Modeling and Forecasting Stock Return Volatility: Level Shift Model with Time Varying Jump Probability and Mean Reversion

Jiawen Xu[*] **Pierre Perron**[†]

Boston University **Boston University**

September 25, 2012

Abstract

We extend the random level shift (RLS) model of Lu and Perron (2010) to model the volatility of asset prices. The original RLS model consists of a short memory process and a level shift component, specified as the cumulative sum of a process that is 0 with probability $1-\alpha$ and is a random variable with probability α. Motivated by features obtained from this model, we extend it in two directions. The first is to speficy a time-varying probability of shifts modeled as a function of the occurrence and magnitude of large negative lagged returns. The second modification is to incorporate a mean reverting mechanism so that the sign and magnitude of the jump component changes according to the deviations of past jumps from their long run mean. It allows a better in-sample description and, more importantly, the possibility of forecasting the sign and magnitude of the jumps. We estimate the model using daily data on four major stock market indices: S&P 500, Nasdaq, Dow Jones Industrial Average and AMEX. We compare the forecasting performance of our model with various competing models. The most striking feature is that the modified RLS model is the only one that belongs to the 10% model confidence set of Hansen et al. (2011) using all comparisons, for all series and all forecasting horizons. The smallest values of the mean squre forecast errors are also obtained with the modified RLS model in 23 out of 24 cases. Overall, this is very strong evidence that our modified RLS model offers important gains in forecasting performance.

Keywords: structural change, time varying probability, mean reversion, forecasting, long-memory.

[*]Department of Economics, Boston University, 270 Bay State Rd., Boston MA 02215 (jwxu@bu.edu).

[†]Department of Economics, Boston University, 270 Bay State Rd., Boston MA 02215 (perron@bu.edu).

1 Introduction

Recently, there has been an upsurge of interest in the possibility of confusing long-memory with structural change in levels. This idea extends that exposited by Perron (1989, 1990) who showed that structural change and unit roots are easily confused. When a stationary process is contaminated by structural change, the estimate of the sum of its autoregressive coefficient is biased towards one and tests of the null hypothesis of a unit root are biased toward non-rejection. This phenomenon has been shown to apply to the long-memory context as well. That is, when a stationary short-memory process is contaminated by structural changes in level, the estimate of the long-memory parameter is biased away from zero and the autocovariance function of the process exhibits a slow rate of decay. Relevant references on this issue include Diebold and Inoue (2001), Engle and Smith (1999), Gourieroux and Jasiak (2001), Granger and Ding (1996), Granger and Hyung (2004), Lobato and Savin (1998), Mikosch and Stărică (2004), Parke (1999) and Teverosovky and Taqqu (1997).

The literature on modeling and forecasting stock return volatility is voluminous. Two approaches that have proven useful are the GARCH and stochastic volatility (SV) models. In their standard forms, the ensuing volatility processes are stationary and weakly dependent with autocorrelations that decrease exponentially. This contrasts with the empirical findings obtained using various proxies for volatility (e.g., daily absolute returns) which indicate autocorrelations that decay very slowly at long lags. In light of this, several long-memory models have been proposed. For example, Baillie, Bollerslev, and Mikkelsen (1996) and Bollerslev and Mikkelsen (1996) considered fractionally integrated GARCH and EGARCH models, while Breidt, Crato and De Lima (1998) and Harvey (1998) proposed long memory SV (LSV) models where the log of volatility is modeled as a fractionally integrated process.

More recently, attempts have been made to distinguish between short-memory stationary processes plus level shifts and long-memory models; see, in particular, Granger and Hyung (2004). They documented the fact that, when breaks determined via some pre-tests are accounted for, the evidence for long-memory is weaker. This evidence is, however, inconclusive since structural change tests are severely biased in the presence of long-memory and log periodogram estimates of the memory parameter are biased downward when sample-selected breaks are introduced. This is an overfitting problem that Granger and Hyung (2004, p. 416) clearly recognized. 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 shifts 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. Mikosch and Stărică (2004) considered the autocorrelation function of the absolute returns of the S&P 500 index for the period 1953-1977. They documented the fact that for the full period, it resembles that of a long-memory process. But, interestingly, if one omits the last fours years of data, the autocorrelation function is very different and looks like one associated with a short-memory process. They explain this finding by arguing that the volatility of the S&P 500 returns has increased over the period 1973-1977. Morana and Beltratti (2004) also argue that breaks in the level of volatility partially explain the long-memory features of some exchange rate series. Perron and Qu (2007) analyzed the time and spectral domain properties of a stationary short memory process affected by random level shifts. Perron and Qu (2010) showed that, when applied to daily S&P 500 log absolute returns over the period 1928-2002, the level shift model explains both the shape of the autocorrelations and the path of log periodogram estimates as a function of the number of frequency ordinates used. Qu and Perron (2012) estimated a stochastic volatility model with level shifts adopting a Bayesian approach using daily data on returns from the S&P 500 and NASDAQ indices over the period 1980.1-2005.12. They showed that the level shifts account for most of the variation in volatility, that their model provides a better in-sample fit than alternative models and that its forecasting performance is better for the NASDAQ and just as good for the S&P 500 as standard short or long-memory models without level shifts.

Lu and Perron (2010) extended the work of Stărică and Granger (2005) by directly estimating a structural model. They 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, they specify a simple mixture model such that the component is the cumulative sum of a process that is 0 with some probability $(1-\alpha)$ and is a random variable with probability α. To estimate such a model, they 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). They restrict the variance of one of the two normal distributions to be zero, allowing a simple but efficient algorithm.

Varneskov and Perron (2011) further extend the random level shift model to combine it with a long memory process, modeled as a $ARFIMA(p,d,q)$ process. They also carry out a comprehensive simulation study to show the precision of the parameter estimates. Their forecasting experiments using six different data series covering both low frequency and high frequency data show that the RLS-ARFIMA model outperforms other competing models with respect to many criteria.

This paper extends Lu and Perron (2010) in several directions. First, we let the jump probability depend on some covariates. This allows a more comprehensive and realistic probabilistic structure for the level shift model. The specification adopted is in the spirit of the "news impact curve" as suggested by Engle and Ng (1993). We model the probability of a shift as a function of the occurrence and magnitude of large negative lagged returns. The second modification is to incorporate a mean reverting mechanism to level shift model so that the sign and magnitude of the jump component changes according to the deviations of past jumps from their long run mean. Apart from being a device that allows a better in-sample description, its advantage is that the sign and magnitude of the jumps can be predicted to some extent. As we shall show this allows much improved forecasts.

We apply the modified level shift model to S&P 500 stock market index data (01/03/1950-10/11/2011; 15543 observations), Dow Jones Industrial Average (DJIA) index data (01/03/1950/-06/15/2012; 15752 observations), AMEX index (01/03/1996-06/18/2012; 4137 observations) and Nasdaq index (02/09/1971-06/18/2012; 10434 observations), using the logarithm of absolute returns as a proxy for volatility. Our point estimate for the average probability of shifts is similar to that of the original model, still a quite small number. But the weight on extreme past negative returns is large enough to result in a significant increase in jump probability when past stock return is taken into account, thereby inducing a clustering property for the jumps. Also, the estimates indicate that a mean reverting mechanism is present, which changes the sign of the jump. When the past jump component deviates from the long run mean by a large amount it is brought back towards the long-run mean.

We compare the forecasting performance of our model with various competing models, the original random level shift model (RLS), RLS with long memory (RLS-ARFIMA), the popular $ARFIMA(1, d, 1)$ and $ARFIMA(0, d, 0)$. The most striking feature is that the modified random level model is the only model that belongs to the 10% model confidence set of Hansen et al. (2010) using all comparisons, for all series and all forecasting horizons. When comparing the mean squared forecast errors, the smallest values are also obtained with the modified random level shift model in 23 out of 24 cases. Overall, this is very strong evidence that our modified random level shift model offers important gains in forecasting performance.

The structure of this paper is as follows. Section 2 briefly describes the data. Section 3 presents the basic random level shift model and discusses key results obtained from estimating it using data on the S&P 500 index in order to motivate subsequent developments. Section 4 discusses the extensions made to the basic model consisting of allowing for time varying

probabilities of jumps and a mean-reverting mechanism. Section 5 presents the estimation methodology. Section 6 presents the full-sample estimates obtained from the extended model. Section 7 presents results for a real-time forecasting experiment, which show that much improved forecasts can be obtained using our extended moddel. Section 8 provides brief concluding remarks.

2 Data and Summary Statistics

The data used to construct the volatility series are based on daily closing prices, say P_t, and the daily returns are computed as $r_t = \ln(P_t) - \ln(P_{t-1})$. The volatility is proxied by log absolute returns. In order to avoid extreme negative volatility, we bound absolute returns away from zero by adding a small constant 0.001, so that the volatility series used is $y_t = \ln(|r_t| + 0.001)$.

We use daily data instead of measures based on high frequency data. First, higher frequency data are not available for long periods and given that level shifts are quite infrequent, a long span of data is desirable; second, the realized volatility constructed using higher frequency data is very sensitive to the window used to construct the series, the treatment of seasonal effects, the method to handle missing observations. The use of daily returns avoids the arbitrariness in the construction of realized volatility.

Table 1 gives summary statistics of those volatility proxies and show their unconditional distribution characteristics. The four stock volatility series have similar characteristics: mean, standard deviation and extreme values. Except for DJIA, the volatility series have positive skewness, i.e., a right-tailed distribution with few high values. The kurtosis values are around 2.7, slightly lower than 3 for the normal distribution.

3 The Basic Random Level Shift Model

The basic random level shift model is:

$$y_t = a + \tau_t + c_t \tag{1}$$

where a is a constant, τ_t is the random level shift component and c_t is a short memory process. The level shift component is specified by

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

where

$$\delta_t = \pi_t \eta_t.$$

Here, π_t follows a Bernoulli distribution that takes value 1 with probability α and value 0 with probability $1-\alpha$. If it takes value 1, then a level shift η_t occurs drawn from a $N(0,\sigma_\eta^2)$ distribution. In general, the short-memory component can be modelled as $c_t = C(L)e_t$, with $e_t \sim i.i.d.\ N(0,\sigma_e^2)$ and $E|e_t|^r < \infty$ for $r > 2$. The polynomial $C(L)$ satisfies $C(L) = \sum_{i=0}^{\infty} c_i L^i$, $\sum_{i=0}^{\infty} i|c_i| < \infty$ and $C(1) \neq 0$. As pointed out by Lu and Perron (2010) and also documented in Section 4, once the level shifts are accounted for, barely any serial correlation remains. Accordingly, we can simply assume c_t to be a white noise process.

The state space representation of this model involves an error term that is a mixture of two normal distributions. With the normality assumption used to construct the quasi-likelihood function, the level shift component τ_t can be represented as a random walk process with errors following mixed normal distributions, namely

$$\begin{aligned} \tau_t &= \tau_{t-1} + \delta_t \\ \delta_t &= \pi_t \eta_t = \pi_t \eta_{1t} + (1-\pi_t)\eta_{2t} \end{aligned}$$

where $\eta_{it} \sim i.i.d.\ N(0,\sigma_{\eta i}^2)$. By specifying $\sigma_{\eta 1}^2 = \sigma_\eta^2$ and $\sigma_{\eta 2}^2 = 0$, we recover our level shift model. To cast the model in state-space form, note that the first differences of y_t are given by:

$$\begin{aligned} \triangle y_t &= \tau_t - \tau_{t-1} + c_t - c_{t-1} \\ &= \delta_t + c_t - c_{t-1} \end{aligned}$$

and, for reasons mentioned above, the short-memory component is simply white noise, so that

$$c_t = e_t.$$

Hence, the state-space representation of the model is

$$\begin{aligned} \triangle y_t &= HX_t + \delta_t \\ X_t &= FX_{t-1} + U_t \end{aligned}$$

where $X_t = [c_t, c_{t-1}]'$, $H = [1, -1]$,

$$F = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

and U_t is a 2-dimensional normally distributed random vector with mean zero and covariance matrix

$$Q = \begin{pmatrix} \sigma_e^2 & 0 \\ 0 & 0 \end{pmatrix}.$$

3.1 Fitted Level Shifts and Autocorrelation Functions

To provide stylized features of the series considered and motivate our subsequent modelling, we consider the last 10,000 observations of the S&P500 series (02/25/1972-10/11/2011). Figure 1 presents a plot of the autocorrelations up to lag 2000, which shows that it displays a slow decay rate, akin to a long-memory process. To see if this long-memory feature can be accounted for by level shifts, we follow Lu and Perron (2010) and estimate the basic random level shift model presented in the previous section in order to extract the fitted level shift component. The method of estimation is described in Lu and Perron (2010). The estimate of the jump probability is 0.0029, so that the estimate of the number of jumps in the series is 29.

To obtain the level shift component of the volatility process, we first need to estimate the dates of the shifts and the means within each regime. Since the smoothed estimate of the level shift component performs poorly in the presence of multiple changes, we use the point estimate of jump probability to get an approximation to the number of level shifts and apply the method of Bai and Perron (2003) to obtain the estimates of the jump dates and regime-specific means as the global minimizers of the following sum of squared residuals:

$$\sum_{i=1}^{m+1} \sum_{t=T_{i-1}+1}^{T_i} \left[y_t - \mu_i\right]^2,$$

where m is the number of breaks (here 29), T_i $(i = 1, ..., m)$ are the candidate break dates with the convention that $T_0 = 0$ and $T_{m+1} = T$ and μ_i $(i = 1, ..., m+1)$ are the means within each regime. Note that since we allow for consecutive level shifts, we set the minimal length of a segment to just one observation. With the estimates of the break dates $\{\hat{T}_i; i = 1, ..., m\}$ and the regime-specific means $\{\hat{u}_i; i = 1, ..., m+1\}$, the level shift component is given by $\sum_{i=1}^{m+1} \hat{u}_i DU_{i,t}$, where $DU_{i,t} = 1$ if $\hat{T}_{i-1} < t \leq \hat{T}_i$ and 0, otherwise. It is plotted in Figure 2 along with a smoothed estimate of the original volatility process (obtained using a nonparametric fit with a standard Gaussian kernel). As can be seen, the general tendency of the fitted level shift component follows the major changes in the volatility process, with a large level shift in both October 1987 and 2008, associated with major financial crises that affected the stock markets.

To see whether the level shift component can explain the long-memory property of the volatility process, we present in Figure 3 the sample autocorrelations the residuals defined as the difference between the original process and the fitted level shift component. A distinctive feature is that now the residuals exhibit barely no serial correlation even at small lags. Hence,

when the level shifts are accounted for, the long-memory property of volatility is no longer present. Although the shifts are rare, they account for almost all the autocorrelations in volatility. As a result, modeling volatility as a short memory process plus a random level shift component appears indeed an attractive avenue.

3.2 Clustering Jumps and Mean Reversion

A close look at the fitted level shift component reveals that some jumps tend to occur within a short period of time. Those time periods are often associated with abnormal price fluctuations, for example financial crashes or important macroeconomics or policy news. There are also few spikes in the level shift process, e.g., 1974-1975, 1987, 1999, 2008-2010. It is indeed expected that volatility jumps should be clustered during periods of financial crises. This clustering phenomenon is interesting and indicates that the level shifts may not be $i.i.d.$ as originally modeled with a constant jump probability for all time periods. On the contrary, the jump probability is likely to change depending on different circumstances. For example, when financial markets are turbulent, it is more likely for the volatility process to jump up. Accordingly, we shall model the probability of a shift as a function of some covariates with the aim at better describing the clustering of jumps.

Another interesting observation is that the jump component seems to follow a mean reverting process. It is indeed implausible that the volatility will jump in an arbitrary manner. Upward shifts are often followed by downward shifts, so that a mean-reverting process is present in the fitted level shift component. Hence, it is highly likely that a proper modeling of this mean reverting mechanism could lead to improved forecasting performance. Accordingly, we shall also introduce a mean-reverting component in the model.

4 Extensions of the Random Level Shift Model

As discussed in the previous section, two features that are likely to improve the fit and the forecasting performance is to allow for changes in the probability of shifts and model explicitly the mean-reverting mechanism of the level shift component. In the first step, we specify the jump probability to be

$$p_t = f(p, x_{t-1})$$

where p is a constant and x_{t-1} are covariates that would allow to better predict the probability of shifts in volatility, and f is a function that ensure that $p_t \in [0, 1]$. Note that x_{t-1} needs

to be in the information set at time t in order for the model to be useful for forecasting. As documented by, e.g., Martens et al. (2004), there is a pronounced relationship between current volatility and lagged returns, sometimes referred to as the leverage effect. A popular way to model this effect is via the "news impact curve" as suggested by Engle and Ng (1993). This usually takes the following form

$$\log(\sigma_t^2) = \beta_0 + \beta_1|r_{t-1}| + \beta_2 I(r_{t-1} < 0) + \beta_3|r_{t-1}|I(r_{t-1} < 0)$$

where σ_t^2 is a measure of volatility and $I(A)$ is the indicator function of the event A. It is typically the case that the estimate of β_1 is not significant (see, e.g., Martens et al, 2004). Hence, we shall ignore this term. Also, since our aim is to model changes in the probability of a shift in volatility and not volatility per se, it is more appropriate to use large negative returns beyond some threshold a, say, stated in relation to the probability that a return exceeds a. In our applications we shall consider negative returns that are at the bottom 1%, 2.5% or 5% of the sample distribution of returns. Hence, the functional form adopted is the following:

$$f(p, x_{t-1}) = \{ \begin{array}{ll} \Phi(p + \gamma_1 1\{x_{t-1} < 0\} + \gamma_2 1\{x_{t-1} < 0\}|x_{t-1}|) & \text{for} \quad |x_{t-1}| > a \\ \Phi(p) & otherwise \end{array} \tag{2}$$

where $\Phi(.)$ is a normal cdf function, so that $f(p, x_{-1})$ is bounded between 0 and 1, as required.

The second step involves building a mean reverting mechanism to the level shift model. The motivation for doing so is that we observe evidence that stock volatility does not jump arbitrarily and that large upward movement tend to be followed by a decrease. This can be seen in Figure 2, where overall the shift component tends to revert back to some long-term mean value. This feature can be beneficial to improve the forecasting performance if explicitly modeled. The specification we adopt is the following:

$$\eta_{1t} = \beta(\tau_{t|t-1} - \bar{\tau}) + \widetilde{\eta}_{1t}$$

where $\widetilde{\eta}_{1t} \sim N(0, \sigma_\eta^2)$, $\tau_{t|t-1}$ is the filtered estimate of the jump component at time t and $\bar{\tau}$ is the mean of all the filtered estimates of the jump component from the beginning of the sample up to time t. This implies a mean-reverting mechanism provided $\beta < 0$. The magnitude of β then dictates the speed of reversion. Note that the specification involves using data only up to time t in order to be useful for forecasting purposes. Note that it will have an impact on forecasts since being in a high (low) volatility state implies that in future

periods volatility will be lower (higher), and more so as the forecasting horizon increases. Hence, this specification has an effect on the forecasts of both the sign and size of future jumps in volatility.

5 Estimation Methodology

The estimation methodology follows Lu and Perron (2010) with appropriate modifications. The main ingredient used 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 then eliminated from the final likelihood function by summing over all possible state realizations.

Let $Y_t = (\Delta y_1, \Delta y_2, \ldots, \Delta y_t)$ be the vector of data available up to time t and denote the vector of parameters by $\theta = [\sigma_\eta^2, p, \sigma_e^2, \gamma_1, \gamma_2, \beta]$. The level shift model is fundamentally different from the popular Markov switching models, especially given the fact that the number of states is determined by the data and none of the states need be revisited. Nevertheless, the two models share similar features when constructing the likelihood function. To illustrate the similarities we adopt the notation in Hamilton (1994), where 1 represents a (4×1) vector of ones, the symbol $\odot$ denotes element-by-element multiplication, $\widehat{\xi}_{t|t-1} = vec(\widetilde{\xi}_{t|t-1})$ with the $(i,j)^{th}$ element of $\widetilde{\xi}_{t|t-1}$ being $\Pr(s_{t-1} = i, s_t = j | Y_{t-1}; \theta)$ and $\omega_t = vec(\widetilde{\omega}_t)$ with the $(i,j)^{th}$ element of $\widetilde{\omega}_t$ being $f(\triangle y_t | s_{t-1} = i, s_t = j, Y_{t-1}; \theta)$ for $i, j \in \{1, 2\}$. Here $s_t = 1$ (resp., 2) when $\pi_t = 1$ (resp., 0), i.e., a level shift occurs (resp., does not occur). The log likelihood function is

$$\ln(L) = \sum_{t=1}^{T} \ln f(\triangle y_t | Y_{t-1}; \theta) \tag{3}$$

where

$$\begin{aligned} f(\triangle y_t | Y_{t-1}; \theta) &= \sum_{i=1}^{2} \sum_{j=1}^{2} f(\triangle y_t | s_{t-1} = i, s_t = j, Y_{t-1}; \theta) \Pr(s_{t-1} = i, s_t = j | Y_{t-1}; \theta) \quad (4) \\ &\equiv 1'(\widehat{\xi}_{t|t-1} \odot \omega_t) \end{aligned}$$

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

$$\begin{aligned} \widetilde{\xi}_{t|t-1}^{ij} &\equiv \Pr(s_{t-1} = i, s_t = j | Y_{t-1}; \theta) \\ &= \Pr(s_t = j) \sum_{k=1}^{2} \Pr(s_{t-2} = k, s_{t-1} = i | Y_{t-1}; \theta) \quad (5) \end{aligned}$$

and

$$\begin{aligned}\widetilde{\xi}^{ij}_{t-1|t-1} &\equiv \Pr(s_{t-2}=k, s_{t-1}=i|Y_{t-1};\theta) \\ &= \frac{f(\triangle y_t|s_{t-2}=k, s_{t-1}=i, Y_{t-2};\theta)\Pr(s_{t-2}=k, s_{t-1}=i|Y_{t-1};\theta)}{f(\triangle y_{t-1}|Y_{t-2};\theta)}\end{aligned}$$

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

$$\begin{aligned}\begin{bmatrix}\widetilde{\xi}^{11}_{t+1|t}\\ \widetilde{\xi}^{21}_{t+1|t}\\ \widetilde{\xi}^{12}_{t+1|t}\\ \widetilde{\xi}^{22}_{t+1|t}\end{bmatrix} &= \begin{bmatrix} p_{t+1}(\widetilde{\xi}^{11}_{t|t}+\widetilde{\xi}^{21}_{t|t}) \\ p_{t+1}(\widetilde{\xi}^{12}_{t|t}+\widetilde{\xi}^{22}_{t|t}) \\ (1-p_{t+1})(\widetilde{\xi}^{11}_{t|t}+\widetilde{\xi}^{21}_{t|t}) \\ (1-p_{t+1})(\widetilde{\xi}^{12}_{t|t}+\widetilde{\xi}^{22}_{t|t})\end{bmatrix} \\ &= \begin{bmatrix} p_{t+1} & p_{t+1} & 0 & 0 \\ 0 & 0 & p_{t+1} & p_{t+1} \\ (1-p_{t+1}) & (1-p_{t+1}) & 0 & 0 \\ 0 & 0 & (1-p_{t+1}) & (1-p_{t+1})\end{bmatrix}\begin{bmatrix}\widetilde{\xi}^{11}_{t|t}\\ \widetilde{\xi}^{21}_{t|t}\\ \widetilde{\xi}^{12}_{t|t}\\ \widetilde{\xi}^{22}_{t|t}\end{bmatrix}\end{aligned}$$

or more compactly by:

$$\widehat{\xi}_{t+1|t} = \Pi\widehat{\xi}_{t|t}$$

with

$$\widehat{\xi}_{t|t} = \frac{\widehat{\xi}_{t|t-1}\odot\omega_t}{\mathbf{1}'(\widehat{\xi}_{t|t-1}\odot\omega_t)}$$

The conditional likelihood for $\triangle y_t$ is the following normal density:

$$\widetilde{\omega}_t^{ij} = f(\triangle y_t|s_{t-1}=i, s_t=j, Y_{t-1};\theta) = \frac{1}{\sqrt{2\pi}}\left|f_t^{ij}\right|^{-\frac{1}{2}}\exp\left\{-\frac{v_t^{ij\prime}(f_t^{ij})^{-1}v_t^{ij}}{2}\right\} \tag{6}$$

where $v_t^{ij} = \triangle y_t - \triangle y^i_{t|t-1}$ is the prediction error and $f_t^{ij} = E(v_t^{ij}v_t^{ij\prime})$ is the prediction error variance. Note that $\triangle y^i_{t|t-1} = E[\triangle y_t|s_{t-1}=i, Y_{t-1};\theta]$ does not depend on the state j at time t because we are conditioning on time $t-1$ information. However, $\triangle y_t$ does depend on $s_t = j$ so that 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

$$\begin{aligned}X^i_{t|t-1} &= FX^i_{t-1|t-1} \\ P^i_{t|t-1} &= FP^i_{t-1|t-1}F' + Q\end{aligned} \tag{7}$$

We have the measurement equation $\triangle y_t = HX_t + \delta_t$, where the measurement error δ_t has mean zero and a variance which can take two possible values: $R_1 = \sigma_\eta^2$, with probability $p_{t,}$ or $R_2 = 0$, with probability $1 - p_t$. Hence, the prediction error is $v_t^{ij} = \triangle y_t - HX_{t|t-1}^i$with associated variance $f_t^{ij} = HP_{t|t-1}^i H' + R_j$. Applying standard updating formulas, we have given $s_t = j$ and $s_{t-1} = i$,

$$\begin{aligned} X_{t|t}^{ij} &= X_{t|t-1}^i + P_{t|t-1}^i H'(HP_{t|t-1}^i H' + R_j)^{-1}(\triangle y_t - HX_{t|t-1}^i) \\ P_{t|t}^{ij} &= P_{t|t-1}^i - P_{t|t-1}^i H'(HP_{t|t-1}^i H' + R_j)^{-1} HP_{t|t-1}^i \end{aligned} \tag{8}$$

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

$$\begin{aligned} X_{t|t}^j &= \frac{\sum_{i=1}^2 \Pr(s_{t-1} = i, s_t = j|Y_t;\theta) X_{t|t}^{ij}}{\Pr(s_t = j|Y_r;\theta)} = \frac{\sum_{i=1}^2 \widetilde{\xi}_{t|t}^{ij} X_{t|t}^{ij}}{\sum_{i=1}^2 \widetilde{\xi}_{t|t}^{ij}} \\ P_{t|t}^j &= \frac{\sum_{i=1}^2 \Pr(s_{t-1} = i, s_t = j|Y_t;\theta)[P_{t|t}^{ij} + (X_{t|t}^j - X_{t|t}^{ij})(X_{t|t}^j - X_{t|t}^{ij})']}{\Pr(s_t = j|Y_r;\theta)} \\ &= \frac{\sum_{i=1}^2 \widetilde{\xi}_{t|t}^{ij}[P_{t|t}^{ij} + (X_{t|t}^j - X_{t|t}^{ij})(X_{t|t}^j - X_{t|t}^{ij})']}{\sum_{i=1}^2 \widetilde{\xi}_{t|t}^{ij}} \end{aligned} \tag{9}$$

By doing so, we make ω_t^{ij} unaffected by the history of states before time $t-1$. Some modifications are needed when including the mean reverting mechanism in the model. In equation (6), the prediction error v_t^{ij} is originally normally distributed with mean 0 and a variance that depends on the particular value of the state. But now the modified model becomes:

$$\begin{aligned} y_t &= a + c_t + \tau_t \\ \triangle y_t &= \tau_t - \tau_{t-1} + c_t - c_{t-1} \\ \tau_t - \tau_{t-1} &= \pi_t[\beta(\tau_{t|t-1} - \overline{\tau}) + \widetilde{\eta}_{1t}] + (1 - \pi_t)\eta_{2t} \end{aligned}$$

At time t when $\pi_t = 1$, we need to subtract the mean reversion term, which is known at

time t and independent from the realization of π_t. Accordingly,

$$\begin{aligned}
\widetilde{\omega}_t^{ij} &= f(\triangle y_t | s_{t-1} = i, s_t = j, Y_{t-1}; \theta) = \frac{1}{\sqrt{2\pi}} \left| f_t^{ij} \right|^{-\frac{1}{2}} \exp\left\{ -\frac{\widetilde{v}_t^{ij\prime} (f_t^{ij})^{-1} \widetilde{v}_t^{ij}}{2} \right\} \qquad (10) \\
\widetilde{v}_t^{ij} &= \begin{Bmatrix} v_t^{11} - \beta(\tau_{t|t-1}^{11} - \bar{\tau}^{11}) \\ v_t^{12} \\ v_t^{21} - \beta(\tau_{t|t-1}^{21} - \bar{\tau}^{21}) \\ v_t^{22} \end{Bmatrix} \\
f_t^{ij} &= E(\widetilde{v}_t^{ij} \widetilde{v}_t^{ij\prime}) = H P_{t|t-1}^{i} H' + R_j
\end{aligned}$$

Since $y_t = a + \tau_t + c_t$, then $\tau_{t|t-1}^{i1} = y_t - a - c_{t|t-1}^{i1} = y_t - a - [0\ 1]'\ X_{t|t-1}^{i1}$. Note that $X_{t|t-1}^{i1}$ being a state variable it can be updated every time period. Therefore, $\tau_{t|t-1}^{i1} - \bar{\tau}^{i1}$ is known at time t. Also $R_1 = \sigma_\eta^2$ with probability p_t and $R_2 = 0$ with probability $1 - p_t$.

6 Full Sample Estimation Results

We first present results from estimating the basic random level shift model using the stock volatility series in order to compare our results with those of Lu and Perron (2010) who used a shorter sample. These are reported in Table 2. Note that the jump probability is quite small, indicating that level shifts are relatively rare events. The point estimates for the jump probability p imply the following number of shifts for each series: 65 for the S&P 500, 34 jumps for Nasdaq, 29 for the DJIA and 29 for the AMEX. Due to the fact that our S&P 500 data covers a longer period than that in Lu and Perron (2010), our point estimate of the number of jumps is also higher. This is especially the case since our sample further include the period 2004 to 2011, a time during which stock markets went through a turbulent period induced by the financial crisis in 2008. A lot of uncertainties existed during that period. Hence, it is not surprising, indeed expected, that level shifts happen more often with this extended sample. The standard error of the short memory component remains the same, while the standard error of the jump variable is smaller compared to the results in Lu and Perron (2010).

In Table 3, we report the estimation results when incorporating a time varying probability into the RLS model. For each series, we consider three different threshold levels to assess the robustness of the results. The threshold level adopted is the value a such that, say, x% of the returns are below a with $x = 1, 2.5$ and 5. The results show that the estimates of both γ_1 and γ_2 are positive. Since we use absolute values of negative returns in the specification,

a positive γ_2 is consistent with the evidence that large negative returns are associated with higher volatility, in our case via a higher probability of a shift occurring. Furthermore, the positive estimate of γ_1 is consistent with the so-called "the news impact" effect. Note that the estimate of p is negative since we use normal cdf functional form for p_t. As the threshold level decreases, we find that γ_1 increases but γ_2 decreases. However, the standard error of γ_1 increases while that of γ_2 decreases, so that γ_2 becomes more significant and γ_1 becomes less significant as the threshold level decreases; see, in particular the results for the Nasdaq series. These results show that extreme bad news do indeed have a significant effect on the jump probability. Note for the AMEX series with a threshold value of 5% or 2.5%, the estimates of γ_1 and γ_2 are negative, though both are insignificant with large standard errors. This may be due to the relatively smaller sample size available for the AMEX series. Figure presents the smoothed estimates of the level shift component for the three threshold values for the case of the S&P 500 index. What transpires from the results is that they are very similar and all equally good in matching the smoothed estimate of the volatility process. Hence, in what follows we shall present results only for the case of a 1% trimming. The same features apply to other stock market indices.

The estimation results obtained when adding only a mean reversion component in the jump process are presented in Table 4. As a first step, we do not include the time varying probabilities in order to assess separately the effect of mean reversion. In all four cases, the estimate of β is significantly negative, indicating that mean reversion is indeed present in the jump process. Note also that by adding a mean-reverting component, the estimate of the probability of shifts increases compared to the that in the basic random level shift model. Also, the standard error of the jump variable is much smaller. This is due to the fact that the mean reversion part account for a large amount of the total variation of the jump process, leaving less to be accounted for by the jump variable itself. Figure 5 presents the smoothed estimate of the level shift component $\tau_{t|T}$, together with the volatility process for the case of the S&P 500. Compared to the smoothed estimate of the level shift component for the basic RLS model it contains more short-term variability, which explains why jumps in the RLS model with mean reversion are estimated to occur more frequently.

Table 5 presents the estimates of the modified RLS model combining both time varying probabilities and mean reversion, using a threshold value of 1%. First, in all cases the estimate of β is significantly negative, again indicating the presence of a mean-reverting property for the level shift component. The estimates are similar to those obtained without allowing for time variation in the probability of shifts, showing some robustness to our findings. The

estimate of γ_2, pertaining to the component $1\{x_{t-1} < 0\}|x_{t-1}|$ in the specification of the functional form for the time-varying probabilities, is significantly positive in all cases, except for the AMEX index. On the other hand, the estimates of γ_1, pertaining to the component $1\{x_{t-1} < 0\}$ are not significant, except for the DJIA. Hence, in the forecasting experiment reported below, we shall omit this component.

The following results were obtained for the case of the S&P 500 series; similar results apply to the other stock market indices and are therefore not reported. Figure 5 presents the smoothed estimates of the volatility and of the level shift component for the four versions of the random level shift model: the basic one, with time-varying probabilities only, with mean reversion only and with time varying probabilities and mean reversion. Note that the smoothed estimate of the level shift component is similar across all models and follows closely the smoothed estimate of the volatility, indicating a good in-sample fit. But as we shall see, even though the models have similar in-sample fit, the out-of-sample fit is not the same with the model incorporating time-varying probabilities and mean reversion performing best. Figure 6 presents the autocorrelation function of the difference between the volatility process and the smoothed level shift component with both time-varying probabilities and mean reversion. It clearly shows that the remaining noise is uncorelated, thereby justifying of specification of the nature of the short-memory component and re-enforcing the conclusion that once level shift are taken into account the long-memory feature of the volatility series is no longer present.

7 Forecasting

We first discuss how to construct out-of-sample forecasts for the random level shift model, assuming the short memory process to be just white noise. According to Varneskov and Perron (2011), the τ-step ahead forecasts of the basic random level shift model is given by

$$\hat{y}_{t+\tau|t} = y_t + HF^{\tau}[\sum_{i=0}^{1}\sum_{j=0}^{1} \Pr(s_{t+1} = j) \Pr(s_t = i|Y_t) X_{t|t}^{ij}]$$

where $E_t(y_{t+\tau}) = \hat{y}_{t+\tau|t}$ is the forecast of volatility at time $t+\tau$, conditional on information at time t. With our modified RLS model, this forecasting formula still holds with appropriate modifications for $X_{t|t}^{ij}$ and $\Pr(s_{t+1} = j)$.

We compare the forecasting performance of our model with the original RLS model, the RLS_ARFIMA(0,d,0), the RLS_ARFIMA(1,d,1), the ARFIMA(1,d,1) and the ARFIMA(0,d,0) models. We concentrate on the ARFIMA class of model as they were shown to provide

better forecasts than GARCH-type models. We use the following forecasting horizons: $\tau = 1, 5, 10, 20, 50$ and 100. The mean square forecast error (MSFE) criterion, proposed by Hansen & Lunde (2006a) and Patton (2011), defined by:

$$MSFE_{\tau,i} = \frac{1}{T_{out}} \sum_{t=1}^{T_{out}} (\bar{\sigma}^2_{t,\tau} - \bar{y}_{t+\tau,i|t})^2$$

where T_{out} is the number of forecasts, $\bar{\sigma}^2_{t,\tau} = \sum_{s=1}^{\tau} y_{t+s}$, and $\bar{y}_{t+\tau,i|t} = \sum_{s=1}^{\tau} \hat{y}_{t+s,i|t}$, with i indexing the model. The relative performance of models i and j at time t is defined as:

$$d_{ij,t} = (\bar{\sigma}^2_{t,\tau} - \bar{y}_{t+\tau,i|t})^2 - (\bar{\sigma}^2_{t,\tau} - \bar{y}_{t+\tau,j|t})^2$$

The different model forecasting performances are evaluated and compared using the 10% model confidence set (MCS) of Hansen et al. (2011).

The forecasting experiment is as follow. We keep the last 1500 observations as the out-of-sample period to be forecasted. Hence the starting date of the forecasts is 10/27/2005 for S&P 500, 07/06/2006 for Nasdaq and AMEX, and 07/03/20-6 for DJIA. The reasons for considering this period is that it contains very different episodes of calm and turbulent periods, mostly as the result of the financial crisis in 2008. Hence, it is ideally suited as a particularly difficult period to forecast volatility. Given that estimating these models is quite time consumin, we estimate the models once withoutthe last 1500 observations. The forecasts are then made conditional on the parameter estimates obtained.

The results are presented in Table 6. The most striking feature is that the modified random level model is the only model that belongs to the 10% model confidence set using all comparisons, for all series and all forecasting horizons. When comparing the MSFE, the smallest values are also obtained with the modified random level shift model in 23 out of 24 cases. The only case in which it does not is for the Nasdaq series with a 100 step horizon, in which case the model that delivers the lowest MSFE is the RLS-ARFIMA(1,d,1). The ARFIMA(1,d,1) and ARFIMA(0,d,0) perform particularly poorly. The p-values that they belong to the model confidence set are 0.00 is 22 out of 24 cases. The improvement in forecacst accuracy of the modifield RLS model relative to the ARFIMA models can be very substantial. For medium term forecasting horizons (5 to 20 days ahead), the MSFEs of the modified RLS can be between 36% and 63% of those of the ARFIMA models. Overall, this is very strong evidence that our modified random level shift model offers important gains in forecasting performance.

8 Conclusion

With the aim of improving the forecasting performance of the random level shift model of Lu and Perron (2010), we proposed two modifications. The first is a structure to allow a time-varying probability of shifts. We modelled the probability of a shift as a function of the occurrence and magnitude of large negative lagged returns. The second modification is to incorporate a mean reverting mechanism so that the sign and magnitude of the jump component changes according to the deviations of past jumps from their long run mean. Apart from being a device that allows a better in-sample description, its advantage is that the sign and magnitude of the jumps can be predicted to some extent. The full sample estimates reveal interesting features usuful to understand the behavior of stock prices volatility. More importantly, the extended model allows much improved forecasts of volatility when applied to stock market indices. Hence, our results provide additional evidence that random level shift models are serious contenders to model volatility and outperforms the popular class of standard long-memory models such as the commonly used ARFIMA model.

References

Bai, J., Perron, P., (2003), "Computation and analysis of multiple structural change models," Journal of Applied Econometrics 18, 1-22.

Baillie, R.T., Bollerslev, T., Mikkelsen, H.O., (1996), "Fractionally integrated generalized autoregressive conditional heteroskedasticity," Journal of Econometrics 73, 3–30.

Bollerslev, T., Mikkelsen, H.O., (1996), "Modeling and pricing long memory in stock market volatility," Journal of Econometrics 73, 151–184.

Breidt, J.F., Crato, N., de Lima, P.J.F., (1998), "On the detection and estimation of long-memory in stochastic volatility," Journal of Econometrics 83, 325–348.

Diebold, F.X., and Inoue, A., (2001), "Long memory and regime switching," Journal of Econometrics 105, 131-159.

Engle, R.F., and Ng, V.K., (1993), "Measuring and testing the impacts of news on volatility," Journal of Finance 48, 1749-1778.

Engle, R.F., and Smith, A.D., (1999), "Stochastic permanent breaks," Review of Economics and Statistics 81, 533–574.

Gourieroux, C., and Jasiak, J., (2001), "Memory and infrequent breaks," Economics Letters 70, 29–41.

Granger, C.W.J., and Ding, Z., (1996), "Varieties of long memory models," Journal of Econometrics 73, 61–77.

Granger, C.W.J., and Hyung, N., (2004), "Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns," Journal of Empirical Finance 11, 399-421.

Hamilton, J.D., (1994), Time Series Analysis. Princeton University Press.

Hansen, P.R., and Lunde, A., (2006a), "Consistent ranking of volatility models," Journal of Econometrics 131, 97-121.

Hansen, P.R., Lunde, A., and Nason, J.M., (2011), "The model confidence set," Econometrica 79, 453-497.

Harvey, A.C., (1998), "Long memory in stochastic volatility," In: Knight, J., Satchell, S. (Eds.), Forecasting Volatility in Financial Markets. Butterworth-Heineman, Oxford, 307–320.

Martens, M., van Dijk, D., and de Pooter, M., (2004), "Modeling and forecasting S&P 500 volatility: Long memory, structural breaks and nonlinearity," Tinbergen Institute Discussion Paper.

Mikosch, T., and Stărică, C., (2004), "Changes of structure in financial time series and the GARCH model," REVSTAT - Statistical Journal 2, 42–73.

Morana, C., and Beltratti, A., (2004), "Structural change and long-range dependence in volatility of exchange rates: either, neither, or both," Journal of Empirical Finance 11, 629-658.

Lobato, I.N., and Savin, N.E., (1998), "Real and spurious long-memory properties of stock-market data," Journal of Business and Economics Statistics 16, 261–268.

Lu, Y.K., and Perron, P., (2010), "Modeling and forecasting stock return volatility using a random level shift model," Journal of Empirical Finance 17, 138-156.

Patton, A., (2011), "Volatility forecast comparison using imperfect volatility proxies," Journal of Econometrics 160, 245-256.

Parke, W.R., (1999), "What is fractional integration?" Review of Economics and Statistics 81, 632–638.

Perron, P., (1989), "The great crash, the oil price shock and the unit root hypothesis," Econometrica 57, 1361-1401.

Perron, P., (1990), "Testing for a unit root in a time series regression with a changing mean," Journal of Business and Economic Statistics 8, 153-162.

Perron, P. and Qu, Z., (2007), "An analytical evaluation of the log-periodogram estimate in the presence of level shifts," Working Paper, Department of Economics, Boston University.

Perron, P., and Qu, Z., (2010), "Long-memory and level shifts in the volatility of stock market return indices", Journal of Business and Economic Statistics 28, 275-290.

Perron, P., and Wada, T., (2009), "Let's take a break: trends and cycles in U.S. real GDP," Journal of Monetary Economics 56, 749-765.

Qu, Z., and Perron, P., (2012), "A stochastic volatility model with random level shifts and its applications to S&P 500 and NASDAQ indices," forthcoming in the Econometrics Journal.

Stărică, C., Granger C.W.J., (2005), "Nonstationarities in stock returns," The Review of Economics and Statistics 87, 503-522.

Teverovsky, V., and Taqqu, M., (1997), "Testing for long-range dependence in the presence of shifting means or a slowly declining trend, using a variance-type estimator," Journal of Time Series Analysis 18, 279–304.

Varneskov, R.T., and Perron, P., (2011), "Combining long memory and level shifts in modeling and forecasting the volatility of asset returns," BU Working paper.

Wada, T., and Perron, P., (2007), "State space models with mixtures of normals: specifications and applications to international data," Working Paper, Boston University.

Table 1 Summary Statistics of the Volatility Series

Volatility	*Mean*	*SD*	*Max*	*Min*	*Skew*	*Kur*
S&P 500	-5.21	0.81	-1.47	-6.91	0.04	2.60
Nasdaq	-5.06	0.87	-2.01	-6.91	0.10	2.63
DJIA	-5.19	0.80	-1.36	-6.91	-0.01	2.58
AMEX	-5.09	0.80	-2.07	-6.91	0.03	2.80

Note "SD","Skew" and "Kur" stand for standard deviation, skewness and kurtosis respectively.

Table 2 Maximum Likelihood Estimates of the Basic RLS Model

	σ_η	p	σ_e
S&P 500	0.49*	0.0042*	0.74*
	(0.09)	(0.002)	(0.004)
Nasdaq	0.66*	0.0031	0.75*
	(0.23)	(0.002)	(0.005)
DJIA	0.84*	0.0018	0.74*
	(0.20)	(0.001)	(0.004)
AMEX	0.54*	0.0071	0.73*
	(0.16)	(0.004)	(0.008)

This table gives the parameter estimates of the basuc random level shift model. Standard errors are in parenthese. Estimates with a (*) are significant at the 5% level.

Table 3 Maximum Likelihood Estimates of RLS Model with time varying Probability

Threshold	σ_η	p	σ_e	γ1	γ2
Panel A: S&P 500					
5%	0.27*	-2.60*	0.74*	1.74*	0.76*
	(0.08)	(0.56)	(0.00)	(0.49)	(0.23)
2.5%	0.24*	-2.40*	0.74*	2.43	0.20*
	(0.07)	(0.54)	(0.00)	(1.37)	(0.03)
1%	0.36*	-2.57*	0.74*	2.27	0.12*
	(0.15)	(0.66)	(0.00)	(1.48)	(0.02)
Panel B: Nasdaq					
	σ_η	p	σ_e	γ1	γ2
5%	0.56*	-2.79*	0.75*	1.02*	0.78*
	(0.10)	(0.36)	(0.01)	(0.21)	(0.32)
2.5%	0.49*	-2.72*	0.75*	1.50*	0.52*
	(0.12)	(0.43)	(0.01)	(0.45)	(0.16)
1%	0.43*	-2.59*	0.75*	2.08	0.35*
	(0.11)	(0.48)	(0.01)	(1.10)	(0.08)
Panel C: DJIA					
	σ_η	P	σ_e	γ1	γ2
5%	0.38*	-2.97*	0.74*	1.76*	0.10*
	(0.13)	(0.49)	(0.00)	(0.66)	(0.01)
2.5%	0.39*	-2.78*	0.74*	1.90*	0.95
	(0.09)	(0.38)	(0.00)	(0.51)	(0.61)
1%	0.48*	-2.82*	0.74*	2.33*	0.10*
	(0.10)	(0.35)	(0.00)	(0.74)	(0.01)
Panel D: AMEX					
	σ_η	p	σ_e	γ1	γ2
5%	0.15*	-1.71*	0.73*	5.06	-4.97
	(0.03)	(0.55)	(0.01)	(47.73)	(32.49)
2.5%	0.58*	-2.50*	0.73*	-1.44	0.58
	(0.17)	(0.50)	(0.01)	(3.72)	(0.46)
1%	0.62*	-2.60*	0.73*	2.31	0.30
	(0.26)	(0.62)	(0.01)	(1.60)	(0.24)

Standard errors are reported in parenthese. Estimates with a (*) are significant at the 5% level.

Table 4 Maximum Likelihood Estimates of the RLS Model with Mean Reversion

	σ_η	p	σ_e	β
S&P 500	0.003	0.05*	0.74*	-0.13*
	(0.01)	(0.02)	(0.004)	(0.003)
Nasdaq	0.098	0.02*	0.75*	-0.20*
	(0.08)	(0.01)	(0.005)	(0.010)
DJIA	0.001	0.06*	0.74*	-0.12*
	(0.003)	(0.02)	(0.004)	(0.002)
AMEX	0.001	0.10*	0.72*	-0.16*
	(0.02)	(0.04)	(0.008)	(0.005)

Standard errors are listed in the parenthese. Estimates with a (*) are significant at the 5% level.

Table 5 Maximum Likelihood Estimates of RLS Model with a time varying Probability of shifts and Mean Reversion

Panel A: S&P 500					
σ_η	p	σ_e	γ1	γ2	β
0.004	-1.46*	0.74*	-2.32	0.67*	-0.12*
(0.01)	(0.21)	(0.00)	(2.34)	(0.16)	(0.002)
Panel B: Nasdaq					
σ_η	p	σ_e	γ1	γ2	B
0.07	-1.88*	0.75*	-2.02	0.31*	-0.19*
(0.13)	(0.36)	(0.01)	(1.54)	(0.09)	(0.01)
Panel C: DJIA					
σ_η	p	σ_e	γ1	γ2	β
0.004	-2.41*	0.74*	1.80*	0.65*	-0.27*
(0.01)	(0.47)	(0.00)	(0.41)	(0.18)	(0.02)
Panel D: AMEX					
σ_η	P	σ_e	γ1	γ2	B
0.0008	-1.12*	0.72*	-4.09	-0.15	-0.14*
(0.01)	(0.29)	(0.01)	(23.87)	(0.47)	(0.004)

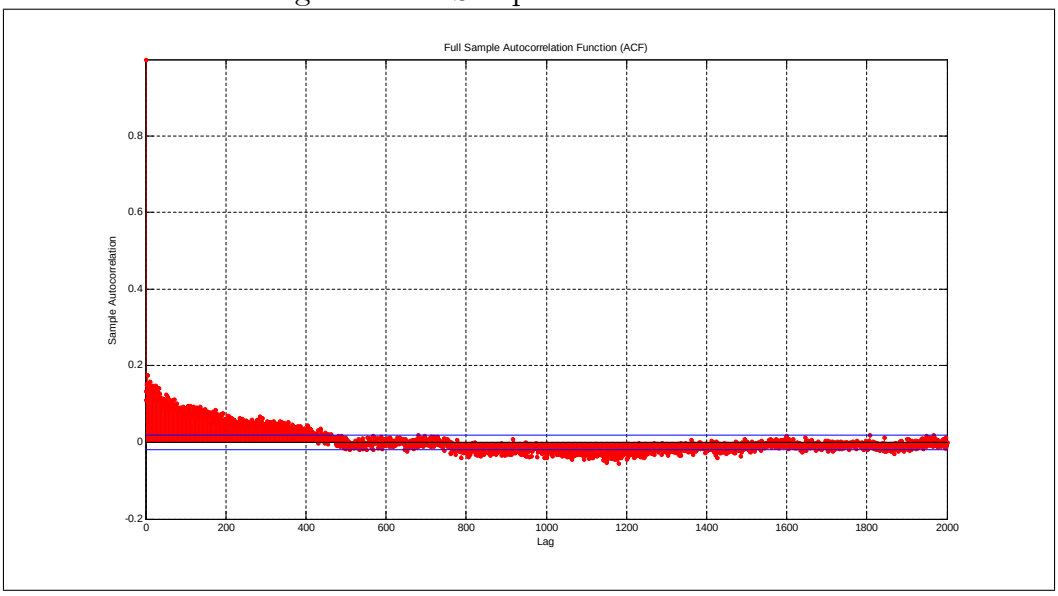
Here we use 1% negative past returns as the threshold. Standard errors are in the parenthese. Estimates with a (*) are significant at the 5% level.

Table 6 Out-of-Sample Forecast Compariosn

	1_step	5_step	10_step	20_step	50_step	100_step
S&P 500						
RLS_modified	0.67	3.95	11.13	37.41	221.74	1027.55
	(1.00*)	(1.00*)	(1.00*)	(1.00*)	(1.00*)	(1.00*)
RLS	0.68	4.11	11.81	40.28	242.11	1140.92
	(0.12*)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
RLS_ARFIMA(1,d,1)	0.68	4.57	13.74	46.73	263.73	1147.98
	(0.14*)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
RLS_ARFIMA(0,d,0)	0.68	4.20	12.12	41.19	242.47	1124.79
	(0.12*)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
ARFIMA(1,d,1)	0.85	8.34	27.83	97.79	516.32	1810.85
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
ARFIMA(0,d,0)	0.87	8.78	29.60	104.92	562.01	2007.72
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Nasdaq						
	1_step	5_step	10_step	20_step	50_step	100_step
RLS_modified	0.74	4.20	11.48	38.01	230.15	1200.17
	(1.00*)	(1.00*)	(1.00*)	(1.00*)	(1.00*)	(0.67*)
RLS	0.742	4.23	11.61	38.44	233.59	1220.22
	(0.93*)	(0.06)	(0.03)	(0.07)	(0.07)	(0.67*)
RLS_ARFIMA(1,d,1)	0.741	4.7	13.7	46.06	266.12	1177.19
	(0.93*)	(0.00)	(0.00)	(0.00)	(0.01)	(1.00*)
RLS_ARFIMA(0,d,0)	0.75	4.31	11.93	39.99	238.76	1223.00
	(0.66*)	(0.03)	(0.02)	(0.01)	(0.07)	(0.67*)
ARFIMA(1,d,1)	0.86	7.13	22.48	76.00	385.10	1284.52
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.14*)
ARFIMA(0,d,0)	0.88	7.70	24.75	85.06	440.86	1503.13
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
DJIA						
	1_step	5_step	10_step	20_step	50_step	100_step
RLS_modified	0.70	4.20	12.14	41.61	271.31	1309.86
	(1.00*)	(1.00*)	(1.00*)	(1.00*)	(1.00*)	(1.00*)
RLS	0.71	4.49	13.32	46.79	311.46	1496.91
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
RLS_ARFIMA(1,d,1)	0.71	5.10	15.97	56.59	351.35	1508.56
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
RLS_ARFIMA(0,d,0)	0.71	4.52	13.43	45.84	309.00	1497.19
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
ARFIMA(1,d,1)	0.87	8.58	26.68	100.80	533.04	1768.50
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
ARFIMA(0,d,0)	0.89	8.95	30.16	106.72	569.54	1916.99
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
AMEX						
	1_step	5_step	10_step	20_step	50_step	100_step
RLS_modified	0.63	3.95	10.88	35.83	198.78	937.34
	(1.00*)	(1.00*)	(1.00*)	(1.00*)	(1.00*)	(1.00*)
RLS	0.63	3.96	11.07	37.27	219.01	1113.03
	(0.99*)	(0.90*)	(0.59*)	(0.26*)	(0.00)	(0.00)
RLS_ARFIMA(1,d,1)	0.64	4.76	14.41	48.21	256.76	962.14
	(0.42*)	(0.01)	(0.00)	(0.00)	(0.00)	(0.55*)
RLS_ARFIMA(0,d,0)	0.64	4.15	11.67	38.01	218.49	999.55
	(0.39*)	(0.21*)	(0.07)	(0.04)	(0.01)	(0.39*)
ARFIMA(1,d,1)	0.73	6.32	19.55	64.64	320.81	1060.45
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.08)
ARFIMA(0,d,0)	0.75	6.76	21.30	71.68	365.16	1239.69
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

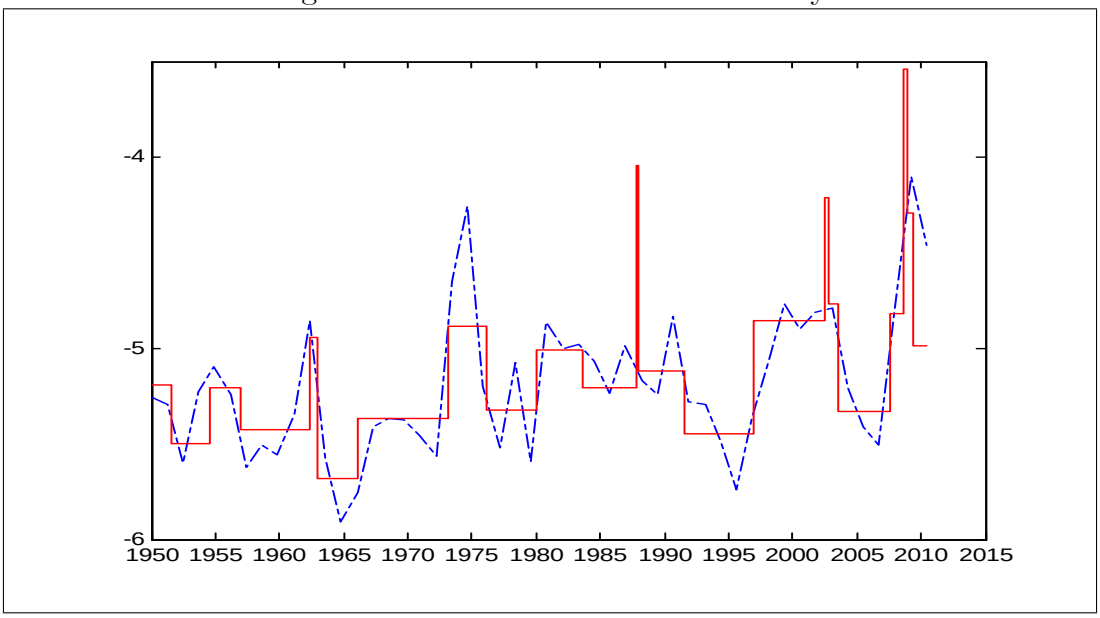
Starting dates for the four series are 10/27/2005(S&P 500), 07/06/2006(Nasdaq,AMEX), 07/03/2006(DJIA). MSFEs are reported, MCS p-values are in the parenthese. The number with a (*) means that the model is within 10% MCS using all comparisons.

Figure 1 Full Sample autocorrelations



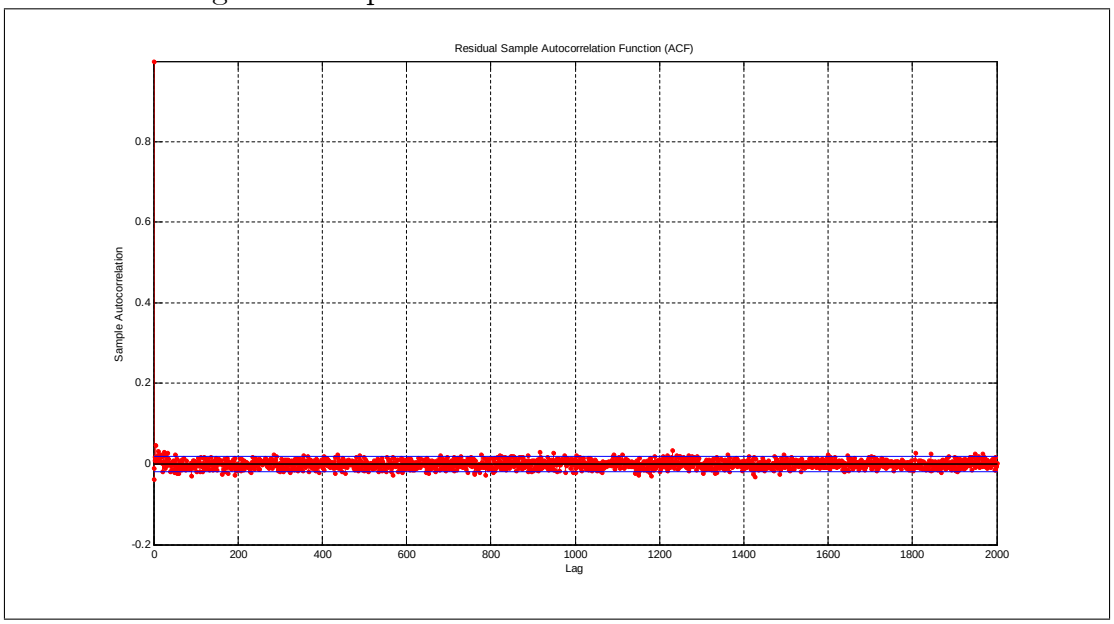
This graph displays the autocorrelation function up to lag 2000 for the last 10,000 observations of the S&P 500 series.

Figure 2 Fitted level shifts and volatility



In this figure, the original process is plotted with the fitted level shift component estimated using Bai and Perron (2003) algorithm.

Figure 3 Sample autocorrelations of S&P 500 residuals



The residuals are calculated as the original process minus the fitted level shift component. This graph shows the autocorrelation function of the residual term up to lag 2000.

Figure 4 S&P 500 Smoothed filter of the level shift components for different thresholds

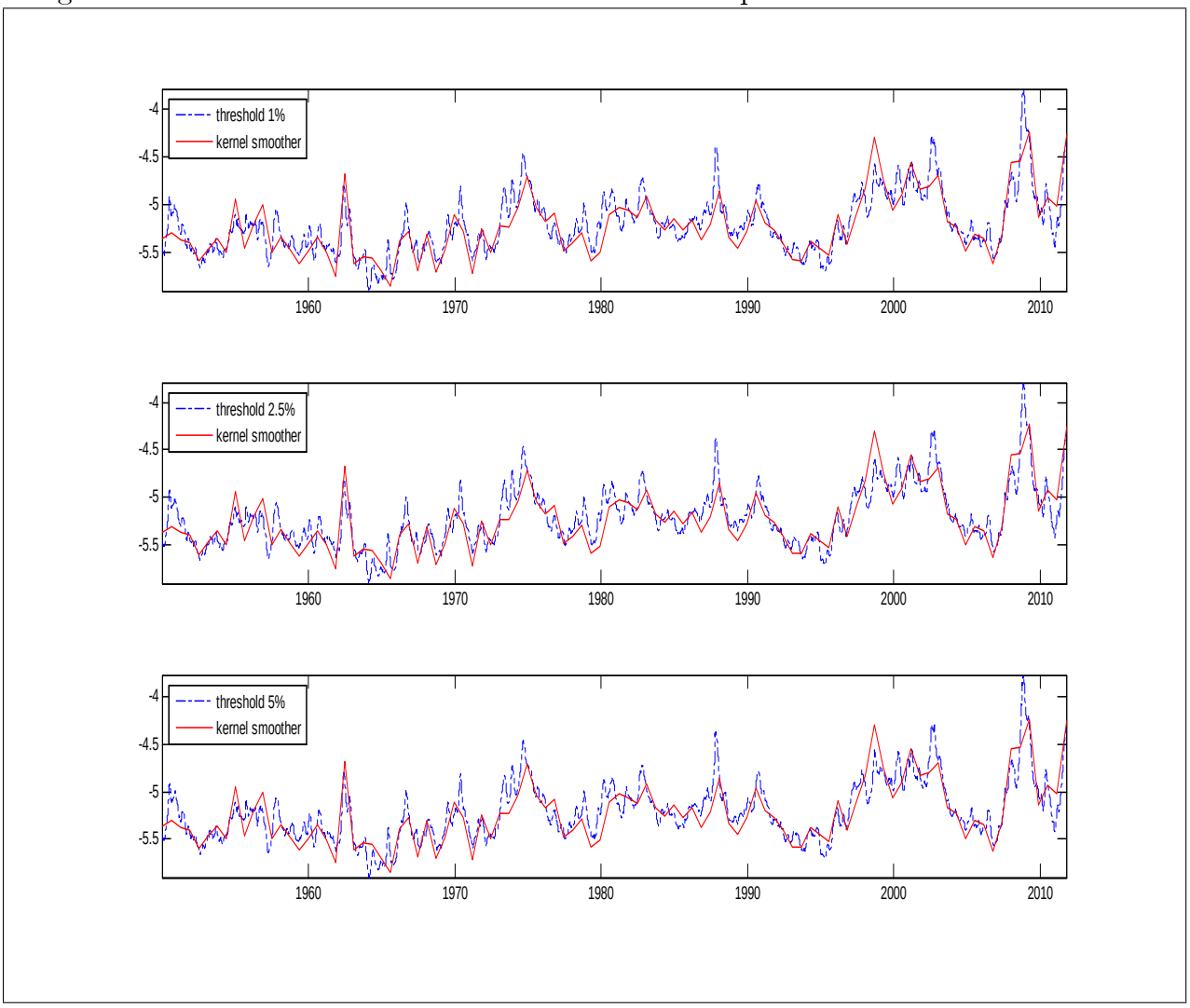


Figure 5 S&P Smoothed filter of the level shift components for different models

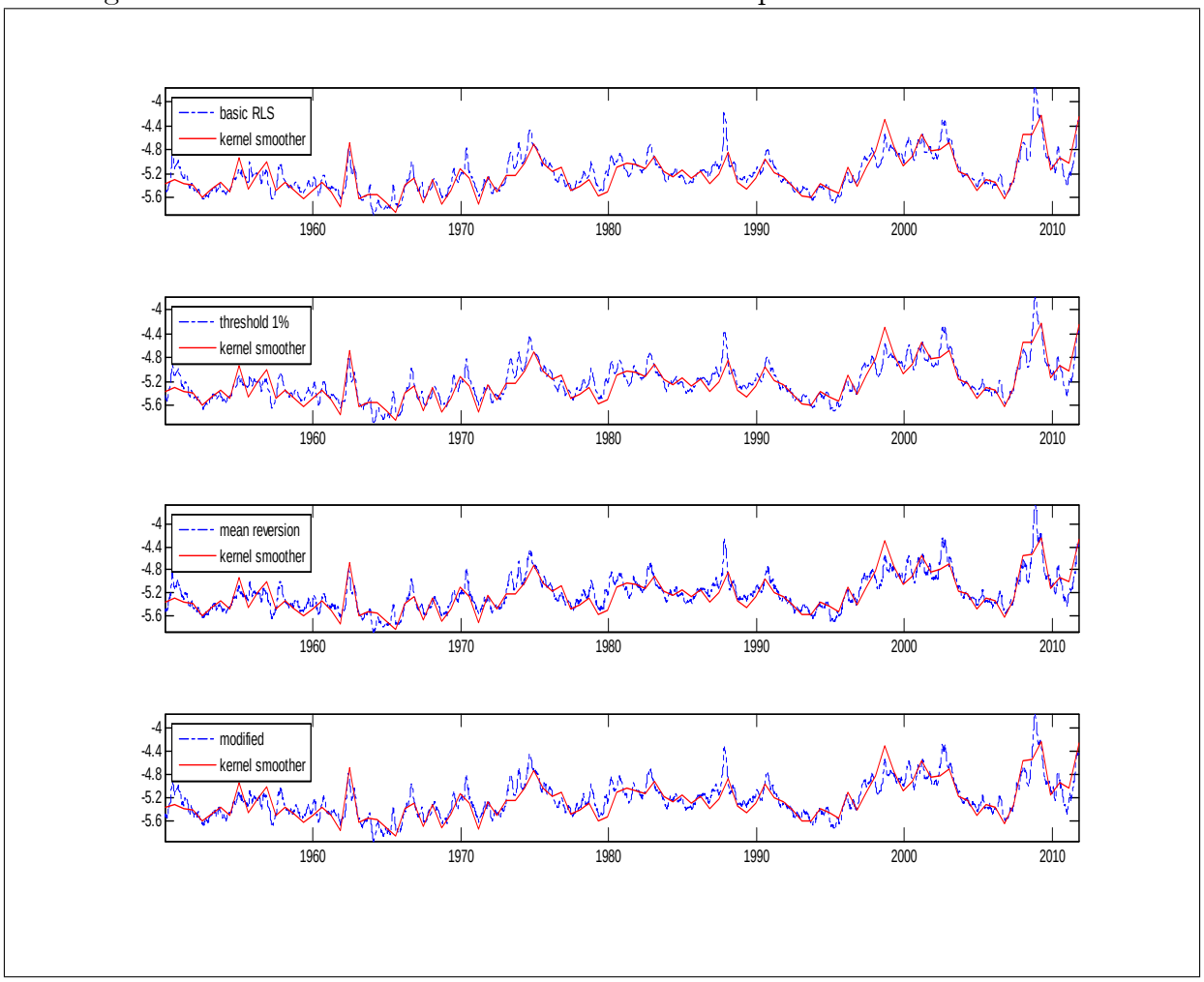
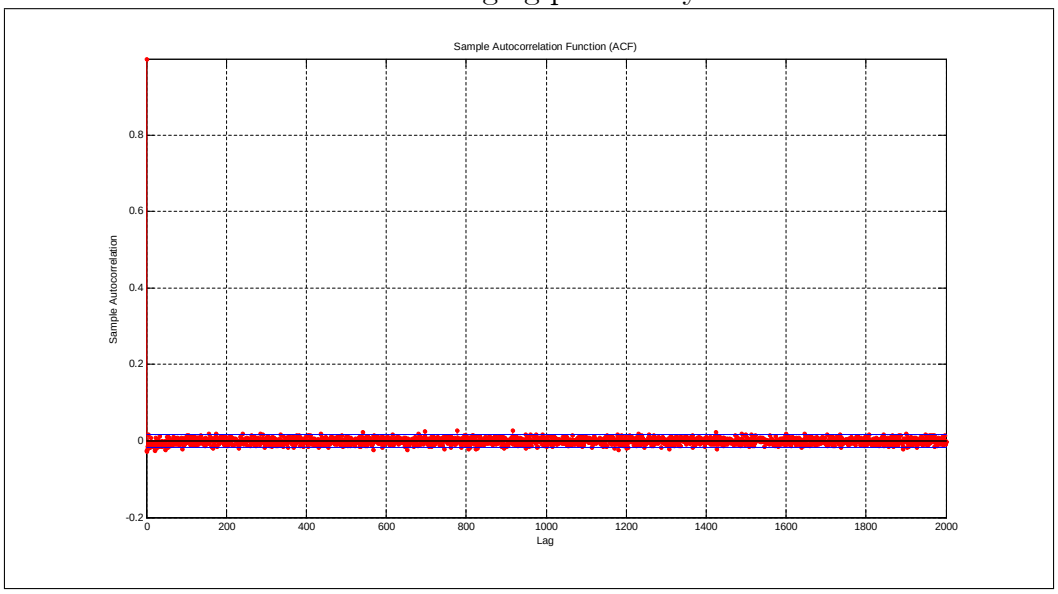


Figure 6 Autocorrelation function of the residual term in RLS with both mean reversion and changing probability



This graph plots the autocorrelation function of the residual term, which is constructed as the difference of volatility series and the smoothed filter of the level shift component.