Modeling Exchange Rate Volatility with Random Level Shifts[*]

Ye Li[†]   **Pierre Perron**[‡]

Citi   **Boston University**

Jiawen Xu[§]

Shanghai University of Finance and Economics

Abstract

Recent literature has shown that the volatility of exchange rate returns displays long memory features. It has also been shown that if a short memory process is contaminated by level shifts, the estimate of the long memory parameter tends to be upward biased. In this paper, we directly estimate a random level shift model to the logarithm of the absolute returns of five exchange rates series, in order to assess whether random level shifts can explain this long memory property. Our results show that there are few level shifts for the five series, but once they are taken into account the long memory property of the series disappears. We also provide out-of-sample forecasting comparisons, which show that, in most cases, the random level shift model outperforms popular models in forecasting volatility. We further support our results using a variety of robustness checks.

JEL Classification Number: C22, F37.
Keywords: Random Level Shifts, Long-Memory, Forecasting, Volatility.

1 Introduction

A vast literature has documented that various measures of the volatility of asset returns display features akin to those of a long-memory process. This is also the case for the volatility of exchange rate series; see, e.g., Anderson et al. (2001) and Anderson and Bollerslev (1997), among others. On the other hand, it has been suggested that the long-memory features present in the data could be due to occasional level shifts; see, e.g., Diebold and Inoue (2001). This follows from similar arguments used in Perron (1989, 1990) who showed that changes in level and/or slope of the trend function of a series causes the estimate of the sum of the autoregressive parameters to be biased towards one, suggesting non-stationarity.

Some recent papers have tried to assess whether random level shifts are indeed responsible for this long-memory feature and not simply a theoretical curiosity. Early attempts to that effect include Stărică and Granger (2005) and Granger and Hyung (2004), who argued that for the volatility of stock market indices the evidence for long-memory is weaker when level shifts are taken into account. 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 showed that this specification has better forecasting performance than the more traditional GARCH(1,1) model and its fractionally integrated counterpart. 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 shifts 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 (2013) estimated a stochastic volatility model with level shifts using a Bayesian approach with 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) considered a random level shifts model for which the series of interest is the sum of a short memory process and a jump or level shifts component, modeled as the cumulative sum of a process which is 0 with some probability $(1 - \alpha)$ and is a random variable with probability α. They applied it to the logarithm of daily absolute returns for the S&P 500, AMEX, Dow Jones and NASDAQ stock market return indices. The point estimates obtained imply few level shifts for all

1

series. But once these are taken into account, there is little evidence of serial correlation in the remaining noise and, hence, no evidence of long-memory. Once the estimated shifts are introduced to a standard GARCH model applied to the returns series, any evidence of GARCH effects disappears. They also considered rolling out-of-sample forecasts of squared returns. In most cases, the simple random level shifts model clearly outperforms a standard GARCH(1,1) model and, in many cases, it also provides better forecasts than a fractionally integrated GARCH model. Varneskov and Perron (2015) extended the analysis to introduce random level shifts in a general ARFIMA (autoregressive fractionally integrated moving-average) model. They showed that random level shifts are an essential component to model adequately the volatility of various series, whether from daily data or from realized volatility series constructed using high frequency data. From a forecasting perspective, they showed that the random level shifts model is the only one that consistently belong to the 10% Model Confidence Set of Hansen et. al. (2011).

Hence, there is growing evidence that a random level shifts model is indeed a genuine contender to explain the long-memory features of volatility. However, most of the results so far pertain to stock market return indices. Little evidence about the adequacy of random level shifts models is available concerning the properties of the volatility of exchange rate series. Our goal is to use some of the methodologies recently developed to address this issue. One exception is Morana and Beltratti (2004) who considered structural changes in the realized variance processes of the DEM/USD and JPY/USD exchange rates. Their results show that the volatility of DEM/USD and JPY/USD exchange rates show clear evidence of genuine long memory and that the structural changes can only partially explain the long memory features. Their forecasting exercises indicate that for short-term forecasting neglecting the structural changes is not that important, but that accounting for them provides substantial improvements for long-term forecasting. However, as noted by Varneskov and Perron (2015), the results obtained are very different when considering historical spans of daily returns compared to shorter spans of realized volatility series constructed from high frequency data.

In this paper, we follow the approach of Lu and Perron (2010). We consider historical series of daily exchange rates for the JPY/USD, DEM/USD, CAD/USD, GBP/USD and EUR/USD. We estimate a random level shifts model for the log absolute return series, adopting the specification that the series is the sum of a short memory process and level shifts component. The level shifts component is specified as a mixture model which takes value 0 with probability α and is some random variable with probability $1 - \alpha$. To estimate the model, we cast it into a generalized state space framework with a mixture of normal

distributions and use the estimation method developed by Perron and Wada (2009). We also evaluate the forecasting performance of the random level shifts model relative to the popular ARFIMA model as well as the time varying parameter and the Markov switching models, which like ours attempt to allow for changes in the mean of the series. We show that the random level shift model indeed provides improved forecasts in all cases. Also, we document that though few level shifts are present, once they are taken into account any evidence of long-memory disappears and what is left is a noise component that is essentially white noise.

We further support our results using a variety of robustness checks: a) the estimate of the long-memory parameter of Hou and Perron (2015) that is robust to random level shifts and noise; b) the test of Qu (2011) for the null hypothesis that a time series is a stationary long memory process against the alternative hypothesis that it is affected by random level shifts (or some other low frequency contamination), which is also robust to noise; c) as in Perron and Qu (2010), we also look at the path of the log-periodogram estimates $\hat{d}$ as the bandwidth m varies, i.e., the number of frequencies used to construct the estimate; d)we present results from applying the model used in Varneskov and Perron (2015), which jointly estimates the random level shifts along with the long-memory parameter (and any short-run dynamics if desired); the ARFIMA(1,d,1) version is robust to noise as discussed in Varneskov and Perron (2015). All these show that the main results remain valid, namely that the exchange rate series are better modelled as random level shifts processes with remaining variations that are essentially white noise. In particular, when accounting for the presence of such random level shifts, the evidence for any remaining long-memory disappears.

The structure of the paper is as follows. Section 2 presents the model and the specifications adopted. Section 3 discusses the method of estimation. The empirical results obtained from estimating the model are presented in Section 4 along with evidence that the level shifts account for all the long-memory features, including various robustness checks. We also show that a simple RLS model with white noise errors forecasts better than ARFIMA models. Section 5 offers brief concluding remarks.

2 Model

The random level shifts model considered is specified by

$$y_t = a + \tau_t + c_t$$

where a is constant term, τ_t is the random level shifts component and c_t is a short memory process to model the remaining noise. The level shifts component is given by:

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

where $\delta_t = \pi_t \eta_t$ with π_t a binomial variable, which takes value 1 with probability α and value 0 with probability $1 - \alpha$. When $\pi_t = 1$, a random level shift η_t happens having a distribution $\eta_t \sim i.i.d$ $N(0, \sigma_\eta^2)$. Furthermore, π_t, η_t and c_t are mutually independent. For the short memory component, in general c_t can be defined by the process $c_t = C(L)e_t$ with $e_t \sim i.i.d.$ $N(0, \sigma_e^2)$, where $C(L) = \Sigma_{i=0}^\infty c_i L^i$, $\sum_{i=0}^\infty i|c_i| < \infty$ and $C(1) \neq 0$. However, as will be shown for the series considered, once the level shifts are accounted for, barely any serial correlation remains. Accordingly, we shall simply specify c_t as an AR(1) process. Hence, the model to be used is:

$$y_t = a + \tau_t + c_t$$
$$\tau_t = \tau_{t-1} + \delta_t$$
$$c_t = \phi c_{t-1} + e_t$$
$$\delta_t = \pi_t \eta_t$$

Note that we can write $\delta_t = \pi_t \eta_{1t} + (1 - \pi_t)\eta_{2t}$, with $\eta_{it} \sim i.i.d$ $N(0, \sigma_{\eta i}^2)$ and $\sigma_{\eta 1}^2 = \sigma_\eta^2$, $\sigma_{\eta 2}^2 = 0$. This allows us to cast the model into a state space framework. More specifically, with the error term being a mixture of two normal distributions, where

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

and

$$\delta_t = \pi_t \eta_{1t} + (1 - \pi_t)\eta_{2t}$$
$$c_t = \phi c_{t-1} + e_t$$

In matrix form,
$$\Delta y_t = HX_t + \delta_t; \quad X_t = FX_{t-1} + U_t$$

In general, when c_t follows an AR(p) process, then

$$X_t = [c_t, c_{t-1}, \cdots, c_{t-p}]'$$

$$F = \begin{pmatrix} \phi_1 & \phi_2 & \cdots & \cdots & \phi_p \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}$$

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

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

Comparing this model with the standard state space model, the difference is that the error term is a mixture of two normal distributions.

3 Estimation Method

We apply the estimation method proposed by Perron and Wada (2009), see also Perron and Wada (2016). In their paper, they generalized the trend cycle decomposition framework based on unobserved components with errors that are mixtures of normal distributions, thereby allowing shifts in the slope and level of the trend functions. The main ingredient that underlies the estimation procedure is that the model can be written as a state space model with normal errors occurring in two different possible states. These states can be described by the combined values of the Bernoulli random variables. From this we can generate the log likelihood function from the decomposition of the prediction errors to obtain estimates. Let $Y_t = [\Delta y_1, \Delta y_2, \cdots, \Delta y_t]$ represents the observations available at time t and $\theta = [\sigma_\eta^2, \alpha, \sigma_e^2, \phi_1, \cdots, \phi_q]$ be the parameter vector to be estimated. The log-likelihood function is

$$\ln(L) = \sum_{t=1}^{T} \ln f(\Delta y_t | Y_{t-1}, \theta),$$

where,

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

Here, s_t is an indicator to represent whether or not a random level shift occurs. That is, when $s_t = 1$, then $\pi_t = 1$ and a random level shift happens; when $s_t = 2$, $\pi_t = 0$ and there

5

is no level shift. Let $\varpi_t^{ij} = f(\Delta y_t | s_{t-1} = i, s_t = j, Y_{t-1}, \theta)$ for $i, j \in 1, 2$, and

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

where,

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

So we have,

$$
\widetilde{\varepsilon}_{t+1|t}^{ki} = Pr(s_{t+1} = i, s_t = k | Y_t, \theta) = Pr(s_{t+1} = i) \sum_{j=1}^{2} \widetilde{\varepsilon}_{t|t}^{jk}.
$$

with

$$
\widetilde{\varepsilon}_{t+1|t}^{11} = \alpha \sum_{j=1}^{2} \widetilde{\varepsilon}_{t|t}^{j1} = \alpha [\widetilde{\varepsilon}_{t|t}^{11} + \widetilde{\varepsilon}_{t|t}^{21}]
$$

$$
\widetilde{\varepsilon}_{t+1|t}^{21} = \alpha \sum_{j=1}^{2} \widetilde{\varepsilon}_{t|t}^{j2} = \alpha [\widetilde{\varepsilon}_{t|t}^{12} + \widetilde{\varepsilon}_{t|t}^{22}]
$$

$$
\widetilde{\varepsilon}_{t+1|t}^{12} = (1 - \alpha) \sum_{j=1}^{2} \widetilde{\varepsilon}_{t|t}^{j1} = (1 - \alpha) [\widetilde{\varepsilon}_{t|t}^{11} + \widetilde{\varepsilon}_{t|t}^{21}]
$$

$$
\widetilde{\varepsilon}_{t+1|t}^{22} = (1 - \alpha) \sum_{j=1}^{2} \widetilde{\varepsilon}_{t|t}^{j2} = (1 - \alpha) [\widetilde{\varepsilon}_{t|t}^{12} + \widetilde{\varepsilon}_{t|t}^{22}]
$$

In matrix form,

$$
\begin{pmatrix} \widetilde{\varepsilon}_{t+1|1}^{11} \\ \widetilde{\varepsilon}_{t+1|t}^{21} \\ \widetilde{\varepsilon}_{t+1|t}^{12} \\ \widetilde{\varepsilon}_{t+1|t}^{22} \end{pmatrix} = \begin{pmatrix} \alpha & \alpha & 0 & 0 \\ 0 & 0 & \alpha & \alpha \\ 1 - \alpha & 1 - \alpha & 0 & 0 \\ 0 & 0 & 1 - \alpha & 1 - \alpha \end{pmatrix} \begin{pmatrix} \widetilde{\varepsilon}_{t|t}^{11} \\ \widetilde{\varepsilon}_{t|t}^{21} \\ \widetilde{\varepsilon}_{t|t}^{12} \\ \widetilde{\varepsilon}_{t|t}^{22} \end{pmatrix}
$$

The conditional likelihood function for Δy_t is:

$$
f(\Delta y_t | s_{t-1} = i, s_t = j, Y_{t-1}, \theta) = \frac{1}{\sqrt{2\pi}} |f_t^{ij}|^{-1/2} \exp\left(-\frac{v_t^{ij'} (f_t^{ij})^{-\frac{1}{2}} v_t^{ij}}{2}\right)
$$

6

where v_t^{ij} is the prediction error,

$$v_t^{ij} = \Delta y_t - \Delta y_{t|t-1}^{ij} = \Delta y_t - E[\Delta y_t | s_t = i, s_{t-1} = j, Y_{t-1}, \theta]$$

and $f_t^{ij} = E(v_t^{ij} v_t^{ij'})$ is the prediction error variance. The prediction $\Delta y_{t|t-1}$ based on past information does not depend on the state of time t, but Δy_t does. The basic inputs are predictions for the state variables and their variances, which are

$$X_{t|t-1}^i = F X_{t|t-1}$$
$$P_{t|t-1}^i = F P_{t|t-1}^i F' + Q$$

The prediction error is $v_t^{ij} = \Delta y_t - H X_{t|t-1}^{ij}$, so that $f_t^{ij} = H P_{t|t-1}^i H' + R_j$, where R_j is the variance of the error term, which takes two possible values: $R_j = \sigma_\eta^2$ with probability α when $\pi_t = 1$, $R_j = 0$ with probability $(1 - \alpha)$ when $\pi_t = 0$. Applying the updating formula, given $s_t = j, s_{t-1} = i$, we obtain:

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

As in Perron and Wada (2009), to reduce the dimension of the estimation problem, we adopt the re-collapsing procedure of Harrison and Stevens (1976), given by:

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

and

$$P_{t|t}^j = \frac{\sum_{i=1}^2 Pr(s_{t-1} = i, s_t = j | Y_t, \theta)[P_{t|t}^{ij} + (X_{t|t}^j - X_{t|t}^{ij})(X_{t|t}^j - X_{t|t}^{ij})']}{Pr(s_t = j | Y_t, \theta)}.$$

4   Empirical Results for Exchange Rate Returns

This section presents the main results of the paper. We first discuss the data used. We then present in Section 4.2 preliminary diagostics suggesting that the various series are better characterized as random level shifts processes with short-memory noise rather than long-memory processes. In Section 4.3, we then present the estimates of the model discussed in the previous section. Using the smoothed estimate of the random level shifts process, we show that once accounting for the shifts the remaining variations are essentially white noise. In order to further substantiate this result, Section 4.4 present results from applying the

7

model used in Varneskov and Perron (2015), which jointly estimates the random level shifts component along with the long-memory parameter (and short-run dynamics if desired). These show that the results remain valid with estimates of d close to 0 and statistically significant parameters for the level shifts process. Finally, in Section 4.5, we show that a simple RLS model with white noise errors forecasts better than ARFIMA, time varying parameter and Markov switching models.

4.1 The data

We consider the random level shifts model for five daily exchange rate returns series: the JPY/USD, DEM/USD (both from 10/11/1983 to 7/30/2010; 6,994 observations; obtained from the CRSP database), CAD/USD (from 01/05/1971 to 09/11/2015; 11224 observations), GBP/USD (from 01/05/1971 to 09/11/2015; 11218 observations) and EUR/USD (from 01/05/1999 to 09/11/2015; 4197 observations). The last three daily series are from the database of the Federal Reserve Bank of St. Louis. We apply our level shifts model to log-absolute returns since they do not suffer from a non-negativity constraint as do, say, absolute or squared returns. There is also no loss relative to using squared returns in identifying level shifts since log-absolute returns are a monotonic transformation. Since we wish to identify the probability of shifts and their locations, the fact that log-absolute returns are quite noisy is not problematic since our methods are robust to the presence of noise. Another reason is the fact that for many asset returns, a log-absolute transformation yields a series that is closer to being normally distributed (see, e.g., Andersen, Bollerslev, Diebold and Labys, 2001). When returns are zero or close to it, the log absolute value transformation implies extreme negative values. Using our method, these outliers would be attributed to the level shifts component and thus bias the probability of shifts upward. To avoid this problem, we bound absolute returns away from zero by adding a small constant, i.e., we use $\log(|r_t| + 0.001)$, a technique introduced by Fuller (1996).

4.2 Preliminary diagnostics

Our aim is to assess whether the long-memory feature is genuine or caused by the presence of level shifts. To that effect, we first discuss some features of the series. We first estimate the long-memory parameter d using the modified local Whittle estimator of Hou and Perron (2014). The most general version of this class of estimators, the LWPLFC estimate, has the advantage of accounting for all kinds of contaminations: low frequency (e.g., random level shifts), additive noise and short memory dynamics. When low frequency contaminations are

8

present, it has, in most cases, the smallest bias and mean-squared error amongst all existing estimators designed to control for low frequency contaminations, whether or not other types of contaminations are present. The results are reported in Table 1, along with the standard Local Whittle estimate (LW) (Kunsch, 1987). First, as expected the estimates of d using the standard LW estimator are high, ranging from 0.23 to 0.46, suggesting long-memory processes. However, the LWPLFC estimates of Hou and Perron (2014), which are robust to noise and random level shifts, are small ranging from -0.02 to 0.11. Comparing the results obtained from both estimator strongly suggests that the apparent long-memory feature in the data is actually due to random level shifts. The bandwidths used in the standard LW and Hou and Perron robust LW are both $m = T^{0.8}$.

To further provide evidence, we applied the test of Qu (2011) for the null hypothesis that a given time series is a stationary long memory process against the alternative hypothesis that it is affected by regime change, random level shifts or a smoothly varying trend (i.e., a low frequency contamination), which is robust to noise. The results are presented in Table 2 for two trimming parameters $\varepsilon = 0.02, 0.05$. Except for the DEM/USD, the test strongly rejects the null hypothesis of a long memory process for all other four volatility series.

Another way to distinguish a long memory process from a short memory process with random level shifts is to look at the path of the log-periodogram estimates $\hat{d}$ as the bandwidth m varies, i.e., the number of frequencies used to construct the log-periodogram estimate of d (Geweke.and Porter-Hudak, 1983). As discussed in Perron and Qu (2010), there is a discontinuity in the asymptotic distribution for small and larger rates of increase of m. First, when m is near or below $T^{1/3}$, $\hat{d}$ will be in a neighborhood of 1. When m is roughly between $T^{1/3}$ and $T^{1/2}$, $\hat{d}$ drops to a new level when the stationary component starts to affect the limiting distribution. As m increases beyond $T^{1/2}$ there is a further gradual decrease in $\hat{d}$ as the short-memory component becomes increasingly more important, relative to the level shifts component, in determining the limiting distribution. The picture is very different if the underlying model is that of a long-memory process, e.g., a fractionally integrated model. Here, the limiting distribution of the log periodogram estimate $\hat{d}$ is the same regardless of the rate of increase of m relative to the sample size T. Hence, we can use the path of the estimates $\hat{d}$ obtained for a wide range of values of m to discriminate between the two models.

In Figure 1, we plot the paths of the log-periodogram estimates as a function of m. The vertical lines in each plot refers to values for the bandwidth of $m = T^a$ for $a = 1/3, 1/2, 2/3$. The pattern of the paths is very similar to what is predicted by the theoretical results if the true underlying structure is a short memory process with level shifts. The estimates $\hat{d}$ reach

9

a peak value near $m = T^{1/3}$, then gradually decrease as the stationary component starts to affect the limiting distribution.

4.3   Results from estimating the basic model

The empirical evidence discussed above indicates that a random level shifts model is a more likely candidate to explain the features of the data rather than a long-memory process. We now present the estimates of the RLS model for the exchange rate volatility series. For the specification of the short memory component, we consider the white noise case: $c_t = e_t$, so that the parameters to be estimated are $(\sigma_\eta, \alpha, \sigma_e)$. The initial value for the state vector is $X'_{0|0} = (0,0)'$ and the initial value for the covariance matrix is set to

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

The estimates are presented in Table 3. The probability of shifts is very small. Given the point estimate of the probability of shifts, one can deduce an implied estimate of the number of shifts in the sample: 5 for JPY/USD, 8 for DEM/USD, 75 for CAD/USD, 47 for GBP/USD and 7 for EUR/USD. As we shall see, even when such few shifts are taken into account the properties of the remaining noise is substantially changed.

We seek to assess whether or not the random level shifts component can explain the long memory property of the exchange rate returns. The first strategy we adopt is the following. Given the estimated number of shifts, we estimate the break dates and regime specific means using the method of Bai and Perron (2003). Once these are obtained, we estimate the noise component as the difference between the original series and the fitted level shifts process [1].

To be more specific, let m be the number of breaks (e.g., 8 for the DEM/USD, 5 for the JPY/USD, etc.), T_i $(i = 1, \cdots, m)$ be the break dates (with the convention that $T_0 = 0, T_{m+1} = T$), and $\{u_i; i = 1, ..., m+1\}$ be the means within each regime. The method of Bai and Perron (2003) allows obtaining estimates of the break dates $\{\hat{T}_i; i = 1, ..., m\}$ and regime-specific means $\{\hat{u}_i; i = 1, ..., m+1\}$ as global minimizers of the objective function $\sum_{i=1}^{m+1} \sum_{t=T_{i-1}+1}^{T_i} (y_t - u_i)^2$. The noise component, say $\hat{c}_t$ is then obtained as $\hat{c}_t = y_t - \sum_{i=1}^{m+1} \hat{u}_i DU_{i,t}$, where $DU_{i,t} = 1$ if $\hat{T}_{i-1} < t \le \hat{T}_i$ and 0, otherwise. To get a better view of the implied level shifts process and its relation to the volatility of the exchange rate series, Figure 2 presents graphs of the fitted level shifts process in conjunction with a smoothed

---

[1] Given the relatively large nuber of breaks due to the long span of data available, note that for CAD/USD we use only the last 4500 observations and for GBP/USD the last 5000.

estimate of the log-absolute returns, obtained using a standard Gaussian kernel. The results reveal that the level shifts capture the main movements of the series.

The autocorrelation functions of y_t and $\hat{c}_t$ are presented in Figure 3. The autocorrelation functions decay slowly for the original series. However, once the level shifts are taken into account, the autocorrelation functions show basically no serial correlations left. Even if the level shifts are few in number, they can fully explain and account for the long-memory features of the exchange rate series.

4.4   Joint estimation of level shifts and long memory

The second strategy we adopt to assess whether or not the random level shifts component can explain the long memory property of the exchange rate returns, is to extend the model to jointly estimate the RLS model along with a long-memory process. We adopt the model and estimation method proposed by Varneskov and Perron (2015) to jointly estimate random level shifts together with an ARFIMA model. We model the volatility of an exchange rate series as following a random level shifts processes with a fractionally integrated noise component. We estimate the model using two specifications, RLS_ARFIMA(0,d,0) and RLS_ARFIMA(1,d,1). The latter incorporates short memory dynamics in the form of an ARMA(1,1) process. Note in particular that the version RLS_ARFIMA(1,d,1) is robust to the presence of noise in the series because of the inclusion of the moving-average component; see Varneskov and Perron (2015) for a discussion and supporting evidence of this feature.

The estimation results are presented in Table 4. In all cases, the estimates of the parameters of the RLS component are similar to those reported in Table 3, in particular the estimated shift probabilities remain small but the overall importance of the RLS component is large. Also of importance is the fact that the estimates of d are all close to 0. This shows that after accounting for random level shifts, there is no evidence for remaining long-memory in the data. Note that the estimate of the AR and MA parameters in the RLS_ARFIMA(1,d,1) specification are close to each other suggesting near-cancellation. The reason they are slightly different is due to the way the RLS_ARFIMA(1,d,1) specification accounts for noise in the series (again, see Varneskov and Perron, 2015, for a discussion).

4.5   Forecasting

We further consider the performance of the random level shifts model with white noise errors in forecasting volatility proxied by squared returns relative to three competing models: the ARFIMA, the time varying parameter (TVP) and the Markov switching models. The reason

to make the comparisons with the ARFIMA model is that it is generally perceived as a good forecasting model for asset volatility. We also include the time varying parameter and Markov switching models because they are popular models that consider smooth changes or regime switches in the unconditional mean, as our model does. The time varying parameter model (TVP) specifies the unconditional mean as a random walk process: $y_t = \tau_t + e_t$ and $\tau_t = \tau_{t-1} + \eta_t$, where e_t, η_t are independent to each other and $e_t \sim N(0, \sigma_e^2), \eta_t \sim N(0, \sigma_\eta^2)$. It is actually a special case of the RLS model proposed in this paper with the binomial variable π_t taking value 1 with probability 1. For the Markov switching model, we use a two-state Markov regime switching model defined by $y_t = \mu_{S_t} + \varepsilon_t$ where $S_t = 1, 2$ and $\varepsilon_t \sim i.i.d.$ $N(0, \sigma_{S_t}^2)$ with an unconstrained transition matrix P for the state variable S_t; see Hamilton (1994) for details [2].

We follow the method adopted by Varneskov and Perron (2015) to assess the relative forecasting performance. For the random level shifts model, we obtain τ-step ahead forecasts directly from the filtered estimates obtained when estimating the state-space model. The τ-step ahead forecast is then given by:

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

where $X_{t|t}^{ij}$ is the filtered estimate of X_t which depends on whether $s_{t+1} = j$ and $s_t = i$ for $i, j \in \{0, 1\}^2$. We consider out-of-sample forecasting of the last 900 ($T^{out} \in [1, 900]$) days of the five exchange rate series. We compare five models, the Random Level Shift, $ARFIMA(1, d, 1)$, $ARFIMA(0, d, 0)$, TVP and markov switching. We consider direct τ-step ahead forecasting for three different horizons $\tau = (1, 5, 10)$. The τ-step ahead forecasts are defined as $\overline{y}_{t+\tau|t} = \sum_{s=1}^{\tau} \widehat{y}_{t+s|t}$. Similarly the cumulative volatility proxy is defined by $\overline{\sigma}_{t,\tau}^2 = \sum_{s=1}^{\tau} y_{t+s}$. We use the mean square forecast error (MSFE) criterion defined as:

$$MSFE_\tau = \frac{1}{T^f} \sum_{t=1}^{T^f} (\overline{\sigma}_{t,\tau}^2 - \overline{y}_{t+\tau|t})^2$$

where T^f is the total number of forecasts produced. The MSFEs of the forecasts are reported in Table 5 for different forecasting horizons. The RLS model performs the best with the smallest MSFEs for all five exchange rate series. In most cases, the TVP model is the second best except for the case of GBP/USD, in which case the $ARFIMA(1, d, 1)$ is the second best

---

model. The forecasting performance of the Markov Switching model is considerably worse than all other competing models. In general, the forecasting performance of the RLS model is robust to different series and different forecasting horizons.

5   Conclusion

We considered series of daily exchange rates for the JPY/USD, DEM/USD, CAD/USD, GBP/USD and EUR/USD. We estimated a random level shifts model for the log absolute return series, adopting the specification that the series is the sum of a short memory process and level shifts component. We documented that though few level shifts are present once they are taken into account any evidence of long-memory disappears and what is left is a noise component that is essentially white noise. We also presented evidence to that effect via various diagnostics and by showing that the long-memory parameter estimate is near zero when estimating a model that accounts for random level shifts and long-memory jointly. Hence, our results are robust. We also evaluated the forecasting performance of the random level shifts model relative to the popular ARFIMA, time varying parameter and Markov switching models. We showed that the forecasting performance of the pure RLS model is superior in the sense that it has the smallest MSFE in all cases and for all forecasting horizons. Our paper therefore adds to the recent literature that considered the volatility of stock market indices by showing that a random level shifts model is indeed a serious contender to explain the long-memory features of the volatility of exchange rate series.

Our results have direct implications that are of relevance for practitioners. First, it points to the random level shifts model as strong candidate for modeling and, especially, forecasting exchange rate volatility. As shown, the superiority in terms of forecast accuracy is important, especially at longer forecasting horizons. Second the timing of the shifts can be of interest in documenting ex-post from a historical perspective the changes in volatility as being related to influential crises, important macroeconomic events or policy announcements. Third, it suggests that an important avenue of future research is to devise ways to forecast such changes by modeling the probability of level shifts and their magnitude as depending on some covariates. This would be helpful not only in enhancing forecast accuracy but could also be beneficial as a valuable indicator for a financial warning tool. For steps in that direction, see Xu and Perron (2015).

References

Anderson, T.G. and T. Bollerslev (1997), "Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long-run in High Frequency Data," *The Journal of Finance* 52, 975-1005.

Anderson, T.G., T. Bollerslev, F.X. Diebold and P. Labys (2001), "The Distribution of Realized Exchange Rate Volatility," *Journal of American Statistical Association* 96, 42-55.

Bai, J. and P. Perron (2003), "Computation and Analysis of Multiple Structural Change Models," *Journal of Applied Econometrics* 18, 1-22.

Diebold, F.X. and A. Inoue (2001), "Long Memory and Regime Switching," *Journal of Econometrics* 105, 131-159.

Fuller, W. A. (1996): *Introduction to Time Series* (2nd ed.), New York: John Wiley.

Geweke, J., and S. Porter-Hudak (1983), "The Estimation and Application of Long Memory Time Series Models," *Journal of Time Series Analysis* 4, 221-238.

Granger, C.W.J. and N. Hyung (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: Princeton University Press.

Hansen, P. R., A. Lunde and J.M. Nason (2011), "The Model Confidence Set," *Econometrica* 79, 453-497.

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

Hou, J., and P. Perron (2014), "Modified Local Whittle Estimator for Long Memory Processes in the Presence of Low Frequency (and Other) Contaminations," *Journal of Econometrics* 182, 309–328.

Kunsch, H. (1987), "Statistical Aspects of Self-similar Processes," In: Prohorov, Y., Sazarov, V. (Eds.), *Proceedings of the First World Congress of the Bernoulli Society.* Vol. 1. VNU Science Press, Utrecht, 67-74.

Lu, Y.K. and P. Perron (2010), "Modeling and Forecasting Stock Return Volatility Using a Random Level Shift Model," *Journal of Empirical Finance* 17, 138-156.

Morana, C. and A. Beltratti (2004), "Structural Change and Long-Range Dependence in Volatility of Exchange Rates: Either, Neither or Both?" *Journal of Empirical Finance* 11, 629-658.

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 with a Changing Mean," *Journal of Business & Economic Statistics* 8, 153-162.

Perron, P. and Z. Qu (2007), "An Analytical Evaluation of the Log-periodogram Estimate in the Presence of Level Shifts," Boston University Working Paper.

Perron, P. and Z. Qu (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 T. Wada (2009), "Let's Take a Break: Trends and Cycles in U.S. Real GDP," *Journal of Monetary Economics* 56, 749-765.

Perron, P. and T. Wada (2016), "Measuring Business Cycles with Structural Breaks and Outliers: Applications to International Data," *Research in Economics* 70, 281-303.

Qu, Z. (2011), "A Test Against Spurious Long Memory," *Journal of Business and Economic Statistics* 29, 423-438.

Qu, Z. and P. Perron (2013), "A Stochastic Volatility Model with Random Level Shifts: Theory and Applications to S&P 500 and NASDAQ Return Indices," *Econometrics Journal* 16, 400-429.

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

Varneskov, R.T. and P. Perron (2015), "Combining Long Memory and Level Shifts in Modeling and Forecasting the Volatility of Asset Returns," Unpublished Manuscript, Department of Economics, Boston University.

Xu, J. and P. Perron (2015), "Forecasting in the Presence of In and Out of Sample Breaks," Unpublished Manuscript, Department of Economics, Boston University.

Table 1: Memory parameter estimates

	JPY	DEM	CAD	GBP	EUR
Standard LW	0.23	0.31	0.46	0.45	0.36
Hou-Perron Robust LW	0.11	0.07	0.11	0.10	-0.02

Table 2: Test statistics for spurious long memory (Qu; 2011)

	JPY	DEM	CAD	GBP	EUR
W(ε=0.02)	1.52***	0.55	1.51**	1.44**	1.47**
W(ε=0.05)	1.52***	0.51	1.11**	1.42**	1.32**

Note: ***, ** and * denote significance at 1%, 5% and 10% level. ε is the trimming proportion.

Table 3: Parameter estimates for the RLS model

Parameter	σ_η	α	σ_e
JPY	1.0430	0.0007	0.7522
	(0.0460)	(0.0003)	(0.0070)
DEM	0.6780	0.0012	0.7480
	(0.2418)	(0.0008)	(0.0065)
CAD	0.3242	0.0067	0.5528
	(0.0622)	(0.0026)	(0.0038)
GBP	0.5600	0.0042	0.6624
	(0.0877)	(0.0012)	(0.0046)
EUR	0.4638	0.0017	0.7016
	(0.2530)	(0.0016)	(0.0078)

Note: Standard errors are in parentheses.

Table 4: Parameter estimates for the RLS_ARFIMA model

Panel A: RLS_ARFIMA(0,d,0)					
Parameter	σ_η	α	σ_e	d	
JPY	0.4331	0.0010	0.7569	0.0582	
	(0.0372)	(0.0024)	(0.0065)	(0.0110)	
DEM	0.1889	0.0097	0.7504	0.0220	
	(0.0351)	(0.0006)	(0.0066)	(0.0129)	
CAD	0.3493	0.0033	0.5571	0.0494	
	(0.0679)	(0.0015)	(0.0039)	(0.0106)	
GBP	0.5906	0.0025	0.6672	0.0460	
	(0.1713)	(0.0012)	(0.0047)	(0.0106)	
EUR	0.4371	0.0019	0.7016	0.00002	
	(0.4526)	(0.0035)	(0.0078)	(0.0005)	
Panel B: RLS_ARFIMA(1,d,1)					
Parameter	σ_η	α	σ_e	d	ϕ	θ
JPY	0.6830	0.0002	0.7588	0.0334	0.2961	0.4021
	(0.8878)	(0.0004)	(0.0065)	(0.0115)	(0.1048)	(0.1117)
DEM	0.0780	0.0043	0.7506	0.0500	0.3962	0.4377
	(0.0128)	(0.0011)	(0.0065)	(0.0236)	(0.0463)	(0.0504)
CAD	0.3858	0.0010	0.5612	0.0123	0.9609	0.9223
	(0.0841)	(0.0005)	(0.0038)	(0.0060)	(0.0166)	(0.0201)
GBP	0.5995	0.0014	0.6715	0.0189	0.9586	0.9292
	(0.1685)	(0.0007)	(0.0048)	(0.0096)	(0.0238)	(0.0289)
EUR	0.3671	0.0029	0.6990	0.0242	0.3021	0.3894
	(0.1216)	(0.0017)	(0.0078)	(0.0024)	(0.1176)	(0.1154)

Note: Standard errors are in parentheses.

Table 5: Forecasting comparisons based on the MSFE

	1_step	5_step	10_step
JPY			
RLS	0.590*	4.295*	11.262*
ARFIMA(0,d,0)	0.639	4.838	12.891
ARFIMA(1,d,1)	0.634	4.645	11.916
TVP	0.622	4.534	12.003
Markov Switching	0.720	6.772	20.277
DEM			
RLS	0.541*	2.901*	7.032*
ARFIMA(0,d,0)	0.603	3.761	10.220
ARFIMA(1,d,1)	0.581	3.306	8.364
TVP	0.577	3.129	7.716
Markov Switching	0.645	4.853	14.382
CAD			
RLS	0.374*	2.206*	5.193*
ARFIMA(0,d,0)	0.419	2.421	5.766
ARFIMA(1,d,1)	0.415	2.459	5.732
TVP	0.413	2.391	5.562
Markov Switching	0.492	4.336	13.269
GBP			
RLS	0.358*	1.826*	4.099*
ARFIMA(0,d,0)	0.400	2.104	5.033
ARFIMA(1,d,1)	0.392	1.978	4.432
TVP	0.393	2.002	4.471
Markov Switching	0.455	3.596	10.850
EUR			
RLS	0.432*	2.200*	5.304*
ARFIMA(0,d,0)	0.475	3.080	8.929
ARFIMA(1,d,1)	0.456	2.394	5.916
TVP	0.450	2.267	5.397
Markov Switching	0.519	3.983	12.136

Note: * indicates the lowest MSFE across all models.

Figure 1: The LP estimates of d with different bandwidth choices
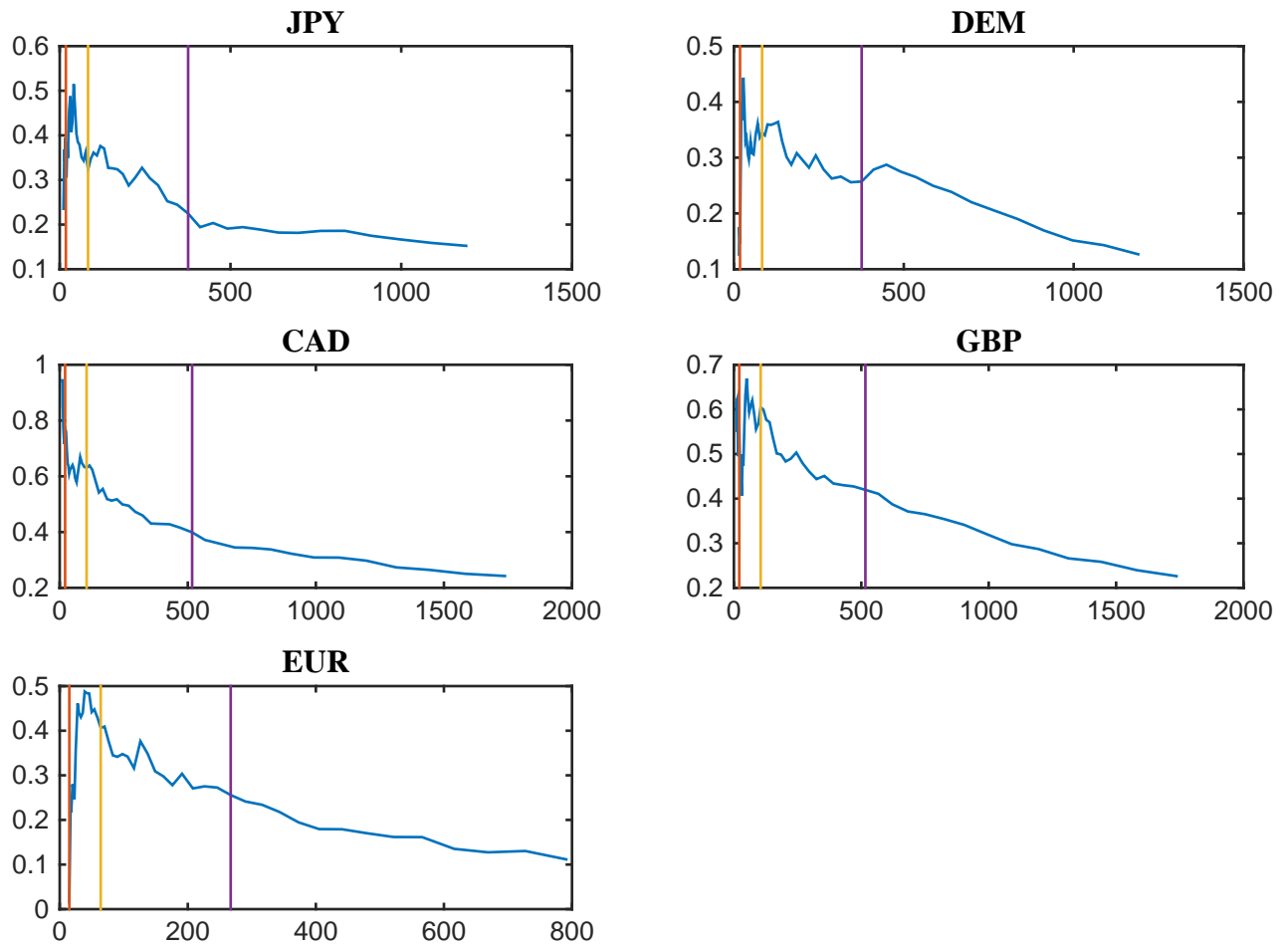
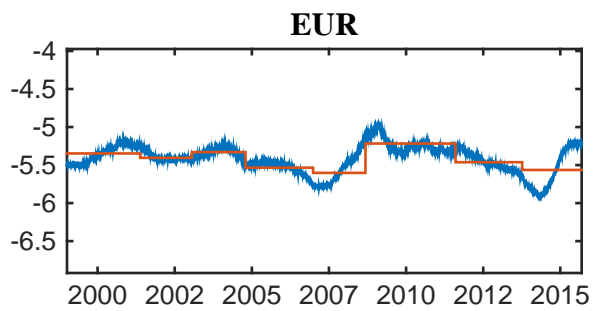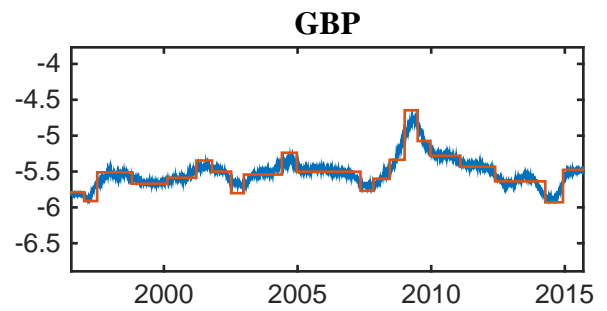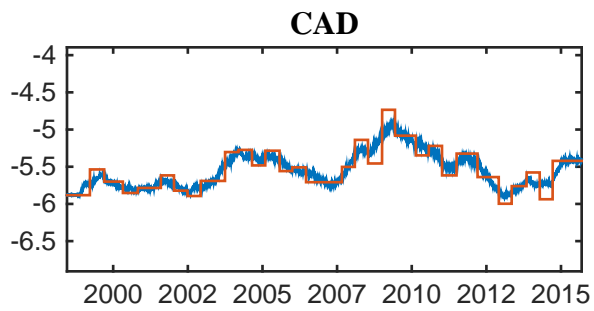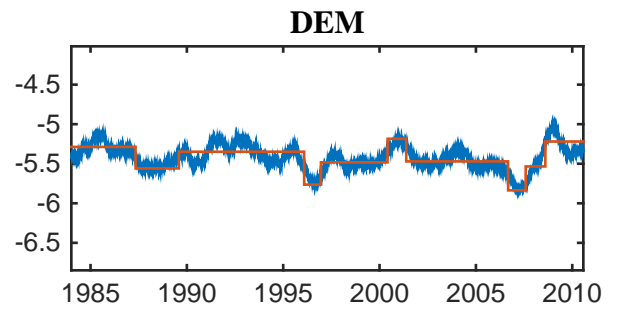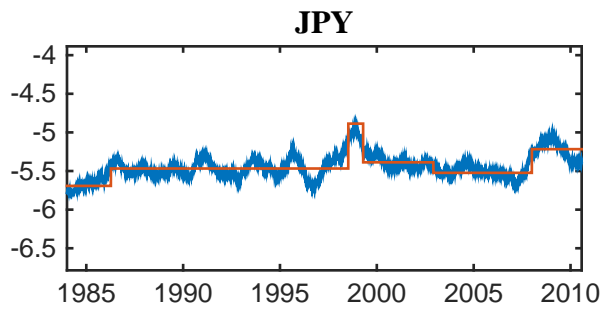Figure 2: Fitted level shifts and smoothed estimates of exchange rate volatilities

Figure 3: Autocorrelations of the original series and the residuals