

Combining Long Memory and Level Shifts in Modeling and Forecasting the Volatility of Asset Returns*

Rasmus T. Varneskov[†]

Aarhus University and CREATES

Pierre Perron[‡]

Boston University

January 30, 2017

Abstract

We propose a parametric state space model of asset return volatility with an accompanying estimation and forecasting framework that allows for ARFIMA dynamics, random level shifts and measurement errors. The Kalman filter is used to construct the state-augmented likelihood function and subsequently to generate forecasts, which are mean and path-corrected. We apply our model to eight daily volatility series constructed from both high-frequency and daily returns. Full sample parameter estimates reveal that random level shifts are present in all series. Genuine long memory is present in most high-frequency measures of volatility, whereas there is little remaining dynamics in the volatility measures constructed using daily returns. From extensive forecast evaluations, we find that our ARFIMA model with random level shifts consistently belongs to the 10% Model Confidence Set across a variety of forecast horizons, asset classes, and volatility measures. The gains in forecast accuracy can be very pronounced, especially at longer horizons.

Keywords: Forecasting, Kalman Filter, Long Memory Processes, State Space Modeling, Stochastic Volatility, Structural Change.

JEL classification: C13, C22, C53

*We wish to thank Asger Lunde for providing cleaned high frequency tick data. Furthermore, we thank Torben G. Andersen, Kim Christensen, Oliver Linton, Morten Ørregaard Nielsen, Anders Rahbek, an anonymous referee, as well as seminar participants at CREATES, Aarhus University, and at the 2nd Humboldt-Copenhagen conference, May 2011, for helpful comments and suggestions. An earlier version of this paper entitled “Combining Long Memory and Level Shifts in Modeling and Forecasting of Persistent Time Series” has been circulated since November 2010. Financial support from Aarhus School of Business and Social Sciences, Aarhus University, and CREATES, funded by the Danish National Research Foundation (DRNF78), is gratefully acknowledged.

[†]Department of Economics and Business, Aarhus University, Fuglesangs Allé 4, 8210 Aarhus V., Denmark. Email: rvarneskov@creates.au.dk.

[‡]Department of Economics, Boston University, 270 Bay State Rd., Boston, MA, 02215. Email: perron@bu.edu.

1 Introduction

The literature on asset return volatility modeling has surged since the introduction of the ARCH model by Engle (1982) due to numerous potential applications in financial economics such as asset and derivative pricing, risk management and portfolio selection. In addition, various volatility-linked derivatives are nowadays being actively traded on the Chicago Board of Options Exchange and in over-the-counter markets. Recently, Andersen, Bollerslev, Diebold & Ebens (2001), Andersen, Bollerslev, Diebold & Labys (2001, 2003), Koopman, Jungbacker & Hol (2005), Deo, Hurvich & Lu (2006), Andersen, Bollerslev & Diebold (2007), Corsi (2009), Chiriac & Voev (2011) and Varneskov & Voev (2013), among others, demonstrate that various realized volatility time series display characteristics compatible with fractionally integrated, or $I(d)$, processes, and that the modeling of such “long memory” features significantly improves the precision of out-of-sample forecasts of future return volatility.

We may formally define fractional integration or, as we will label it throughout, *genuine long memory* as follows; let $e_t = C(L)\epsilon_t$ with $\epsilon_t \sim \text{i.i.d.}(0, \sigma_\epsilon^2)$ and $\mathbb{E}[|\epsilon_t|^r] < \infty$ for some $r > 2$ be a short memory process with lag polynomial $C(L) = \sum_{i=0}^{\infty} c_i L^i$ satisfying $\sum_{i=0}^{\infty} i|c_i| < \infty$ and $C(1) \neq 0$, then $h_t = (1 - L)^d e_t$ for $t = 1, 2, \dots$ is fractionally integrated of order d , or $I(d)$, with autocovariance function

$$R_h(\tau) = g(\tau)\tau^{2d-1}, \quad \text{as } \tau \rightarrow \infty, \quad (1)$$

where $g(\tau)$ is a slowly varying function as τ increases. The properties of such processes depend critically on the magnitude of the fractional integration order, d . In this paper, we shall mainly be concerned with the case $0 \leq d < 1/2$, that is, with a stationary process that exhibits genuine long memory whenever $d > 0$, and which is characterized by having hyperbolically decaying autocovariances. However, we will also make references to the non-stationary case $d \geq 1/2$. The fractional ARIMA, or ARFIMA, model, independently introduced by Granger & Joyeux (1980) and Hosking (1981), is a flexible time series specification that captures genuine long memory and, as a result, has become popular for volatility modeling and forecasting, e.g., Andersen, Bollerslev, Diebold & Labys (2003).

Recently, however, a parallel literature has studied the possibility of genuine long memory being confused with a short memory process contaminated by random level shifts, spurred by the expositions in Perron (1989, 1990), who show that unit roots ($d = 1$) and structural changes are easily confused in the sense that the sum of the autoregressive coefficients is biased towards one if a stationary process is contaminated by level shifts. Applying this concept to the context of genuine long memory modeling, Lobato & Savin (1998), Diebold & Inoue (2001), Granger & Hyung (2004), and Perron & Qu (2007, 2010), among others, show theoretically and through simulations that if a short memory process is contaminated by random level shifts, the resulting time series will display many of the same characteristics as one of genuine long memory; for example, hyperbolically decaying autocovariances.¹ Motivated by these findings, Lu & Perron (2010) and Qu & Perron (2013), extending earlier work by Chen & Tiao

¹Related findings are made by Bhattacharya, Gupta & Waymire (1983), Mikosch & Střaričá (2004), Střaričá & Granger (2005), Ohanissian, Russell & Tsay (2008), and Christensen & Varneskov (2016).

(1990) and McCulloch & Tsay (1993), propose parametric models of asset return volatility, which allow for both random level shifts and short memory dynamics. They perform empirical analyses using daily stock index returns and argue that the (genuine) long memory properties of the volatility in such series are, indeed, spurious. These findings are corroborated in Xu & Perron (2014). Similar conclusions arise from another branch of the literature, which consider semi-parametric estimation and testing for genuine long memory; see, e.g., Smith (2005), Perron & Qu (2010), Qu (2011), McCloskey & Perron (2013), and McCloskey & Hill (2015). However, the proposed semi-parametric frameworks have a disadvantage in that random level shifts are not identified, making them unsuitable for forecasting.

As such, we face a dual problem. The presence of random level shifts may bias the parameter estimates for genuine long memory models and, consequently, lead to misspecified dynamics of asset return volatility. However, the presence of genuine long memory may also cause spurious detection of random level shifts in the series; see, e.g., Nunes, Newbold & Kuan (1995) and Granger & Hyung (2004). As a solution to this problem, we propose a parametric framework for asset return volatility modeling, which allows volatility to exhibit *both* random level shifts and ARFIMA dynamics. Furthermore, we allow for measurement errors in the observable volatility proxies such that we may analyze series constructed from daily as well as high-frequency data. The idea of combining random level shifts with a fractionally integrated component for time series modeling resembles the strategy in Ray & Tsay (2002). However, we introduce a framework that augments their Bayesian approach in four different directions; by allowing for a short memory ARMA component, by allowing for measurement errors, by allowing random level shifts to occur at each time t , and not in larger blocks, and, finally, we extend their analysis by providing a forecasting framework for the general class of models considered.

In particular, we propose a parametric state space framework to estimate the class of models and perform out-of-sample forecasting. The estimation procedure is similar to the one introduced by Perron & Wada (2009) and Lu & Perron (2010) where the basic principle is to augment the probability of states by the realizations of a mixture of normally distributed processes and apply the Kalman filter to construct the likelihood function conditional on the realization of states. However, an additional challenge arises since there exists no exact finite state space representation if the underlying process contains a genuine long memory component. We argue and show through simulations that this problem may be solved by using a relatively smaller order truncation of lags, which makes estimation feasible in practice, largely without loss of precision in the parameter estimates.² In addition to analyzing the truncation order, our simulation study demonstrates the adequacy of the estimation methodology as well as compares the estimated memory parameters from our random level shift ARFIMA, or RLS-ARFIMA, model with standard ARFIMA parameter estimates, which, as we illustrate, are severely affected by random level shifts. The recursive structure of the Kalman filter allows us to introduce a new forecasting framework for general parametric random level shift models, which utilizes the information in the Kalman recursions to generate forecasts for a given state and, then, weight them with the probability of being on a given

²In principle, one always needs to truncate the number of included lags when estimating models with ARFIMA dynamics. Our contribution comes from showing that this truncation order can be relatively small for autoregressive representations of the latter in a state space context with level shifts. This eases the computational burden considerably.

transition path. Hence, the forecasts are both mean and path-corrected.

We apply the proposed reduced form modeling framework to eight daily asset return volatility series, which differ, not only with respect to the sampling frequency with which they are constructed, using either daily or high-frequency data, but also according to time span and asset class. We compare the full sample parameter estimates and out-of-sample forecasting performance of our RLS-ARFIMA model to six popular models in the literature and uncover some novel empirical findings. First, the random level shift component is important for all series, delivering more frequent shifts for all volatility proxies constructed from high-frequency data, but with less variability for most compared to those associated with the daily return series. Second, once level shifts are taken into account, most high-frequency volatility measures are characterized by a large genuine long memory component, whereas the remaining dynamics of the volatility proxies, constructed as log-absolute returns, may be described as a combination of short memory dynamics and measurement errors. As such, it is not surprising that the measurement errors are larger for the log-absolute return series than for the volatility measures constructed from high-frequency data, which are known to be more efficient, but the differences in terms of persistence of the remaining dynamics are striking. Third, we show that if one fails to take both genuine long memory and random level shifts into account, the resulting parameter estimates will reflect either spurious long memory or spurious breaks. Most importantly, however, from our out-of-sample forecasting analysis, we show that the RLS-ARFIMA model is, by far, the most frequent member of the 10% Model Confidence Set (MCS) proposed by Hansen, Lunde & Nason (2011). It delivers good out-of-sample performance across various forecast periods, forecast horizons, asset classes, and volatility measures. The forecast gains can be very pronounced, especially at longer horizons.

The outline of the paper is as follows. Section 2 introduces the discrete time volatility model, and Section 3 describes the data as well as provides motivational evidence. Section 4 re-casts the model in a state space framework and introduces the forecasting procedure. The simulation study is presented in Section 5, while Section 6 has the empirical analysis and robustness checks. Finally, Section 7 concludes. The appendix in Section A details our treatment of measurement errors, and the web appendix, Varneskov & Perron (2017), contains additional theory, evidence, and proofs.

2 A Discrete Time Volatility Model

We aim to provide a unified *discrete time* framework for capturing the dynamics of *daily* volatility measures, constructed from either daily or high-frequency data. Hence, we need to specify a general time series model that not only accommodates some of the extensively documented empirical regularities of such processes such as volatility clustering, genuine long memory and/or random level shifts, but also allows for measurement errors in the volatility proxies. The inclusion of such features will allow us to assess which components are the most important contributors to the variation in different volatility series without taking a stance on modeling paradigm, and we will, thus, nest them within the same parametric framework. Specifically, let $x_t \in \mathbb{R}$ denote the latent, univariate *logarithmic* volatility process, then we

assume that the *observable* log-volatility proxy, $y_t \in \mathbb{R}$, follows the signal-plus-noise model:

$$x_t = a + h_t + v_t, \quad (2)$$

$$y_t = x_t + u_t \quad \text{where} \quad u_t \sim \text{i.i.d.N}(0, \sigma_u^2) \quad (3)$$

is the measurement error in the volatility proxy, a is a constant, h_t is a stationary long memory process, and v_t is the random level shift component.³ The simple decomposition of the model (2)-(3) encompasses many parametric volatility models in the extant literature as well as nests all subsequent models developed in this paper. Next, we impose a parametric structure on both v_t and h_t . First, we assume that the random level shift process is given by

$$v_t = \sum_{j=1}^t \delta_{T,j} \quad \text{where} \quad \delta_{T,j} = \pi_{T,j} \eta_j, \quad \eta_j \sim \text{i.i.d.N}(0, \sigma_\eta^2), \quad \pi_{T,j} \sim \text{i.i.d.Bernoulli}(\gamma/T),$$

for some $\gamma \in [0, T]$, that is, the process is modeled as the sum of level shifts of magnitudes η_j , drawn from a Gaussian distribution, and whose frequency are determined by $\pi_{T,j}$. Second, the genuine long memory component of the model, h_t , is assumed to obey ARFIMA dynamics,

$$\Phi(L)(1-L)^d h_t = \Theta(L)\epsilon_t, \quad \text{where} \quad \epsilon_t \sim \text{i.i.d.N}(0, \sigma_\epsilon^2),$$

and $\Phi(L) = (1 - \phi_1 L - \dots - \phi_p L^p)$ and $\Theta(L) = (1 - \theta_1 L - \dots - \theta_q L^q)$ are autoregressive and moving average lag ($Lh_t = h_{t-1}$) polynomials of orders p and q , respectively. The component h_t captures the transitory part of the model. Moreover, its stationarity and uniqueness, thus allowing identification of the parameters, are assured by assuming $0 \leq d < 0.5$ and that the roots of $\Phi(x)=0$ and $\Theta(x) = 0$ are outside the unit circle and distinct; see, e.g., Brockwell & Davis (1991, p. 525). Last, we assume that the components $\pi_{T,t}$, η_t , u_t and h_t are mutually independent.⁴

Before proceeding, several features of the model should be highlighted. First, by imposing either $\gamma = 0$ or $\sigma_\eta = 0$, we recover the long memory stochastic volatility (LMSV) model and, if $\sigma_u = 0$ is additionally imposed, the stationary ARFIMA model, which are advanced by Deo et al. (2006) and Andersen et al. (2003) in the context of realized volatility modeling and forecasting. This implies that if either $\gamma = 0$ or $\sigma_\eta = 0$, the other parameter affecting the random level shift process is not identified. This feature is evident in our simulation study in Section 5. However, and as we will elaborate upon in later sections, the likelihood function for the ARFIMA parameters are unaffected by this boundary case. Also, since we find both $\gamma > 0$ and $\sigma_\eta > 0$ for all series considered, and the main emphasis is on forecasting, the possibility of non-identified parameters is innocuous for the present analysis.

³In the supplementary appendix, Varneskov & Perron (2017), we briefly discuss how the volatility in discrete time return models relate to the quadratic variation from continuous time return models. Moreover, we make a direct comparison of the *discrete* signal-plus-noise model in (2)-(3) to a contemporaneous *continuous time* stochastic volatility model.

⁴Following a previously circulated draft of this paper, Grassi & de Magistris (2014) study the small sample properties of estimators of the integration order, d , using a simplified version of the proposed model (2)-(3) in a simulation setup.

Second, if we impose $d = 0$, we recover a short memory stochastic volatility model with ARMA dynamics and random level shifts in the mean. We note that even this restricted version of the model generalizes the corresponding model in Qu & Perron (2013) by allowing for an MA component and, similarly, Lu & Perron (2010) by accommodating both an MA component and measurement errors in the series. Hence, our framework in (2)-(3) offers substantial flexibility when modeling the dynamics of various daily log-volatility measures. In particular, it allows us to remain agnostic as to whether the persistent features of the series are better described by genuine long memory, random level shifts, or both, and it may be applied to daily as well as high-frequency measures of volatility.

Third, we impose normality on ϵ_t and u_t , which may be restrictive considering that measurement errors for daily volatility proxies, in particular, can be highly non-Gaussian. The assumption, however, should be interpreted in a quasi-maximum likelihood (QML) sense. That is, we use it to derive the predictive likelihood function via the Kalman filter to estimate different versions of the model, similarly to the strategy devised by, e.g., Harvey & Shephard (1996) for short memory stochastic volatility models who show that consistency and asymptotic normality still hold when the measurement errors deviate from Gaussianity for a related QML estimator based on the Kalman filter. Moreover, as we analyze logarithmic transformations of the volatility proxies, we do not expect to see dramatic violations of Gaussianity, cf. the distributional results in Andersen, Bollerslev, Diebold & Ebens (2001), Andersen, Bollerslev, Diebold & Labys (2001) as well as the summary statistics provided below.

Fourth, the accommodation of measurement errors in the signal-plus-noise model has implications for the *reduced form* dynamics of the observable log-volatility proxy, y_t . In particular, and similarly to the analyses in Meddahi (2003) and Hansen & Lunde (2014), who assume that realized volatility proxies obey ARMA dynamics, we may reformulate the model as

$$(1 - L)^d \Phi(L)(y_t - a - v_t) = \Theta(L)\epsilon_t + (1 - L)^d \Phi(L)u_t. \quad (4)$$

This representation has implications for how we treat measurement errors and interpret the estimated MA parameters. A detailed discussion of these issues is deferred to Section 4 and Appendix A.

Finally, we stress that the Bernoulli probability of a random level shift is dependent on the sample size, T , to make the expected number of shifts constant and equal to γ . This is needed to model structural changes in mean (or rare events), which affect the properties of the series until the next shift (event) occurs. The long memory component allows the process to have transitory shocks that are long-lasting in periods between structural changes. For example, in the context of volatility modeling, this may potentially capture volatility clustering between financial crises (which may be seen as rare events). If only one persistent component is present in the log-volatility series, our model is able to assess whether it is better described by genuine long memory or random level shifts. If level shifts are present, however, as clearly seen from equation (4), our model is non-stationary with level $a + v_t$, thus devoid of long-run mean reversion (recall, the breaks are i.i.d.). While the latter is generally not accepted when volatility hits extreme levels, our model should be viewed as a (better) finite sample approximation to the log-volatility dynamics. In our model, “mean reversion” from extreme levels will be captured by

another (downward) level shift when the volatility level has decreased sufficiently.

Remark 1. *The local level model $y_t = x_t + u_t$, $x_t = x_{t-1} + \epsilon_t$ put forth in, e.g., Harvey (1989), is embedded in our framework by imposing a level shift in each period, i.e., $\gamma = T$. However, in general, we require $\gamma \in (0, T)$ fixed such that $\gamma/T \rightarrow 0$ as $T \rightarrow \infty$ for the level shift component to generate autocorrelations akin to genuine long memory, see, e.g., Perron & Qu (2010). Furthermore, we find that $\gamma = T$ is strongly rejected for all series in our empirical analysis.*

Remark 2. *For many applications in economics and finance, it is the volatility in either its standard deviation or variance form (and its corresponding forecast) that is of interest to researchers. We focus on the log-volatility process in (2)-(3) to allow for both positive and negative breaks of unknown and random magnitudes as well as to alleviate the parameter biases, which often plague estimation techniques when the innovations are highly non-Gaussian, as found for the alternative volatility transformations. For example, Haldrup & Nielsen (2007) find that outliers cause a substantial negative bias in different estimators of the fractional integration order, d . We will, however, discuss how to extrapolate volatility and variance forecasts from our log-volatility model in Section 6.4 and give an empirical illustration.*

3 Empirical Volatility Measures and Preliminary Evidence

This section describes the data, the empirical volatility proxies, and it provides preliminary summary statistics as well as some initial tests of particular dynamic features of the series to motivate the proposed volatility modeling framework.

3.1 Data and Empirical Volatility Measures

We consider eight daily log-volatility series in our empirical analysis, which differ, not only according to the sampling frequency of the data with which they are constructed, but also according to time span and asset class: (1) For three stocks, Bank of America Corp. (BAC), Merck & Co., Inc. (MRK), and the Standard & Poor's Depository Receipts (SPY), we have tick-by-tick trades available with observations stamped to the nearest second from January 1997 through July 2008; (2) For futures contracts on the S&P 500 and 10-year Treasury bonds, we have one-minute observations available for every trading day from January 1983 through May 2009; (3) For the three exchange rates, USD-AUD, USD-CHF, and USD-JPY, we have daily observations available from January 4th 1971 through April 10th 2009.⁵

The number of trading days, hence the time span, is considerably smaller for the volatility measures constructed from intra-daily data than for the daily volatility proxies. However, from the theory of quadratic variation, it is well-known that, under mild conditions on the *efficient* price process, we may utilize high-frequency data to get a precise estimate of the whole return variance trajectory over a (trading) day. In particular, if the applied estimator is able to account for an array of market frictions that are inherent to *observable* intra-daily log-prices, then high-frequency data-based estimates

⁵We are grateful to Asger Lunde for providing cleaned tick data.

of quadratic variation make unbiased and efficient proxies, thus having measurement errors that are vanishingly small, which has been shown to improve out-of-sample forecasting in, e.g., Andersen et al. (2003), Koopman et al. (2005), Deo et al. (2006), and Varneskov & Voev (2013).

The volatility for the three daily exchange rate series is proxied by log-absolute returns.⁶ The daily quadratic variation, on the other hand, for the remaining series with high-frequency data available is estimated using the flat-top realized kernel approach, put forth in Varneskov (2016a, 2016b), since it is robust to general forms of market microstructure noise and has optimal asymptotic, as well as good finite sample, properties.⁷ Each flat-top realized kernel estimate is subsequently square-root and log-transformed such that its unit is comparable to that of log-absolute returns.⁸ We provide a few unconditional and conditional summary statistics of the eight volatility proxies in Table 1.

From the unconditional summary statistics, we see that the three exchange rate volatility series display slightly more left-skewed distributions with slightly higher excess kurtosis relative to the remaining series based on high-frequency data. However, it is clear that the logarithmic transformation has removed the pronounced right-skew and excess kurtosis, which usually characterize volatility proxies in their standard deviation or variance form. These distributional results are in line with prior findings, e.g., Andersen, Bollerslev, Diebold & Ebens (2001) and Andersen, Bollerslev, Diebold & Labys (2001).

3.2 Preliminary Evidence on Volatility Dynamics

As an initial gauge of the conditional properties of the series, we present log-periodogram (LP) and local Whittle (LW) estimates of the fractional integration order using a bandwidth $m = \lfloor T^{1/2} \rfloor$. Furthermore, we include results from the testing procedure of Perron & Qu (2010) for the null hypothesis that the volatility series have genuine long memory against the alternative of being comprised of level shifts and short memory dynamics, and a similar test by Qu (2011), which shares the same null hypothesis, but also allows the alternative to be a combination of genuine long memory and level shifts.⁹

The point estimates of the fractional integration order, d , from the LP and LW estimators suggest that all volatility series have $d > 1/2$, that is, are fractionally integrated within the non-stationary range. At first glance, this makes our assumption that $0 \leq d < 1/2$ seem erroneous. However, if the series contain level shifts, these will dominate the periodogram behavior at the very lowest frequencies, causing an upward bias in the LP and LW estimators. Moreover, in the supplementary appendix, we show that the two estimators are very sensitive to the number of frequency ordinates included, showing almost monotonically declining d estimates as m increases, in addition to a steep pole near the origin, and we argue that this can be interpreted as evidence of level shifts (see the web appendix for details).

⁶Strictly speaking, we use $\ln(|r_t| + 0.001)$, r_t being the daily log-return, to bound zero daily returns away from minus infinity. This follows, e.g., Hurvich & Ray (2003), Stărică & Granger (2005), Perron & Qu (2010) and references therein.

⁷We provide details on the flat-top realized kernel estimator and its implementation in the supplementary appendix.

⁸Note that the measures based on high-frequency data account for the quadratic variation over one *trading* day. As a result, they differ from the daily exchange rate series, which include holiday, overnight, and weekend effects.

⁹We detail the testing procedures and the LP and LW estimators in the supplementary appendix, where we also provide a more in-depth analysis of the conditional properties of the volatility series. This includes theoretical and empirical results on the autocorrelation function for time series with genuine long memory, random level shifts, and measurement errors.

As a more formal check of whether the seemingly non-stationary fractional integration in the series is generated (exclusively) by a genuine long memory component, we apply the tests of Perron & Qu (2010) and Qu (2011). From these results in Table 1, we find clear evidence against the null hypothesis of no level shifts for the USD-AUD and USD-JPY series, no significant evidence against it for the MRK and SPY series, and mixed evidence against it for the remaining series. Hence, since we are unable to make definitive statements about the underlying data generating process for the eight log-volatility series using these semi-parametric tests, this suggest to incorporate both genuine long memory and random level shift components to capture the low-frequency variation in the daily volatility measures and, subsequently, for generating competitive volatility forecasts.

4 Econometric Methodology

In this section, we re-cast the reduced form model (4) in state space form to provide a feasible estimation and forecasting framework, generalizing the estimation methods in Perron & Wada (2009) and Lu & Perron (2010) by allowing for genuine long memory. Additionally, we provide a forecasting procedure, which is easy to implement and may also be used for previously proposed short memory random level shift models. From (4), we see that an RLS-ARFIMA(p, d, ∞) structure is generally needed to describe the reduced form dynamics of the log-volatility, y_t . Going forward, however, y_t is treated as having an MA component of (finite) order q . As such, this may be seen as restricting $y_t = x_t$. However, we emphasize that since an ARMA(p, q) process plus noise has ARMA($p, \max(p, q)$) representation, the procedures developed here do, indeed, accommodate measurement errors. To further support this claim, we provide empirical evidence and a detailed discussion in Appendix A, arguing that all series are appropriately described by an RLS-ARFIMA($1, d, 1$) parameterization. Hence, with a slight abuse of notation, we use $\Theta(L)$ to describe the MA lag structure in (4) in the remaining part of the paper, despite encompassing both the case with and without measurement errors in the volatility proxy. Moreover, we are careful when interpreting the estimated MA parameters, which may reflect either measurement errors, an MA component, or a combination of the two.

4.1 State Space Representation

First, redefine the random level shift component, v_t , as a random walk with innovations that obey a mixture of two normally distributed processes,

$$v_t = v_{t-1} + \delta_{T,t} \quad \text{where} \quad \delta_{T,t} = \pi_{T,t}\eta_{1t} + (1 - \pi_{T,t})\eta_{0t}$$

and $\eta_{jt} \sim \text{i.i.d.N}(0, \sigma_{\eta_j}^2)$ for $j = (0, 1)$. We impose the restrictions $\sigma_{\eta_1}^2 = \sigma_{\eta}^2$ and $\sigma_{\eta_0}^2 = 0$ to recover the representation in (2). The intuition for reducing the two components of the model to one is the following; if a structural change occur, it will have a long-lasting impact on the volatility level, at least until the next structural change. However, writing v_t using this “two-component-form” allows us to adopt a state

space representation that resembles the corresponding one for Markov regime switching models see, among others, Hamilton (1994b), and this is helpful in developing the estimation procedure. Moreover, this specification also highlights that level shifts are modeled as independent random events, which are invariant to past realizations of the data. Next, under the conditions of Section 2, the ARFIMA long memory component, h_t , in (2)-(3) may be written as an AR(∞) process,

$$h_t = \sum_{i=1}^{\infty} \psi_i h_{t-i} + \epsilon_t, \quad \text{where} \quad \sum_{i=0}^{\infty} \psi_i L^i = \frac{\Phi(L)}{\Theta(L)} (1-L)^d, \quad (5)$$

and for which the contribution of the fractional difference filter may be written as a binomial expansion $(1-L)^d = \sum_{i=0}^{\infty} \pi_i L^i$ with $\pi_i = \Gamma(i-d)/(\Gamma(i+1)\Gamma(-d))$ where $\Gamma(\cdot)$ is the gamma function. Using this representation, we may rewrite y_t in first differences as $\Delta y_t = h_t - h_{t-1} + \delta_{T,t}$ for $t = 2, \dots, T$. Similarly to the frameworks for ARFIMA models in Chan & Palma (1998) and Beran (1995), Δy_t does not have an *exact* finite dimensional state space representation unless $d = 0$ and $p, q < \infty$. Hence, we follow the literature and *approximate* the AR(∞) process by an AR(M) where M must be chosen suitably. We discuss theoretical as well as finite sample guidance for M in Sections 4.2 and 5 below. Now, by combining the mixture of normals formulation for v_t above with (5), the approximate state space matrix representation of Δy_t is given by

$$\Delta y_t = \mathbf{F} \mathbf{H}_t + \delta_{T,t}, \quad \mathbf{H}_t = \mathbf{G} \mathbf{H}_{t-1} + \mathbf{E}_t \quad (6)$$

where $\mathbf{F} = (1, -1, 0, \dots, 0)'$, $\mathbf{H}_t = (h_t, h_{t-1}, \dots, h_{t-M+1})$, and $\mathbf{E}_t = (\epsilon_t, 0, \dots, 0)$ are $M \times 1$ vectors, $\mathbf{E}_t \sim \text{i.i.d.N}(\mathbf{0}_{M \times 1}, \mathbf{Q})$ and $\mathbf{0}_{M \times 1}$ denotes a $M \times 1$ vector of zeros. Here, \mathbf{G} and \mathbf{Q} are both $M \times M$ matrices of parameters and identifying terms,

$$\mathbf{G} = \begin{pmatrix} \Psi_{M-1} & \psi_M \\ \mathbf{I}_{M-1} & \mathbf{0}_{(M-1) \times 1} \end{pmatrix}, \quad \mathbf{Q} = \begin{pmatrix} \sigma_\epsilon^2 & \mathbf{0}_{1 \times (M-1)} \\ \mathbf{0}_{(M-1) \times 1} & \mathbf{0}_{(M-1) \times (M-1)} \end{pmatrix},$$

where $\Psi_M = (\psi_1, \dots, \psi_M)$ is $1 \times M$ and \mathbf{I}_M is an M -dimensional identity matrix. The added challenge relative to the genuine long memory state space framework of Chan & Palma (1998) is due to the state-dependent error in the measurement equation, whereas relative to Lu & Perron (2010), it is the presence of $(1-L)^d/\Theta(L)$ in the representation of h_t such that no finite state space representation exists.

4.2 Maximum Likelihood Estimation

The basic principle behind the estimation procedure is to augment the probability of states (or different level regimes) by the realizations of a mixture of normally distributed processes at time t and apply the Kalman filter to construct the likelihood function conditional on the realization of states. Since we truncate the AR(∞) representation of h_t in (5) at lag M , the resulting estimation method becomes similar to the corresponding procedures in Perron & Wada (2009) and Lu & Perron (2010), despite significantly generalizing their modeling frameworks. Hence, details on the construction of the log-

likelihood function are deferred to the supplementary appendix.

It is important to note, however, that if either $\gamma = 0$ or $\sigma_\eta = 0$, the other parameter affecting the level shift process is not identified, and the maximum estimation procedure collapses to the genuine long memory state space framework analyzed by Chan & Palma (1998). In this case, and if defining the parameter vector $\boldsymbol{\Sigma} = (\sigma_\eta, \gamma, \sigma_\epsilon, d, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)'$, then we know from their Theorems 3.1 and 3.2 that the estimates of the ARFIMA parameters $\boldsymbol{\Pi} = \boldsymbol{\Sigma} \setminus \{\gamma, \sigma_\eta\}$, denoted $\hat{\boldsymbol{\Pi}}$, are consistent when $M = T^\beta$ with $\beta > 0$, and when $\beta \geq 1/2$, $\sqrt{T}(\hat{\boldsymbol{\Pi}} - \boldsymbol{\Pi}) \xrightarrow{\mathbb{D}} N(\mathbf{0}, \boldsymbol{\Lambda}^{-1}(\boldsymbol{\Pi}))$ where $\boldsymbol{\Lambda}(\boldsymbol{\Pi})$ is the usual information matrix. In other words, the ARFIMA parameter estimates have the usual maximum likelihood properties and are unaffected by the possible event of non-identification of the level shift parameters. These asymptotic results, thus, provide theoretical guidance for the selection of M , and we compare and discuss the choice $M = T^{1/2}$ to other rule-of-thumb selections in Section 5.

While the ARFIMA parameter estimates have the usual maximum likelihood properties in the event of non-identification of the level shift parameters, the constant, a , will not be identified by our estimation procedure irrespective of whether random level shifts occur in the series, or not. If level shifts are present, that is, if we have $\gamma > 0$ and $\sigma_\eta > 0$, the constant a may simply be absorbed into the initial value of the level shift process, v_0 , without loss of generality. If there are no level shifts in the series, however, the proposed estimation procedure will *not* identify a since we consider Δy_t . Hence, we suggest to estimate the model parameters using the following 2-step procedure:

- (1) Estimate $\boldsymbol{\Sigma}$ using the Kalman filter maximum likelihood procedure discussed above and in the supplementary appendix. If the estimates have $\hat{\gamma} > 0$ and $\hat{\sigma}_\eta > 0$ and are significant, stop here.
- (2) If either of the estimates $\hat{\gamma}$ or $\hat{\sigma}_\eta$ is insignificantly different from 0, estimate an ARFIMA(p, d, q) model with non-zero mean, a , using the conditional sum-of-squares (CSS) estimator, cf. Beran (1995) and Nielsen (2015), and the first-stage estimates $\hat{\boldsymbol{\Pi}}$ as initial values.

Whereas this 2-step procedure is, indeed, applied to all RLS-ARFIMA specifications considered in the empirical analysis below, we note that all models stop after the first step. That is, we find significant level shifts in all series. When considering ARFIMA specifications *without* modeling random level shifts in the empirical analysis, we use the CSS estimator directly, as this not only allows for the inclusion of a constant, but this estimator is also valid for stationary as well as non-stationary values of d . The latter is important since failure to model random level shifts may bias the estimate of the fractional integration order into the non-stationary range, as for the LP and LW estimators.

4.3 Forecasting with the RLS-ARFIMA Model

In addition to a unified framework for model parameter estimation, covering both the case with and without random level shifts in Steps 1 and 2 above, respectively, two corresponding forecast procedures are needed. For the case without random level shifts in Step 2, we apply standard ARFIMA forecasts; see (8) below and, among others, Brockwell & Davis (1991) and Doornik & Ooms (2004).¹⁰ However,

¹⁰This forecasting method is also applied to all ARFIMA(p, d, q) specifications that does *not* model random level shifts.

as mentioned above, since we find evidence of random level shifts in all series, it is important to develop a forecasting procedure for Step 1, i.e., for the proposed state space model in (6).

Hence, let us first define $\mathbf{Y}_t = (\Delta y_2, \Delta y_3, \dots, \Delta y_t)'$, and denote the filtered state vector and its associated covariance matrix by $\mathbf{H}_{t|t}^{ij}$ and $\mathbf{P}_{t|t}^{ij}$, respectively, both of which depend on whether a random level shift occurs at either time $t - 1$, time t , neither, or both, as indexed by the (ij) superscript. Specifically, they signify that $\pi_{T,t-1} = i$ and $\pi_{T,t} = j$ for $(i, j) \in \{0, 1\}^2$. The dependence on only one transition between regimes arises naturally from the Kalman predictions.¹¹ Since level shifts are modeled as independent events, thus invariant to past realizations of the data, one may apply standard rules for conditional probabilities in each time period to construct the most likely path of the state vector. Hence, at time $t - 1$, the Kalman filter constructs the prediction,

$$\mathbf{H}_{t|t-1}^i = \mathbf{G}\mathbf{H}_{t-1|t-1}^i, \quad \mathbf{H}_{t-1|t-1}^i = \frac{\sum_{k=0}^1 \Pr(\pi_{T,t-2} = k, \pi_{T,t-1} = i | \mathbf{Y}_{t-1}; \boldsymbol{\Sigma}) \mathbf{H}_{t-1|t-1}^{ki}}{\Pr(\pi_{T,t-1} = i | \mathbf{Y}_{t-1}; \boldsymbol{\Sigma})} \quad (7)$$

where $\mathbf{H}_{t-1|t-1}^i$ is the probability weighted value of the state vector based on the likelihood of a level shift having occurred in time $t - 2$, and conditional on being in regime i at time $t - 1$. Hence, as this predictive relation can be traced backwards, $\mathbf{H}_{t-1|t-1}^i$ reflects the history of the state vector, recursively weighted to fit the most likely realization of regimes (or random level shifts). However, since a level shift may also occur at time t , the predictive updating from $\mathbf{H}_{t|t-1}^i$ to $\mathbf{H}_{t|t}^{ij}$ reflects this. Now, to obtain a prediction of Δy_{t+1} , we are interested in the value of the state vector $\mathbf{H}_{t+1|t+1}^{ij}$, conditional on time t information, with (ij) referring to the same transition between regimes as above, but occurring between times t and $t + 1$. The forecasts of this state vector as well as the conditional probability of a level shift transition may be obtained from the Kalman filter recursions and Bayesian probability updating. Multi-step-ahead predictions may, then, be generated by applying the updating algorithm sequentially, in conjunction with the probability of future random level shifts being invariant to past realizations.¹² Hence, the state space structure of the RLS-ARFIMA(p, d, q) model in (6) allows us to obtain τ -step-ahead forecasts, for some integer $\tau > 0$, which is summarized by the following proposition:

Proposition 1. *Let y_t satisfy the conditions of Section 2 and let $\mathbb{E}_t[y_{t+\tau}] = \hat{y}_{t+\tau|t}$ denote the expected value of the process at time $t + \tau$, conditional on the information available at time t , then the τ -step-ahead forecast is*

$$\hat{y}_{t+\tau|t} = y_t + \mathbf{F}\mathbf{G}^\tau \sum_{i=0}^1 \sum_{j=0}^1 \Pr(\pi_{T,t+\tau} = j) \Pr(\pi_{T,t} = i | \mathbf{Y}_t; \boldsymbol{\Sigma}) \mathbf{H}_{t|t}^{ij}.$$

Proof. See the supplementary appendix, Varneskov & Perron (2017). □

Proposition 1 illustrates two key differences between forecasts from the RLS-ARFIMA model and

¹¹We only convey the intuition behind the forecast construction here. For technical details on the Kalman recursions as well as the estimation procedure, we refer to the supplementary appendix.

¹²Our forecasting procedure is related to corresponding methods from the state space and Markov regime switching forecasting literature; see, e.g., Brockwell & Davis (1991), Hamilton (1994a), and Gabriel & Martins (2004).

standard ARFIMA forecasts, which, using the notation in (4) and (5), may be written on the form

$$\tilde{y}_{t+\tau|t} = a + \sum_{i=1}^{\infty} \psi_i^\tau h_{t+1-i}, \quad \text{where } h_t = y_t - a \quad (8)$$

when there are no random level shifts in the series. The first of these is a *mean* correction where the usual unconditional mean parameter, a , is replaced by y_t plus the innovations to a probability weighted history of the state vector, as reflected by past values of $\mathbf{H}_{t|t}^{ij}$ used in the predictive updating. Here, the filtered history of $\{\mathbf{H}_{k|t}^{ij}\}_{k=1}^t$, with (ij) referring to the transition between consecutive regimes at any given time k , captures both the dynamics of h_t as well as the random level shifts. That is, the generated forecast from $\hat{y}_{t+\tau|t}$ has time t updated information about which regime the process is currently in, its transition history, whereas no such information is conveyed in a . Of course, this information is also reflected in the lagged history $\{y_k\}_{k=1}^t$ on the right-hand-side of (8), but since $\tilde{y}_{t+\tau|t}$ is anchored by a , this can generate large prediction errors, in particular for medium horizon forecasting where there may be large discrepancies between the unconditional mean and the regime specific mean.

The second difference is a *path* correction. A τ -step-ahead forecast for the state vector realization is computed conditional on being in regime i at time t and regime j at time $t + \tau$, $\mathbf{H}_{t+\tau|t}^{ij} = \mathbf{G}^\tau \mathbf{H}_{t|t}^{ij}$, and then weighted by the probability of being on a given transition path between regimes at the respective time points, $\Pr(\pi_{T,t+\tau} = j) \Pr(\pi_{T,t} = i | \mathbf{Y}_t; \boldsymbol{\Sigma})$, which has been updated to reflect time t information. Since level shifts are invariant to past data, $\Pr(\pi_{T,t+\tau} = j) = \Pr(\pi_{T,t+1} = j)$ for integers $\tau \geq 1$. The path correction may be viewed as a predictive tilt of the state vector dynamics from h_t in $\mathbf{H}_{t|t}^{ij}$ relative to lagged ARFIMA dynamics in (8), where there are no such transitions between regimes.

Remark 3. *The proposed forecasting framework encompasses multiple types of forecasting schemes; recursive estimation using an expanding window of observations, rolling window of observations, and a one-time estimation of the parameters, which, in conjunction with the Kalman recursions, may be used to generate forecasts conditional on the parameter estimates.*

5 Simulation Study

In this section, we investigate the accuracy of the parameter estimates from the state space estimation methodology. To show the validity of our proposed estimation method, and to get an indication of how to select M , the order of truncation of the $\text{AR}(M)$ representation, we set up a simulation study to examine whether the RLS-ARFIMA model can distinguish between time series persistence generated by random level shifts, genuine long memory, or both. Additionally, we compare the parameter estimates to ones obtained from fitting $\text{ARFIMA}(p, d, q)$ models to gauge how the latter are affected by level shifts. Finally, we analyze estimation of the (RLS-)ARFIMA(p, d, q) parameters when level shifts are absent, since the presence of two non-identified parameters may lead to efficiency losses.

5.1 Simulation Setup and Implementation

We consider a Monte-Carlo study with $N = 100$ replications, sample size $T = 3000$, and four different truncation lengths $M = \{5, 10, 20, T^{1/2}\}$.¹³ The theoretically consistent selection $T^{1/2} \simeq 55$ is much larger than the remaining truncation orders, which are motivated by Chan & Palma (1998) and Martin & Wilkins (1999), who find that smaller truncation orders suffice to capture the dynamics of an ARFIMA process in related settings, albeit without random level shifts. The choice of sample size is motivated by the typical length of financial time series. We examine data generating processes (DGP's) that are simulated from an RLS-ARFIMA(1, d , 1) model,

$$y_t = x_t, \quad x_t = h_t + v_t, \quad (1 - L)^d(1 - \phi L)h_t = (1 - \theta L)\epsilon_t,$$

with $\phi = 0.2$, $\theta = -0.1$, $\sigma_\epsilon = 0.5$ as well as for **(DGP 1)** $d = 0$, $\gamma/T = 0.02$, $\sigma_\eta = 3\sigma_\epsilon$; **(DGP 2)** $d = 0.35$, $\gamma/T = 0$, $\sigma_\eta = 0$; **(DGP 3)** $d = 0.35$, $\gamma/T = 0.02$, $\sigma_\eta = 3\sigma_\epsilon$; and for **(DGP 4)** $d = 0.6$, $\gamma/T = 0.02$, $\sigma_\eta = 3\sigma_\epsilon$. The choice of parameters for the first three DGP's are based on estimates from the level shift literature for DGP 1, e.g., Qu & Perron (2013), from the long memory volatility modeling literature for DGP 2, e.g., Andersen et al. (2003), and from one that combines them for DGP 3, e.g., this paper's estimates for the S&P 500 series. We include DGP 4 as a robustness check to ensure that our empirical detection of random level shifts is not spuriously caused by a non-stationary fractionally integrated component. Moreover, note that these DGP's also correspond well with the discussion and motivational empirical evidence on measurement errors in Sections A.1 and A.2 of the appendix. Specifically, DGP 1 may capture the case where the residual dynamics - the dynamics once level shifts are taken into account - consist of a short memory process and measurement errors (again, AR(1) plus noise has an ARMA(1,1) representation), and DGP's 2-4 to the case where there are no measurement errors and the residual dynamics are of the ARFIMA(1, d , 1) form.¹⁴

The RLS-ARFIMA models are estimated as described in Section 4.2. Since all components of the state vector in (6) are stationary, we initialize the Kalman filter updating equations using their unconditional expected values, $\mathbf{H}_{0|0}^{ij} = \mathbf{0}_{M \times 1}$ and $\mathbf{P}_{0|0}^{ij} = \mathbf{Q}$. To start the probability weighting of the likelihood function, we set $\Pr(\pi_{T,0} = 1 | \mathbf{Y}_0; \boldsymbol{\Sigma}) = \gamma/T$. Lastly, we draw the initial values of the parameters from a uniform distribution five times and select the optimized estimates with the highest associated log-likelihood value. The ARFIMA(p, d, q) models in (5) are estimated using the CSS estimator (see also Section 4.2), where the residual standard deviation is computed as $\hat{\sigma}_\epsilon = \sqrt{(T-1)^{-1} \sum_{t=1}^T \hat{\epsilon}_t^2}$ with $\hat{\epsilon}_t$ being the model-implied residuals. For all (RLS-)ARFIMA models, we restrict attention to the (0, d , 0) and (1, d , 1) parameterizations, in line with the discussion and empirical results in the previous section and the appendix, and since simpler models are often advocated for out-of-sample forecasting.

¹³We also performed some simulations for sample sizes $T = 1000$ and $T = 5000$, which showed proportionally worse/better results. Ideally, we would carry out the simulations for $N \gg 100$. However, this presents a computation challenge, in particular for large M . Hence, the results should be interpreted as indicative rather than definitive.

¹⁴We have also fitted an RLS-ARFIMA(1, d , 1) model to a simulated long memory stochastic volatility model with random level shifts, the RLS-LMSV model, which is considered in Section A.2 of the appendix. The simulation results for the key persistence parameters are similar to the ones reported in Table 9 of the appendix and are, thus, omitted.

5.2 Simulation Results

The bias and RMSE of the parameter estimates for all estimators and DGP's are presented in Table 2. For DGP 1, we observe that the RLS-ARFIMA(1, d , 1) estimate of σ_η is slightly upward biased, and that the model provides precise estimates of d and γ/T . The relative difference in the estimate of d obtained from the ARFIMA(1, d , 1) model, on the other hand, is quite suggestive, and while the evidence is provided in a stylized setup, we observe exactly the same pattern in our empirical analysis below. As documented by Perron & Qu (2010), if random level shifts are present in the series, the resulting estimate of d obtained from an ARFIMA(1, d , 1) model will be inflated to capture the large estimates of d obtained from a log-periodogram regression with few frequency ordinates. In order to capture the smaller estimates when more frequency ordinates are included, the fitted MA parameter is biased towards a large negative value to accentuate the short-run mean reversion. Similarly, we find the ARFIMA(0, d , 0) estimate of d to be upward biased, yet the bias is not as dramatic as for the ARFIMA(1, d , 1) model since the former lacks an MA parameter to help fit movements at higher frequencies. Last, for the RLS-ARFIMA(0, d , 0) model, we see that the inclusion of positive short-run dynamics in the DGP causes d to be overestimated. This holds true for all DGP's considered.

The results for DGP 2 verify that the ARFIMA(1, d , 1) model parameters are precisely estimated, as expected. What is particularly interesting for the present analysis, however, is that the ARFIMA parameters of the RLS-ARFIMA(1, d , 1) model are estimated with the same precision. As emphasized in Section 4.2, this may be explained by the fact that when $\sigma_\eta \rightarrow 0$ (which occurs in the table when the truncation order, M , increases), the estimation procedure collapses to the genuine long memory state space framework of Chan & Palma (1998), who show that the ARFIMA parameter estimates have maximum likelihood properties. Hence, the non-identification of the level shift parameters has no impact on the ARFIMA parameters. Interestingly, the results for DGP 2 show only modest efficiency losses when using the RLS-ARFIMA(1, d , 1) model and truncation $M = 20$ compared to the ARFIMA(1, d , 1) estimates using the CSS methodology, which includes the entire lag history, suggesting that a smaller-order truncation suffices to recover the parameters of the transitory part of the RLS-ARFIMA model. This is in line with the findings in Chan & Palma (1998) and Martin & Wilkins (1999).

For DGP 3, we observe that when the specification is tailored to the reduced form model, all the parameter estimates are unbiased and precise, while the corresponding estimates for the ARFIMA(1, d , 1) model display exactly the same bias as for DGP 1. The almost identical results for DGP 4 document that the RLS-ARFIMA model does not confuse random level shifts with non-stationary fractional integration. As such, the proposed model is able to distinguish between the proportion of persistence attributed to random level shifts and genuine long memory. The bias in the estimates of the various memory parameters are generally decreasing in M . However, we find only smaller gains in precision when going from $M = 20$ to the theoretically consistent choice $M = T^{1/2}$. Due to the size of these gains, and since there is a tradeoff with computational speed, especially for the longer series of daily returns, we select $M = 20$ for the empirical analysis. The choice of truncation is important, however, and, as a robustness check, we have experimented with selections $M = \{30, 40\}$ in both the simulation study

and in the empirical analysis below. The results are almost identical to those obtained for $M = 20$. Similarly, we have estimated the parameters for the volatility measures based on high-frequency data using the theoretically consistent choice $M = \lfloor T^{1/2} \rfloor$. Again, they are very similar to those reported below. Finally, note that we will increase the number of draws of the initial values to 10 in the empirical analysis to ensure that we do not report results from a local maximum.

6 Empirical Analysis of Asset Return Volatility

We proceed demonstrating the relevance of the proposed reduced form (log-)volatility modeling and forecasting framework by comparing the full-sample parameter estimates and out-of-sample forecasting performance of specific RLS-ARFIMA models to other widely applied models in the discrete time volatility literature. Initially, we consider parameter estimates from the RLS-ARFIMA(0, d , 0), RLS-ARFIMA(1, d , 1), RLS-ARMA(0, 0), RLS-ARMA(1, 1), ARFIMA(0, d , 0), and ARFIMA(1, d , 1) models for three reasons. First, it allows us to assess whether the most persistent component in the series is better described by random level shifts and/or genuine long memory and the impact of neglecting either one on the parameter estimates. Second, less parameterized models are often advocated for forecasting, see, e.g., Andersen et al. (2003). Third, as argued in the appendix, smaller order parameterizations suffice to capture both the short-run dynamics and measurement errors in the volatility proxies.

In the forecasting exercise, we also include the six models mentioned above. The ARFIMA class of models has recently received much attention in the volatility prediction literature. For example, it has been shown in, among others, Andersen et al. (2003), Koopman et al. (2005), Deo et al. (2006), Chiriac & Voev (2011), and Varneskov & Voev (2013) to outperform the popular class of GARCH models in terms of out-of-sample forecasting when applied to high-frequency measures of volatility. Similarly, Lu & Perron (2010) and Qu & Perron (2013) find that short memory-style random level shift models provide forecasts, which are, at least, on par with those obtained from (FI)GARCH and discrete time SV models when applied to volatility proxies constructed from daily data. Hence, to examine the usefulness of the proposed RLS-ARFIMA model in different settings, we compare its out-of-sample forecasting performance to these state-of-the-art competitors. In addition, we include the HAR model introduced by Corsi (2009), which has been shown to provide accurate forecasts for realized volatility measures, and a benchmark GARCH(1, 1) model in our out-of-sample analysis.

Finally, note that we will describe the results for the SPY and USD-JPY series in details throughout since they represent two different groupings of the series (SPY: BAC, MRK, S&P 500) and (USD-JPY: USD-AUD, USD-CHF), which share similar characteristics within each group. The T-bond series, on the other hand, is harder to classify as it sometimes shares characteristics with the SPY group and sometimes with the USD-JPY group. We will make the distinction clear when necessary.

6.1 Full-Sample Parameter Estimates

We report parameter estimates for the eight log-volatility series in Tables 3-4.¹⁵ In particular, note that the results for the SPY series are presented in Panel C of Table 3, and those for the USD-JPY series in Panel D of Table 4. We first discuss the results for the SPY series. The estimated persistence parameters of the RLS-ARFIMA(0, d , 0) model are $d = 0.4181$ and $\gamma/T = 0.0177$, which suggests a joint presence of genuine long memory and random level shifts. The estimated probability of level shifts indicates that they occur with an average duration of 56 days. Said duration is fairly low compared to the results in Lu & Perron (2010) for daily log-absolute returns on the S&P 500, AMEX, Dow Jones, and NASDAQ. We obtain similar estimates of the corresponding persistence parameters for the RLS-ARFIMA(1, d , 1) model, in addition to large and significant estimates of the two ARMA parameters. The latter, however, seem to characterize a common factor, they have fairly high standard errors, and their inclusion hardly increases the log-likelihood value. This clearly suggests that the most important sources of variation are captured by the joint modeling of genuine long memory and random level shifts. The estimation results for the ARFIMA(0, d , 0) model, similarly, indicate the presence of a stationary genuine long memory component, while the corresponding estimate $d = 0.5965$ for the ARFIMA(1, d , 1) model suggests that the series is a non-stationary fractionally integrated process. Furthermore, we observe that the estimated ARMA parameters of the ARFIMA(1, d , 1) model are large and distinct, however insignificant. As explained in Section 5, this particular difference between the RLS-ARFIMA and ARFIMA parameter estimates is exactly what we expect when a random level shift component is present; the estimate of d is biased upwards to capture movements at the lower frequencies, while the MA parameter is biased towards a large negative value to accentuate the short-run mean reversion. When accounting for random level shifts, such biases are no longer present, and the genuine long-memory component is seen to be stationary with the remaining short-run variation close to being serially uncorrelated. Finally, the estimated probabilities of random level shifts using the RLS-ARMA(0, 0) and RLS-ARMA(1, 1) models are $\gamma/T = 0.2082$ and $\gamma/T = 0.0797$, respectively, suggesting that level shifts, which are assumed to be rare events, occur with very low durations. This is clearly empirical evidence of spurious breaks. That is, when a genuine long memory component is present in the log-volatility series, the RLS-ARMA models are attempting to fit the additional persistence by overestimating the number of shifts.

Next, consider the parameter estimates for the RLS-ARFIMA(0, d , 0) model and the USD-JPY series. The persistence parameters are $d = 0.0532$ and $\gamma/T = 0.0027$, both statistically significant. The former, however, while deemed statistically significant indicates that the genuine long memory component is essentially irrelevant for characterizing persistent movements in the series. The estimated probability of random level shifts suggests that they are rare (26 in 9600 days) and occur with an average duration of 370 days. However, their magnitude $\sigma_\eta = 3.0657$, in comparison with the residual standard deviation $\sigma_\epsilon = 1.2765$, demonstrates that they are large contributors to the total variation in the series. The results for the RLS-ARFIMA(1, d , 1) model are similar; the impact of random level shifts is almost

¹⁵The associated standard errors are computed using the (inverse) numerical Hessian matrix.

identical, and the estimate of d is even smaller with a value of 0.0000.¹⁶ Moreover, we observe that the ARMA coefficient estimates are both high and of similar magnitude, which is consistent with the interpretation in Appendix A that the daily log-volatility measure exhibits a combination of AR(1) residual dynamics and measurement errors. Unlike the results for the SPY series, the ARMA parameter estimates for the USD-JPY series are seen to have small standard errors, and their inclusion increases the log-likelihood value, especially relative to the RLS-ARMA(0,0) case. For the latter, we observe an estimated probability of random level shifts that is twice as high, which is, again, suggestive of positively dependent residual dynamics, though not as strong as for the SPY series.

Given the evidence from the RLS-ARFIMA models that random level shifts describe the low-frequency movements in the USD-JPY series, it is interesting to consider the estimated integration orders from the ARFIMA(0, d , 0) and ARFIMA(1, d , 1) models, which are, in contrast, but as expected, much higher and significant. Again, we observe interesting differences between the two models. The estimate of d is much higher for the ARFIMA(1, d , 1) model since it has a large negative MA component to induce strong mean reversion. These features are similar to the ones obtained for the SPY series, along with those in the simulation study, and they support the findings of random level shifts in the series.

The results are similar within each of the two groups, and the T-bond series seems to be better characterized by those obtained for the SPY group.¹⁷ Thus, we may draw some conclusions from our analysis so far. The random level shift component is important for all series, being more frequent for all high-frequency measures of volatility, but with less variability for most. Once this is taken into account, the SPY group still contains a large genuine long memory component. The remaining dynamics for the USD-JPY group, on the other hand, may be described using a positively dependent short memory component in combination with measurement errors. The difference between the reduced form dynamics of the return volatility series constructed from daily and high-frequency data is puzzling, and we, thus, continue with a robustness check using high-frequency data for the USD-JPY exchange rate.

6.2 Robustness Check for the USD-JPY Series

As a gauge of whether the striking parameter differences in Tables 3-4 are either sampling frequency or asset class specific, neither, or both, we carry out a robustness check for the USD-JPY series using high-frequency data, which spans the period from January 2000 through April 10th 2009, corresponding, approximately, to the last quarter of the daily sample. Specifically, we have one-minute observations available for each trading day from both pit and electronic trading, and we estimate the daily quadratic variation using the flat-top realized kernel approach.¹⁸ The estimates are square-root and

¹⁶A lower bound of zero is imposed on d in the estimation.

¹⁷The d estimates from the RLS-ARFIMA models are slightly larger for the USD-AUD and USD-CHF series compared to those for the USD-JPY series. However, their small magnitudes still make them largely irrelevant for characterizing the low-frequency variation in the series.

¹⁸For comparability with the daily series, we add here the close-to-open squared return from the preceding trading day to the flat-top realized kernel estimate, thereby accounting for holiday, overnight, and weekend effects. As a robustness check, however, we also estimated an RLS-ARFIMA(1, d , 1) model without correcting for overnight returns; since the ARFIMA parameters are very close to those reported ($d = 0.19$ vs. $d = 0.17$ below), and the level shift parameters signify slightly

log-transformed, leaving a series with $T = 2458$ observations. We report the full-sample parameter estimates using the same set of dynamic models, as in the previous section, in Table 5.

Table 5 contain some interesting results. First, for the RLS-ARFIMA(0, d , 0) model, we find $\gamma/T = 0.0234$, which, similarly to the estimates for the remaining high-frequency volatility measures, implies that level shifts occur with a much shorter duration, on average, than what is suggested by the volatility proxies based on daily returns. Moreover, we find $d = 0.0000$, that is, no evidence of genuine long memory in the residual dynamics. Second, if we compare these estimates with the corresponding ones for the RLS-ARFIMA(1, d , 1) model, we find similar level shift parameters, an integration order $d = 0.1735$, as well as significant AR and MA components. However, the standard errors for d , ϕ , and θ are large, even deeming d insignificant, and the model generates almost no additional gains in log-likelihood value. This clearly suggests that level shifts are an important source of variation in the series, while genuine long memory, if present, is largely irrelevant for capturing its dynamics. Lastly, the ARFIMA(0, d , 0) and ARFIMA(1, d , 1) models find, not surprisingly, spurious evidence of genuine long memory.¹⁹

This robustness check, thus, verifies that more frequent breaks occur in the volatility measures based on high-frequency data, but it also suggests that the lack of a genuine long memory in the daily exchange rate series cannot be attributed to the sampling period nor the sampling frequency, but rather seems to be series specific. A detailed study of this result, however, is left for further research.

6.3 Forecasting Performance Evaluation

The class of RLS-ARFIMA models allows for a more flexible description of the low-frequency variation in log-volatility series. However, whether such flexibility improves out-of-sample forecast performance remains to be determined. Hence, we investigate the usefulness of the RLS-ARFIMA approach by comparing its forecasting performance to that from each of the competing dynamic models presented earlier along with the HAR model and a GARCH(1,1) benchmark, whose specifications and implementation procedures are briefly described in the supplementary appendix. This section proceeds by laying out the forecast evaluation framework before presenting the results from the out-of-sample exercise.

6.3.1 Forecast Evaluation Framework

We consider out-of-sample forecasting over the last $T_{out} = 900$ days for the eight series.²⁰ The various model parameters are estimated once, without the last 900 days in the sample, and the forecasts are computed conditional on these estimates.²¹ The out-of-sample period spans 3.6 years (assuming 250

more frequent shifts of slightly larger magnitude than those reported, these estimates are omitted for brevity.

¹⁹As an additional robustness check, we have estimated an RLS-ARFIMA(1, d , 1) model for similar high-frequency volatility series on the USD-CHF and USD-AUD using data from 2000-2009 and 2005-2014, respectively. Whereas the estimated impact of random level shifts is similar to that reported for the USD-JPY series, there is slightly weaker evidence of genuine long memory in the USD-CHF series and slightly stronger evidence of it in the USD-AUD series.

²⁰Not including the shorter high-frequency data-based USD-JPY series, which was used as a robustness check.

²¹This approach is chosen due to the heavy computational task of re-estimating parameters in each step for the group of (RLS-)ARFIMA models. As robustness checks, however, both recursive and rolling window estimation procedures have been used for some of the series; the numerical results are similar, and the model rankings are identical. This is explained

trading days per year) and, for each series, covers interesting and diverse market conditions such as the calm 2006 as well as the turbulent financial crisis of 2008, or the run-up to it. Hence, we will also perform robustness checks of the relative forecasting performance for the SPY and USD-JPY series using three non-overlapping sub-samples to examine how the models perform under different market conditions.

As we seek to evaluate the performance of direct τ -step-ahead forecasts for three different horizons, $\tau = \{1, 5, 10\}$, let the cumulative forecast be defined as $\bar{y}_{t+\tau, i|t} = \sum_{s=1}^{\tau} \hat{y}_{t+s, i|t}$ for model $i \in \mathcal{M}^0$ where \mathcal{M}^0 is the initial finite set of models and, similarly, let the cumulative log-volatility proxy be denoted by $\bar{\sigma}_{t, \tau} = \sum_{s=1}^{\tau} y_{t+s}$. Then, we apply the mean squared forecast error (MSFE) criterion for the out-of-sample evaluation, $\mathcal{MSFE}_{\tau, i} = \frac{1}{T_{out}} \sum_{t=1}^{T_{out}} (\bar{\sigma}_{t, \tau} - \bar{y}_{t+\tau, i|t})^2$, which has been shown by Hansen & Lunde (2006) and Patton (2011) to be robust against measurement errors in the (log-)volatility proxy. To facilitate model comparison, define the relative performance of models $i, j \in \mathcal{M}^0$ at time t as $d_{ij, t} = (\bar{\sigma}_{t, \tau} - \bar{y}_{t+\tau, i|t})^2 - (\bar{\sigma}_{t, \tau} - \bar{y}_{t+\tau, j|t})^2$, for which, we assume the sequence $(d_{ij, t}), \forall i, j \in \mathcal{M}^0, t = 1, \dots, T_{out}$ satisfies the following conditions: For some $r > 2$ and $\gamma > 0$, $\mathbb{E}[|d_{ij, t}|^{r+\gamma}] < \infty$, and $(d_{ij, t})$ is strictly stationary with variance $\mathbb{V}[d_{ij, t}] > 0$ and α -mixing of order $-r/(r-2)$.

Remark 4. *These conditions impose restrictions on the sequences of relative forecast performances, $(d_{ij, t})$, **not** directly on the loss function, which is allowed to exhibit structural breaks, genuine long memory, etc. They seem to be satisfied by plots of the loss differentials and the robustness of our results to the use of recursive and rolling estimation windows. Even in the event that the conditions for the validity of the MCS evaluation procedure are violated, the numerical MSFE's will provide a strong indication of the relative model performance.*

Under the stated conditions on the sequence of loss differentials, we may assess the relative forecast accuracy of the models using the 10% MCS of Hansen et al. (2011), see the supplementary appendix for a review. It is important for our application that the MCS is based on a bootstrap implementation, which is robust against comparisons of nested models when the parameters are estimated once using the same in-sample period for all models, see, e.g., the discussions in Giacomini & White (2006) and Hansen et al. (2011). The MSFE's and accompanying MCS p -values (in parentheses) are reported in Tables 6-8, where we use **boldface** notation to indicate whether a model belongs to the 10% MCS. The results for the robustness checks where the out-of-sample period is divided into three non-overlapping sub-samples for the SPY and USD-JPY series are reported in Tables 6 and 7.

Remark 5. *Whereas this forecasting framework may readily be applied to test the relative quality of log-volatility predictions, the conclusions from this exercise do not necessarily pertain to volatility forecasts in their standard deviation or variance form. We discuss this issue in Section 6.4.*

6.3.2 Out-of-Sample Results

First, to assist interpretation of the results, we illustrate how to read Tables 6-8 by considering the relative forecasting performance of the HAR model over the whole out-of-sample period of 900 days for

by the parameter estimates being fairly robust to the choice of estimation window.

the SPY series, which is reported in the bottom-right panel of Table 6. In this case, we observe that the HAR model belongs to the 10% MCS for one-step-ahead predictions, but not for five nor ten-step-ahead forecasts, which implies that the model is significantly worse than the best set of dynamic models at predicting log-volatility for horizons of five and ten days.

In general, we find that when considering the SPY series and the whole out-of-sample period it is only the RLS-ARFIMA(1, d , 1) model that belongs to the 10% MCS for all forecast horizons, thus ranking as the best overall model. The RLS-ARFIMA(0, d , 0) ranks as the second best in terms of numerical MSFE's. The RLS-ARMA(0, 0) and RLS-ARMA(1, 1) models also perform well for longer horizons, whereas the ARFIMA(0, d , 0), ARFIMA(1, d , 1) and HAR models do well for one-step-ahead predictions, but display MSFE's of, at least, a factor three larger for ten-step-ahead predictions. This clearly shows the value of applying the proposed forecast procedure, which leads to significant gains in terms out-of-sample precision with the largest gains attributed to the mean correction. When the forecast performance is decomposed into three non-overlapping sub-periods, the RLS-ARFIMA(1, d , 1) model performs well in all cases, and its relative superiority over the remaining RLS-AR(FI)MA specifications is driven, in part, by the last 300 days of the sample. Note, however, that the forecast errors for the RLS-AR(FI)MA models are also the largest in this sub-period, while the discrepancy to the remaining models is the smallest, suggesting that, not surprisingly, it is difficult to pin down the mean of the series during the period covering the financial turmoil of late 2007 through July 2008. When the mean-behavior of the series is slightly less erratic, as during the first 600 out-of-sample days, the RLS-ARFIMA models performs much better than models that do not allow for random level shifts in the mean.

We proceed to evaluate the out-of-sample performance of the eight dynamic models using the USD-JPY series in Table 7 and readily observe a similar model ranking; the RLS-ARFIMA(1, d , 1) model is significantly the best forecasting model for all horizons, followed by the RLS-ARFIMA(0, d , 0), RLS-ARMA(0, 0) and RLS-ARMA(1, 1) models, which comprise a clear second tier.²² If we consider the evidence from Table 4 that the ARFIMA class of models display severely upward biased estimates of the (fractional) integration order, it is not surprising that we find these - along with the HAR and GARCH(1, 1) - models to display *much* larger forecast errors, especially for longer horizons. In particular, this follows since they are not flexible enough to adequately describe the low-frequency variation in the volatility series and, thus, mistakenly summarizes the persistence as determined by a large genuine long memory component. Moreover, when decomposing the relative forecast performance into three non-overlapping samples, we see that all models, not surprisingly, deliver the largest forecast errors during the last 300 days, which cover most of the recent financial crises of 2008, and we observe that the RLS-ARFIMA(1, d , 1) model consistently exhibits the smallest MSFE's across sub-periods.

Finally, we may generalize the conclusions from the SPY and USD-JPY series by considering the out-of-sample results for the six remaining series in Table 8. Aggregating the results across the volatility series and forecast horizons, the RLS-ARFIMA class of models belong to the MCS in 21/24 cases, the

²²The difference between the RLS-ARMA(1, 1) and RLS-ARFIMA(1, d , 1) models may, given the parameter estimates in Table 4, seem surprising. However, when we remove the last 900 days to avoid using in-sample information for estimation of the parameters, we observe minor differences between the parameter estimates from the two models.

RLS-ARMA class in 16/24 cases, the HAR model in 8/24 cases, the ARFIMA class in 7/24 cases, and the GARCH model never belongs to the MCS. Furthermore, we observe large reductions in the MSFE's with models that explicitly capture random level shifts. The comparatively poor out-of-sample performance of dynamic models that do not explicitly model random level shifts, is, in itself, indirect evidence of their presence. As discussed previously, if level shifts are present, they bias the estimate of d upwards for the ARFIMA models (often in the non-stationary region) and the estimate of the MA parameter towards a large negative value. Similar biases affect the HAR and GARCH models, and they are responsible for the deterioration of the out-of-sample performance.

In general, we observe a good correspondence between in-sample fit and out-of-sample performance. The only exception arises if we contrast the parameter estimates for the RLS-AR(FI)MA models and the T-bond series, as shown in Table 4, with their respective out-of-sample results, where we see that the inclusion of ARMA parameters improves the in-sample fit, but leads to deteriorating out-of-sample performance as the forecast horizon increases. To elaborate on this observation, we depict the ten-step-ahead out-of-sample volatility for the T-bond series in Figure 1 together with the corresponding loss differentials from a bivariate comparison of the RLS-ARFIMA(0, d , 0) model against the RLS-ARFIMA(1, d , 1) model and a comparison of the RLS-ARMA(0, 0) model against the RLS-ARMA(1, 1) model. From the three series, we observe a distinct pattern; after an abrupt change around day 400, the log-volatility level is gradually increasing until, approximately, day 750. The less parameterized RLS-ARFIMA(0, d , 0) and RLS-ARMA(0, 0) models are better at capturing this increase, suggesting that the inclusion of ARMA parameters, in particular a strongly mean-reverting MA component, induces over-smoothing of the log-volatility series. This eventually leads to the deterioration in forecast performance as the mean-reverting log-volatility level deviates from its increasing out-of-sample counterpart for 350 observations. On the other hand, the inclusion of ARMA parameters seemingly improves the forecast performance of the models during the first part of the sample. This suggests that further out-of-sample gains may potentially be extracted by constructing forecast combinations of the dynamics models. However, a deeper investigation of this potential is beyond the scope of the paper.

In sum, there is overwhelming evidence in favor of using the RLS-ARFIMA class of models, which is not only able to distinguish between the contributions from random level shifts and genuine long memory to the low-frequency variation of the log-volatility series, but also delivers consistently good out-of-sample performance across a variety of forecast periods, forecast horizons, asset classes, and volatility proxies with varying degrees of measurement errors.

6.4 Log-volatility versus Volatility Forecasting

The favorable forecasting results for log-volatility series do not necessarily imply that our RLS-ARFIMA model is better at forecasting volatility in its standard deviation or variance form. There are a number of advantages to work with logs: (1) It reduces non-Gaussian features in the innovations, as seen by the results in Table 1, which not only alleviate concerns about finite sample parameter biases, but also improves the properties of forecast significance tests, e.g. Patton (2011); and (2) it allows the volatility

process to display both positive and negative level shifts without the need to impose non-negativity constraints. Additionally, it is important to note that the concave logarithmic transformation has likely reduced the dispersion of the level shifts, which would suggest that models taking the latter into account should perform equally well, if not even better, under more convex transformations.

Finally, we note that our log-volatility model may readily be used to forecast volatility in its standard deviation and variance forms since, conditional on the realization of states, the model is Gaussian and we may apply a log-normal-type correction. This is summarized by the following proposition:

Proposition 2. *Suppose the conditions of Proposition 1. Moreover, let $Y_{t+\tau} = \exp(y_{t+\tau})$, then the conditional expectation of $Y_{t+\tau}$ at time t is given by $\mathbb{E}_t[Y_{t+\tau}] = \exp(\hat{y}_{t+\tau|t} + \hat{\zeta}_{t+\tau|t}/2)$, where $\hat{y}_{t+\tau|t}$ is provided by Proposition 1, and with the variance correction, $\hat{\zeta}_{t+\tau|t}$, defined as*

$$\hat{\zeta}_{t+\tau|t} = \sum_{s=1}^{\tau} \sum_{i=0}^1 \sum_{j=0}^1 \Pr(\pi_{T,t+s} = j) \Pr(\pi_{T,t} = i | \mathbf{Y}_t; \boldsymbol{\Sigma}) \left(\mathbf{F} \mathbf{G}^{s-1} \left(\mathbf{G} \mathbf{P}_{t|t}^{ij} \mathbf{G}' + \mathbf{Q} \right) \left(\mathbf{F} \mathbf{G}^{s-1} \right)' + \sigma_{\eta j}^2 \right).$$

Proof. See the supplementary appendix, Varneskov & Perron (2017). □

Proposition 2, similarly to the forecasting procedure in Proposition 1, shows that the convexity correction when switching from log-volatility to volatility depends on the transition path between regimes. We illustrate the use of this procedure by forecasting the cumulate volatility of the SPY series in its standard deviation form for horizons of 1, 5 and 10 days (which is equivalent to the average volatility) using either of the RLS-ARFIMA(1, d , 1) and ARFIMA(1, d , 1) models in Figure 2. Whereas the figure shows that both models forecast the volatility well one-step-ahead, there is a clear difference between their respective forecasting performances at longer horizons; the RLS-ARFIMA model predicts the volatility well, and the ARFIMA model clearly gets the volatility level and path wrong. This may be explained by the biased MA parameter for the ARFIMA model (discussed above), inducing too strong mean reversion to the “wrong” volatility level, especially for the first part of the sample. When evaluating whether these differences in forecasting performance are significant using the MCS, the results are similar to those reported in Table 6, that is, there is no significant difference between their one-step-ahead forecasts, but the RLS-ARFIMA model is significantly better at multi-step-ahead predictions.

This illustrates that the potential forecasting gains from using the RLS-ARFIMA models are not confined to the logarithmic transformation. A detailed study of volatility forecasts in their standard deviation and variance forms, however, using Proposition 2 is left for further research.

7 Conclusion

We propose a reduced form framework for modeling the volatility of asset returns, which allows for the presence of random level shifts, genuine long memory and measurement errors. In particular, we advocate a parametric state space model where the underlying dynamics is decomposed into a simple level shift component and ARFIMA dynamics. This allows both long and short memory parameters to

be estimated together with the probability and magnitude of random level shifts. Measurement errors are accounted for by careful modeling and interpretation of the ARMA parameters. We provide an estimation procedure and a forecasting framework to construct mean and path-corrected forecasts.

We perform an empirical analysis using eight daily return volatility series, which differ, not only according to the sampling frequency of the data with which they are constructed, but also with respect to time span and asset class. In particular, we demonstrate the usefulness of the proposed modeling framework by comparing the full sample parameter estimates and out-of-sample forecasting performance of specific RLS-ARFIMA models relative to that from other popular models in the literature.

The full sample parameter estimates reveal that random level shifts are important components of all series and that a genuine long memory component is present in most volatility series constructed using high-frequency data. The remaining dynamics in volatility proxies constructed as log-daily absolute returns, on the other hand, may be described as a combination of short memory dynamics and measurement errors. Finally, we show that the RLS-ARFIMA model display consistently good out-of-sample performance across forecast periods, forecast horizons, asset classes, and volatility measures, by being the most frequent model in the 10% MCS of Hansen et al. (2011). The forecast gains can be very pronounced at longer horizons. This shows that there is substantial statistical value in distinguishing between random level shifts and genuine long memory for forecasting.

Summary Statistics of the Volatility Proxies											
	Descriptive Statistics					Fractional Integration					
	Max	Min	Skew	Ekur	#obs	d_{LP}	d_{LW}	$W_{0.02}$	$W_{0.05}$	$S_d(1/3)$	$S_d(1/2)$
BAC HF	2.25	-1.22	0.20	-0.39	2913	0.539 (0.100)	0.566 (0.069)	1.16*	1.16**	1.40	-0.90
MRK HF	2.17	-1.16	0.51	1.25	2913	0.585 (0.100)	0.590 (0.069)	0.64	0.64	1.11	1.56
SPY HF	1.64	-1.67	0.25	0.02	2914	0.548 (0.100)	0.547 (0.069)	0.76	0.76	0.84	-0.31
S&P 500 HF	3.26	-2.89	0.42	2.21	6691	0.571 (0.078)	0.604 (0.056)	0.99	0.75	1.78*	1.13
T-Bonds HF	0.93	-3.56	0.27	0.42	6640	0.763 (0.078)	0.731 (0.056)	0.85	0.49	4.13***	4.60***
USD-AUD	2.96	-6.91	-1.35	1.44	9612	0.857 (0.070)	0.828 (0.051)	3.54***	3.33***	5.30***	7.76***
USD-CHF	1.76	-6.91	-1.52	3.73	9606	0.669 (0.070)	0.623 (0.051)	1.02	1.02*	2.99***	5.32***
USD-JPY	2.25	-6.91	-1.50	2.90	9600	0.622 (0.070)	0.622 (0.051)	1.53***	1.53***	3.87***	3.23***

Table 1: The first half of this table provides some unconditional summary statistics for the eight volatility series. For the three exchange rates, the daily volatility is proxied by $\ln(|r_t| + 0.001)$. For the remaining series, the daily quadratic variation is estimated using the flat-top realized kernel approach of Varneskov (2016a, 2016b), see the supplementary appendix for details, and subsequently square-root and log-transformed. Here, skewness and *excess* kurtosis (compared to 3) are denoted “Skew”, and “Ekur”, respectively. The number of observations after deleting missing entries are denoted “#obs”. The second half of the table presents summary statistics describing the conditional properties of the series. In particular, d_{LP} and d_{LW} denote log-periodogram and local Whittle estimates, respectively, using a bandwidth $\lfloor T^{1/2} \rfloor$. Furthermore, $S_d(a, 4/5) \equiv S_d(a)$ and W_ϵ denote different implementations of the testing procedures proposed by Perron & Qu (2010) and Qu (2011), respectively, of the null hypothesis that the series are genuine long memory series against an alternative data generating process with level shifts and short memory dynamics. Qu (2011) also allows for genuine long memory under the alternative. In particular, the tests are implemented with $a = \{1/3, 1/2\}$ and $\epsilon = \{0.02, 0.05\}$, respectively. See the supplementary appendix for details. Finally, (*), (**), and (***) denote rejection at a 10%, 5%, and 1% significance level, respectively.

Simulation Results												
DGP 1	Bias						RMSE					
	d	ϕ	θ	σ_ϵ	γ/T	σ_η	d	ϕ	θ	σ_ϵ	γ/T	σ_η
RLS-ARFIMA(1, d , 1, 5)	0.01	-0.06	-0.04	0.00	-0.00	0.18	0.03	0.10	0.08	0.01	0.01	0.27
RLS-ARFIMA(1, d , 1, 10)	0.01	-0.05	-0.04	0.00	-0.00	0.18	0.03	0.09	0.08	0.01	0.00	0.27
RLS-ARFIMA(1, d , 1, 20)	0.01	-0.06	-0.04	0.00	-0.00	0.18	0.03	0.10	0.08	0.01	0.01	0.27
RLS-ARFIMA(1, d , 1, $T^{1/2}$)	0.01	-0.05	-0.04	0.00	-0.00	0.18	0.03	0.10	0.08	0.01	0.01	0.27
RLS-ARFIMA(0, d , 0, $T^{1/2}$)	0.29	-	-	0.02	-0.01	0.26	0.29	-	-	0.02	0.01	0.33
ARFIMA(1, d , 1)	1.01	0.11	0.86	0.10	-	-	1.01	0.11	0.86	0.10	-	-
ARFIMA(0, d , 0)	0.65	-	-	0.11	-	-	0.66	-	-	0.11	-	-
DGP 2	d	ϕ	θ	σ_ϵ	γ/T	σ_η	d	ϕ	θ	σ_ϵ	γ/T	σ_η
RLS-ARFIMA(1, d , 1, 5)	-0.03	-0.00	-0.02	-0.00	0.04	0.18	0.07	0.11	0.08	0.01	0.05	0.21
RLS-ARFIMA(1, d , 1, 10)	-0.03	0.02	-0.01	-0.00	0.04	0.12	0.06	0.11	0.07	0.01	0.04	0.16
RLS-ARFIMA(1, d , 1, 20)	-0.02	0.01	-0.01	-0.00	0.04	0.06	0.05	0.10	0.07	0.01	0.04	0.08
RLS-ARFIMA(1, d , 1, $T^{1/2}$)	-0.01	0.00	-0.01	-0.00	0.03	0.02	0.04	0.09	0.07	0.01	0.04	0.04
RLS-ARFIMA(0, d , 0, $T^{1/2}$)	0.15	-	-	0.01	0.04	0.00	0.15	-	-	0.04	0.04	0.03
ARFIMA(1, d , 1)	-0.00	-0.00	-0.01	0.00	-	-	0.03	0.08	0.07	0.01	-	-
ARFIMA(0, d , 0)	0.21	-	-	0.01	-	-	0.21	-	-	0.01	-	-
DGP 3	d	ϕ	θ	σ_ϵ	γ/T	σ_η	d	ϕ	θ	σ_ϵ	γ/T	σ_η
RLS-ARFIMA(1, d , 1, 5)	-0.11	0.07	-0.01	-0.01	0.00	-0.02	0.13	0.14	0.07	0.01	0.01	0.22
RLS-ARFIMA(1, d , 1, 10)	-0.09	0.07	0.00	-0.00	0.00	0.03	0.12	0.14	0.07	0.01	0.00	0.23
RLS-ARFIMA(1, d , 1, 20)	-0.05	0.04	-0.00	-0.00	-0.00	0.05	0.10	0.14	0.08	0.01	0.00	0.25
RLS-ARFIMA(1, d , 1, $T^{1/2}$)	-0.03	0.01	-0.02	-0.00	-0.00	0.06	0.08	0.12	0.08	0.00	0.00	0.25
RLS-ARFIMA(0, d , 0, $T^{1/2}$)	0.15	-	-	0.01	0.00	0.01	0.15	-	-	0.01	0.01	0.26
ARFIMA(1, d , 1)	0.66	0.37	0.93	0.06	-	-	0.66	0.37	0.93	0.07	-	-
ARFIMA(0, d , 0)	0.39	-	-	0.06	-	-	0.39	-	-	0.07	-	-
DGP 4	d	ϕ	θ	σ_ϵ	γ/T	σ_η	d	ϕ	θ	σ_ϵ	γ/T	σ_η
RLS-ARFIMA(1, d , 1, 5)	-0.06	0.00	-0.04	-0.01	0.01	-0.15	0.11	0.15	0.10	0.01	0.01	0.30
RLS-ARFIMA(1, d , 1, 10)	-0.07	0.05	-0.01	-0.01	0.01	-0.07	0.12	0.15	0.07	0.01	0.01	0.28
RLS-ARFIMA(1, d , 1, 20)	-0.05	0.03	-0.01	-0.00	0.00	-0.01	0.10	0.13	0.07	0.01	0.01	0.27
RLS-ARFIMA(1, d , 1, $T^{1/2}$)	-0.02	0.00	-0.01	-0.00	0.00	0.03	0.08	0.12	0.08	0.01	0.01	0.27
RLS-ARFIMA(0, d , 0, $T^{1/2}$)	0.22	-	-	0.01	-0.00	0.11	0.22	-	-	0.01	0.01	0.31
ARFIMA(1, d , 1)	0.62	0.17	0.52	0.05	-	-	0.65	0.38	0.75	0.05	-	-
ARFIMA(0, d , 0)	0.26	-	-	0.05	-	-	0.26	-	-	0.06	-	-

Table 2: Simulation results using the following configurations: RLS-ARFIMA(1, d , 1) with $\phi = 0.2$, $\theta = -0.1$, $\sigma_\epsilon = 0.5$ and **(DGP 1)** $d = 0$, $\gamma/T = 0.02$, $\sigma_\eta = 3\sigma_\epsilon$, **(DGP 2)** $d = 0.35$, $\gamma/T = 0$, $\sigma_\eta = 0$, **(DGP 3)** $d = 0.35$, $\gamma/T = 0.02$, $\sigma_\eta = 3\sigma_\epsilon$, **(DGP 4)** $d = 0.6$, $\gamma/T = 0.02$, $\sigma_\eta = 3\sigma_\epsilon$. The bias and root mean squared error (RMSE) are computed for different values of M (the last entry for RLS-ARFIMA), $T = 3000$ and $N = 100$ replications.

Full Sample Parameter Estimates								
BAC HF	a	ϕ	θ	d	γ/T	σ_η	σ_ϵ	KMLE
ARFIMA(0, d , 0)	0.3111 (0.5181)	-	-	0.5287 (0.0400)	-	-	-	-
ARFIMA(1, d , 1)	0.1827 (0.7091)	0.2941 (0.4300)	0.4537 (0.5120)	0.6445 (0.1507)	-	-	-	-
RLS-ARFIMA(0, d , 0)	-	-	-	0.4509 (0.0223)	0.0162 (0.0091)	0.3024 (0.0877)	0.2239 (0.0037)	137.029
RLS-ARFIMA(1, d , 1)	-	-0.2979 (0.5959)	-0.2699 (0.6127)	0.4738 (0.0170)	0.0139 (0.0082)	0.3115 (0.0955)	0.2246 (0.0037)	137.577
RLS-ARMA(0, 0)	-	-	-	-	0.4225 (0.0861)	0.1612 (0.0176)	0.1758 (0.0039)	83.9218
RLS-ARMA(1, 1)	-	0.6373 (0.0736)	0.2844 (0.0609)	-	0.0864 (0.0354)	0.2073 (0.0406)	0.2112 (0.0047)	129.527
MRK HF	a	ϕ	θ	d	γ/T	σ_η	σ_ϵ	KMLE
ARFIMA(0, d , 0)	0.3357 (0.2431)	-	-	0.4041 (0.0378)	-	-	-	-
ARFIMA(1, d , 1)	0.3198 (0.3441)	0.5300 (0.2270)	0.6439 (0.2348)	0.5075 (0.1433)	-	-	-	-
RLS-ARFIMA(0, d , 0)	-	-	-	0.3481 (0.0207)	0.0134 (0.0040)	0.8991 (0.1492)	0.2405 (0.0041)	-186.491
RLS-ARFIMA(1, d , 1)	-	0.6184 (0.1424)	0.6991 (0.1166)	0.4361 (0.0446)	0.0135 (0.0042)	0.8945 (0.1417)	0.2405 (0.0042)	-185.957
RLS-ARMA(0, 0)	-	-	-	-	0.1034 (0.0146)	0.4015 (0.0364)	0.2048 (0.0042)	-248.768
RLS-ARMA(1, 1)	-	0.6217 (0.0844)	0.3074 (0.0835)	-	0.0216 (0.0055)	0.7432 (0.1063)	0.2340 (0.0045)	-193.244
SPY HF	a	ϕ	θ	d	γ/T	σ_η	σ_ϵ	KMLE
ARFIMA(0, d , 0)	-0.2260 (0.4263)	-	-	0.4944 (0.0403)	-	-	-	-
ARFIMA(1, d , 1)	-0.3184 (0.5840)	0.3921 (0.3369)	0.5206 (0.3891)	0.5965 (0.1455)	-	-	-	-
RLS-ARFIMA(0, d , 0)	-	-	-	0.4181 (0.0241)	0.0177 (0.0087)	0.3864 (0.1027)	0.2272 (0.0039)	56.7714
RLS-ARFIMA(1, d , 1)	-	0.5844 (0.1602)	0.6525 (0.1540)	0.4903 (0.0296)	0.0172 (0.0090)	0.3937 (0.1051)	0.2274 (0.0040)	57.1266
RLS-ARMA(0, 0)	-	-	-	-	0.2082 (0.0279)	0.2487 (0.0206)	0.1793 (0.0040)	19.7437
RLS-ARMA(1, 1)	-	0.5244 (0.0967)	0.2100 (0.0799)	-	0.0797 (0.0254)	0.2454 (0.0395)	0.2128 (0.0053)	51.2194
S&P 500 HF	a	ϕ	θ	d	γ/T	σ_η	σ_ϵ	KMLE
ARFIMA(0, d , 0)	-0.2297 (0.2282)	-	-	0.4090 (0.0189)	-	-	-	-
ARFIMA(1, d , 1)	-0.0898 (0.3758)	0.3098 (0.1305)	0.4978 (0.1551)	0.5442 (0.0640)	-	-	-	-
RLS-ARFIMA(0, d , 0)	-	-	-	0.2604 (0.0198)	0.0285 (0.0046)	0.8436 (0.0726)	0.2746 (0.0039)	-1620.95
RLS-ARFIMA(1, d , 1)	-	-0.4904 (0.5315)	-0.4700 (0.5467)	0.2800 (0.0150)	0.0273 (0.0046)	0.8586 (0.0790)	0.2755 (0.0040)	-1620.05
RLS-ARMA(0, 0)	-	-	-	-	0.0773 (0.0063)	0.5658 (0.0320)	0.2448 (0.0031)	-1675.72
RLS-ARMA(1, 1)	-	0.6837 (0.0619)	0.4612 (0.0594)	-	0.0323 (0.0046)	0.8034 (0.0639)	0.2703 (0.0039)	-1624.31

Table 3: Parameter estimates of the various dynamic models with standard errors in parentheses for the high-frequency log-volatility proxies on BAC, MRK, SPY and S&P 500. “KMLE” denotes the predictive log-likelihood value from the Kalman filter. Here, a refers to the constant in an ARFIMA model. The standard errors are computed using the (inverse) numerical Hessian matrix.

Full Sample Parameter Estimates (continued)								
T-Bonds HF	a	ϕ	θ	d	γ/T	σ_η	σ_ϵ	KMLE
ARFIMA(0, d , 0)	-1.1777 (0.0815)	-	-	0.2721 (0.0142)	-	-	-	-
ARFIMA(1, d , 1)	-1.1050 (0.2589)	0.2566 (0.0487)	0.6455 (0.0659)	0.5513 (0.0648)	-	-	-	-
RLS-ARFIMA(0, d , 0)	-	-	-	0.0965 (0.0169)	0.0161 (0.0057)	0.2759 (0.0463)	0.3611 (0.0036)	-2905.11
RLS-ARFIMA(1, d , 1)	-	0.3311 (0.0416)	0.5863 (0.0528)	0.3874 (0.0294)	0.0046 (0.0028)	0.3260 (0.0776)	0.3690 (0.0036)	-2897.65
RLS-ARMA(0, 0)	-	-	-	-	0.0332 (0.0089)	0.2479 (0.0337)	0.3527 (0.0034)	-2919.84
RLS-ARMA(1, 1)	-	0.9429 (0.0203)	0.8409 (0.0214)	-	0.0053 (0.0034)	0.3132 (0.0721)	0.3689 (0.0038)	-2904.26
USD-AUD	a	ϕ	θ	d	γ/T	σ_η	σ_ϵ	KMLE
ARFIMA(0, d , 0)	-2.2685 (0.0610)	-	-	0.2469 (0.0026)	-	-	-	-
ARFIMA(1, d , 1)	-3.3750 (0.1954)	0.2924 (0.0100)	0.7511 (0.0122)	0.5520 (0.0151)	-	-	-	-
RLS-ARFIMA(0, d , 0)	-	-	-	0.0414 (0.0111)	0.0025 (0.0006)	1.0443 (0.0420)	1.3842 (0.0104)	-16932.5
RLS-ARFIMA(1, d , 1)	-	0.4494 (0.1136)	0.5043 (0.1368)	0.0983 (0.0290)	0.0021 (0.0007)	1.0408 (0.0332)	1.3878 (0.0110)	-16931.7
RLS-ARMA(0, 0)	-	-	-	-	0.0041 (0.0008)	0.9795 (0.0090)	1.3758 (0.0101)	-16939.3
RLS-ARMA(1, 1)	-	0.8135 (0.1103)	0.7806 (0.1121)	-	0.0024 (0.0007)	1.0305 (0.0158)	1.3852 (0.0108)	-16931.5
USD-CHF	a	ϕ	θ	d	γ/T	σ_η	σ_ϵ	KMLE
ARFIMA(0, d , 0)	-1.3256 (0.0298)	-	-	0.1624 (0.0035)	-	-	-	-
ARFIMA(1, d , 1)	-2.2695 (0.1797)	0.3296 (0.0126)	0.7099 (0.0181)	0.4507 (0.0203)	-	-	-	-
RLS-ARFIMA(0, d , 0)	-	-	-	0.0448 (0.0110)	0.0032 (0.0017)	0.6209 (0.0927)	1.2481 (0.0094)	-15872.7
RLS-ARFIMA(1, d , 1)	-	0.3611 (0.0919)	0.4799 (0.1135)	0.1615 (0.0210)	0.0012 (0.0007)	0.6960 (0.0850)	1.2542 (0.0092)	-15866.9
RLS-ARMA(0, 0)	-	-	-	-	0.0078 (0.0031)	0.5492 (0.0780)	1.2400 (0.0093)	-15880.1
RLS-ARMA(1, 1)	-	0.9444 (0.0214)	0.9059 (0.0243)	-	0.0010 (0.0007)	0.6996 (0.1004)	1.2547 (0.0093)	-15866.2
USD-JPY	a	ϕ	θ	d	γ/T	σ_η	σ_ϵ	KMLE
ARFIMA(0, d , 0)	-1.7860 (0.0522)	-	-	0.2280 (0.0032)	-	-	-	-
ARFIMA(1, d , 1)	-2.7731 (0.1810)	0.2785 (0.0145)	0.6034 (0.0190)	0.4490 (0.0134)	-	-	-	-
RLS-ARFIMA(0, d , 0)	-	-	-	0.0532 (0.0117)	0.0027 (0.0006)	3.0657 (0.5819)	1.2765 (0.0096)	-16297.2
RLS-ARFIMA(1, d , 1)	-	0.7776 (0.0783)	0.7281 (0.0793)	0.0000 (0.0002)	0.0024 (0.0007)	3.1717 (0.6447)	1.2786 (0.0101)	-16295.4
RLS-ARMA(0, 0)	-	-	-	-	0.0049 (0.0010)	2.3171 (0.3802)	1.2641 (0.0098)	-16306.4
RLS-ARMA(1, 1)	-	0.7778 (0.0635)	0.7282 (0.0646)	-	0.0024 (0.0007)	3.1715 (0.6472)	1.2786 (0.0099)	-16295.4

Table 4: Parameter estimates of the various dynamic models with standard errors in parentheses for the log-volatility proxies on the T-bonds, USD-AUD, USD-CHF and USD-JPY. “KMLE” denotes the predictive log-likelihood value from the Kalman filter. Here, a refers to the constant in an ARFIMA model. The standard errors are computed using the (inverse) numerical Hessian matrix.

Robustness Check for the USD-JPY								
USD-JPY HF	a	ϕ	θ	d	γ/T	σ_η	σ_ϵ	KMLE
ARFIMA(0, d , 0)	-0.6287 (0.0413)	-	-	0.2019 (0.0137)	-	-	-	-
ARFIMA(1, d , 1)	-0.5825 (0.1572)	0.2385 (0.0545)	0.7610 (0.0646)	0.6108 (0.1069)	-	-	-	-
RLS-ARFIMA(0, d , 0)	-	-	-	0.0000 (0.0001)	0.0234 (0.0042)	1.9009 (0.2084)	0.3964 (0.0081)	-1592.22
RLS-ARFIMA(1, d , 1)	-	0.5121 (0.2165)	0.6848 (0.1163)	0.1735 (0.0975)	0.0236 (0.0066)	1.8783 (0.2541)	0.3972 (0.0134)	-1590.97
RLS-ARMA(0, 0)	-	-	-	-	0.0234 (0.0042)	1.9008 (0.2071)	0.3964 (0.0081)	-1592.22
RLS-ARMA(1, 1)	-	0.7737 (0.1683)	0.8065 (0.1761)	-	0.0264 (0.0055)	1.8115 (0.2064)	0.3903 (0.0108)	-1591.68

Table 5: Parameter estimates of the various dynamic models with standard errors in parentheses for the high-frequency log-volatility series on the USD-JPY. “KMLE” denotes the predictive log-likelihood value from the Kalman filter. Here, a refers to the constant in an ARFIMA model. The standard errors are computed using the (inverse) numerical Hessian matrix.

Forecast Evaluations for SPY						
	$t_{out} \in [1, 300]$			$t_{out} \in [301, 600]$		
	1-step	5-step	10-step	1-step	5-step	10-step
RLS-ARFIMA(0, d , 0)	0.0306 (0.96)	0.4693 (0.13)	1.7937 (0.02)	0.0553 (0.57)	1.0168 (1.00)	3.7569 (1.00)
RLS-ARFIMA(1, d , 1)	0.0306 (1.00)	0.4607 (1.00)	1.7389 (1.00)	0.0557 (0.16)	1.0326 (0.16)	3.8333 (0.11)
RLS-ARMA(0, 0)	0.0321 (0.32)	0.4878 (0.72)	1.9532 (0.34)	0.0588 (0.08)	1.2068 (0.02)	4.9780 (0.00)
RLS-ARMA(1, 1)	0.0309 (0.32)	0.4732 (0.72)	1.8978 (0.34)	0.0559 (0.32)	1.0649 (0.21)	4.1045 (0.11)
ARFIMA(0, d , 0)	0.0307 (0.96)	2.9345 (0.00)	15.697 (0.00)	0.0549 (1.00)	4.1251 (0.00)	21.030 (0.00)
ARFIMA(1, d , 1)	0.0307 (0.96)	2.6416 (0.00)	13.778 (0.00)	0.0554 (0.57)	3.8275 (0.00)	18.959 (0.00)
HAR	0.0317 (0.18)	3.1097 (0.00)	34.656 (0.00)	0.0575 (0.08)	4.3144 (0.00)	43.279 (0.00)
log-GARCH	0.0572 (0.00)	2.5295 (0.00)	18.409 (0.00)	0.0847 (0.00)	3.2677 (0.00)	21.836 (0.00)
	$t_{out} \in [601, 900]$			$t_{out} \in [1, 900]$		
	1-step	5-step	10-step	1-step	5-step	10-step
RLS-ARFIMA(0, d , 0)	0.0618 (0.63)	1.6347 (0.00)	7.3594 (0.00)	0.0493 (0.50)	1.0370 (0.14)	4.2690 (0.01)
RLS-ARFIMA(1, d , 1)	0.0618 (0.51)	1.5992 (1.00)	7.1319 (1.00)	0.0494 (0.50)	1.0277 (1.00)	4.2021 (1.00)
RLS-ARMA(0, 0)	0.0657 (0.08)	1.5982 (1.00)	7.6520 (0.61)	0.0522 (0.00)	1.0948 (0.29)	4.8297 (0.04)
RLS-ARMA(1, 1)	0.0628 (0.06)	1.6251 (0.99)	7.6460 (0.43)	0.0499 (0.02)	1.0512 (0.44)	4.5146 (0.04)
ARFIMA(0, d , 0)	0.0606 (1.00)	2.1894 (0.00)	9.3499 (0.08)	0.0488 (1.00)	3.0880 (0.00)	15.426 (0.00)
ARFIMA(1, d , 1)	0.0613 (0.71)	2.3486 (0.00)	10.127 (0.00)	0.0492 (0.50)	2.9425 (0.00)	14.335 (0.00)
HAR	0.0608 (0.91)	1.6102 (1.00)	8.9436 (0.43)	0.0500 (0.19)	3.0192 (0.00)	29.184 (0.00)
log-GARCH	0.0636 (0.71)	1.6480 (0.99)	8.1086 (0.61)	0.0685 (0.00)	2.4864 (0.00)	16.208 (0.00)

Table 6: Forecast evaluations of the eight dynamic models. We use mean squared forecast errors (MSFE's) and consider MCS comparisons with all models included in the initial set. Here, **boldface** notation indicate whether a model belongs to the 10% MCS. The MCS p -values are in parentheses. See the main text for details.

Forecast Evaluations for USD-JPY						
	$t_{out} \in [1, 300]$			$t_{out} \in [301, 600]$		
	1-step	5-step	10-step	1-step	5-step	10-step
RLS-ARFIMA(0, d , 0)	1.2418 (0.26)	5.9518 (0.02)	13.781 (0.01)	1.7523 (0.06)	12.100 (0.07)	30.393 (0.01)
RLS-ARFIMA(1, d , 1)	1.2273 (1.00)	5.1191 (1.00)	10.526 (1.00)	1.6613 (1.00)	10.530 (1.00)	22.730 (1.00)
RLS-ARMA(0, 0)	1.2329 (0.67)	6.1010 (0.01)	14.548 (0.00)	1.7596 (0.05)	12.538 (0.05)	32.031 (0.01)
RLS-ARMA(1, 1)	1.2401 (0.32)	5.9229 (0.04)	13.666 (0.02)	1.7472 (0.06)	12.022 (0.07)	30.068 (0.01)
ARFIMA(0, d , 0)	1.3017 (0.00)	9.2743 (0.00)	30.938 (0.00)	1.7086 (0.20)	12.935 (0.07)	36.010 (0.00)
ARFIMA(1, d , 1)	1.2517 (0.01)	42.195 (0.00)	199.33 (0.00)	1.6895 (0.14)	43.850 (0.00)	194.57 (0.00)
HAR	1.2624 (0.01)	9.4700 (0.00)	53.660 (0.00)	1.6905 (0.20)	15.801 (0.00)	66.003 (0.00)
log-GARCH	1.9869 (0.07)	30.249 (0.00)	126.68 (0.00)	2.5164 (0.00)	38.101 (0.00)	148.62 (0.00)
	$t_{out} \in [601, 900]$			$t_{out} \in [1, 900]$		
	1-step	5-step	10-step	1-step	5-step	10-step
RLS-ARFIMA(0, d , 0)	1.3759 (0.56)	11.051 (0.45)	31.781 (0.38)	1.4566 (0.03)	9.6937 (0.01)	25.245 (0.00)
RLS-ARFIMA(1, d , 1)	1.3155 (1.00)	10.272 (1.00)	28.001 (1.00)	1.4013 (1.00)	8.6313 (1.00)	20.334 (1.00)
RLS-ARMA(0, 0)	1.3808 (0.44)	11.202 (0.45)	32.453 (0.38)	1.4578 (0.03)	9.9401 (0.01)	26.275 (0.00)
RLS-ARMA(1, 1)	1.3770 (0.56)	11.102 (0.12)	32.006 (0.06)	1.4548 (0.03)	9.6745 (0.01)	25.171 (0.00)
ARFIMA(0, d , 0)	1.3681 (0.35)	30.555 (0.00)	128.80 (0.00)	1.4594 (0.00)	17.516 (0.00)	64.536 (0.00)
ARFIMA(1, d , 1)	1.3224 (0.88)	82.939 (0.00)	398.10 (0.00)	1.4212 (0.03)	56.179 (0.00)	262.50 (0.00)
HAR	1.3247 (0.88)	16.070 (0.00)	93.273 (0.00)	1.4258 (0.03)	13.767 (0.00)	70.728 (0.00)
log-GARCH	1.4978 (0.07)	15.351 (0.00)	50.342 (0.00)	2.0003 (0.00)	27.970 (0.00)	109.200 (0.00)

Table 7: Forecast evaluations of the eight dynamic models. We use mean squared forecast errors (MSFE's) and consider MCS comparisons with all models included in the initial set. Here, **boldface** notation indicate whether a model belongs to the 10% MCS. The MCS p -values are in parentheses. See the main text for details.

Forecast Evaluations for the Remaining Series, $t_{out} \in [1, 900]$						
	BAC HF			MRK HF		
	1-step	5-step	10-step	1-step	5-step	10-step
RLS-ARFIMA(0, d , 0)	0.0491 (1.00)	1.2855 (0.01)	5.5562 (0.12)	0.0776 (0.01)	1.1819 (0.05)	4.3179 (0.06)
RLS-ARFIMA(1, d , 1)	0.0491 (0.66)	1.3982 (0.00)	6.0759 (0.00)	0.0775 (0.01)	1.1671 (1.00)	4.2600 (1.00)
RLS-ARMA(0, 0)	0.0509 (0.27)	1.1056 (1.00)	5.1127 (0.84)	0.0795 (0.00)	1.2315 (0.38)	4.9083 (0.02)
RLS-ARMA(1, 1)	0.0495 (0.51)	1.1276 (0.59)	5.0695 (1.00)	0.0779 (0.00)	1.2127 (0.05)	4.5891 (0.00)
ARFIMA(0, d , 0)	0.0496 (0.51)	4.8718 (0.00)	25.428 (0.00)	0.0732 (1.00)	1.4327 (0.00)	5.5504 (0.00)
ARFIMA(1, d , 1)	0.0493 (0.51)	4.5018 (0.00)	22.706 (0.00)	0.0733 (0.83)	1.3486 (0.00)	5.0752 (0.06)
HAR	0.0498 (0.39)	2.4582 (0.00)	28.671 (0.00)	0.0743 (0.07)	1.2103 (0.58)	5.0140 (0.06)
log-GARCH	0.0697 (0.00)	2.6312 (0.00)	18.741 (0.00)	0.0876 (0.00)	2.0187 (0.00)	10.621 (0.00)
	S&P 500 HF			T-Bonds HF		
	1-step	5-step	10-step	1-step	5-step	10-step
RLS-ARFIMA(0, d , 0)	0.1491 (0.01)	2.4664 (0.71)	9.6852 (1.00)	0.1133 (0.37)	0.7731 (0.84)	2.3151 (1.00)
RLS-ARFIMA(1, d , 1)	0.1495 (0.01)	2.5329 (0.01)	9.9817 (0.01)	0.1155 (0.11)	1.4219 (0.00)	5.4489 (0.00)
RLS-ARMA(0, 0)	0.1538 (0.00)	2.5262 (0.68)	10.089 (0.37)	0.1127 (1.00)	0.7703 (1.00)	2.3161 (0.98)
RLS-ARMA(1, 1)	0.1502 (0.01)	2.4788 (0.71)	9.7703 (0.38)	0.1157 (0.08)	1.6141 (0.00)	6.4042 (0.00)
ARFIMA(0, d , 0)	0.1407 (0.29)	6.6100 (0.00)	31.096 (0.00)	0.1244 (0.00)	3.6058 (0.00)	16.195 (0.00)
ARFIMA(1, d , 1)	0.1391 (1.00)	6.7864 (0.00)	31.943 (0.00)	0.1147 (0.21)	3.7387 (0.00)	17.308 (0.00)
HAR	0.1394 (0.69)	2.3750 (1.00)	11.648 (0.10)	0.1151 (0.22)	1.0908 (0.00)	3.5153 (0.00)
log-GARCH	0.1833 (0.00)	5.0218 (0.00)	30.807 (0.00)	0.1682 (0.00)	2.5371 (0.00)	11.378 (0.00)
	USD-AUD			USD-CHF		
	1-step	5-step	10-step	1-step	5-step	10-step
RLS-ARFIMA(0, d , 0)	1.5891 (1.00)	9.6999 (1.00)	22.646 (1.00)	1.4567 (0.54)	7.4997 (1.00)	15.768 (1.00)
RLS-ARFIMA(1, d , 1)	1.5914 (0.64)	9.7033 (0.83)	22.657 (0.85)	1.4625 (0.49)	7.6311 (0.23)	16.478 (0.03)
RLS-ARMA(0, 0)	1.5904 (0.93)	9.7476 (0.58)	22.814 (0.56)	1.4539 (1.00)	7.5353 (0.62)	15.807 (0.87)
RLS-ARMA(1, 1)	1.6240 (0.24)	10.381 (0.03)	25.381 (0.01)	1.4616 (0.49)	7.6531 (0.05)	16.594 (0.00)
ARFIMA(0, d , 0)	1.6814 (0.00)	42.801 (0.00)	188.88 (0.00)	1.4992 (0.02)	9.3991 (0.00)	24.561 (0.00)
ARFIMA(1, d , 1)	1.5903 (0.93)	119.32 (0.00)	582.26 (0.00)	1.4632 (0.20)	135.47 (0.00)	666.52 (0.00)
HAR	0.1611 (0.38)	12.982 (0.00)	129.49 (0.00)	1.4681 (0.48)	8.7184 (0.00)	19.999 (0.00)
log-GARCH	2.0832 (0.00)	22.307 (0.00)	73.953 (0.00)	2.1949 (0.00)	28.801 (0.00)	113.70 (0.00)

Table 8: Forecast evaluations of the eight dynamic models. We use mean squared forecast errors (MSFE's) and consider MCS comparisons with all models included in the initial set. Here, **boldface** notation indicate whether a model belongs to the 10% MCS. The MCS p -values are in parentheses. See the main text for details.

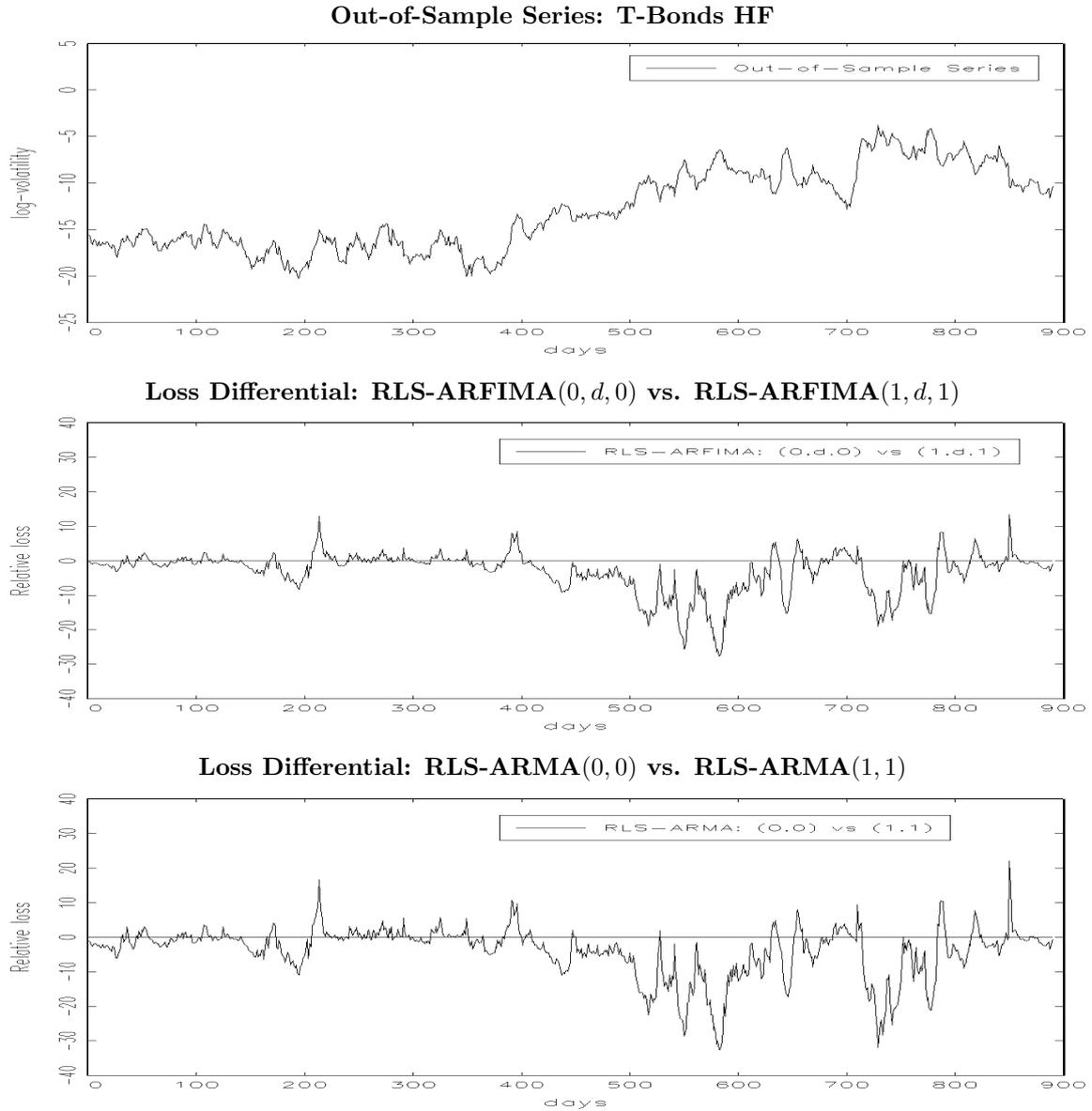
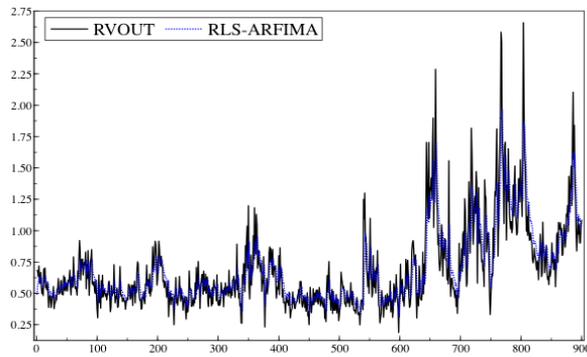
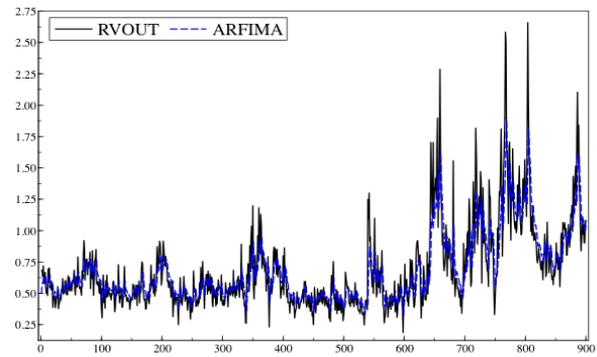


Figure 1: The upper panel displays the cumulative ten-step-ahead log-volatility proxy for the T-bond series. The middle and lower panels display the corresponding loss differentials, $d_{ij,t} = (\bar{\sigma}_{t,\tau} - \bar{y}_{t+\tau,i|t})^2 - (\bar{\sigma}_{t,\tau} - \bar{y}_{t+\tau,j|t})^2$, from the comparisons of the RLS-ARFIMA(0, d , 0) model against the RLS-ARFIMA(1, d , 1) model and the RLS-ARMA(0, 0) model against the RLS-ARMA(1, 1) model.

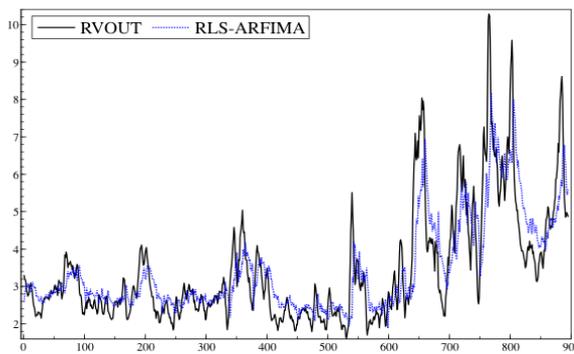
1-day Forecasts: RLS-ARFIMA



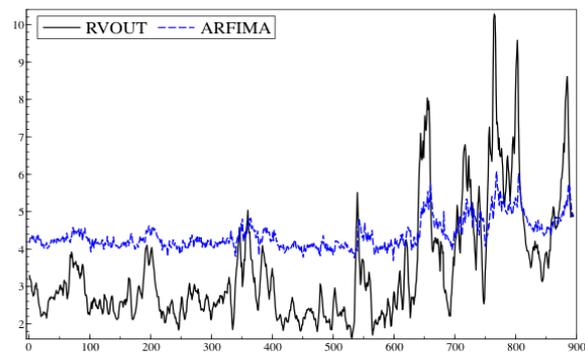
1-day Forecasts: ARFIMA



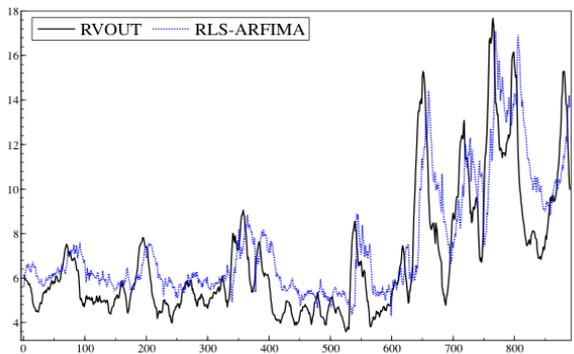
5-day Forecasts: RLS-ARFIMA



5-day Forecasts: ARFIMA



10-day Forecasts: RLS-ARFIMA



10-day Forecasts: ARFIMA

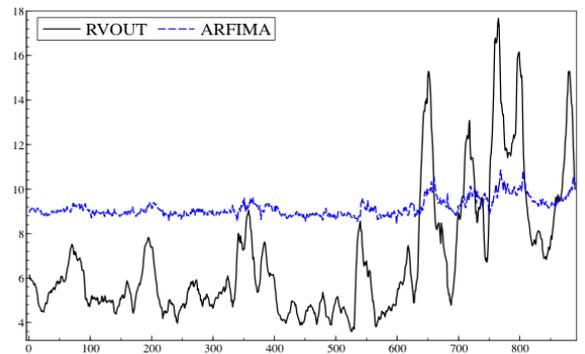


Figure 2: The upper left panel displays the out-of-sample 1-day volatility proxy (line, black) together with the RLS-ARFIMA(1, d , 1) forecast (dotted, blue) for the last 900 days of the SPY sample; the upper right panel displays the corresponding plot for the ARFIMA(1, d , 1) forecast (dashed, blue). The two middle and lower panels have the same left and right split, only the forecasts are cumulative direct 5 and 10-step-ahead, respectively.

A Handling Measurement Errors

The econometric methodology in the main text relies on an RLS-ARFIMA(p, d, q) approximation of the reduced form dynamics, that is, on a finite-order ARMA representation to account for both the underlying short memory dependencies and measurement errors in the series. We discuss the validity of our approach and provide evidence in favor of the specific parameterizations used.

A.1 Measurement Errors and the (RLS-)ARFIMA Representation

From the reduced form of y_t in (4), one observes that by allowing for measurement errors in the log-volatility proxy, it is generally necessary to have an RLS-ARFIMA(p, d, ∞) structure to fully capture its dynamics. However, we will emphasize two different cases, which empirically describe our volatility series well (evidence will follow) and render a finite order MA structure appropriate.

First, if $d = 0$, then (4) illustrates that we may model the residual dynamics, that is, the dynamics once level shifts are taken into account, by a finite ARMA($p, \max(p, q)$) specification, see Granger & Morris (1976), with an MA component determined by $\Theta(L)\epsilon_t + \Phi(L)u_t$. Second, if $d > 0$ and there are *no* measurement errors in the series, then an ARFIMA(p, d, q) model with MA component $\Theta(L)\epsilon_t$ will fully capture the residual dynamics. Hence, both cases allow for short memory dynamics of finite order. It is important to emphasize, however, that we do not impose these restrictions on the parameters from the outset. Rather, it is an empirical observation that the first case pertains to our daily FX volatility series, and the second case to the remaining high-frequency volatility measures. In fact, we find that AR(1) and ARMA(1, 1) specifications adequately capture the short memory dependencies of the former, respectively, the latter. A model that encompasses both these cases is the RLS-ARFIMA(1, d , 1) specification, which we analyze in detail in the simulation study as well as the empirical analysis of the main text. We will below add empirical evidence from a filtered long memory stochastic volatility model with random level shifts, or the RLS-LMSV model, to support these claims.

A.2 Motivational Evidence from an RLS-LMSV(1, d) Model

We consider an RLS-LMSV(1, d) model, that is, a SV model that allows for random level shifts, genuine long memory, and first-order AR dynamics. This specification readily extends the respective discrete-time SV models in Deo et al. (2006) and Qu & Perron (2013). The model is also closely related to the reduced form RLS-ARFIMA(1, d , 1) specification considered, but allows for structural inference on the standard deviation of the measurement error, σ_u , in addition to the fractional integration order and level shifts parameters. Hence, it facilitates a direct assessment of whether $d = 0$ and/or $\sigma_u = 0$ is appropriate, corresponding to the two cases discussed in Section A.1. Before proceeding, however, we note that estimation of the RLS-LMSV(1, d) model is computational intensive, and the model generates likelihood values similar to those for the RLS-ARFIMA(1, d , 1) model as well as delivers slightly worse forecasts, on average. As a result, we use the former only to motivate the latter.

A.2.1 RLS-LMSV: State Space Formulation and Estimation

Rather than working in first differences, as for the RLS-ARFIMA model in Section 4, let the observable log-volatility proxy, $y_t = x_t + u_t$, be written in a modified, yet still truncated, state space form as

$$y_t = \tilde{\mathbf{F}}\tilde{\mathbf{H}}_t + u_t, \quad \text{with} \quad \tilde{\mathbf{H}}_t = \tilde{\mathbf{G}}\tilde{\mathbf{H}}_{t-1} + \tilde{\mathbf{T}}\tilde{\mathbf{E}}_{t,\pi} \quad (9)$$

where $\tilde{\mathbf{F}} = (1, 0, \dots, 0, 1)'$ and $\tilde{\mathbf{H}}_t = (\mathbf{H}_t, v_t)$ are $(M + 1) \times 1$ vectors, and the 2×1 vector containing the state vector innovations, $\tilde{\mathbf{E}}_{t,\pi} \sim \text{i.i.d.N.}(\mathbf{0}_{2 \times 1}, \tilde{\mathbf{Q}}_\pi)$, depends on the particular regime of the process at time t through

$$\tilde{\mathbf{Q}}_1 = \begin{pmatrix} \sigma_\epsilon^2 & 0 \\ 0 & \sigma_\eta^2 \end{pmatrix} \quad \text{and} \quad \tilde{\mathbf{Q}}_0 = \begin{pmatrix} \sigma_\epsilon^2 & 0 \\ 0 & 0 \end{pmatrix},$$

corresponding to $\pi_{T,t} = 1$ and $\pi_{T,t} = 0$. Furthermore, since by defining the matrices

$$\tilde{\mathbf{G}} = \begin{pmatrix} \mathbf{G} & \mathbf{0}_{M \times 1} \\ \mathbf{0}_{1 \times M} & 1 \end{pmatrix} \quad \text{and} \quad \tilde{\mathbf{T}} = \begin{pmatrix} 1 & \mathbf{0}_{M \times 1} \\ \mathbf{0}_{M \times 1} & 1 \end{pmatrix}$$

of dimensions $(M + 1) \times (M + 1)$ and $(M + 1) \times 2$, respectively, the model has a state space structure that resembles the one in (6), and we can apply an estimation procedure that is very similar to the one described in the supplementary appendix for the RLS-ARFIMA model. Before proceeding to the empirical estimation results, however, we assess the accuracy of the RLS-LMSV parameter estimates from the proposed state space methodology in a small simulation exercise.

A.2.2 RLS-LMSV: Preliminary Numerical Results

We assess the accuracy of the RLS-LMSV parameter estimates by simulating an RLS-LMSV(1, d) process,

$$y_t = x_t + u_t, \quad x_t = h_t + v_t, \quad (1 - L)^d(1 - \phi L)h_t = \epsilon_t,$$

with $d = 0.35$, $\gamma/T = 0.02$, $\sigma_\epsilon = 0.5$, $\sigma_\eta = 3\sigma_\epsilon$, $\phi = 0.2$ and two different levels of measurement errors specified through the noise-to-signal ratio $\xi = \sigma_u^2/\sigma_\epsilon^2(1 - \phi)^2$, specifically $\xi = \{1, 2\}$. We compute the bias and root mean squared error (RMSE) of the parameter estimates for sample sizes $T = \{3000, 6000\}$, truncations $M = \{20, 30, T^{1/2}\}$ of the AR(M) representation, and $N = 100$ replications. The results of this exercise are presented in Table 9 below. We refer to the main simulation study in Section 5 for a discussion of the specific setup, i.e., of the choice of truncation and implementation details. Here, we are mainly interested in whether we can identify the key parameters γ/T , σ_η , d , and σ_ϵ for fairly high levels of measurement noise. The latter, in particular, is chosen higher than what our empirical estimates suggest, except for the USD-JPY series, to conservatively assess the inference procedure.

Table 9 illustrates two important points. First, we observe that the RLS-LMSV(1, d) model estimates the random level shift parameters γ/T and σ_η with no or a vanishingly small bias. Second, the estimates of the genuine long memory parameter, d , is slightly downward biased (not surprisingly, given the high

level of measurement errors). However, this bias diminishes when increasing the sample size, truncation length, or decreasing the noise-to-signal ratio. As stressed above, the bias is to be interpreted as a conservative estimate. Importantly, the RLS-LMSV(1, d) model is able to reliably recover information about the key persistence parameters in the underlying process.

A.2.3 RLS-LMSV: Empirical Estimates

To provide support for the claim that the eight log-volatility series can be categorized into the two cases referenced in Section A.1, as well as for our RLS-ARFIMA specifications, we report the parameter estimates from an RLS-LMSV(1, d) model for all series in Table 10 below using truncation $M = 20$.²³

From Table 10, we make a few noteworthy observations. First, for the three log-volatility series constructed from tick-by-tick trades, BAC, MRK, and SPY, we see that the impact of measurement errors is negligible, and similarly for the S&P 500 series. For the three exchange rate series, on the other hand, we observe non-negligible measurement noise. However, the estimated noise-to-signal ratios for the USD-AUD and USD-CHF series are still (much) smaller than the corresponding simulated values, but similar for the USD-JPY series. Moreover, note that we hardly estimate any ARFIMA dynamics for the former two. In this case, we cannot separately identify σ_ϵ and σ_u since the parameters will collectively measure the noise level in the series.²⁴ As seen in Section 5, if we simply interpret the noise as coming from one source, here σ_ϵ , the RLS-ARFIMA model precisely recovers this parameter.

In general, and as shown in Section 6, the estimated parameters in Table 10 present a striking pattern across the volatility series. Random level shifts are present in all series, occurring more frequently for all volatility proxies constructed from high-frequency data, but with less variability for most compared to those associated with the daily return series. In addition, the high-frequency volatility measures contain a large genuine long memory component, whereas there are seemingly little ARFIMA dynamics remaining in the exchange rate volatility series once level shifts have been accounted for. Moreover, we also observe a combination of measurement errors, random level shifts, and genuine long memory for the high-frequency-based T-bond series. However, given equivalent representations of AR(1) plus noise and ARMA(1, 1) dynamics and our empirical findings for the RLS-ARFIMA model in Section 6, we cannot exclude that the former is caused by a negative MA(1) component, which appears prominently in our empirical analysis. We refer to Sections 6.1 and 6.2 for a thorough discussion of these findings.

Finally, we have also estimated an RLS-LMSV model where we allow the latent short memory dynamics to follow an ARMA(1, 1) process. These results, though not reported here, provide no qualitative changes to the conclusions emerging from Table 10, thus supporting our claim in Section A.1 that an RLS-ARFIMA(1, d , 1) specification will encompass the two cases with $d = 0$ and $\sigma_u = 0$, respectively. Although the RLS-ARFIMA model does not facilitate structural inference in σ_u , treating instead measurement errors as an MA component, it is numerically much preferable to the RLS-LMSV(1, d) model,

²³Our choice of a relatively smaller truncation order $M = 20$, rather than $M = T^{1/2}$ as the theory in Section 4.2 dictates, reflects the conclusions from our main simulation study in Section 5, which we refer to for details. It is, however, worth noting that we have performed robustness checks using truncations $M = \{30, 40\}$, providing similar results.

²⁴The lack of identification in this case follows from $\epsilon_t + u_t \sim \text{i.i.d.N}(0, \sigma_\epsilon^2 + \sigma_u^2)$ by the assumptions in Section 2.

and it generates slightly better forecasts, on average, with no loss in terms of in-sample fit. Hence, this leads us to focus on the RLS-ARFIMA modeling strategy in the main text, implementing it using the estimation and forecasting procedures developed in Sections 4.2 and 4.3, respectively.

Simulations for an RLS-LMSV(1, d) Process								
	Bias				RMSE			
	γ/T	σ_η	d	σ_ϵ	γ/T	σ_η	d	σ_ϵ
$\xi = 2, T = 3000, M = 20$	-0.00	0.11	-0.15	0.10	0.01	0.30	0.18	0.19
$\xi = 2, T = 3000, M = T^{1/2}$	0.00	0.09	-0.14	0.09	0.01	0.29	0.18	0.19
$\xi = 1, T = 3000, M = 20$	0.00	0.06	-0.12	0.05	0.01	0.25	0.11	0.16
$\xi = 1, T = 3000, M = T^{1/2}$	0.00	0.06	-0.09	0.06	0.01	0.26	0.12	0.13
$\xi = 2, T = 6000, M = 20$	-0.00	0.08	-0.14	0.03	0.00	0.18	0.16	0.18
$\xi = 2, T = 6000, M = 30$	-0.00	0.08	-0.11	0.08	0.00	0.18	0.16	0.18

Table 9: Simulation results for an RLS-LMSV(1, d) model fitted to an RLS-LMSV(1, d) process with parameters $d = 0.35$, $\gamma/T = 0.02$, $\sigma_\epsilon = 0.5$, $\sigma_\eta = 3\sigma_\epsilon$, $\phi = 0.2$ and two different levels of measurement errors specified through the noise-to-signal ratio, $\xi = \sigma_u^2/\sigma_\epsilon^2(1 - \phi)^2$, where we, specifically, consider $\xi = (1, 2)$. Furthermore, we vary the sample size $T = (3000, 6000)$, truncations $M = (20, 30, T^{1/2})$, and consider $N = 100$ replications.

Full Sample Parameter Estimates							
RLS-LMSV(1, d)	ϕ	d	γ/T	σ_η	σ_ϵ	σ_u	KMLE
BAC HF	-0.0329	0.4795	0.0169	0.2721	0.2248	0.0004	135.934
MRK HF	0.0470	0.3063	0.0152	0.8466	0.3063	0.0000	-188.044
SPY HF	0.0063	0.4106	0.0193	0.3794	0.2266	0.0000	57.2792
S&P 500 HF	-0.0266	0.3168	0.0263	0.8738	0.2602	0.0891	-1620.01
T-Bonds HF	0.0254	0.4936	0.0051	0.3275	0.1355	0.3294	-2902.42
USD-AUD	0.0146	0.0277	0.0028	1.0459	1.3567	0.2651	-16933.4
USD-CHF	-0.0492	0.0874	0.0017	0.6769	1.2413	0.1638	-15872.1
USD-JPY	0.6903	0.0000	0.0028	2.9687	0.2516	1.2301	-16297.7

Table 10: Parameter estimates of the RLS-LMSV(1, d) model for the eight log-volatility series. “KMLE” denotes the predictive log-likelihood value from the Kalman filter.

References

- Andersen, T. G., Bollerslev, T. & Diebold, F. X. (2007), ‘Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility’, *The Review of Economics & Statistics* **89**, 701–720.
- Andersen, T. G., Bollerslev, T., Diebold, F. X. & Ebens, H. (2001), ‘The distribution of realized stock return volatility’, *Journal of Financial Economics* **61**, 43–76.
- Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2001), ‘The distribution of exchange rate volatility’, *Journal of the American Statistical Association* **96**, 42–55.
- Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2003), ‘Modeling and forecasting realized volatility’, *Econometrica* **71**, 579–625.
- Beran, J. (1995), ‘Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models’, *Journal of the Royal Statistical Society* **57**, 659–672.
- Bhattacharya, R., Gupta, V. & Waymire, E. (1983), ‘The Hurst effect under trends’, *Journal of Applied Probability* **20**, 649–662.
- Brockwell, P. J. & Davis, R. A. (1991), *Time Series: Theory and Methods*, 2nd ed. Springer Verlag, New York.
- Chan, N. H. & Palma, W. (1998), ‘State space modeling of long-memory processes’, *The Annals of Statistics* **26**, 719–740.
- Chen, C. & Tiao, G. C. (1990), ‘Random level shift time series models, ARIMA approximations, and level-shift detection’, *Journal of Business & Economic Statistics* **8**, 83–97.
- Chiriac, R. & Voev, V. (2011), ‘Modelling and forecasting multivariate realized volatility’, *Journal of Applied Econometrics* **28**, 922–947.
- Christensen, B. J. & Varneskov, R. T. (2016), ‘Medium band least squares estimation of fractional cointegration in the presence of low-frequency contamination’, *Journal of Econometrics* **forthcoming**.
- Corsi, F. (2009), ‘A simple approximate long-memory model of realized volatility’, *Journal of Financial Econometrics* **7**, 174–196.
- Deo, R. S., Hurvich, C. M. & Lu, Y. (2006), ‘Forecasting realized volatility using a long memory stochastic volatility model: estimation, prediction, and seasonal adjustment’, *Journal of Econometrics* **131**, 29–58.
- Diebold, F. X. & Inoue, A. (2001), ‘Long memory and regime switching’, *Journal of Econometrics* **105**, 131–159.
- Doornik, J. A. & Ooms, M. (2004), ‘Inference and forecasting for ARFIMA models with an application to US and UK inflation’, *Studies in Nonlinear Dynamics & Econometrics* **8**, Article 14.
- Engle, R. (1982), ‘Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation’, *Econometrica* **50**, 987–1007.
- Gabriel, V. J. & Martins, L. F. (2004), ‘On the forecasting ability of ARFIMA models when infrequent breaks occur’, *Econometrics Journal* **7**, 455–475.
- Giacomini, R. & White, H. (2006), ‘Tests of conditional predictive ability’, *Econometrica* **74**, 1545–1578.

- 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.
- Granger, C. W. J. & Joyeux, R. (1980), ‘Long memory relationships and the aggregation of dynamic models’, *Journal of Time Series Analysis* **1**, 15–29.
- Granger, C. W. J. & Morris, M. J. (1976), ‘Time series modelling and interpretation’, *Journal of the Royal Statistical Society. Series A.* **139**, 246–257.
- Grassi, S. & de Magistris, P. S. (2014), ‘When long memory meets the Kalman filter: A comparative study’, *Computational Statistics and Data Analysis* **76**, 301–319.
- Haldrup, N. & Nielsen, M. O. (2007), ‘Estimation of fractional integration in the presence of data noise’, *Computational Statistics and Data Analysis* **51**, 3100–3114.
- Hamilton, J. D. (1994a), ‘State space models’, in R. F. Engle & D. L. McFadden, eds, ‘*Handbook of Econometrics, Volume IV*’, Elsevier, pp. 3041–3080.
- Hamilton, J. D. (1994b), *Time Series Analysis*, Princeton University Press, Princeton, New Jersey.
- Hansen, P. R. & Lunde, A. (2006), ‘Consistent ranking of volatility models’, *Journal of Econometrics* **131**, 97–121.
- Hansen, P. R. & Lunde, A. (2014), ‘Estimating the persistence and the autocorrelation function of a time series that is measured with error’, *Econometric Theory* **30**, 60–93.
- Hansen, P. R., Lunde, A. & Nason, J. M. (2011), ‘The model confidence set’, *Econometrica* **79**, 453–497.
- Harvey, A. C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge, United Kingdom.
- Harvey, A. C. & Shephard, N. (1996), ‘Estimation of an asymmetric stochastic volatility model for asset returns’, *Journal of Business & Economic Statistics* **14**, 429–434.
- Hosking, J. R. M. (1981), ‘Fractional differencing’, *Biometrika* **68**, 165–176.
- Hurvich, C. M. & Ray, B. K. (2003), ‘The local whittle estimator of long-memory stochastic volatility’, *Journal of Financial Econometrics* **1**, 445–470.
- Koopman, S. J., Jungbacker, B. & Hol, E. (2005), ‘Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements’, *Journal of Empirical Finance* **12**, 445–475.
- Lobato, I. & Savin, N. (1998), ‘Real and spurious memory long-memory properties of stock market data’, *Journal of Business & Economic 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.
- Martin, V. L. & Wilkins, N. P. (1999), ‘Indirect estimation of ARFIMA and VARFIMA models’, *Journal of Econometrics* **93**, 149–175.
- McCloskey, A. & Hill, J. B. (2015), ‘Parameter estimation robust to low frequency contamination’, *Journal of Business & Economic Statistics* **forthcoming**.
- McCloskey, A. & Perron, P. (2013), ‘Memory parameter estimation in the presence of level shifts and deterministic trends’, *Econometric Theory* **29**, 1196–1237.

- McCulloch, R. & Tsay, R. (1993), ‘Bayesian inference and prediction for mean and variance shifts in autoregressive time series’, *Journal of the American Statistical Association* **88**, 968–978.
- Meddahi, N. (2003), ‘ARMA representation of integrated and realized variances’, *Econometrics Journal* **6**, 334–379.
- Mikosch, T. & Stărică, C. (2004), ‘Nonstationarities in financial time series, the long range dependence, and the IGARCH effects’, *Review of Economics & Statistics* **86**, 378–390.
- Nielsen, M. O. (2015), ‘Asymptotics for the conditional-sum-of-squares estimator estimator in multivariate fractional time-series models’, *Journal of Time Series Analysis* **36**, 154–188.
- Nunes, L. C., Newbold, P. & Kuan, C.-M. (1995), ‘Spurious breaks’, *Econometric Theory* **11**, 555–577.
- Ohanissian, A., Russell, J. R. & Tsay, R. S. (2008), ‘True or spurious long memory? A new test’, *Journal of Business & Economic Statistics* **26**, 161–175.
- Patton, A. (2011), ‘Volatility forecast comparison using imperfect volatility proxies’, *Journal of Econometrics* **160**, 246–256.
- 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 changing mean’, *Journal of Business & Economic Statistics* **8**, 153–162.
- Perron, P. & Qu, Z. (2007), An analytical evaluation of the log-periodogram estimate in the presence of level shifts. Unpublished Manuscript, 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 & Economic Statistics* **28**, 275–290.
- Perron, P. & Wada, T. (2009), ‘Let’s take a break: Trends and cycles in US real DGP’, *Journal of Monetary Economics* **56**, 749–765.
- Qu, Z. (2011), ‘A test against spurious long memory’, *Journal of Business & Economic Statistics* **29**, 423–438.
- Qu, Z. & Perron, P. (2013), ‘A stochastic volatility model with random level shifts and its applications to S&P 500 and NASDAQ return indices’, *Econometrics Journal* **16**, 300–339.
- Ray, B. K. & Tsay, R. S. (2002), ‘Bayesian methods for change-point detection in long-range dependent processes’, *Journal of Time Series Analysis* **23**, 687–705.
- Smith, A. (2005), ‘Level shifts and the illusion of long memory in economic time series’, *Journal of Business & Economic Statistics* **23**, 355–389.
- Stărică, C. & Granger, C. W. J. (2005), ‘Nonstationarities in stock returns’, *The Review of Economics & Statistics* **87**, 503–522.
- Varneskov, R. T. (2016a), “Estimating the quadratic variation spectrum of noisy asset prices using generalized flat-top realized kernels”, *Econometric Theory* **forthcoming**.
- Varneskov, R. T. (2016b), ‘Flat-top realized kernel estimation of quadratic covariation with non-synchronous and noisy asset prices’, *Journal of Business & Economic Statistics* **34**, 1–22.

- Varneskov, R. T. & Perron, P. (2017), 'Combining long memory and level shifts in modeling and forecasting the volatility of asset returns: Supplementary appendix'. Unpublished Manuscript, Boston University.
- Varneskov, R. T. & Voev, V. (2013), 'The role of realized ex-post covariance measures and dynamic model choice on the quality of covariance forecasts', *Journal of Empirical Finance* **20**, 83–95.
- Xu, J. & Perron, P. (2014), 'Forecasting return volatility: Level shifts with varying jump probability and mean reversion', *International Journal of Forecasting* **30**, 449–463.