

QF5206 TOPICS IN QUANTITATIVE FINANCE II

Group Project

Modeling and Forecasting Stock Return Volatility Using a Random Level Shift Model

Yang K.Lu Pierre Perron

Yu Lu A0075290

Lu Hongda A0075446

Meng Shuangshuang A0078753

4/11/2011

Contents

1 Introduction 2

2 The model 3

3 Estimation Method 4

4 Empirical results for returns on stock market indices 8

4.1 Estimation results **8**

4.2 The effect of level shifts on long-memory and conditional heteroskedasticity **9**

5.Forecasting ..... 错误!未定义书签。

5.1 Construction of the forecasts **12**

5.2 Forecasting Comparisons **13**

6. Conclusions 14

References 15

1 Introduction

More recently, attempts have been made to distinguish between stationary noise plus level shift and long-memory models: Stărică and Granger (2005) presented evidence that log-absolute returns of the S&P 500 index is an i.i.d series affected by occasional shift in the unconditional variance and show that this specification has better forecasting performance than the more traditional GARCH (1, 1) model and its fractionally integrated counterpart.

Our approach extends the work of Stărică and Granger (2005) by directly estimating a structural model. We adopt a specification for which the series of interest is the sum of a short-memory process and a jump or level shift component. For the latter, we specify a simple mixture model such that the component is the cumulative sum of a process which is 0 with some probability $(1 - \alpha)$ and is a random variable with probability α. To estimate such a model, we transform it into a linear state space form with innovations having a mixture of two normal distributions and adopt an algorithm similar to the one used by Perron and Wada (2009) and Wada and Perron (2007).

We restrict the variance of one of the two normal distributions to be zero, allowing us to achieve a simple but efficient algorithm. We apply this random level shift model to the logarithm of absolute returns for the following stock market return indices: S&P 500 (1962/07/03 to 2004/03/25; 10504 observations), AMEX (1962/07/03 to 2006/12/31; 11201 observations), Dow Jones (1957/03/04 to 2002/10/30; 11534 observations) and NASDAQ (1972/12/15 to 2006/12/31; 8592 observations).

Our point estimates imply few level shifts for all series. But once these are taken into account, there is little evidence of serial correlation in the remaining noise and, hence, no evidence of long-memory. Furthermore, once the estimated shifts are introduced to a standard GARCH model, any evidence of GARCH effects disappears. We also produce recursive out-of-sample forecasts of squared returns. In most cases, our simple random level shifts model clearly outperforms a standard GARCH (1, 1)

model and, in many cases, it also provides better forecasts than a fractionally integrated GARCH model.

2 The model

In this project, we apply regime level shift model to log-absolute returns is given by:

$$y_t = a + \tau_t + c_t \quad (1)$$

Where α is a constant, τ_t is the random level shift component and c_t is a short-memory process, which is used to capture the remaining noise of the model. The level shift component is specified by $\tau_t = \tau_{t-1} + \delta_t$, where $\delta_t = \pi_t \eta_t$. Here, π_t is a binomial variable that takes value 1 with probability α and value 0 with probability (1 −α). If it takes value 1, then a random level shift η_t occurs, specified by $\eta_t \sim i.i.d.N(0,\ \sigma_\eta^2)$. In its most general form, we should make the following assumptions. Firstly, we shall for simplicity assume that$c_t = \varphi c_t + e_t$, secondly, we also assume that the componentsτ_t, $\eta_t$and c_t are mutually independent. Thirdly, the normality assumption for e_t is needed to construct the likelihood function.

In order to estimate the model, we shall embed it in a state space framework involving errors that have a mixture of two normal distributions. The level shift component τ_t can be specified as a random walk process with innovations distributed according to a mixture of two normally distributed processes:

$$\tau_t = \tau_{t-1} + \delta_t$$

$$\delta_t = \pi_t \eta_{1t} + (1 - \pi_t)\eta_{2t}$$

Where $\eta_{it} \sim i.i.d.N(0, \sigma_{\eta i}^2)$, and $\pi_t$is a Bernoulli random variable that takes value one with probability α and value 0 with probability$1 - \alpha$. After specifying $\sigma_{\eta 1}^2 = \sigma_\eta^2$ and $\sigma_{\eta 2}^2 = 0$, we can cover equation (1). We next specify the model in terms of first-differences of the data:

$$\Delta y_t = \tau_t - \tau_{t-1} + c_t - c_{t-1} = \delta_t + c_t - c_{t-1}$$

Where $\delta_t = \pi_t \eta_{1t} + (1 - \pi_t)\eta_{2t}$.We then have the following state space form:

$$\Delta y_t = c_t - c_{t-1} + \delta_t$$

$$c_t = \varphi c_{t-1} + e_t$$

or more generally,

$$\Delta y_t = HX_t + \delta_t,$$

$$X_t = FX_{t-1} + U_t,$$

where in the case of an AR(p) process, $X_t = \left[c_t, c_{t-1}, \dots \dots, c_{t-p-1}\right]'$

$$F = \begin{bmatrix} \phi_1 & \phi_2 & \phi_p \\ 1 & & \\ & 1 & \\ \vdots & \ddots & \vdots \\ & \dots & 1 \end{bmatrix}$$

$H = [1, -1, 0, \dots ., 0]$ And U_t is a p-dimensional normally distributed random vector with mean zero and covariance matrix

$$Q = \begin{pmatrix} \sigma_e^2 & 0_{1\times(p-1)} \\ 0_{(p-1)\times 1} & 0_{(p-1)(p-1)} \end{pmatrix}$$

3 Estimation Method

We previously described the model which is a special case of the class of models considered by Wada and Perron (2007). They modeled the trend-cycle decomposition of a macroeconomic time series with shocks affecting the level, slope and cyclical components of the time series. Here, we only focus on shocks affecting the level of the series and simply add the restriction that the variance of one component of the mixture of normal distributions is zero. Then, we briefly describe the method, and more details can be found in Wada and Perron (2007). We should note that despite their fundamental differences, our model and the Markov regime switching models are really similar in estimation methodology. The basic ingredient for estimation is the augmentation of the states by the realizations of the mixture at time t so that the Kalman filter can be used to generate the likelihood function, conditional on the realizations of the states. The latent states are eliminated from

the final likelihood expression by summing over all possible state realizations. As we shall show, our model takes a structure that is similar to a version of the Markov regime switching model Let $Yt = (\Delta y1, \ldots, \Delta yt)$ be the vector of data available up to time t and denote the vector of parameters by $\theta = \left[\sigma_\eta^2, \alpha, \sigma_e^2, \phi_1, \ldots \phi_q\right]$. To illustrate the similarities we adopt the notations in Hamilton (1994), where 1 represents a (4 × 1) vector of ones, the symbol Θ denotes element-by-element multiplication, $\hat{\xi}_{t|t-1} = vec(\bar{\xi}_{t|t-1})$ With (i, j^{th}) element $\bar{\xi}_{t|t-1} = P(s_{t-1} = j, s_t = i|Y_{t-1};\ \theta)$ and $\omega_t = vec(\overline{\omega_t})$ with the $(\mathrm{i,j})^{th}$ element $\overline{\omega_t} = f(\Delta y_t \left| s_{t-1} = i, s_t = j|Y_{t-1};\ \theta\right.)$, for $i, j \in \{1,2\} s_t = 1(resp, 0) when\, \pi_t = 1(resp, 0)$ i.e., a level shift occurs (resp., does not occur). The log likelihood function is:

$$ln(L) = \sum_{t=1}^{T} lnf(\Delta y_t|Y_{t-1};\theta) \quad (2)$$

Where

$$f(\Delta y_t|Y_{t-1};\theta) = \sum_{i=1}^{2}\sum_{j=1}^{2} f(\Delta y_t|s_{t-1} = i, s_t = j, Y_{t-1};\theta)\Pr(s_{t-1} = i, s_t = j, Y_{t-1};\theta)$$

$$= 1'(\hat{\xi}_{t|t-1}\Theta\overline{\omega}_t) \quad (3)$$

We first focus on the evolution of $\hat{\xi}_{t|t-1}$ Applying rules for conditional probabilities, Bayes' rule and the independence of s_t with past realizations, we have

$$\tilde{\xi}^{ij}_{t|t-1} \equiv \Pr(s_{t-1} = i, s_t = j|Y_{t-1};\theta)$$

$$= \Pr(s_t = j)\sum_{k=1}^{2} Pr(s_{t-2} = k, s_{t-1} = i|Y_{t-1};\theta)$$

and $\tilde{\xi}^{ki}_{t|t-1} \equiv \Pr(s_{t-2} = k, s_{t-1} = i|Y_{t-1};\theta)$

$$= \frac{f(\Delta y_{t-1}|s_{t-2} = k, s_{t-1} = i|Y_{t-2};\theta)\Pr(s_{t-2} = k, s_{t-1} = i|Y_{t-2};\theta)}{f(\Delta y_{t-1}|Y_{t-2};\theta)}$$

Therefore, the evolution of $\hat{\xi}_{t|t-1}$ is given by:

$$\begin{bmatrix} \tilde{\xi}^{11}_{t+1|t} \\ \tilde{\xi}^{21}_{t+1|t} \\ \tilde{\xi}^{12}_{t+1|t} \\ \tilde{\xi}^{22}_{t+1|t} \end{bmatrix} \begin{bmatrix} \alpha(\tilde{\xi}^{11}_{t|t} + \tilde{\xi}^{21}_{t|t}) \\ \alpha(\tilde{\xi}^{12}_{t|t} + \tilde{\xi}^{22}_{t|t}) \\ (1-\alpha)(\tilde{\xi}^{11}_{t|t} + \tilde{\xi}^{21}_{t|t}) \\ (1-\alpha)(\tilde{\xi}^{12}_{t|t} + \tilde{\xi}^{22}_{t|t}) \end{bmatrix} \begin{bmatrix} \alpha & \alpha & 0 & 0 \\ 0 & 0 & \alpha & \alpha \\ (1-\alpha) & (1-\alpha) & 0 & 0 \\ 0 & 0 & (1-\alpha) & (1-\alpha) \end{bmatrix} \begin{bmatrix} \tilde{\xi}^{11}_{t|t} \\ \tilde{\xi}^{21}_{t|t} \\ \tilde{\xi}^{12}_{t|t} \\ \tilde{\xi}^{22}_{t|t} \end{bmatrix}$$

Or more compactly by: $\hat{\xi}_{t+1|t} = \pi\hat{\xi}_{t|t}$ (4)

With $\hat{\xi}_{t|t} = \frac{\hat{\xi}_{t|t-1}\Theta\omega_t}{1'(\hat{\xi}_{t|t-1}\Theta\omega_t)}$

The conditional likelihood for Δy_t is the following Normal density:

$$\widetilde{\omega}_t^{ij} = f(\Delta y_t | s_{t-1} = i, s_t = j, Y_{t-1}; \theta) = \frac{1}{\sqrt{2\pi}} \left| f_t^{ij} \right|^{-\frac{1}{2}} exp\left\{ -\frac{v_t^{ij\prime} f_t^{ij-1} v_t^{ij}}{2} \right\} \quad (5)$$

Where $v_t^{i,j} = \Delta y_t - \Delta y_{t|t-1}^i$, the prediction error and $f_t^{i,j} = E(v_t^{i,j} v_t^{i,j\prime})$ is the prediction error variance. Note that $\Delta y_{t|t-1}^i = E[\Delta y_t s_{t-1} = i | Y_{t-1};\ \theta]$, not depend on the state j at time t because we are conditioning on time (t−1) information. However, Δy_t does depend on $s_t = j$. So the prediction error and its variance depend on both i and j. The best forecast for the state variable and its associated variance conditional on past information and $s_{t-1} = i$. Are

$$X_{t|t-1}^i = FX_{t-1|t-1}^i$$

$$P_{t|t-1}^i = FP_{t-1|t-1}^i + Q \quad (6)$$

We have the measurement equation $\Delta y_t = HX_t + \delta_t$ where the measurement error δ_t has mean zero and a variance which can take two possible var $(\delta_t) = (\sigma_\eta^2, p = \alpha; 0, p = 1 - \alpha)$ hence, the prediction error and associated variance are:

$v_t^{ij} = \Delta y_t - HX_{t|t-1}^i \qquad f_t^{i,j} = HP_{t|t-1}^i H' + R_j,$

Applying standard updating formulae, we have (given $s_t = j, \text{and}, s_{t-1} = i$)

$$X_{t|t}^{ij} = X_{t|t-1}^i = + P_{t|t-1}^i H'(HP_{t|t-1}^i H' + R_j)^{-1}(\Delta y_t - H\ X_{t|t-1}^i) \quad (7)$$

$$P_{t|t}^{ii} = P_{t|t-1}^i - P_{t|t-1}^i H'(HP_{t|t-1}^i H' + R_j)^{-1} HP_{t|t-1}^i$$

To reduce the dimension of the estimation problem, we adopt the re-collapsing procedure suggested by Harrison and Stevens (1976), given by

$$X_{t|t}^j = \frac{\sum_{i=1}^2 \Pr(s_{t-1}=i, s_t=j | Y_t; \theta) X_{t|t}^{ij}}{\Pr(s_t=j | Y_t; \theta)} = \frac{\sum_{i=1}^2 \tilde{\xi}_{t|t}^{ij} X_{t|t}^{ij}}{\sum_{i=1}^2 \tilde{\xi}_{t|t}^{ij}} \quad (8)$$

$$P_{t|t}^{j} = \frac{\sum_{i=1}^{2} \Pr(s_{t-1} = i, s_t = j|Y_t;\theta)\left[P_{t|t}^{ij} + \left(X_{t|t}^{j} - X_{t|t}^{ij}\right)\left(X_{t|t}^{j} - X_{t|t}^{ij}\right)\right]}{\Pr(s_t = j|Y_t;\theta)}$$

$$= \frac{\sum_{i=1}^{2} \tilde{\xi}_{t|t}^{ij}[P_{t|t}^{ij} + (X_{t|t}^{j} - X_{t|t}^{ij})(X_{t|t}^{j} - X_{t|t}^{ij})]}{\sum_{i=1}^{2} \tilde{\xi}_{t|t}^{ij}}$$

By doing so, we make $\omega_t^{i,j}$ unaffected by the history of states before time t-1.

If we define$S_t \equiv (s_t, s_{t-1})$, we then have four possible states:$\mathrm{S_t} = 1, \text{when } (\mathrm{s_t} = 1, \mathrm{s_{t-1}} = 1)$ $S_t = 2, \text{when } (\mathrm{s_t} = 1, \mathrm{s_{t-1}} = 2)$ and $\mathrm{S_t} = 3, \text{when } (\mathrm{s_t} = 2, \mathrm{s_{t-1}} = 1), \mathrm{S_t} = 4, \text{when } (\mathrm{s_t} = 2, \mathrm{s_{t-1}} = 2)$ with the transition matrix Π as defined in (4) the vector of conditional densities $\omega_t = (\omega_t^l \dots \omega_t^4)$, thus have a more compact representation given by:

$$\omega_t^l = f(\Delta y_t|s_t = l, Y_{t-1};\theta) = \frac{1}{\sqrt{2\pi}}\left|f_t^l\right|^{-\frac{1}{2}} exp\left\{-\frac{v_t^{l\prime} f_t^{l^{-1}} v_t^l}{2}\right\} \qquad (9)$$

where v_t^l and f_t^l are as defined in (5) with the values of $\mathrm{s_t}$ and $\mathrm{s_{t-1}}$ corresponding to $s_t = i$ This definition of ω_t together with $\hat{\xi}_{t|t}$ and $\hat{\xi}_{t+1|t}$, the collection of conditional probabilities $Pr(s_t = l|Y_t\,;\,\theta)$ *for l=1,…,4* and it's the one-period ahead forecasts, evolving as in (4), give us the same structure as a version of the Markov regime switching model used by Hamilton (1994). However, there are two extra complexities here. Firstly, the mean and variance in the conditional density function are nonlinear functions of the fundamental parameters θ and past realizations *{*$\Delta y_{t-j}, j \geq 1$ *}.* This non-linearity and time dependence complicates the maximization of the log-likelihood function since we cannot separate out some elements of θ in the first order conditions. Accordingly, the standard EM algorithm does not apply. Secondly, the conditional probability of being in a given regime $\hat{\xi}_{\mathrm{t|t}}$ is not separable from the

Conditional densities ω_t since $\tilde{\xi}_{t|t}^{i,j}$ enters in its construction (see (8)).

Next, we apply our model and estimation method to the returns of four major market indices: the S&P 500, AMEX, Dow Jones and NASDAQ. The daily returns are computed by first differencing the logarithm of the index price series $r_t = \ln(P_t) -$

$\ln(P_{t-1})$. And to avoid the possible extreme negative values we bound absolute returns away from zero by adding a small constant, i.e., we use log ($|r_t|$ + 0.001). For specifying the short-memory componentc_t, we start by setting $c_t = e_t$ following Stărică and Granger (2005) who report that the short-memory component in such series is just white noise. As a robustness check, we also report the estimates for the specification that c_t follows an AR (1) process $c_t = \varphi c_t + e_t$ since all components of the state vector are stationary, we can initialize the state vector and its covariance matrix by their unconditional expected values, i.e.$X_{(0|0)} = (0,0)'$ and

$$P_{(0|0)} = \begin{bmatrix} \sigma_e^2 & 0 \\ 0 & 0 \end{bmatrix}$$

We obtain estimates by directly maximizing the likelihood function $\ln(L) = \sum_{t=1}^{T} lnf(\Delta y_t|Y_{t-1};\theta)$. In order to avoid the problem of local maxima, we re-estimate with different initial values of θand pick the estimates associated with the largest likelihood value upon convergence.

4 Empirical results for returns on stock market indices

4.1 Estimation results

The parameter estimates are presented in Table 1 for both the cases in which the short memory component c_t is specified to be white noise and an AR (1) process. To assess the relative importance of the two components to the total variation of the series, we also report the standard deviation of the original series$log(|r_t|)$. The estimates reported are the standard deviation of level shift component $\sigma_\eta$the probability of a shift α, the standard deviation of the stationary component σ_e and the autoregressive coefficient φ when considering the AR(1) specification for c_t.

These results exhibit obvious two features. Firstly, for the short memory component, when considering the AR (1) specification forc_t, the estimate of φ is very small and close to zero. Also, in all cases, adopting either an AR (1) or white noise specification for c_t yields very similar results. So in subsequent sections we shall only consider results based on the white noise specification for the short-memory component. The

second apparent feature is the fact that in all cases, the probability of level shifts is very small. Given that the shifts occur so infrequently, the noise component accounts for the bulk of total variation.

4.2 The effect of level shifts on long-memory and conditional heteroskedasticity

Given the features documented above, we will then determine whether the level shifts are important to dictating the overall behavior of the series. We shall address this issue by investigating whether the shifts can explain a) the well-documented feature of long-memory and b) the presence of conditional heteroskedasticity. Here we only illustrate the case of S&P 500.

(a) The first issue we handle is whether the level shift can account for the long-memory feature of the series $\log|r_t|$. To do so, we plot the autocorrelation functions (up to 300 lags) of the original series $\log|r_t|$ of S&P 500 and its short-memory component, c_t obtained by subtracting the fitted trend from$\log|r_t|$. The results are presented in Figure 2. One can see that for all cases, the log-absolute returns clearly display an autocorrelation function that resembles that of a long-memory process: it decays very slowly and the values remain important even at lag 300. Additionally from the ACF of residuals, for all practical purposes, we can view the short-memory component as being nearly white noise.

(b) Furthermore, we investigate the effects of level shift on conditional heteroskedasticity. Here we use the standard GARCH(1,1) model and in order to demonstrate that the stock returns exhibit conditional heteroskedasticity, we also use the popular components GARCH model:

GARCH (1, 1) model with Student-t errors given by, for the demeaned returns process $\tilde{r}_t$

$$\tilde{r}_t = \sigma_t \varepsilon_t$$

$$\sigma_t{}^2 = \mu + \beta_r \tilde{r}_{t-1}^2 + \beta_\sigma \sigma_{t-1}^2$$

Where ε_t $\sim$*i.i.d.* Student-t distributed with mean 0 and variance 1 and the parameters of interest β_r and β_σ measures the extent of conditional heteroskedasticity present in the data.

CGARCH model:

$$\tilde{r_t} = \sigma_t \varepsilon_t$$

$$n_t = \mu + \rho(n_{t-1} - \mu) + \varphi(\tilde{r}_{t-1}^2 - \sigma_{t-1}^2) + \sum_{i=2}^{m+1} D_{i,t}\gamma_i$$

$$({\sigma_t}^2 - n_t) = \beta_r(\tilde{r}_{t-1}^2 - n_{t-1}) + \beta_\sigma(\sigma_{t-1}^2 - n_{t-1})$$

Where $D_{i,t} = 1$ if t is in regime i, i.e., $t \in \{T_{i-1} + 1, \dots, T_i\}$and 0 otherwise, with $T_i\,(\mathrm{i} = 1 \dots \mathrm{m})$ being the break dates documented in Figure 1 (again $T_0 = 0$ and $T_{m+1} = T$). The coefficients γ_i, which index the magnitude of the shifts, are treated as unknown and are estimated with the remaining parameters, while the number of breaks is obtained from the point estimate of α.

For no level shifts circumstances, using the standard GARCH (1, 1) model, both estimates are highly significant for all series. In particular, the value of β_σ is quite high with estimated value 0.938. We also estimate the components GARCH model without level shifts and the results are similar for the estimates of β_r and β_σ. Additionally, the estimates of ρ is very close to one: 0.9996, which is due to the fact that the model attempts to capture features akin to long memory so that a value of ρ close to one implies highly persistent shocks.

On the other hand, once the level shifts are accounted for, the picture is completely different. Neither the estimates of β_σ or β_r is significant. Therefore, introducing few level shifts imply a markedly different interpretation of the data. These results indicate that level shifts in log-absolute or squared returns account for nearly all of the documented evidence of conditional heteroskedasticity in these stock returns series.

Thus, in summary of the results of part 4, we conclude that the level shifts model with white noise errors appears to provide an accurate description of the data. Also,

the level shift component is an important feature that explains both the long-memory and conditional heteroskedasticity features generally perceived as stylized facts.

5. Forecasting

We follow Stărică and Granger (2005) to design the forecasting experiment in order to evaluate the performance of the level shift model in forecasting volatility relative to the current leading model, the GARCH (1, 1) model. We view the short memory component as white noise and the log-absolute returns are proxied by squared returns. Considering the smoothed estimates of the level shift components are erratic and not in accord with the postulated model, we shall turn to another method. We initially forecast at observations 2,000 and then re-estimate the models every 20 days, at which point forecasts of up to 200 days are constructed. Since the realized squared returns as proxy for volatility are quite noisy, it's necessary to reduce the effect of sampling variability. We also follow Stărică and Granger (2005) in the construction of a metric to get the relative performance. Let $\widehat{\sigma}_{t\,n}^{2}$ be a p-step ahead forecast of σ_{t+n}^{2}, the variance of returns r_t at time t + p, proxied by the squared demeaned returns. Let n be the number of forecasts produced, and then construct the estimated MSE as following:

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(\overline{r}_{t,p}^{2} - \sigma_{t,p}^{2})^{2} \qquad (12)$$

Where $\sigma_{t,p}^{2} = \sum_{k=1}^{p}\tilde{\sigma}_{t+k}^{2}$ and $\bar{r}_{t,p}^{2}$=$\sum_{k=1}^{p} r_{t+k}^{2}$ is the realized volatility over [t+1, t+p]. This estimate of the MSE is better than the simpler version $\sum_{\mathrm{t=1}}^{\mathrm{n}}(\mathrm{r}_{\mathrm{t+p}}^{2} - \sigma_{t,p}^{2})^{2}$, because the latter uses a poor measure of realized return volatility (see Anderson and Bollerslev, 1998). The squared returns are not quite noisy because some of the idiosyncratic noise in the high frequency is canceled out by averaging. Throughout, the relative forecasting performance of the two models is evaluated by the ratio of their MSEs as defined above.

5.1 Construction of the forecasts

The level shift has an important feature that explains the presence of both the long-memory and conditional heteroskedasticity features and it can also provide important improvements in forecasting volatility when using squared returns as a proxy. We construct the forecasts for each method in the following way. We begin with the level shift model applied to log-absolute returns of which the short-memory component is simply white noise. Since our model and estimates pertain to log-absolute returns, we need to take an appropriate transformation to generate forecasts of the variance of returns. Recall that our model specifies the following process for log absolute returns: $\log(|\mathrm{r_t}| + \mathrm{C}) = \tau_t + \mathrm{C_t}$

Where a positive constant C is added to bound returns away from zero. Then we get the following model for returns: $|\mathrm{r_t}| + \mathrm{C} = \mathrm{h_t^{1/2}}\tilde{\epsilon}_t$ where $\mathrm{h_t} = \mathrm{e}^{2\tau_t}\mathrm{E}(\mathrm{e}^{2c_t})$ and $\tilde{\epsilon}_t = \mathrm{e}^{c_t}/[\mathrm{E}(\mathrm{e}^{2c_t})]^{1/2}$, so that $\mathrm{E}\ \tilde{\epsilon}_t^{\,2}$ =1. $\tilde{\epsilon}_t$ is independent of $\mathrm{h_t}$and $2\mathrm{c_t} \sim \mathrm{i.i.d}, \text{hence } E(e^{2c_t}) = e^{0+4\sigma_e{}^2/2} = e^{2\sigma_e{}^2}$. Since there is considerably uncertain about the timing and magnitudes of the rare level shifts, we ignore them when forecasting. We then have

$$E_t(|r_t| + C)^2 = E_t h_{t+k} = exp(\,2\tau_t + 2\sigma_e{}^2)$$

So that the k-period ahead forecast of the squared returns $\mathrm{r_{t+k}}^2$ is

$$E_t r_{t+k}^2 = exp(2\tau_t + 2\sigma_e{}^2) - 2CE_t|r_{t+k}| - C^2$$

Where $E_t\,|r_{t+k}\,| = E_t exp\,(\tau_t + C_{t+k})$ - C=$exp\,(\tau_t + 0.5\sigma_e{}^2) - C$

For the Student-t GARCH (1, 1) model, it can be written as:

$$\widetilde{R_t} = r_t - u = \sigma_t \varepsilon_t \qquad (13)$$

$$\sigma_t{}^2 = \alpha_1 + \alpha_2 \tilde{r}_{t-1} + \alpha_3 \sigma_{t-1}{}^2 \qquad (14)$$

Where the innovation ε_t is *i.i.d* N (0, 1) and $\widetilde{\mathrm{R_t}}$ is demeaned returns. Take the following transformation:

$$\tilde{r}_t^2 = \alpha_1 + (\alpha_2 + \alpha_3)\,\tilde{r}_{t-1}^2 + \omega_t - \alpha_3\omega_{t-1} \qquad (15)$$

Where $\omega_t = \tilde{\mathrm{r}}_t^2 - \sigma_t{}^2$ is the forecast error for the forecast of $\tilde{\mathrm{r}}_t^2$, thus ω_t is white noise process that is fundamental to$\tilde{\mathrm{r}}_t^2$. Assuming α2 + α3 < 1, by $\mathrm{E}(\tilde{\mathrm{r}}_t^2) = \sigma^2 = \alpha_1/(1 - \alpha_2 - \alpha_3)$, we can compute the unconditional mean of $\tilde{\mathrm{r}}_t^2$ easily from (15). The

recursive form for the squared demeaned returns

$E_t\tilde{r}^2_{t+k}=\sigma^2+(\alpha_2+\alpha_3)^{k-1}(E_t\tilde{r}^2_{t+1}-\sigma^2)=\sigma^2+(\alpha_2+\alpha_3)^{k-1}({\sigma_{t+1}}^2-\sigma^2)$

${\sigma_{t+1}}^2=\alpha_1+\alpha_2\tilde{r}_t^2+\alpha_3{\sigma_t}^2$. Due to the fact that the time variation in the conditional mean of returns is quantitatively negligible, we make the following adjustment, adding μ^2 to $\mathrm{E_t\tilde{r}^2_{t+k}}$ and get the forecasts for the squared returns:

$$E_t r^2_{t+k}=E_t\tilde{r}^2_{t+k}+(E_t r_{t+k})^2\approx E_t\tilde{r}^2_{t+k}+\mu^2$$

5.2 Forecasting Comparisons

With the above construction of two models, we should make some comparisons of the forecasting results to highlight the performance of the level shift model. We make an unfair but instructive comparison. For the GARCH (1, 1) Model, we begin forecast at observations 2,000 and re-estimate every 20 observations. However, for the random level shifts model, we use the fitted means obtained from the full sample. The fitted level shift process depicted in Figure 1 is obtained using the Bai and Perron (2003) algorithm. The probability α of a level shift at each period implies the number of breaks for the full sample. The results are presented in Figure 3 (level shifts versus GARCH). It is apparent that the random level shift model performs much better forecasting at all horizons and for almost all series. Some exceptions appear in very short horizons for the AMEX and NASDAQ indices, in which case the GARCH has a slight advantage. Although it is an unfair comparison, we can obtain much better forecasts from the level shift model than the other models, if we have a precise estimate of the mean of log-absolute returns at a given date.

Thus how to obtain a good estimate of the current mean of a regime at a given date becomes the major issue, at which the forecasts are made without using information after that date. This turns out to be a delicate issue. Though we get the filtered estimates of τ_t, the level shift component by the Kalman filter algorithm, they are too volatile to be useful. There is no choice but to repeat the whole process we did previously. The problem here is that if a possible change at each day is allowed, as in the theoretical model, the fitted values often indicate that a change occurred at the

end of the sample when, ex-post, no such long lasting change have occurred. In light of these issues, we resort to a procedure, the CUSUM, which can effectively indicate the date when a forecast failure occurs and is, accordingly, the best suited for forecasting (see, e.g., Pesaran and Timmerman, 1999)

Overall, we get the encouraging results showing that the random level shift model is a serious contender as a forecasting model and is substantially accurate in forecasting. That is, the level shift model does provide a better description of the data in sample, and improve forecasting performance. Hence, level shifts appear to be genuinely present in the data and not simply a modeling convenience that allows for a better in-sample fit. This contrasts with the common perception that structural change models are not useful for forecasting.

6. Conclusions

In this paper, firstly, a simple estimation method was proposed for level shift model of a series which is consisting of a random level shift and a short-memory component. The model yields impressive results when applied to log-absolute returns of the S&P 500, AMEX, Dow Jones and NASDAQ indices. The short-memory component in such series is set to be white noise with which the level shift model appears to provide an accurate description of the data. The level shift component can explain the well-documented feature of long-memory and the presence of heteroskedasticity. When taking into account the level shifts, the evidence in favor of long-memory and conditional heteroskedasticity disappears. Important improvements in forecasting volatility has been provided by the model using squared returns as a proxy. These are especially noticeable with well estimated mean of the regime in effect. However, it is difficult to obtain precise estimates of this mean in real time. Many cases have proved that forecasting improvements can still be obtained from our model by using a backward recursive CUSUM test to determine the length of the last regime.

References

Andersen, T.G., and T. Bollerslev (1998), “Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts,” International Economic Review 39, 885-905.

Andersen, T.G., and T. Bollerslev (1998), “Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts,” International Economic Review 39, 885-905.

Anderson, T.G., T. Bollerslev, F.X. Diebold, and P. Labys (2003): “Modeling and Forecasting Realized Volatility,” Econometric, 71, 579-626.

Bollerslev, T. (1986), “Generalized Autoregressive Conditional Heteroskedasticity,” Journal of Econometrics 31, 307-327.234.

Mikosch, T., and C. Stărică (2005), “Changes of Structure in Financial Time Series and the GARCH Model,” Unpublished manuscript.

Stărică, C., and C. Granger (2005), “Nonstationarities in Stock Returns,” The Review of Economics and Statistics 87, 503-522.

Table 1: Maximum Likelihood Estimates

	σ_η	α	σ_e	ϕ
S&P 500	0.7524	0.00198	0.73995	
SD(0.7221)	0.74290	0.00082	0.73920	0.00030
AMEX	0.84579	0.00157	0.72346	
SD(0.7773)	0.85127	0.00097	0.70671	-0.0146
Dow Jones	0.97270	0.00120	0.73291	
SD(0.7888)	0.94899	0.00121	0.73196	-0.00087
NASDAQ	1.53121	0.00081	0.74280	
SD(0.8528)	1.53283	0.00079	0.74286	-0.00029

Table 2: S&P 500: Parameter Estimates; GARCH and CGARCH models

S&P 500					
	Coefficient	Estimate	Std. Error	t-statistic	p-value
No level shift In GARCH	β_r	0.057	0.004	13.55	0
	β_σ	0.938	0.004	214.75	0
No level shift In CGARCH	β_r	0.058	0.007	7.98	0
	β_σ	0.939	0.009	95.34	0
	ρ	0.9996	0.0003	3250.19	0
	$\emptyset$	0.027	0.005	5.36	0
Level shift in CGARCH	β_r	-0.085	0.069	4.73	0.082
	β_σ	0.525	0.217	1.54	0.088
	ρ	0.622	0.012	106.98	0
	$\emptyset$	0.138	0.089	2.47	0.031

Figure 1 S&P 500; fitted level shift component (right axis) and series (left axis)

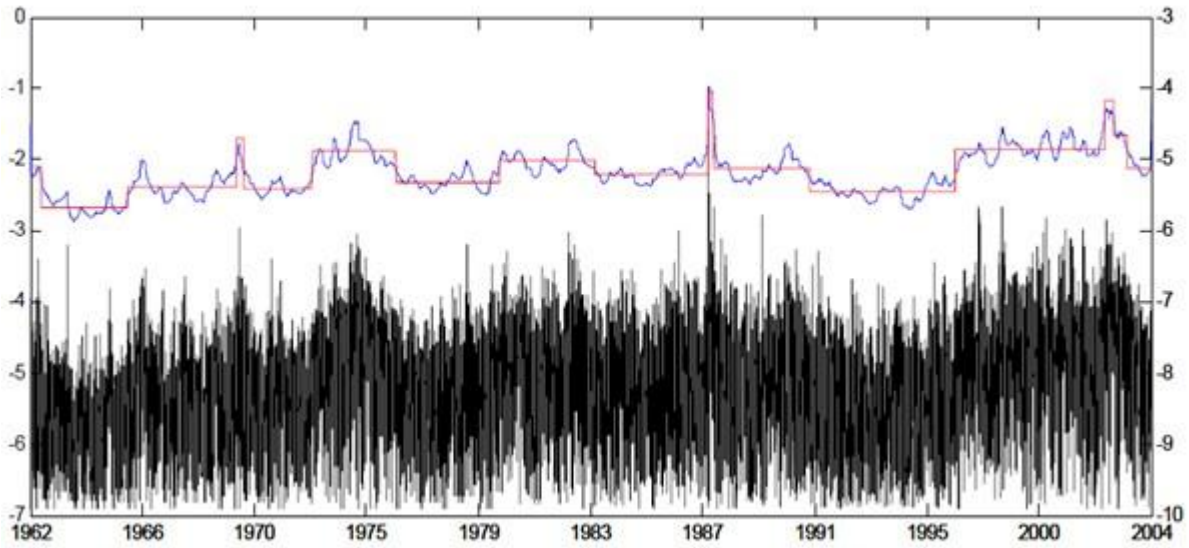
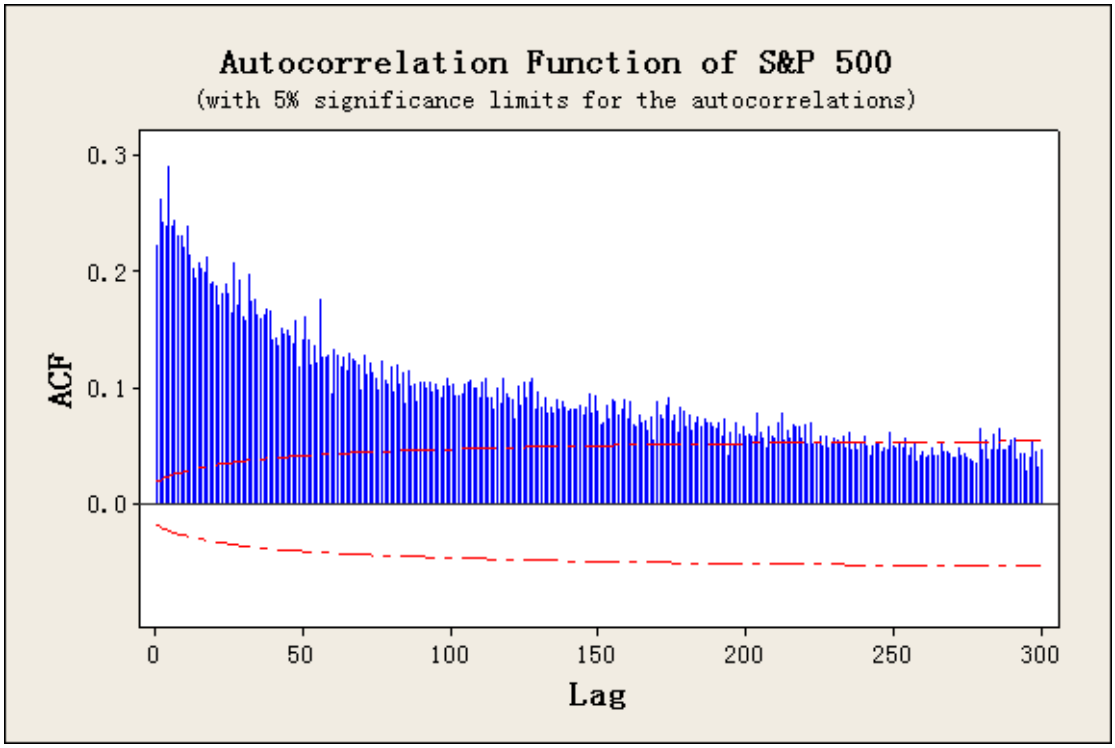


Figure 2 S&P 500; Autocorrelations



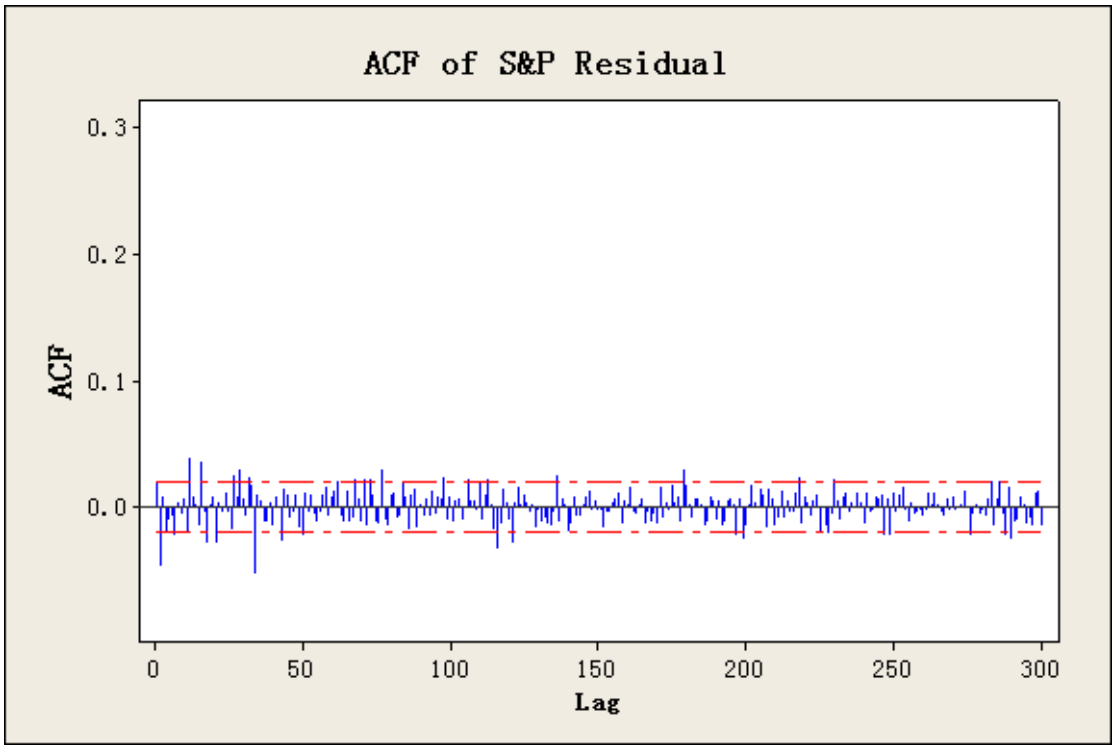


FIGURE 3

horizon	$MSE^{LS}(P)$ $/MSE^{GARCH}(P)$	horizon	$MSE^{LS}(P)/MSE^{GARCH}(P)$
P days	S&P 500	P days	S&P 500
20	0.5	120	0.51
40	0.44	140	0.54
60	0.44	160	0.56
80	0.46	180	0.59
100	0.48	200	0.62