Forecasting Return Volatility: Level Shifts with Varying Jump Probability and Mean Reversion*

Jiawen Xu†
Shanghai University of Finance and Economics and Key Laboratory of Mathematical Economics

Pierre Perron‡
Boston University

March 1, 2013; Revised November 28, 2013.

Abstract

We extend the random level shift (RLS) model of Lu and Perron (2010) for the volatility of asset prices, which consists of a short memory process and a random level shift component. Motivated by empirical features a) we specify a time-varying probability of shifts as a function of large negative lagged returns; b) we incorporate a mean reverting mechanism so that the sign and magnitude of the jump component change according to the deviations of past jumps from their long run mean. This allows the possibility of forecasting the sign and magnitude of the jumps. We estimate the model using twelve different series. We compare its forecasting performance with a variety of competing models at various horizons. A striking feature is that the modified RLS model has the smallest mean square forecast errors in 64 out of the 72 cases, while it is a close second for the other 8 cases. The improvement in forecast accuracy is often substantial, especially for medium to long-horizon forecasts. This is strong evidence that our modified RLS model offers important gains in forecasting performance.

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

---

*We are grateful to Rasmus Varneskov and Shinsuke Ikeda for providing the realized volatility series used and for useful comments. We also thank an associate editor and two referees for helpful constructive comments.

†School of Economics, Shanghai University of Finance and Economics, 777 Guoding Road, Shanghai, China 200433, and Key Laboratory of Mathematical Economics (SUFE), Ministry of Education, Shanghai 200433, China (xu.jiawen@mail.shufe.edu.cn).

‡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 changes in mean, the estimate of the sum of its autoregressive coefficients 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 a white noise series affected by occasional shifts in the unconditional variance and showed 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 four 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 adopted 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 specified 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 transformed it into a linear state space form with innovations having a mixture of two normal distributions and adopted an algorithm similar to the one used by Perron and Wada (2009) and Wada and Perron (2007). They restricted the variance of one of the two normal distributions to be zero, allowing a simple but efficient algorithm.

Varneskov and Perron (2013) further extended the random level shift model to combine it with a long memory process, modeled as a $ARFIMA(p,d,q)$ process. They provided a forecasting framework for a class of long-memory models with level shifts. Their forecasting experiments using six different data series covering both low frequency and high frequency data showed that the RLS-ARFIMA model outperforms other competing models.

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 change 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 the following daily return series using absolute returns as a proxy for volatility and a logarithmic transformation to have series closer to being normally distributed and also not bounded below by zero: S&P 500 stock market index, Dow Jones Industrial Average (DJIA) index, AMEX index, Nasdaq index, Nikkei 225 index, IBM stock prices, Crude Oil prices, Treasury Bond Futures, Trade Weighted U.S. Dollar Index. To assess the sensitivity of our results, we also consider three realized volatility series, also in logarithmic form, constructed from 5 minutes returns on the S&P 500 and Treasury Bond Futures, as well as a realized volatility series constructed from tick-by-tick trades on the SPY, an exchange traded fund that tracks the S&P 500. 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 eight competing models: the original random level shift model (RLS), the popular $ARFIMA(1, d, 1)$ and $ARFIMA(0, d, 0)$, a GARCH(1,1), a fractionally integrated GARCH model (FIGARCH(1,d,1)), the HAR model, a Multiple Regime Smooth Transition Heterogeneous Autoregressive Model (HARST) and a Markov Regime Switching model. We consider forecast horizons of 1, 5, 10, 20, 50 and 100 days. This gives 72 cases in total. A striking feature is that the modified RLS model has the smallest mean square forecast errors in 64 out of the 72 cases, while it is a close second for the other 8 cases. The improvement in forecast accuracy is often substantial, especially for medium to long-horizon forecasts. 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 extensions of the basic model to allow 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 model. Section 8 provides brief concluding remarks.

2 Data and Summary Statistics

The series used consist of nine daily returns series and three realized volatility series. The daily returns series are: S&P 500 stock market index (01/03/1950-10/11/2011; 15,543 observations), Dow Jones Industrial Average (DJIA) index (01/05/1950/-06/15/2012; 15,752 observations), AMEX index (01/03/1996-06/18/2012; 4,137 observations), Nasdaq index (02/09/1971-06/18/2012; 10,434 observations), Nikkei 225 index (05/18/1949-08/12/2013; 16,000 observations), IBM stock prices (01/06/1970-06/05/2007; 9,444 observations), Crude Oil Prices [1] (01/06/1986-08/06/2013; 6,960 observations), Treasury Bond Futures (01/05/1983-06/11/2009; 6,639 observations), Trade Weighted U.S. Dollar Index: Major Currencies (DTWEXM) [2], (01/04/1973-08/16/2013; 10,180 observations). For these daily series, 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 absolute returns and a logarithmic transformation is applied to have series closer to being normally distributed and also not bounded below by zero. In order to avoid extreme negative values, we bound absolute returns away from zero by adding a small constant 0.001, so that the volatility series used is $y_t = \ln(|r_t| + 0.001)$. The percentage of zero values is indicated in Table 1. They are quite small, the highest being 3.8% for the Treasury Bond Futures [3].

[1] West Texas Intermediate (WTI) - Cushing, Oklahoma (DCOILWTICO); units are dollars per barrel.

[2] A weighted average of the foreign exchange value of the U.S. dollar against a subset of the broad index currencies that circulate widely outside the country of issue. The major currencies index includes the Euro Area, Canada, Japan, United Kingdom, Switzerland, Australia, and Sweden. Unit are normalized with the value in March 1973 set to 100.

[3] The series for S&P 500, Nasdaq, Dow Jones, AMEX, and IBM are from Yahoo finance (http://finance.yahoo.com/). The series for Nikkei 225 index, Oil price, and Trade Weighted U.S. Dollar Index (DTWEXM) are from the Federal Reserve Bank of St. Louis (http://research.stlouisfed.org/). The Treasury Bond Futures series was provided by Rasmus Varneskov and is the daily version of the corresponding realized volatility series also used. The data was obtained from "Tickdata" (http://www.tickdata.com/).

To assess the robustness of our results, we also consider three realized volatility series, also in logarithmic form, constructed from 5 minutes returns on the S&P 500 (04/22/1982-03/02/2007; 6,261 observations) and Treasury Bond Futures (01/05/1983-06/11/2009; 6,639 observations), as well as a realized volatility series constructed from tick-by-tick trades on the SPY, an exchange traded fund that tracks the S&P 500 (01/03/1997-07/02/2008; 2,913 observations). The realized volatility series for the S&P 500 Futures was provided by Shinsuke Ikeda; see Ikeda (2013) for details about how the original data was cleaned. The realized volatility series on the Treasury Bond Futures and the SPY were provided by Rasmus Varneskov. The realized volatility series for the S&P 500 and Treasury Bond Futures were constructed from 5-minutes returns, i.e., $\log([\sum_{t=1}^{n} r_t^2]^{1/2})$ where r_t are 5-minutes returns and n is the number of such returns within a day. The construction of the realized volatility series for the SPY is more involved as it uses all tick-by-tick data, obtained from Asger Lunde and cleaned using the procedure in Barndorff-Nielsen et al. (2009). It is based on the modulated realized volatility approach of Podolskij and Vetter (2009a,b) which accounts for microstructure noise and jumps. The obtained realized volatility series, say $C_{MRV,t}$, is given by $C_{MRV,t} = \log[\sqrt{C_{MRV,t}^2}]$ where $C_{MRV,t}^2 = MRV_t - J_t$, with $MRV_t = (c_{1,t}c_2/\nu_{1,t})MBV(2,0)_t - (\nu_{2,t}/\nu_{1,t})\hat{\omega}_t^2$, $MBV(2,0)_t = \sum_{j=1}^{M_t} |\bar{p}_{t,j}^{(K_t)}|^2$ and $\bar{p}_{t,j}^{(K_t)} = (n_t/M_t - K_t + 1)^{-1} \sum_{i=(j-1)n/M_t}^{jn_t/M_t - K_t} (p_{\tau_{t,i+K_t}} - p_{\tau_{t,i}})$ is the j-th averaged return, $\tau_{t,i} = i/n_t$ ($i = 0, \cdots, n_t$) is the intra-daily time stamp, $K_t = c_{1,t}n_t^{1/2}$, and $M_t = n_t^{1/2}/(c_{1,t}c_2)$ for some optimally determined coefficients $c_{1,t}$, c_2, $\nu_{1,t}$ and $\nu_{2,t}$ (as defined in Podolskij and Vetter, 2009a). Also, $\hat{\omega}_t^2$ is the $\sqrt{n_t}$-consistent noise variance estimator $\hat{\omega}_t^2 = (2n_t)^{-1} \sum_{i=1}^{n_t} (p_{\tau_{t,i}} - p_{\tau_{t,i-1}})^2$. The jump component is J_t and its construction follows the method recommended by Podolskij and Vetter (2009a).

Table 1 gives summary statistics of those volatility proxies and shows their unconditional distribution characteristics. The daily series have similar characteristics: mean, standard deviation and extreme values. The skewness is small in absolute value (from -.58 to .22) and the kurtosis ranges from 2.42 to 2.94, slightly lower than 3 for the normal distribution. One exception is the Treasury Bond Futures series, which has a high kurtosis value of 7.47. For the realized volatility series, the skewness is more positive and the kurtosis somewhat higher (between 2.72 and 4.70).

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

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

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 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 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 the 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 $\sum_{i=1}^{m+1} \sum_{t=T_{i-1}+1}^{T_i} [y_t - \mu_i]^2$, 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 events 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 of the residuals defined as the difference between the original process and the fitted level shift component. A distinctive feature is that now the residuals essentially exhibit 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 are to allow for changes in the probability of shifts and to 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 ensures $p_t \in [0, 1]$. 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 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:

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

where $\Phi(.)$ is a normal cdf function, so that $f(p, x_{t-1})$ is 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 movements 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:

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

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 $\bar{\tau}_t$ 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 dictating the speed of reversion. Note that the specification involves using data only up to time t in order to be useful for forecasting. Also 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 Markov switching models, especially since 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

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

and

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

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

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

or more compactly by $\hat{\xi}_{t+1|t} = \Pi \hat{\xi}_{t|t}$ with $\hat{\xi}_{t|t} = [\hat{\xi}_{t|t-1} \odot \omega_t]/\mathbf{1}'(\hat{\xi}_{t|t-1} \odot \omega_t)$. 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 since we condition 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 variance conditional on past information and $s_{t-1} = i$ are

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

The measurement equation is $\triangle y_t = H X_t + \delta_t$, where the measurement error δ_t has mean zero and a variance taking two possible values: $R_1 = \sigma^2_\eta$, with probability $p_{t,}$ or $R_2 = 0$, with probability $1 - p_t$. Hence, the prediction error is $v_t^{ij} = \triangle y_t - H X^i_{t|t-1}$with variance $f_t^{ij} = H P^i_{t|t-1} H' + R_j$. From the updating formulas, we have given $s_t = j$ and $s_{t-1} = i$,

$$\begin{aligned} X^{ij}_{t|t} &= X^i_{t|t-1} + P^i_{t|t-1} H' (H P^i_{t|t-1} H' + R_j)^{-1} (\triangle y_t - H X^i_{t|t-1}) \\ P^{ij}_{t|t} &= P^i_{t|t-1} - P^i_{t|t-1} H' (H P^i_{t|t-1} H' + R_j)^{-1} H P^i_{t|t-1} \end{aligned} \tag{8}$$

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

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

By doing so, ω^{ij}_t is unaffected by the history of states before time $t-1$. Some modifications are needed when including the mean reverting mechanism. In equation (6), the prediction error v^{ij}_t 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:

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

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,

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

Since $y_t = a + \tau_t + c_t$, then $\tau^{i1}_{t|t-1} = y_t - a - c^i_{t|t-1} = y_t - a - [0\ 1]'\ X^{i1}_{t|t-1}$. Note that $X^{i1}_{t|t-1}$ being a state variable it can be updated every time period. Therefore, $\tau^{i1}_{t|t-1} - \bar{\tau}^{i1}_t$ is known at time t. Also $R_1 = \sigma^2_\eta$ with probability p_t and $R_2 = 0$ with probability $1-p_t$. The standard errors were computed from the numerical Hessian.

6 Full Sample Estimation Results

We first consider full sample estimation results for sub-cases of the general model to highlight the contributions of each components and compare our results with those of Lu and Perron (2010). For this part, we consider only the same series as they do, namely S&P 500, Nasdaq, DJIA and AMEX. We then consider the estimation results of the full model for all series.

6.1 Results for sub-models

We first present results from estimating the basic random level shift model using the U.S. stock indices 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 S&P 500, 32 for Nasdaq, 28 for DJIA and 29 for AMEX. Since our S&P 500 data covers a longer period, our point estimate of the number of jumps is also higher. This is especially the case since our sample further includes the period 2008 to 2011, a time during which stock markets went through a turbulent period induced by the financial crisis in 2008. 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 a 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 that 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 4 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 the other U.S. 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 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 accounts 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.

6.2 Results for the full model

Table 5 presents the estimates of the modified RLS model combining both time varying probabilities and mean reversion, using a threshold value of 1% for the full set of 12 series considered [4]. First, in all cases the estimate of β is significantly negative, indicating the presence of a mean-reverting property for the level shift component for all series. In the case of the S&P 500, Nasdaq, DJIA and AMEX series, 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

[4] For the realized volatility series of the SPY and S&P 500 Futures, only the absolute value of the daily returns was available to us. Hence, for these two series we used extreme absolute values of past returns instead of extreme negative returns to forecast the probability of shifts. This should not affect the results given that we use the extreme 1% of the distribution of returns and very large negative returns are more common than very large positive returns.

positive, except for the AMEX index, the T-Bond and DTWEXM, though for the latter two the values are very small. On the other hand, the estimates of γ_1, pertaining to the component $1\{x_{t-1} < 0\}$ are not significant, except for DJIA, IBM and T-Bond (for Nikkei 225, it is significant but very small). 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 series 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 uncorrelated, thereby justifying the specification of the nature of the short-memory component and re-enforcing the conclusion that once level shifts are taken into account the long-memory feature 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 (2013), the τ-step ahead forecasts of the basic random level shift model is given by

$$\hat{y}_{t+\tau|t} = y_t + HF^{\tau}[\sum_{i=1}^{2}\sum_{j=1}^{2}\Pr(s_{t+1} = j)\Pr(s_t = i|Y_t)X_{t|t}^{ij}]$$

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 $\Pr(s_{t+1} = j)$; see Section 5.

We compare the forecasting performance of our model with eight other models: 1) the original RLS model; 2) the popular ARFIMA(1,d,1); 3) the ARFIMA(0,d,0); 4) the GARCH(1,1) (Bollerslev, 1986); 5) the FIGARCH(1,1) (Baillie et al., 1996); 6) a two-state

Markov regime switching model (Hamilton, 1994) defined by [5] $y_t = \mu_{S_t} + \varepsilon_t$ where $S_t = 1, 2$ and $\varepsilon_t \sim i.i.d.\ N(0, \sigma^2_{S_t})$ with an unconstrained transition matrix P; 7) the HAR model (Corsi, 2009, Chiriac and Voev, 2011):

$$z^{(d)}_{t+1} = \alpha + \beta_1 z^{(d)}_t + \beta_2 z^{(w)}_t + \beta_3 z^{(bw)}_t + \beta_4 z^{(m)}_t + \varepsilon^{(d)}_t$$

where d, w, bw and m stand for a daily, weekly (5 days), biweekly (10 days), and monthly (22 days) sampling frequency, respectively, α is a constant and $\varepsilon^{(d)}_t \sim i.i.d. N(0, \sigma^2_\varepsilon)$. The regressors on the right hand side are constructed as averages of past values, e.g., $z^{(w)}_t = (1/5)\sum_{i=0}^{4} z_{t-i}$; 8) Multiple-Regime Smooth Transition Heterogeneous Autoregressive (HARST) model (McAleer & Medeiros, 2008) [6]. Let $y_{t,h} = (y_t + y_{t-1} + \cdots + y_{t-h+1})/h$, and $\iota = (\iota_1, ..., \iota_p)'$ be a set of indices where $\iota_1 < \cdots < \iota_p$, and $x_t = (1, y_{t-1,\iota_1}, ..., y_{t-1,\iota_p})'$. A time series $\{y_t\}_{t=1}^{T}$ follows a HARST model with $M + 1$ limiting regimes if

$$y_t = \beta_0' x_t + \sum_{m=1}^{M} \beta_m' x_t f(z_t; \gamma_m, c_m) + \varepsilon_t$$

where $f(z_t; \gamma_m, c_m) = [1 + \exp(-\gamma_m(z_t - c_m))]^{-1}$ is the logistic function and z_t is a covariate that affects the transitions. Here, we use $x_t = (1, y_{t-1}, y_{t-1,5}, y_{t-1,22})$ and $z_t = r_{t-1}$. The value of M is selected according to a sequential testing procedure as described in McAleer and Madeiros (2008, Section 4).

We use the following forecasting horizons: $\tau = 1, 5, 10, 20, 50$ and 100. For the multi-step forecasts, we use the indirect or iterative method for all eight competing models. The mean square forecast error (MSFE) criterion, analyzed by Hansen & Lunde (2006) and Patton (2011), is:

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

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 the models is assessed using the relative MSFEs, which provides a consistent ranking. For the realized volatility series, we do not include the GARCH(1,1) and FIGARCH(1,1) models as these are estimated from returns series.

The forecasting experiment is as follows. We keep the last 1500 observations as the out-of-sample period to be forecasted. The reasons for considering this period is that it contains very

[5] For the estimation, we used the Matlab codes based on MS_Regress, the MATLAB Package for Markov Regime Switching Models by Marcelo Perlin.

[6] We used the code for estimation available on Marcelo Medeiros' website.

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 the RLS and RLS-modified models is quite time consuming, we opted for a fixed pre-forecast window whereby we estimate these two models once without the last 1500 observations. The forecasts are then made conditional on the parameter estimates obtained. For the other models, we re-estimated them every period in order to make the results as strong as possible. It is indeed the case that many models can adapt to changing structures and provide better forecasts when estimated using all available observations. This was indeed the case here. For the competing models, we tried three forecasting schemes: fixed (as for the RLS and RLS-modified), recursive and a 4-years rolling window. For all models, except the GARCH(1,1), the recursive scheme provided the best forecast. The forecasts obtained using the rolling window were noticeably inferior. These results concur with those of Brownlees et al. (2012). In any event, the qualitative conclusions remained unchanged irrespective of the forecasting scheme used for the competing models. For example, when using the fixed scheme for all models, one can then compare their relative forecasting performance using the Model Confidence Set of Hansen et al. (2011). The results showed that our modified RLS belong to the 10% MCS using all comparisons for all assets and all forecast horizons. It also yielded the smallest MSFE in 67 of 72 cases. Since we re-estimate the competing models every period for the competing models, we cannot use the MCS of Hansen et al. (2012). Hence, we simply report the MSFE and compare the various models using this criterion. Accordingly, the results should be viewed as providing a lower bound on the relative advantage of our modified RLS model.

The results are presented in Table 6. The most striking feature is that the smallest MSFE values are obtained with the modified random level shift model in 64 out of the 72 cases. The cases for which our modified RLS does not have the smallest MSFE are: 1) the T-Bond at forecast horizons 5 and 10 days, the DTWEXM at horizon 1 day, the RV-S&P 500 at horizon 5 days and the SPY at horizon 1 day, in which cases the basic RLS model performs best; 2) the Nasdaq at horizon 100 days and the RV-S&P 500 at horizons 50 and 100 days, in which case the HAR performs best. In these eight cases, the RLS-modified is a close second best.

Amongst the non-RLS models, the HAR performs best in 62 out of 72 cases, the HARST in 8 and the ARFIMA(1,d,1) in 2. For the daily series, the worst performing models are the GARCH(1,1) in 25 out of 54 cases and the FIGARCH in 29 cases. For the realized volatility series, the worst performing models are the ARFIMA(1,d,1) in 14 cases, the HARST in 2 cases and the Regime-Switching in 2 cases. Hence, a useful benchmark to assess the improvement in forecasting performance is to compare the relative MSFE of the RLS-modified and the HAR. For medium term forecasting horizons (5 to 20 days ahead), the MSFEs of the modified RLS is between 67% and 98% of those of the HAR models. This is strong evidence that the RLS-modified model provides substantial improvements in forecast accuracy over a range of competing models. It does so for a wide variety of assets and a wide range of forecast horizons. It also provides substantial improvements over the original RLS model, which in turn outperforms all other models by a considerable margin, and more so as the forecast horizon increases.

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 useful to understand the behavior of various volatility series. More importantly, the extended model allows much improved forecasts of volatility when applied to a variety of assets. Hence, our results provide additional evidence that random level shift models are serious contenders to model volatility and outperform 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.

Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., & Shephard, N. (2009). Realized kernels in practice: trades and quotes. *Econometrics Journal*, 12, C1-C32.

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

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.

Brownlees, C., Engle, R. F., & Kelly, B. (2012). A practical guide to volatility forecasting through calm and storm. *Journal of Risk*, 14 (2), 3-22.

Chiriac, R., & Voev, V. (2011). Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics*, *28*, 922-947.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, *7*, 174-196.

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

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

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

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

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

Granger, C. W. J., & 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., & Lunde, A. (2006). Consistent ranking of volatility models. *Journal of Econometrics*, *131*, 97-121.

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

Harrison, P. J., & Stevens, C. F. (1976). Bayesian forecasting. *Journal of the Royal Statistical Society, Series B, 38*, 205-247.

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.

Ikeda, S. (2013). Two scale realized kernels: a univariate case. Forthcoming in *Journal of Financial Econometrics* (doi: 10.1093/jjfinec/nbt024).

Lobato, I. N., & 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., & Perron, P. (2010). Modeling and forecasting stock return volatility using a random level shift model. *Journal of Empirical Finance*, *17*, 138-156.

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

McAleer, M., & Medeiros, M. C. (2008). A multiple regime smooth transition heterogeneous autoregressive model for long memory and asymmetries. *Journal of Econometrics*, *147*, 104-119.

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

Morana, C., & 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.

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

Podolskij, M., & Vetter, M. (2009a). Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli*, 15, 634-658.

Podolskij, M., & Vetter, M. (2009b). Bipower-type estimation in a noisy diffusion setting. *Stochastic Processes and their Applications*, 119, 2803-2831.

Qu, Z., & 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., & 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., & Perron, P. (2013). Combining long memory and level shifts in modeling and forecasting the volatility of asset returns. Manuscript, Department of Economics, Boston University.

Wada, T., & Perron, P. (2007). State space models with mixtures of normals: specifications and applications to international data. Manuscript, Department of Economics, Boston University.

Table 1: Summary Statistics of the Volatility Series

Series	Mean	SD	Max	Min	Skew	Kur	Period	Number of Observations	percentages of zero returns
S&P 500	-5.21	0.81	-1.47	-6.91	0.04	2.57	01/05/1950-10/11/2011	15543	0.8%
Nasdaq	-5.06	0.87	-2.01	-6.91	0.10	2.63	02/09/1971-06/18/2012	10434	0.35%
DJIA	-5.19	0.80	-1.36	-6.91	-0.01	2.58	01/03/1950-06/15/2012	15752	0.64%
AMEX	-5.09	0.80	-2.07	-6.91	0.03	2.80	01/03/1996-06/18/2012	4137	0.12%
Nikkei 225	-5.03	0.86	-1.82	-6.91	-0.05	2.55	05/18/1949-08/12/2013	16000	0.58%
IBM	-4.73	0.90	-1.34	-6.91	-0.34	2.89	01/06/1970-06/05/2007	9444	3.24%
Oil	-4.41	0.97	-0.90	-6.91	-0.38	2.94	01/06/1986-08/06/2013	6960	1.78%
T-bond	-1.82	1.38	1.30	-6.91	-1.80	7.47	01/05/1983-06/11/2009	6639	3.8%
DTWEXM	-5.72	0.65	-2.89	-6.91	0.22	2.42	01/04/1973-08/16/2013	10180	0.04%
RV T-bond	-2.16	0.81	1.83	-6.32	0.45	3.69	01/05/1983-06/11/2009	6639	0.00%
RV S&P500	-0.56	0.87	6.41	-5.63	0.52	4.70	04/22/1982-03/02/2007	6261	0.00%
SPY	-0.34	0.47	1.44	-1.89	0.10	2.72	01/03/1997-07/02/2008	2913	0.00%

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)

Table 3: Maximum Likelihood Estimates of the RLS Model with Time Varying Probability

Threshold	σ_η	p	σ_e	γ_1	γ_2
			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	γ_1	γ_2
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	γ_1	γ_2
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	γ_1	γ_2
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)

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)

Table 5: Maximum Likelihood Estimates of RLS Model with Time Varying Probability of Shifts and Mean Reversion

	σ_η	p	σ_e	γ_1	γ_2	β
S&P 500	0.004	-1.46*	0.74*	-2.32	0.67*	-0.12*
	(0.01)	(0.21)	(0.00)	(2.34)	(0.16)	(0.002)
Nasdaq	0.07	-1.88*	0.75*	-2.02	0.31*	-0.19*
	(0.13)	(0.36)	(0.01)	(1.54)	(0.09)	(0.01)
DJIA	0.004	-2.41*	0.74*	1.80*	0.65*	-0.27*
	(0.01)	(0.47)	(0.00)	(0.41)	(0.18)	(0.02)
AMEX	0.0008	-1.12*	0.72*	-4.09	-0.15	-0.14*
	(0.01)	(0.29)	(0.01)	(23.87)	(0.47)	(0.004)
Nikkei 225	0.10*	-1.74*	0.77*	0.01*	0.72*	-0.19*
	(0.04)	(0.28)	(0.00)	(0.00)	(0.14)	(0.01)
IBM	0.04	-2.15*	0.85*	0.70*	0.31*	-0.15*
	(0.05)	(0.73)	(0.01)	(0.26)	(0.05)	(0.01)
Oil	0.09*	-1.37*	0.91*	0.31	0.23*	-0.07*
	(0.03)	(0.05)	(0.01)	(0.46)	(0.04)	(0.00)
T-bond	0.57*	-2.84*	1.35*	-0.41*	-0.03*	-0.21*
	(0.09)	(0.38)	(0.01)	(0.14)	(0.00)	(0.01)
DTWEXM	0.06*	-1.47*	0.60*	-1.98	-0.001*	-0.08*
	(0.02)	(0.28)	(0.00)	(3.07)	(0.00)	(0.00)
RV T-bond	0.30*	-1.57*	0.57*	-1.99	0.01*	-0.22*
	(0.04)	(0.19)	(0.01)	5.74	(0.00)	(0.01)
RV S&P500	0.64*	-1.45*	0.40*	0.30	0.52*	-0.16*
	(0.05)	(0.11)	(0.01)	(0.22)	(0.04)	(0.00)
SPY	0.15*	-0.39*	0.20*	0.17	0.59	-0.18*
	(0.02)	(0.14)	(0.01)	(2.36)	(0.49)	(0.01)

Table 6: Out-of-Sample Forecast Comparisons based on Mean Square Forecast Errors

			S&P 500			
	1_step	5_step	10_step	20_step	50_step	100_step
RLS_modified	0.673*	3.95*	11.1*	37*	222*	1028*
RLS	0.679	4.11	11.8	40	242	1141
ARFIMA(1,d,1)	0.851	8.34	27.8	98	516	1811
ARFIMA(0,d,0)	0.869	8.78	29.6	105	562	2008
Regime Switching	0.810	7.40	24.3	86	465	1715
HAR	0.686	4.41	13.1	46	301	1407
HARST	0.683	4.21	12.5	51	418	1666
GARCH(1,1)	1.015	12.75	46.5	180	1135	4771
FIGARCH(1,1)	0.977	11.90	44.0	109	585	2283
			Nasdaq			
	1_step	5_step	10_step	20_step	50_step	100_step
RLS_modified	0.741*	4.20*	11.5*	38.0*	230*	1200
RLS	0.742	4.23	11.6	38.4	234	1220
ARFIMA(1,d,1)	0.858	7.13	22.5	76.0	385	1285
ARFIMA(0,d,0)	0.881	7.70	24.7	85.1	441	1503
Regime Switching	0.827	6.37	19.5	64.3	337	1292
HAR	0.746	4.47	12.7	42.6	236	1072*
HARST	0.755	4.56	13.2	49.4	361	1363
GARCH(1,1)	1.037	11.90	43.0	169.0	1116	5191
FIGARCH(1,1)	1.017	11.59	42.2	166.3	1063	4472
			DJIA			
	1_step	5_step	10_step	20_step	50_step	100_step
RLS_modified	0.696*	4.20*	12.1*	42*	271*	1310*
RLS	0.706	4.49	13.3	47	311	1497
ARFIMA(1,d,1)	0.875	8.58	28.7	101	533	1768
ARFIMA(0,d,0)	0.889	8.95	30.2	107	570	1917
Regime Switching	0.815	7.17	23.5	83	470	1700
HAR	0.707	4.59	13.7	48	324	1420
HARST	0.708	4.61	15.0	68	608	2545
GARCH(1,1)	1.022	12.50	45.4	175	1102	4428
FIGARCH(1,1)	1.016	12.51	46.5	185	1221	5051
			AMEX			
	1_step	5_step	10_step	20_step	50_step	100_step
RLS_modified	0.6333*	3.95*	10.9*	36*	199*	937*
RLS	0.6334	3.96	11.1	37	219	1113
ARFIMA(1,d,1)	0.7321	6.32	19.5	65	321	1060
ARFIMA(0,d,0)	0.7499	6.76	21.3	72	365	1240
Regime Switching	0.8089	8.11	26.3	90	475	1736
HAR	0.6448	4.33	12.7	44	287	1305
HARST	0.6374	4.47	14.6	59	440	1795
GARCH(1,1)	0.8537	9.54	33.0	119	662	2486
FIGARCH(1,1)	0.8618	10.06	36.1	140	869	3540

Note: A * indicates the smallest Mean Squared Forecast Error across models for a given forecast horizon.

Table 6 (cont'd): Out-of-Sample Forecast Comparisons based on Mean Square Forecast Errors

	1_step	5_step	10_step	20_step	50_step	100_step
Nikkei 225						
RLS_modified	0.753*	4.56*	12.6*	39*	232*	1023*
RLS	0.756	4.70	13.4	43	269	1172
ARFIMA(1,d,1)	0.857	7.10	21.8	70	348	1137
ARFIMA(0,d,0)	0.888	7.88	24.9	83	428	1475
Regime Switching	0.829	6.38	19.0	61	328	1348
HAR	0.758	4.93	14.2	46	290	1336
HARST	0.868	5.76	17.0	56	879	1687
GARCH(1,1)	1.059	12.27	44.0	168	1115	4768
FIGARCH(1,1)	1.014	11.57	41.5	158	981	3771
IBM						
	1_step	5_step	10_step	20_step	50_step	100_step
RLS_modified	0.659*	3.80*	9.5*	23.7*	102*	417*
RLS	0.659*	3.81	9.6	24.0	108	466
ARFIMA(1,d,1)	0.830	7.99	26.0	87.9	464	1604
ARFIMA(0,d,0)	0.807	7.43	23.7	79.3	417	1449
Regime Switching	0.774	6.62	20.6	67.3	360	1339
HAR	0.678	4.34	12.0	35.5	210	950
HARST	0.676	4.49	14.2	56.5	381	1350
GARCH(1,1)	1.043	13.53	48.8	186.8	1224	5454
FIGARCH(1,1)	1.087	15.06	56.2	220.8	1469	6418
Oil						
	1_step	5_step	10_step	20_step	50_step	100_step
RLS_modified	0.811*	4.94*	11.3*	29*	149*	673*
RLS	0.812	4.97	11.4	30	161	761
ARFIMA(1,d,1)	0.944	8.18	24.1	80	428	1583
ARFIMA(0,d,0)	0.921	7.60	21.8	70	367	1324
Regime Switching	0.919	7.55	21.5	69	361	1307
HAR	0.826	5.24	12.5	36	211	937
HARST	0.820	5.26	13.3	44	289	1158
GARCH(1,1)	1.197	14.70	51.3	197	1298	5819
FIGARCH(1,1)	1.218	15.68	56.5	225	1615	8505
T-Bond						
	1_step	5_step	10_step	20_step	50_step	100_step
RLS_modified	1.309*	6.86	13.74	30.8*	143*	517*
RLS	1.309*	6.84*	13.71*	31.0	145	536
ARFIMA(1,d,1)	1.364	8.20	19.33	54.0	274	954
ARFIMA(0,d,0)	1.366	8.26	19.55	55.0	280	990
Regime Switching	1.369	8.35	19.93	56.5	292	1048
HAR	1.327	7.13	14.85	38.0	219	884
HARST	1.325	7.33	16.03	43.5	268	1036
GARCH(1,1)	1.850	20.52	69.63	262.2	1679	7151
FIGARCH(1,1)	1.857	20.59	69.66	262.8	1692	7200

Table 6 (cont'd): Out-of-Sample Forecast Comparisons based on Mean Square Forecast Errors

	DTWEXM					
	1_step	5_step	10_step	20_step	50_step	100_step
RLS_modified	0.4531	2.258*	4.98*	14.2*	75*	317*
RLS	0.4529*	2.264	5.04	14.4	77	324
ARFIMA(1,d,1)	0.5033	3.563	10.19	33.7	176	618
ARFIMA(0,d,0)	0.5041	3.581	10.26	33.9	178	620
Regime Switching	0.4901	3.256	9.02	29.5	159	599
HAR	0.4574	2.351	5.59	17.1	104	470
HARST	0.4579	2.407	6.01	23.5	190	722
GARCH(1,1)	0.6442	7.165	25.03	97.4	644	2855
FIGARCH(1,1)	0.6283	6.643	22.52	83.9	513	2161
	RV T-bond					
	1_step	5_step	10_step	20_step	50_step	100_step
RLS_modified	0.389*	2.83*	8.3*	29.9*	203*	1012*
RLS	0.391	2.86	8.4	30.4	214	1103
ARFIMA(1,d,1)	0.930	16.19	61.0	235.6	1418	5455
ARFIMA(0,d,0)	0.829	13.64	50.9	195.2	1169	4464
Regime Switching	0.748	11.67	43.2	168.0	1064	4370
HAR	0.403	3.14	9.6	37.4	319	1935
HARST	0.397	4.25	20.7	109.7	898	3694
	RV SP500					
	1_step	5_step	10_step	20_step	50_step	100_step
RLS_modified	0.197*	3.14	12.95*	59*	481	2266
RLS	0.198	3.12*	13.05	60	496	2349
ARFIMA(1,d,1)	1.074	23.86	92.16	353	2046	7549
ARFIMA(0,d,0)	1.035	22.90	88.36	338	1957	7224
Regime Switching	0.431	8.00	30.13	119	816	3740
HAR	0.242	3.70	14.16	60	439*	2126*
HARST	0.242	5.86	34.26	186	2737	90535
	RV SPY					
	1_step	5_step	10_step	20_step	50_step	100_step
RLS_modified	0.044	0.80*	3.5*	17*	132*	632*
RLS	0.043*	0.81	3.6	18	142	689
ARFIMA(1,d,1)	0.242	5.39	20.7	79	444	1613
ARFIMA(0,d,0)	0.241	5.38	20.7	79	443	1609
Regime Switching	0.189	4.26	17.1	70	484	2149
HAR	0.060	1.09	4.5	21	169	968
HARST	0.065	1.68	9.0	46	346	1496

Figure 1: Full Sample autocorrelations; S&P 500

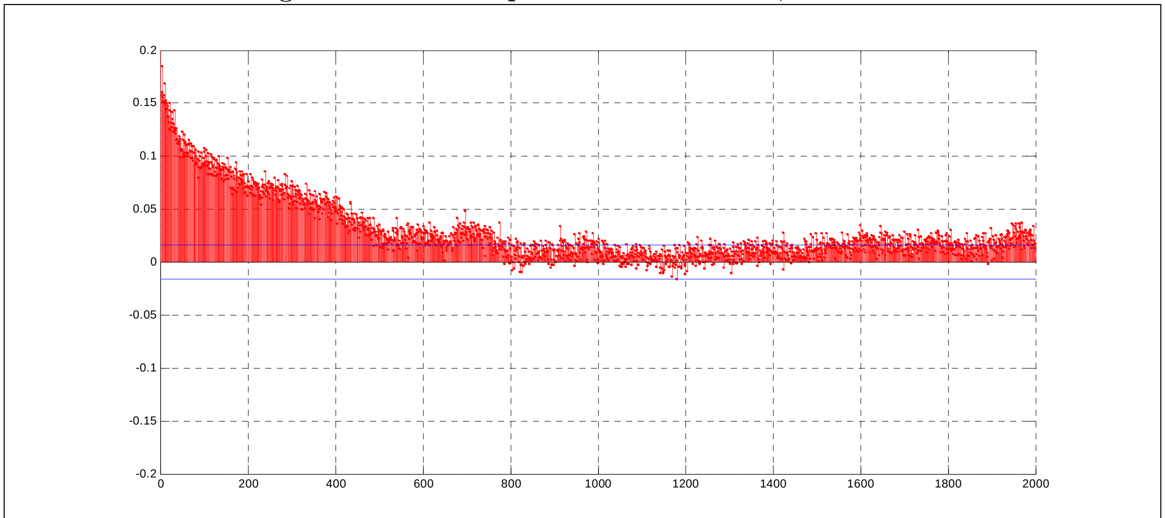

Figure 2: Fitted level shifts and volatility; S&P 500

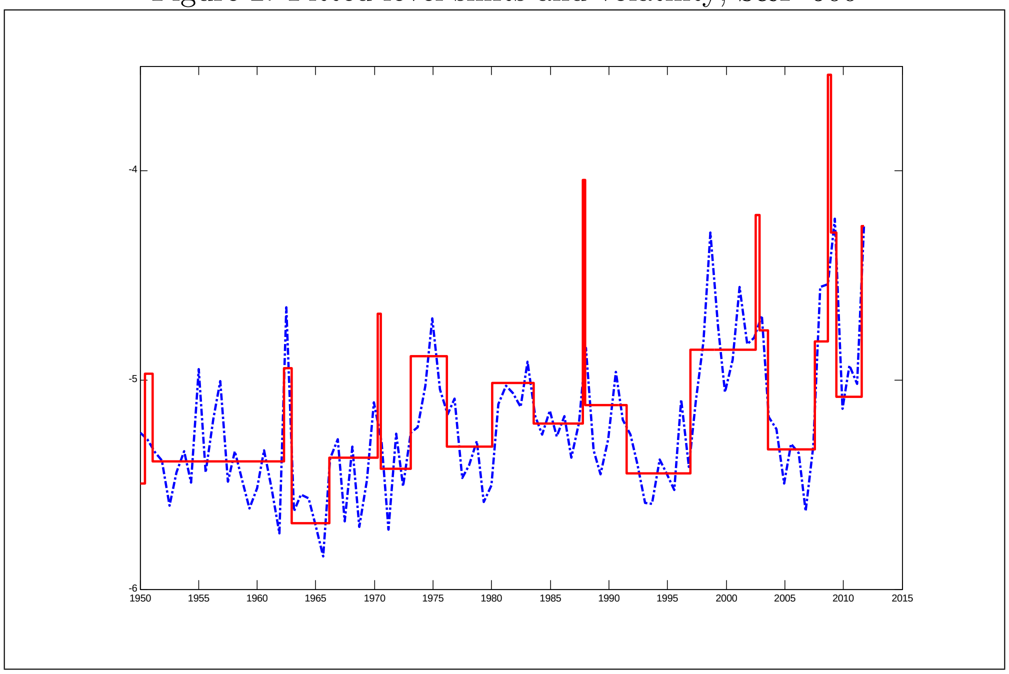

Figure 3: Sample autocorrelations of S&P 500 residuals

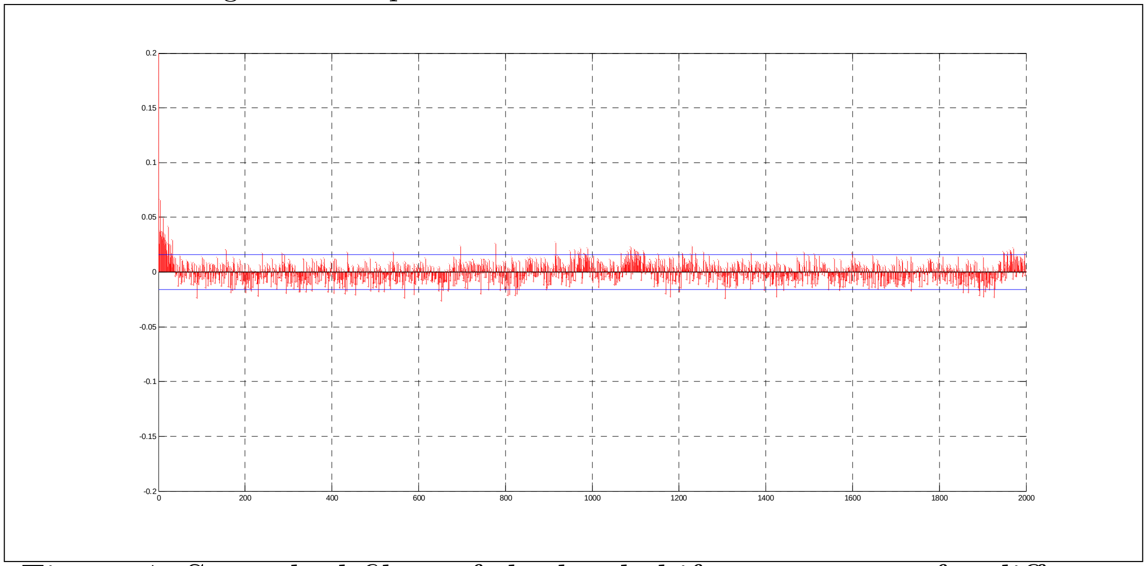

Figure 4: Smoothed filter of the level shift components for different thresholds; S&P 500

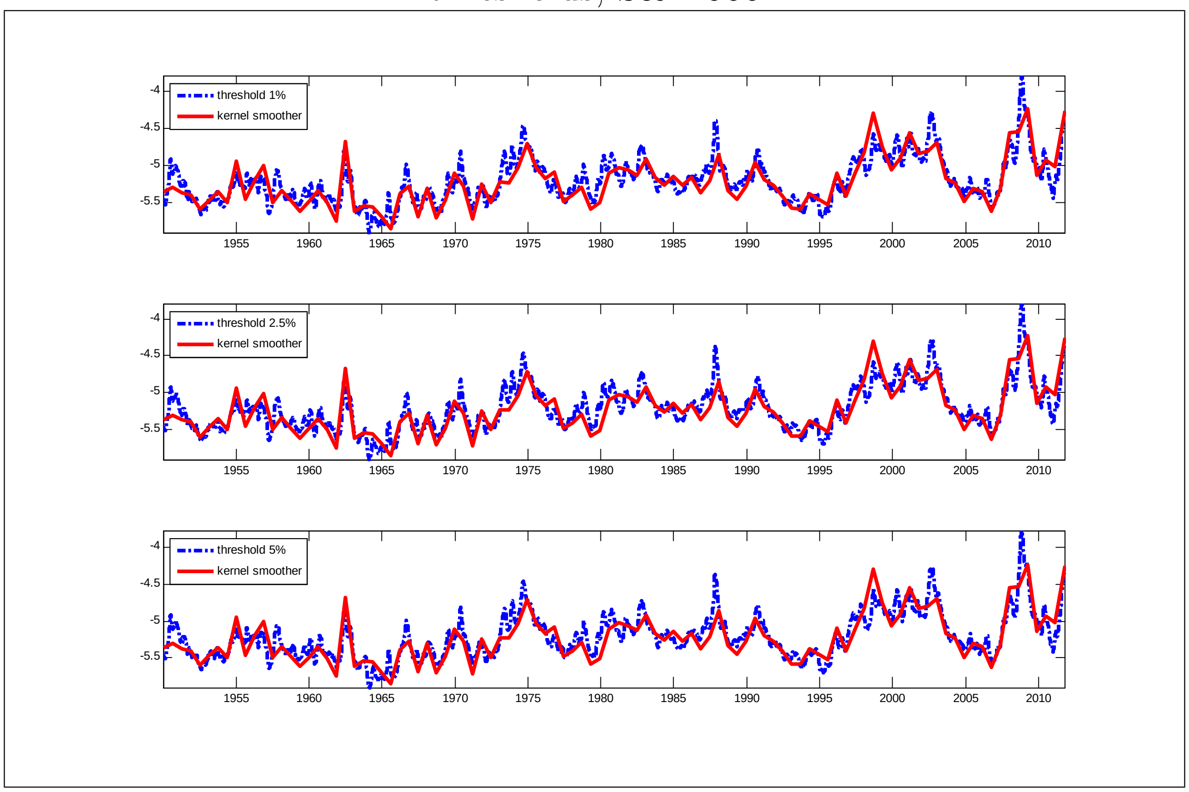

Figure 5: S&P Smoothed filter of the level shift components for different models

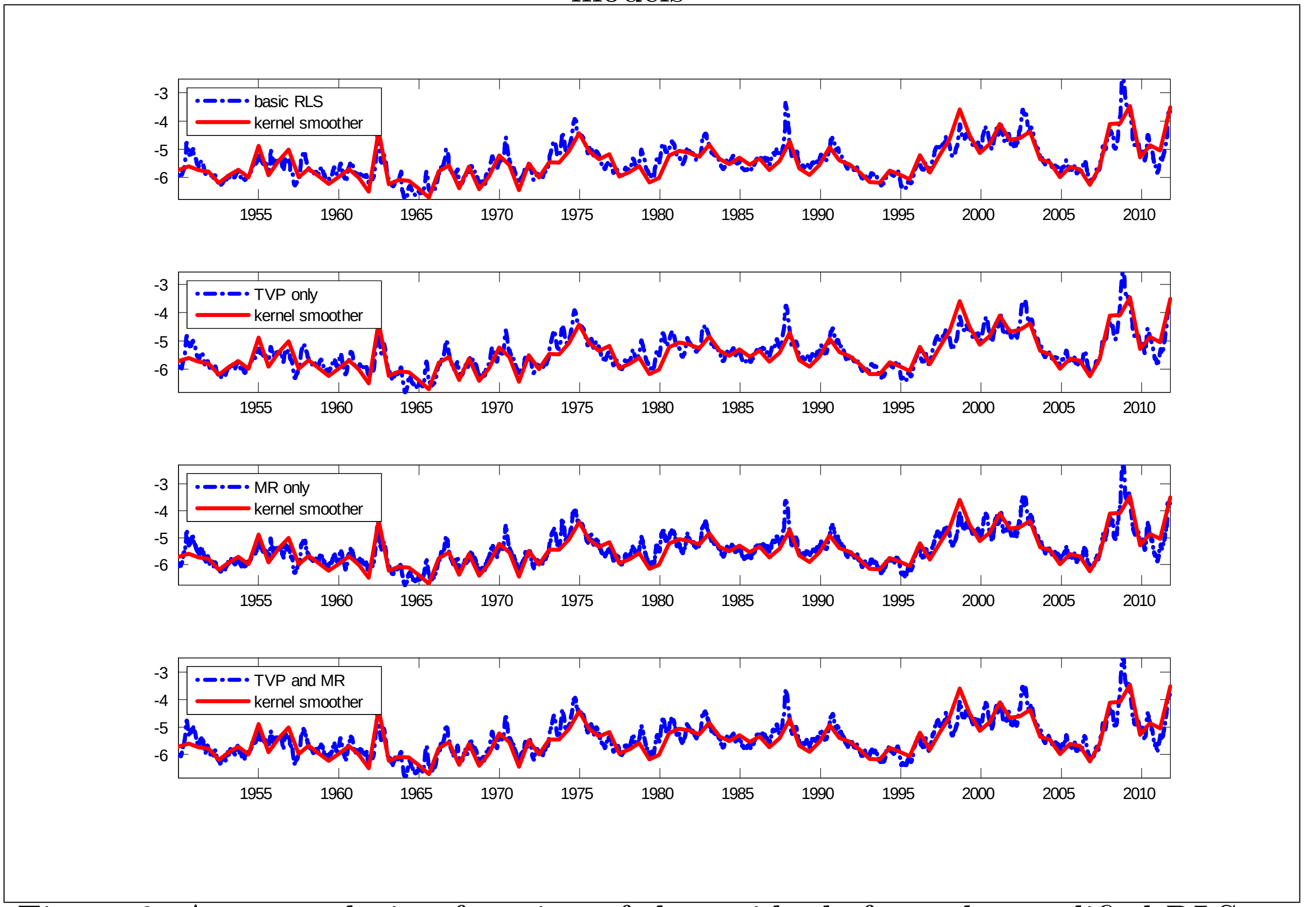

Figure 6: Autocorrelation function of the residuals from the modified RLS with both mean reversion and changing probability; S&P 500