore, h_t is identical to the estimator based on perfect correlation with $\varepsilon_t = \kappa \eta_t$ which is given in (20). □

References

Alexander, C.O. (2001), *Orthogonal GARCH*, in *Mastering Risk*, Volume II, edited by C.O. Alexander, pp. 21–38, Prentice Hall.

Andersen, T. and Bollerslev, T. (1998), Answering the skeptics: Yes, standard volatility models do provide accurate forecasts, *International Economic Review*, 39, 885-905.

Bollerslev, T. (1986), Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, 31, 307-327.

Brännäs, K. and De Gooijer, J.G. (1994), Autoregressive-asymmetric moving average models for business cycle data, *Journal of Forecasting* 13, 529-544.

Engle, R.F. (1982), Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, 50, 987-1007.

Francq, C. and Zakoian, J.-M. (2006), Linear-representation based estimation of stochastic volatility models, *Scandinavian Journal of Statistics* 33, 785-806.

Geweke, J. (1986). Modelling the Persistence of Conditional Variance: A Comment. *Econometric Reviews* 5, 5761.

Harvey, A. (1989), *Forecasting, structural time series models and the Kalman filter*, Cambridge University Press.

Harvey, A. and Koopman, S.J. (2000), Signal extraction and the formulation of unobserved components models, *Econometrics Journal* 3, 84-107.

Harvey, A. and Shephard, N. (1996), Estimation of an asymmetric stochastic volatility model for asset returns, *Journal of Business & Economic Statistics* 14, 429-434.

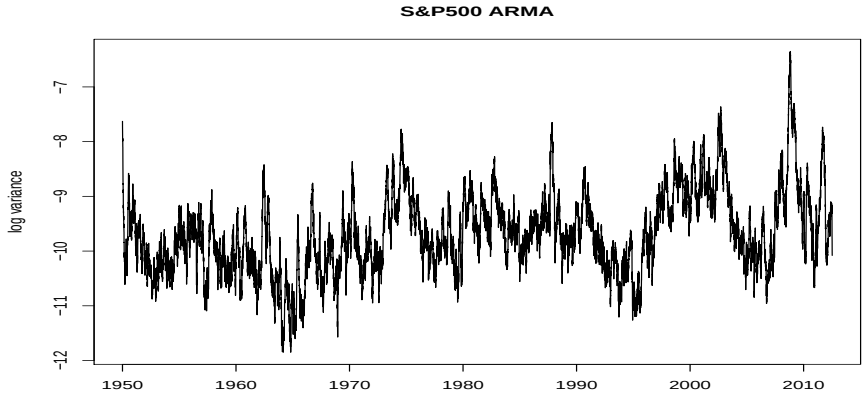
Nelson, D. (1991), Conditional heteoskedasticity in asset returns: A new approach. *Econometrica* 59, 347-370.

Pantula, S. (1986). Modelling the Persistence of Conditional Variance: A Comment. Econometric Reviews 5, 7173.

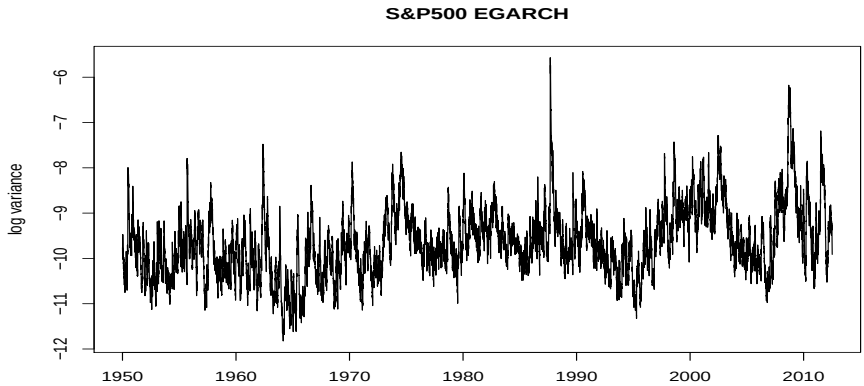
Politis, D.N. (2007), Model-free versus Model-based Volatility prediction, *Journal of Financial Econometrics* 5, 358–359.

Proietti, T. (2006) Trend-Cycle Decompositions with Correlated Components, *Econometric Reviews*, 25, 61-84.

Sandmann, G. and Koopman, S.J. (1998). Estimation of stochastic volatility models via Monte Carlo maximum likelihood. *Journal of Econometrics*, 87, 271-301.



(a) Nowcast log volatility using the ARMA-L model



(b) Predicted log volatility using EGARCH

Figure 1: Volatility estimates for daily S&P 500 returns.

Table 1: Performance under the SV model (29)

		EGARCH		SV		ARMA	
β	CV	R^2	s.d.	R^2	s.d.	R^2	s.d.
				$n = 500$			
0.9	1	0.360	(0.143)	0.528	(0.115)	0.588	(0.095)
0.98	1	0.746	(0.165)	0.883	(0.066)	0.874	(0.079)
0.9	10	0.271	(0.199)	0.626	(0.071)	0.641	(0.061)
0.98	10	0.590	(0.321)	0.863	(0.072)	0.859	(0.074)
				$n = 2000$			
0.9	1	0.314	(0.099)	0.392	(0.078)	0.438	(0.047)
0.98	1	0.737	(0.097)	0.776	(0.061)	0.771	(0.064)
0.9	10	0.297	(0.149)	0.579	(0.048)	0.592	(0.044)
0.98	10	0.679	(0.232)	0.807	(0.051)	0.806	(0.051)

Note: Pseudo-R^2 of fitted volatility models for h_t compared with true, simulated stochastic volatility series. The standard deviation of the sample R^2 is indicated as s.d. (in parentheses). The coefficient of variation is denoted by $CV = \mathbb{V}\text{ar}[\exp(h_t)]/\mathbb{E}[\exp(h_t)]^2$.

Table 2: Performance under the ARMA model (30)

		EGARCH		SV		ARMA	
β	CV	R^2	s.d.	R^2	s.d.	R^2	s.d.
				$n = 500$			
0.9	1	0.453	(0.170)	0.843	(0.092)	0.917	(0.091)
0.98	1	0.794	(0.176)	0.959	(0.108)	0.953	(0.120)
0.9	10	0.263	(0.222)	0.956	(0.034)	0.971	(0.037)
0.98	10	0.623	(0.337)	0.982	(0.054)	0.978	(0.057)
				$n = 2000$			
0.9	1	0.430	(0.151)	0.857	(0.124)	0.958	(0.049)
0.98	1	0.808	(0.141)	0.976	(0.101)	0.969	(0.104)
0.9	10	0.317	(0.181)	0.964	(0.097)	0.981	(0.099)
0.98	10	0.673	(0.299)	0.987	(0.055)	0.985	(0.059)

Note: Pseudo-R^2 of fitted volatility models for h_t compared with true, simulated ARMA series. Remaining notes as in Table 1.

Table 3: Performance under the EGARCH model (31)

		EGARCH		SV		ARMA	
β	CV	R^2	s.d.	R^2	s.d.	R^2	s.d.
				$n = 500$			
0.9	1	0.865	(0.035)	0.810	(0.041)	0.816	(0.058)
0.98	1	0.956	(0.013)	0.965	(0.014)	0.958	(0.030)
0.9	10	0.854	(0.019)	0.827	(0.031)	0.808	(0.048)
0.98	10	0.713	(0.129)	0.966	(0.012)	0.964	(0.014)
				$n = 2000$			
0.9	1	0.880	(0.022)	0.763	(0.022)	0.743	(0.042)
0.98	1	0.912	(0.010)	0.943	(0.011)	0.937	(0.024)
0.9	10	0.837	(0.010)	0.794	(0.018)	0.769	(0.028)
0.98	10	0.786	(0.049)	0.947	(0.010)	0.945	(0.011)

Note: Pseudo-R^2 of fitted volatility models for h_t compared with true, simulated EGARCH series. Remaining notes as in Table 1.

Table 4: Parameter estimates of alternative volatility models

	EGARCH	SV	ARMA	ARMA-L
α	-0.2666 (0.0100)	-0.0755 (0.0204)	-0.0822 (0.0166)	-0.0792 (0.0148)
β	0.9839 (0.0009)	0.9932 (0.0018)	0.9926 (0.0015)	0.9930 (0.0013)
γ	0.1475 (0.0033)			
θ	-0.0647 (0.0019)		0.9552 (0.0038)	
θ^+				0.9590 (0.0036)
θ^-				0.9359 (0.0088)
σ_η^2		0.0097 (0.0022)		
σ_ε^2		5.3156 (0.0950)		
$\tilde{R}_x^2$	0.1020	0.1512	0.1536	0.1562

Note: The $\tilde{R}_x^2$ criterion is given by (43), where h_t^* is either the predicted volatility using EGARCH, the updated volatility $h_{t|t}$ using SV, or the estimated h_t^* using the ARMA model. ARMA-L is the asymmetric ARMA model of section 6. Standard errors are reported in parentheses.