Spurious Long Memory, Uncommon Breaks and the Implied-Realized Volatility Puzzle

Neil M. Kellard, Ying Jiang, Mark Wohar

PII: S0261-5606(15)00064-9
DOI: 10.1016/j.jimonfin.2015.04.003
Reference: JIMF 1552

To appear in: *Journal of International Money and Finance*

Spurious Long Memory, Uncommon Breaks and the Implied-Realized Volatility Puzzle

Neil M. Kellard[*†], Ying Jiang [**] and Mark Wohar[***]

[*]Essex Business School and Essex Finance Centre, University of Essex
[**]Nottingham University Business School, China
[***]Department of Economics, University of Nebraska at Omaha

October 2014

[†] Corresponding author: Essex Business School, Essex Finance Centre, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, United Kingdom. Tel: +44 1206 874153, Fax: +44 1206 873429. E-mail: nkellard@essex.ac.uk. We are grateful to Tim Bollerslev, Marcus Chambers, Jerry Coakley, David Harvey, David Hendry, John Nankervis and Rob Taylor for helpful comments and suggestions. Earlier versions of this paper were presented at the International Symposium on Forecasting in San Diego, 2010, and a seminar at the University of Oxford, 2013, and we thank participants for their comments. Any errors remain our own.

Spurious Long Memory, Uncommon Breaks and the Implied-Realized Volatility Puzzle

Abstract

One of the puzzles in international finance is the frequent finding that implied volatility is a biased predictor of realized volatility. However, given asset price volatility is often characterized as possessing long memory, recent literature have shown that allowing for long-range dependence removes this bias. Of course, the appearance of long memory can be generated by the presence of structural breaks. This paper discusses the effect of structural breaks on the implied-realized volatility relation. Simulations show that if significant structural breaks are omitted, testing can spuriously show the typical patterns of fractional cointegration found in the literature. Next, empirical results show that foreign exchange implied and realized volatility contain structural breaks. The breaks in the implied series never closely anticipate or co-occur with those of the realized series, suggesting the market has no ability to forecast structural change. When breaks are accounted for in the bi-variate framework, the point estimate of the slope parameter falls and the null of unbiasedness can be rejected. Allowing for structural breaks, suggests the implied-realized volatility puzzle might not be solved after all.

JEL classification: C14, C22, F31, G14.

Keywords: implied-realized relation, unbiasedness, uncommon structural change, foreign exchange, Monte Carlo simulation.

1. Introduction

Optimal modelling and forecasting of volatility is essential for a variety of risk assessment and trading purposes. However, standard market efficiency tests in the extant literature (see, *inter alia*, Christensen and Prabhala, 1998, and Poteshman, 2000) have routinely led to the conclusion that option implied volatility (IV) is a biased forecast of realized volatility (RV). Specifically, given the regression below

$$\sigma_{t+\tau}^{R} = \alpha + \beta\sigma_{t}^{IV} + u_{t+\tau} \qquad (1)$$

where σ_{t}^{IV} is IV over a time period τ and $\sigma_{t+\tau}^{R}$ represents RV over that same period, least squares estimation typically finds $\hat{\beta} < 1$, violating the joint unbiasedness restrictions of $\alpha = 0$, $\beta = 1$ and $u_{t+\tau}$ being serially uncorrelated. This *bias* occurs across a number of asset markets (see Neely, 2009) and has therefore inspired the search for an appropriate rationale. Common suggestions include that volatility risk is not priced (Chernov, 2007), computing RV with low-frequency data (Poteshman, 2000) and that the standard estimation with overlapping observations produces inconsistent parameter estimates (Christensen, Hansen and Prabhala, 2001). However, Neely (2009), shows the conditional bias in IV is robust to these potential solutions.

The optimality of the approach applied to the estimation of (1) relies crucially on the order of integration (d) of the covariates. Given the extant literature suggests individual volatility series are appropriately represented as long memory, fractionally integrated processes with $0 < d < 1$ (Anderson *et al.*, 2001a and 2001b), least squares estimates of (1) will be inconsistent when $d < 0.5$, and although consistent when $d > 0.5 > 1$, converges slowly[1] at the rate $O_p(T^{1-d-d_{\min}})$.

[1] See Marinucci and Robinson (2001; p.231).

Employing either foreign exchange or stock market data, Kellard *et al.* (2010, hereafter KDS), Nielsen (2007), Bandi and Perron (2006), Christensen and Nielsen (2006) show that IV and RV are fractionally cointegrated series wherein equation (1), $u_{t+\tau} \sim I(d-b)$ and $b \leq d$. Moreover, this literature suggests that estimators, such as narrow band least squares[2] (NBLS), account for the fractional character of volatility and find a unity slope parameter in equation (1) cannot be rejected. In other words, the traditional slope bias disappears. However, KDS also show that the frequency of data used for measuring RV within a fractionally cointegrating framework is important for the results of unbiasedness tests. Specifically, for some popular exchange rates, the use of less noisy intra-day rather than daily data reveals the possibility a different bias, as evidence of a latent fractionally integrated risk premium is detected.

For the sake of clarity, consider augmenting regression (1) with a time-varying risk premium term rp_t

$$\sigma_{t+\tau}^{RV} = \alpha + \beta\sigma_t^{IV} + \delta rp_t + u_{t+\tau}. \qquad (2)$$

A corollary of finding fractional cointegration between RV and IV is that any risk premium will be of a lower order of (fractional) integration than the original volatilities. In this context (see Bandi and Perron, 2006), spectral methods like NBLS will still estimate regression (1) consistently. Re-arranging (2) leads to

$$\sigma_{t+\tau}^{RV} - \alpha - \beta\sigma_t^{IV} = \delta rp_t + u_{t+\tau}. \qquad (3)$$

If daily data is relatively noisy, KDS posit any long memory behaviour of the risk premium[3] is swamped[4] and therefore hidden by $u_{t+\tau}$ in finite samples. Contrastingly,

[2] See Robinson and Marinucci (2003).
[3] Evidence for a fractionally integrated risk premium in forward foreign exchange markets is provided by Kellard and Sarantis (2008). Further discussion of volatility risk premia in other markets can be found in Almeida and Vicente (2009) and Doran and Ronn (2008).

the use of a less noisy intra-day derived RV may lead to a smaller $u_{t+\tau}$ and therefore the detection of the time series properties of a time-varying risk premium. Following Bandi and Perron (2006), KDS purposely avoid modelling a specific functional form for a risk premium, arguing that long memory behaviour in the residual of (1) presents *prima facie* evidence for concealed risk premia. Indeed, fractionally integrated behaviour where $b < d$, is found in the estimated residual of (1) when intra-day data is employed to construct RV.

In any case, it is important to note that other recent work has proposed that long memory is an illusory feature of volatility series. In particular, Granger and Hyung (2004) and Mikosch and Starica (1999) demonstrate that ignoring significant structural change causes the appearance of persistence in individual time series. Additionally, Choi *et al*. (2010) show that allowing for structural breaks in daily realized volatility of three foreign exchange rates partially explains their persistence whereas Li and Perron (2013) suggests that the long memory property disappears altogether. Even more pertinently for our study, Christensen and de Magistris (2010), using S&P 500 futures from April 1988 to October 2007, show that a common level shift process appropriately fits RV and IV data. When the common shift process is removed from the data, a simple VAR model subsequently shows that IV has no explanatory power for future RV. Finally, Monte Carlo evidence shows that a latent common level shift process in the DGP of the volatilities can spuriously lead to the finding of fractional cointegration. On the other hand, it should be noted that Garvey and Gallagher (2012), employing a sample of 16 FTSE 100 stocks from October 1997 to December 2003, suggest that the long memory property of volatility series are not due to breaks occurring over their chosen sample period.

[4] See Maynard and Phillips (2001), Kellard (2006) and Kellard and Sarantis (2008) for other discussions of swamping and its effect in finite time series.

In a similar vein, this paper examines the effect of structural change on the IV-RV relation. Christensen and de Magistris (2010) argue that common breaks in series lead to long memory persistence in volatility and fractional cointegration in (1) with $u_{t+\tau} \sim I(0)$ where $b = d$. However, KDS find approximately $u_{t+\tau} \approx I(0.3)$, a result that as noted above, can be driven by the presence of a risk premium. Another possible explanation is the existence of *uncommon* structural breaks in RV and IV, which therefore create level breaks in the residual of (1) and spurious fractional cointegration. Christensen and de Magistris (2010) suggest that co- breaking is likely because of common responses to movements between booms and busts in financial markets. Of course, allowing peso-type problems to exist or some other time-varying risk premia it is quite possible that traders' trade implied volatility at a different level to realized volatility. On the other hand, it is also possible that IV is a relatively poor forecast of RV. In either case, it is quite possible that market-specific breaks in RV are not contemporaneously mirrored in IV.

This paper therefore extends the extant literature in five steps. Firstly, Monte Carlo experiments show that uncommon structural breaks can cause the finding of fractional cointegration with $b < d$. Secondly, the Bai and Perron (1998, 2003a, 2003b, 2004) method is employed to test for uncommon multiple breaks in the mean levels of foreign exchange volatility series. Thirdly, we explicitly examine the time series properties of break-free individual volatility series and fourthly, we examine a break-free version of regression (1).

Our fifth contribution to the literature is derived from noting that the testing procedure suggested above relies on two notions. Firstly, (i) that any structural breaks identified are the (at least) partial cause of spurious long memory and (ii) that uncertainty over the estimation of structural breaks is ignored in the construction of confidence intervals of break-free version of regression (1). To address the former point we carry out the popular Ohanissian *et al*. (2008) test which uses the self-similarity property of true long memory processes to assess whether any long memory in our individual series is true or due to breaks. Simulations are also undertaken to assess whether the Ohanissian *et al.* test can be usefully used when non-stationary long memory is suspected. To address point (ii) we note that uncertainty in any first-stage break estimation comes from three sources: the number, date and the size of the breaks. It would appear difficult to deal with these 3 sources of uncertainty in a formal, rigorous way and previous literature has not considered the issue. To assess the extent of the issue, we provide Monte Carlo evidence on the effect of first-stage uncertainty on second-stage confidence interval coverage. This allows us to suggest a new sequential estimation procedure for the demeaned version of regression (1) which is more robust to break estimation uncertainty.

The empirical results are interesting; using data on three currencies for the period 1991-2007, evidence is provided by both the Bai and Perron procedure and the Ohanissian *et al*. tests, that RV and IV contain breaks. The estimated breaks in the implied series never occur *just* before or contemporaneously with those of the realized series, suggesting the market has no ability to forecast or mirror structural change. Moreover, when breaks are accounted for, the magnitude of fractional integration parameter d drops towards zero for individual volatility series. However, moving to the bi-variate framework, the point estimate of the slope parameter falls away from unity and the null of unbiasedness can often be rejected. The rejection of unbiasedness is particularly acute when RV is constructed by intra-day rather than daily data. In summary, explicitly modelling structural breaks suggests the implied-realized volatility puzzle is not resolved by using econometric techniques that allow for long memory.

The paper is divided into six sections: Section 2 presents the empirical methodology; section 3 describes the data; section 4 provides the simulation results; section 5 analyses the empirical results and, finally, section 6 concludes.

2. Empirical methodology

(i) Testing for long memory and fractional cointegration

To estimate the order of integration of RV and IV series and subsequently estimate (1), the recent literature has employed techniques that account for long memory. For example, there are several approaches to estimating d the memory parameter for individual series. Perhaps the most commonly used, partly due to its semi-parametric nature, is the log periodogram estimator (see Geweke and Porter-Hudak, 1983; Robinson, 1995a; Velasco, 1999a) typically known as the GPH statistic. This involves the least squares regression

$$\log I(\lambda_j) = \beta_0 - d \log\{4\sin^2(\lambda_j / 2)\} + u_j, \quad j = 1, 2, ..., m \tag{4}$$

where $I(\lambda_j)$ is the sample spectral density of y_t evaluated at the $\lambda_j = 2\pi j / T$ frequencies, T is the number of observations and m is small compared to T. However, a more contemporary alternative that has been used recently in the estimation of long memory in volatility series is the Gaussian semiparametric[5] estimate (GSP) of Robinson (1995b) shown below

$$\hat{d} = \arg \min_{d \in \theta} R(d) \tag{5}$$

[5] Semiparametric estimation is typically preferred in the long memory estimation of volatility (see Bandi and Perron, 2006). For example, Christensen and Nielsen (2006) employ GSP estimation. Fully parametric estimation of the ARFIMA model is more efficient but inconsistent if the order of *p* and *q* are incorrectly selected.

where

$$R(d) = \log \hat{G}(d) = 2d \frac{1}{m}\sum_{j=1}^{m} \log \lambda_j \qquad \hat{G}(d) = \frac{1}{m}\sum_{1}^{m} \lambda_j^{2d} I(\lambda_j)$$

Velasco (1999b) shows the GSP estimate is consistent over $d \in (-1/2, 1)$ and asymptotically distributed $d \in (-1/2, 3/4)$. Furthermore, the GSP estimator is shown to be more efficient than the GPH regression estimator. Therefore, in the later results section, the GSP estimator will be employed to assess long memory in observed volatility time series.

Following Christensen and Nielsen (2006), this paper adopts a multi-step methodology where the concluding step estimates the GSP statistic, δ, for the narrow band least squares (NBLS) residual of the equilibrium relationship. Here the β slope coefficient in (1) is estimated by

$$\hat{\beta}_z = \left[\sum_{j=0}^{z} I_{\sigma_{IV}}(\lambda_j)\right]^{-1} \sum_{j=0}^{z} I_{\sigma_{IV}\sigma_{RV}}(\lambda_j), \quad 0 \le z \le T-1 \tag{6}$$

where $I_{\sigma_{IV}}(\lambda_j)$ is the sample spectral density of IV and $I_{\sigma_{IV}\sigma_{RV}}(\lambda_j)$ is the cross-spectrum between IV and RV[6]. Furthermore, band spectrum regression is NBLS given

$$\frac{1}{z} + \frac{z}{T} \to 0 \;\; as\, T \to \infty \tag{7}$$

In the non-stationary case where $d > 0.5$, Velasco (2003) shows that when the cointegrating relationship has significantly less memory than the observed series, and is derived from consistent[7] estimates of the parameters, the GSP estimate is asymptotically normal. Subsequently, Christensen and Nielsen (2006) examining the

---

[6]Hence, $\hat{\beta}_{T-1}$ is a special case, equal to the OLS estimate of β in (1).

[7] Both OLS and NBLS estimates are consistent and converge at appropriate rates in the non-stationary region.

stationary case where $d < 0.5$, assume that δ for the NBLS[8] residual can be estimated as if the residuals are observed.

Finally, to construct 90% and 95% confidence intervals for the slope coefficient in (1), and following KDS and Gerolimetto (2006), a wild bootstrap procedure is employed. In the frequency domain, this involves resampling the NBLS residuals with replacement and subsequently constructing a bootstrapped dependent variable. The new dependent variable is then regressed on the original frequency domain regressors to obtain the bootstrapped coefficient vector. Repeating this procedure by using the bootstrap class in OX, 1000 bootstrapped slope coefficients were generated.

(ii) Detecting spurious long memory and estimating multiple structural breaks

As commented on earlier, fractional integration can be spuriously identified in the presence of latent structural breaks (see Granger and Hyung, 2004) or regime switches (Diebold and Inoue, 2001). To assess this issue, some techniques for distinguishing between true long memory and level shifts have been suggested in the recent literature. For example, Ohanissian *et al*. (2008)[9] notes that if data are from a true long memory process, the fractional differencing parameter is the same across all temporal aggregation levels (see Chambers, 1998). From this observation, a test is proposed that does not require the estimation of the number of structural breaks and has the null hypothesis, $H_0 : d_{m_1} = d_{m_2} = ... = d_{m_M} = d$, where m_j represents the level of temporal aggregation[10] and $m_1 < m_2 < ... < m_M$. The test statistic can be written

[8] OLS parameter estimates for the cointegrating vector are inconsistent in the stationary region.
[9] Although the Ohanissian *et al*. (2008) is perhaps the most popular test for true long memory, other published alternatives include Qu (2011). These tests typically assume $d < 0.5$.
[10] Following Ohanissian *et al*. (2008) for daily frequency data (see their Table 6), we set $m_j = 2^{j-1}$, $j = 1, 2, ..., M$; $M = 4$. Note that all $\hat{d}$ are estimated using the GPH estimator.

$$\hat{W} = (Z\hat{d})'(Z\Lambda Z')^{-1}(Z\hat{d}), \tag{8}$$

where $\hat{d}$ is an M-dimensional vector of estimated memory parameters, $(\hat{d} = \hat{d}_{m_1} = \hat{d}_{m_2} = \ldots = \hat{d}_{m_M})'$, Λ is the asymptotic covariance matrix of $\hat{d}$ and Z represents an $(M-1, M)$ matrix allowing Λ to be invertible[11]. Under the null, $\hat{W}$ has an asymptotic χ^2_{M-1} distribution.

A frequently adopted approach (see Coakley *et al*., 2011, Choi *et al.*, 2010, Kellard and Sarantis, 2008, and Choi and Zivot, 2007) to estimate multiple structural breaks is due to Bai and Perron (1998, 2003a, 2003b, 2004). To explain, consider the m -breaks in mean model below

$$y_t = \mu_j + \varepsilon_t \tag{9}$$

where $j = 1,\ldots, m+1$ and μ_j is the mean level of y_t in the j^{th} regime. Moreover, let the m-partition $(T_{1,\ldots,}T_m)$ be the breakpoints for the different regimes and conventionally, $T_0 = 0$ and $T_{m+1} = T$. To estimate the breakpoints, the objective function below is employed

$$(\hat{T}_1,\ldots,\hat{T}_m) = \arg\min_{T_1,\ldots,T_m} S_T(T_1,\ldots,T_m) \tag{10}$$

where for each m-partition $(T_{1,\ldots,}T_m)$, the least squares estimates of μ_j are generated by minimising the sum of the squared residuals

$$S_T(T_1,\ldots,T_m) = \sum_{j=1}^{m+1} \sum_{t=T_{j-1}+1}^{T_j} (y_t - \mu_j)^2 \tag{11}$$

That is, the breakpoint estimators correspond to the global minimum of the sum of the squares objective function. To solve the minimization problem in (10), Bai and Perron

[11] See Ohanissian *et al*. (2008, p. 166). Additionally, note we estimate Λ using the approximated variance-covariance matrix from equation (3) in Ohanissian *et al*. The test is programmed using Ox version 7.00.

(2004) propose the use of a specific dynamic programming algorithm. Obviously, after estimating the breakpoints, it is straightforward to obtain the corresponding least-squares regression parameter estimates $\hat{\mu}_j(\hat{T}_1,...,\hat{T}_m)$.

A useful attribute of the Bai and Perron (1998, 2003a) method is that their test statistics[12] can be generated under reasonably general specifications. Specifically, specifications can allow for (i) autocorrelation and heteroskedasticity in the regression model residuals and (ii) different moment matrices for the regressors in the different regimes. To incorporate all these features, we employ the most general Bai and Perron (1998, 2003a) specification[13]. Finally, given computed structural breaks in RV or IV, we estimate the following 'break-free' version of regression (1):

$$\sigma_{t+\tau}^{R*} = \alpha + \beta\sigma_t^{IV*} + u_{t+\tau} \quad (12)$$

where $\sigma_{t+\tau}^{*} = \sigma_{t+\tau} - \hat{\mu}_j$, and $\hat{\mu}_j(j = 1,...,m+1)$ is the estimated mean level of volatility in the j^{th} regime.

3. Data

Monthly time series of foreign exchange RV and IV were constructed from daily data for the period January 1991 to December 2007[14]. IV is measured by at-the-money, one-month forward, over-the-counter (OTC) market quoted volatilities[15] for European options at close of business in London, obtained from brokers by Reuters. By contrast,

---

[12] Such as *UD*max and *WD*max that test the null hypothesis of no structural breaks versus the alternative of an unknown number of breaks and $SupF_T(b+1|b)$ to test the null hypothesis of b breaks against the alternative of $b+1$.

[13] Specifically, using the notation of Bai and Perron (2004), we set cor_u = 1, het_u = 1 and $\pi = 0.15$. Following Choi and Zivot (2007), we set $M = 5$. Note that the Bai and Perron (1998, 2003a,b) statistics are computed using the GAUSS program available from Pierre Perron's home page at http://econ.bu.edu/perron/.

[14] Our paper uses the same data as Kellard *et al*. (2010) to maintain consistency with previous literature.

[15] Also see Dunis and Keller (1995), Dunis and Huang (2002) and Sarantis (2006).

many of the studies in the literature employ IV backed out from exchange-traded option prices. However, the OTC FX market is greatly more liquid than its exchange-traded counterpart[16].

As Covrig and Low (2003) describe, although participants in exchange-traded markets quote prices in terms of the familiar option premium, OTC prices are given in terms of volatility. In other words, an option could be quoted at 12% p.a. and subsequently converted into the appropriate option premium by using the Garman–Kohlhagen model. Therefore, given currency volatility has become a traded quantity in financial markets, it is therefore directly observable on the marketplace[17] and the use of these volatilities avoids the potential biases (i.e., errors in the choice of option pricing model and the measurement of model inputs) associated with backing out data from an option pricing model. These 'traded' IVs measure[18] the market's expectation about the future volatility of the spot exchange rate for three currencies: US dollar Sterling, Swiss Franc/US dollar and US dollar/Yen.

To match with the IV data for each day, two versions of RV are calculated over the remaining one month of the option. Firstly, employing intra-day data[19], the sum of the 5-minute squared logarithmic returns for each foreign exchange rate series is used to compute the daily variance (h_t) and then the RV quantity

$$\sigma_{t+\tau}^{RV^h} = \sqrt{\frac{252}{\tau-1}\sum_{i=1}^{\tau} h_{t+i}} \; . \tag{13}$$

[16] Consider that at the end of June 2012, the Bank of International Settlement (2012) reported the notional amount outstanding in the OTC currency option market stood at $11.1 trillion, compared with $111 billion for the exchange-traded market. Moreover, the US Dollar (i.e., $8.7 trillion outstanding) and the Euro (i.e., $4.1 trillion outstanding) are the two most heavily traded currencies within the option OTC market.

[17] This data was obtained from CIBEF at Liverpool John Moores University. Unfortunately, the databank is no longer updated.

[18] Implied volatilities are also annualised rates so that a quoted volatility of 5 per cent typically translates to a monthly variance rate of $(0.05^2)(21/252)$. The calculations assume that annualised rates refer to a 252 trading day year.

[19] From Olsen Associates.

Secondly, using daily returns data

$$\sigma_{t+\tau}^{RV^d} = \sqrt{\frac{252}{\tau - 1}\sum_{i=1}^{\tau}(r_{t+i} - \bar{r}_t)^2} \, , \tag{14}$$

where τ is the relevant number of trading days[20], S_t is the closing (London time) average of bid and ask quotes for the spot exchange rates and $r_t = \ln(S_t / S_{t-1})$. The constructed daily dataset contains 4348 time series observations for each volatility series.

As argued by Christensen and Prabhala (1998), the estimation of equation (1) will suffer from overlapping data problems if daily datasets are employed. To avoid this, a monthly dataset from the daily version by selecting an IV observation from the next trading day after the final day used in the calculation of the prior RV figure. Continuing in this manner, the data cycles through the calendar and the sampled dataset presents 198 non-overlapping observations for each volatility series. As an example, the logarithm of each monthly volatility series for US dollar Sterling are plotted in Figures 1 to 3.

[Insert Figures 1 to 3]

4. Simulations

Before moving to the later empirical sections that apply the above structural break methodology to the RV-IV relation, we carry out the following simple simulations over $t = 1,\ldots,T$, to gauge the effect of level structural breaks on any bivariate relationship.

(i) Spurious fractional cointegration and common breaks

[20] Assumed to be 21 days.

Initially, let x_t be generated by a short memory $z_t \sim$ AR (1) process with two level breaks

$$z_t = \rho z_{t-1} + \eta\varepsilon_{1t}, \qquad \varepsilon_{1t} \sim N(0,1) \qquad (15)$$
$$x_t = z_t + \sum_{j=1}^{2}\delta_j DU_{jt}(T_j)$$

where $DU_{jt}(T_j) = 1(t > T_j)$, 1(.) denoting the indicator function, T_j the break dates and η is a scaling parameter for calibration purposes. In our model 1, with common breaks in each series, the true regression model is therefore

$$y_t = x_t + \kappa\varepsilon_{2t}, \qquad \varepsilon_{2t} \sim N(0,1) \qquad (16)$$

where κ is a second scaling parameter. Based on an estimated AR(1) models[21] for our actual monthly data, we set $\rho = 0.6$, $\delta_{j=1} = -0.25$, $\delta_{j=2} = -0.25$, $\eta = 0.15$, $\kappa = 0.10$, $T = 200$, $T_{j=1} = 50$ and $T_{j=2} = 150$. After running a 1000 replications, we used NBLS to estimate

$$y_t = \hat{\alpha}_z + \hat{\beta}_z x_t + \hat{u}_t. \qquad (17)$$

GSP statistics with bandwidths $m = T^i$ $(i = 0.5, 0.6, 0.7)$ are computed for the individual simulated series y_t and x_t, and also for $\hat{u}_t$ and the mean of each of those statistics shown in rows 2 to 4 of Table 1.

[Insert Table 1]

The results above clearly confirm previous work that suggests that spurious fractional cointegration can be created in the presence of level structural breaks; specifically, the semi-parametric GSP statistic estimates orders of integration for the individual series of around 0.5, whereas the common break process allows the mean integration order of approximately zero for the residual series.

[21] Results are omitted to save space but available from the authors on request. Please note that as can be seen in our later data section, our primary dataset has a monthly frequency with approximately 200 observations. This is why we set T = 200 in this simulation.

(ii) Uncommon breaks

In the introduction, we suggested that uncommon, rather than common, breaks might provide an explanation for some of the results found in the prior literature. To simulate uncommon breaks for our model 2, x_t is defined as before in (15) but y_t can now be written

$$y_t = z_t + \sum_{k=1}^{2} \delta_k DU_{kt}(T_k) + \kappa\varepsilon_{2t} \tag{18}$$

However, in (18) only, we now allow $T_{k=1} = 100$ and $T_{k=2} = 150$. In other words, we allow x_t to present the first level shift 50 observations before y_t. Again GSP statistics with bandwidths $m = T^i$ $(i = 0.5, 0.6, 0.7)$ are computed for the individual simulated series y_t and x_t, and also for $\hat{u}_t$ and the mean of each of those statistics shown in rows 5 to 7 of Table 1. The mean order of integration for the residual series in (17) is now positive across all estimated bandwidths, a finding generated by the uncommon breaks in the individual series, resulting in a level break in the residual process of the bivariate relationship. This detection of spurious long memory in the residual series may be the rationale for the suggested risk premia in extant work like KDS. In any case, the Monte Carlos above indicate that, whether in the presence of common or uncommon structural breaks, such time series behaviour should be modelled explicitly to appropriately assess the bivariate volatility relation. This is how we will proceed in the later empirical section.

(iii) Power and size of the Ohanissian et al. (2008) test

As noted earlier, a number of tests have been developed recently to try and distinguish between true long memory and structural breaks. However, tests like the popular

Ohanissian *et al*. (2008) test for spurious long memory formally require $d \in (0, 1/2)$. However, in the extant literature, the estimated order of integration for foreign exchange volatility is typically close to the non-stationary boundary (i.e. $\hat{d} \approx 0.5$), with some estimates located in the non-stationary region (i.e. $\hat{d} > 0.5$). To assess the effect of non-stationarity on the $\hat{W}$ test statistic in (8), we initially let x_t be generated by an ARFIMA $(0, d, 0)$ series

$$(1-L)^d x_t = \varepsilon_{1t}, \quad (19)$$

where the fractional difference operator is defined by the Maclaurin series

$$(1-L)^d = \sum_{j=0}^{\infty} \frac{\Gamma(-d+j)L^j}{\Gamma(-d)\Gamma(j+1)} = \sum_{j=0}^{\infty} d_j L^j, \ \ d_j = \frac{(j-1-d)d_{j-1}}{j}, \ \ d_0 = 1 \quad (20)$$

and $\Gamma(.)$ is the gamma function. To avoid the initial conditions effect, sample sizes $t = 1,...,T+w$ are generated and the first $w = 1000$ observations removed. Additionally, $\sum_{j=0}^{\infty} d_j L^j$ is approximated by allowing $d_j = 0$ when $j > 1000$. In the experiments of a 1000 replications, tests at the 5% and 10% level are calculated. Again we use bandwidths $m = T^i$ $(i = 0.5, 0.6, 0.7)$, although the Ohanissian et *al*. test uses the GPH statistic from (4) rather than the GSP alternative, and we set $M = 4,5,6$ in (8) [22]. To begin with Table 2 shows the rejection frequency of the $\hat{W}$ tests when we allow[23] $d = 0.6$ and $T = 4000$.

[Insert Table 2]

[22] Following Ohanissian *et al*. (2008) for daily data (see their Table 6), we initially set $M = 4$. However, to assess the performance of higher levels of aggregation we also allow $M = 5$ and $M = 6$.
[23] Although most of the latter empirical analysis is carried out using a monthly dataset (see section 5), tests for spurious long memory typically need the greater degrees of freedom achieved at a higher frequency or daily dataset. This is because the levels of temporal aggregation required, combined with applying a typical bandwidth used in semi-parametric estimators of *d*, removes a large number of observations before estimating the test statistic. To mimic this requirement for a higher frequency dataset we set *T* = 4000.

The results above clearly show that the methodology considered typically produces a reasonably sized test statistic for the sample size and possible order of integration we will encounter in our later empirical exercise. Tests become marginally oversized when *i*, the bandwidth exponent, or *M*, the level of temporal aggregation are increased. To assess the power of the approach we simulate x_t as in (15) but now[24] with $T = 4000$, $T_{j=1} = 1000$, $T_{j=2} = 3000$ and $\eta = 0.05$. Table 3 shows the rejection frequency of the $\hat{W}$ tests.

[Insert Table 3]

Importantly Table 3 shows a marked difference in the power of the Ohanissian *et al.* test conditional on the bandwidth exponent *i*. Only a higher exponent (i.e. $i = 0.7$) produces reliable power scores and does so though producing a smaller standard deviation around the estimates of *d*, making it easier for the test procedure to distinguish between the different estimates of *d* associated with different levels of temporal aggregation. As such, in the later empirical analysis we shall place more weight on results employing this higher exponent[25].

(iv) Bootstrap coverage for the NBLS confidence interval

As noted in the methodology section, if level breaks are detected by the Bai and Perron procedure, the volatility series are demeaned accordingly, before NBLS estimation of the bi-variate regression (12). However, the NBLS confidence interval around the slope coefficient in (12), does not allow for the inherent uncertainty from the 'first-stage' structural break estimation procedure. This uncertainty derives from three sources: (i) uncertainty around the number of breaks (ii) uncertainty around the

[24] Based on an estimated AR(1) for our daily dataset.
[25] We also tried replacing the GPH statistic with the more efficient GSP alternative in the simulations. Although this produced a more powerful version of the Ohanissian *et al.* (2008) test, it also produced a test that was greatly oversized. Results available from the authors on request.

dates of the breaks and (iii) uncertainty around the magnitude and sign of the breaks. Typically, the extant literature does not account for this uncertainty and it appears difficult to account for these three sources in the variance of the 'second-stage' slope coefficient or confidence interval in a rigorous manner.

Of course, an alternative approach is to assess how robust the typical approach is to the uncertainties described. As previously, (15) and (16) are used to simulate common break series, whilst (15) and (18) are used for uncommon break series. For each of the 1000 replications, we estimate bi-variate regression (17) and calculate a bootstrap confidence interval, which itself is generated from a 1000 bootstrapped slope coefficients. The bootstrap coverage measure represents the proportion that the 95% confidence interval includes the unity coefficient.

[Insert Tables 4a, b, and c]

As can be seen from Table 4, when one break in each series exists, the coverage of the bootstrap confidence interval is often close to the nominal value specified in the second-stage. For example, Table 4(b), when $\delta_{j=k=1} = -0.15$, $T_{j=1} = 100$ and $T_{k=1} = 105$, then the coverage of the 95% bootstrap confidence interval is 0.934. In other words, the real coverage interval is reasonably close to the 95% nominal value; the uncertainty in the first-stage does not seriously affect the efficacy of the bootstrap confidence interval proposed. However, when the size and distance between the single breaks in the two series increases, (e.g., when $\delta_{j=k=1} = -0.25$, $T_{j=1} = 100$ and $T_{k=1} = 150$) the bootstrap confidence interval falls to 0.835. It should be noted that this latter value is still much higher than the simulated coverage if we do not demean (i.e. 0.531).

As might be expected, the bootstrap confidence interval typically declines when we move from modelling one break to two breaks per series. However, as long as breaks occur closely together and/or the break sizes are relatively modest, then the real confidence interval reasonably approximates the nominal value. On the other hand, one needs to be careful when interpreting results, when the distance between breaks and the break size are both relatively large. For example, in Table 4(c), when $\delta_{j=k=1} = -0.25$, $\delta_{j=k=2} = 0.25$, $T_{j=1} = 50$, $T_{k=1} = 100$, $T_{j=2} = 145$, and $T_{k=1} = 150$, then the coverage of the 95% bootstrap confidence interval declines to 0.740. Again, it should be noted that this value is still much higher than the simulated coverage if we do not demean (i.e. 0.442).

Of course, in reality, even if breaks are present, we do not know the true number of them or their magnitude. Given the extant literature suggests that the Bai and Perron procedure may detect 'spurious breaks' in the presence of long memory, we suggest a robustness-check sequential procedure for estimating the breaks and the demeaned regressions. Specifically, we suggest a specific-to-general approach, estimating the demeaned regression allowing initially for only one break in each series. Subsequently, the number of breaks allowed is then increased by 1 and the demeaned regression re-estimated. This procedure continues until we reach the number of breaks indicated by the Bai and Perron technique. This approach also seems sensible given the simulations above show that negative effects of uncertainty from the structural break 'first-stage' estimation on the 'second-stage' bootstrap confidence intervals, increases as the number of breaks to be estimated rises.

5. Empirical results

(i) No-break analysis

Table 5 shows the GSP statistics for the logarithm[26] of monthly[27] volatility series estimated using Ox version 7.0 (see Doornik, 2001). Columns 3-5 give the results for the GSP statistic with $m = T^{0.5}$, $m = T^{0.6}$ and $m = T^{0.7}$ respectively[28]. Although this type of semi-parametric approach is typical in the literature (see, *inter alia*, Nielsen, 2007), a substitute procedure would be to estimate fully parametric ARFIMA (p, d, q) models. However, given the short-run dynamics are poorly specified[29] the latter approach will be inconsistent.

[Insert Table 5]

Table 5 shows analogous results to those in the extant literature. Specifically, the GSP point estimates indicate that foreign exchange volatility is fractionally integrated with $0 < d < 1$ and presents standard errors that cannot easily distinguish between either stationary (i.e., $0 < d < 0.5$) or non-stationary (i.e., $0.5 \le d < 1$) processes. Notably, RV and IV series show comparable orders of integration.

To estimate regression (1), the possible long-run fractionally cointegrating relationship, and analogously to KDS we employ NBLS with bandwidth $z = T^{0.75}$ and a wild bootstrap procedure to generate confidence intervals. These results are produced in Table 6.

[Insert Table 6]

The results are representative of those papers using estimators that allow for long memory behaviour in volatility, providing convincing evidence of a unity slope coefficient in the implied-realized volatility relation. Out of the 12 confidence

---

[26]Christensen and Hansen (2002) show that taking natural logarithms of volatility series aids in minimising the possibility of non-normality.
[27]As in Christensen and de Magistris (2010), the monthly dataset is used in the empirical work to circumvent overlapping data problems discussed by, inter alios, Christensen and Prabhala (1998).
[28] The use of diverse bandwidths is to assess the stability of the estimated parameter to different inputs as the optimal bandwidth is typically difficult to ascertain.
[29]Recent work all employ semi-parametric estimation of the long memory parameter.

intervals presented in Table 6, only one does not include unity. Moreover, Table 7 below displays the GSP statistics $\hat{\delta}$ for the NBLS residuals.

[Insert Table 7]

The results in Table 7 imply that RV and IV are fractionally cointegrated given that the point estimate of δ is typically lower than *d*. As with KDS, $\hat{\delta}$ appears to be higher with RV^h than with RV^d and KDS posit this is *prima facie* evidence for a fractionally integrated risk premium.

(ii) Structural break analysis

As discussed earlier, uncommon structural breaks may be present in individual volatility series with time series behaviour implications for the resulting cointegrating residual. For a preliminary test, to examine whether the volatility data contains true long memory we use the test of Ohanissian *et al*. (2008) with our daily dataset; see Table 8.

[Insert Table 8]

Interestingly, Table 8 shows that the Ohanissian *et al*. test is far more likely to reject the null hypothesis of no spurious long memory when $m = T^{0.7}$. Given our simulations in section 4(iii), where this higher exponent produces far more reliable power scores, it seems appropriate to place more weight on results employing $m = T^{0.7}$. At this bandwidth, of our nine volatility series, only the US dollar/Yen IV series cannot reject true long memory and therefore it appears at least a reasonable possibility that structural breaks are present in the individual volatility series. To assess this, we next employ the Bai and Perron (1998, 2003a) estimation procedure and, switching back to our monthly dataset, Table 9 reports test statistics of structural change in the mean series of all volatility.

[Insert Table 9]

The *UD*max and *WD*max statistics provide evidence that structural breaks are clearly an important component of both RV and IV. Specifically, the $SupF_T(b+1|b)$ statistics suggest a range of 1 to 3 breaks for our volatility series. Of course, under the assumption of common breaks, the results in Table 9 should show that RV and IV within the same currency present the same number of breaks. However, in the Swiss Franc/US dollar case, RV^d contains 1 break to the 2 given by RV^h and IV. Moreover, the US dollar/Yen IV contains 3 breaks to the 2 presented by RV^h and RV^d. This provides the first evidence that level breaks are uncommon. To further investigate this point, the estimated coefficients and break points for each volatility series are reported in Table 10.

[Insert Table 10]

Table 10 reports the dates for the structural breaks in the mean level of monthly volatility series and their 90% and 95% confidence intervals for each of the break dates. The break dates correspond to the end of each regime. In addition, the average (mean) value of volatility is reported for each regime. These level breaks are also superimposed graphically on volatility figures 1 to 3. In cases where the IV and RV break relatively closely together, the point value of the breaks in the implied series never occur before or contemporaneously with those of the realized series, suggesting the market has no ability to forecast or mirror structural change. For example, the first break in US dollar/Yen RV^h occurs at observation 72 whereas a corresponding break in IV can be found at observation 76 perhaps reflecting learning period by market participants. Likewise, the second break in US dollar/Yen RV^h is at observation 109, and although closer, the break in IV is at observation 110.

Of course, one might note that for the US dollar/Yen examples given above, the magnitude of the confidence intervals do not allow for a formal statistical rejection of common breaks. However, in the case where IV traders follow movements in RV closely and react with small delays, it is clearly going to be difficult to discern uncommon structural breaks in finite samples of single exchange rates. Here it is instructive to re-emphasise that across our 'panel of exchange rates', IV never closely anticipates or co-occurs with breaks in RV. However, occasionally IV appears to break well before RV^h; for example, consider that the first break in US dollar/Sterling IV occurs at observation 35, whereas the initial break in RV^h occurs at observation 94. Amongst other reasons, one might reasonably attribute this to (i) either traders follow daily rather than high frequency RV or (ii) traders perceive the risk of a possible jump in RV^h at some point in the future and the current price of IV reflects a type of peso problem. In any case, the weight of evidence in Tables 5 and 6 points towards uncommon breaks across RV and IV.

Now we assess whether allowing for the estimated level breaks reveals spurious long memory in our volatility series. Specifically, each series is demeaned employing the estimate μ_j from the OLS regression of (9) on the estimated break points $(T_{1,\ldots,}T_m)$ according to the method followed in Coakley *et al.* (2011) amongst others. Table 11 reports the GSP estimates of the integration order for break-free volatility series, that is estimates of *d* for the series. $\sigma^*_{t+\tau} = \sigma_{t+\tau} - \hat{\alpha}_j$.

[Insert Table 11]

Allowing for multiple structural mean-breaks clearly accounts for at least some, if not all, long memory behaviour in foreign exchange RV and IV. For example, in the case of US dollar/Sterling IV when $m = T^{0.5}$, the GSP point estimate $\hat{d}$ has fallen from

0.638 to 0.023 when the series are demeaned. Overall, in Table 11, the values of $\hat{d}$ are lower than those presented in Table 5, and typically now found to be either zero or in the covariance stationary (i.e., $0 < d < 0.5$) region, emphasising the importance of allowing for level breaks when assessing the time series properties of volatility.

For the next test, we examine the order of integration of the NBLS residual from demeaned regression (12) and present the results in Table 12.

[Insert Table 12]

We now find an order of integration close to or zero in many cases, greatly weakening the evidence for a fractionally integrated risk premium. For example, in Table 12, the GSP estimate for the demeaned, high frequency US dollar/Yen (bandwidth $m = T^{0.7}$) is -0.048. Contrastingly, the non-demeaned equivalent in Table 7 is 0.216 and significantly different from zero. Clearly, empirically modelling uncommon structural breaks, as suggested in the simulations of section 4(ii), not only removes persistence from the individual volatility series but also from the bivariate volatility regression residual. Prior findings of fractional cointegration in the literature may well be predominantly spurious.

For the final analysis, Table 13 shows the estimated coefficients and bootstrapped confidence intervals obtained when applying NBLS to (12).

[Insert Table 13]

When breaks are accounted for in the bi-variate framework, the point estimate of the slope parameter always falls and the null of unbiasedness can now be rejected in 4 out of 6 cases. It would appear that by identifying and then modelling structural breaks, the implied-realized volatility puzzle re-emerges.

(iii) Robustness-check sequential procedure

The simulations in section 4(iv) suggest the negative effect of uncertainty in the 'first-stage' break estimation on the bootstrap confidence interval, increases as the number of breaks to be estimated and the magnitude of those breaks both rise. As such, we propose a specific-to-general approach, estimating demeaned regression (12) allowing initially for only one break in each series. Subsequently, the number of breaks allowed is then increased by one and the demeaned regression re-estimated. This procedure continues until we reach the number of breaks indicated by the Bai and Perron technique.

[Insert Table 14]

Table 14 shows[30] the one break analog of Table 13. Clearly, the results for the US dollar/Sterling series in Table 14 are the same as in Table 13; the individual series only presented one break each during our prior analysis. However, the parameter estimates shown for the other two exchange rates, now restricted to one break, are consequently different. However, a similar interpretation can be placed on these new results - when a single break in each series is accounted for in the demeaned bi-variate framework, the point estimate of the slope parameter is less than unity and the null of unbiasedness is rejected in the majority (i.e., 4 out of 6) cases. This finding provides more support to the prior reinstatement of the implied-realized volatility puzzle given the use of more reliable bootstrapped confidence intervals.

[Insert Table 15]

For completeness, the two break maximum version of Table 13 is provided by Table 15 – again, a similar story is told. It should also be noted that in every case where the realized volatility is measured by the more accurate high frequency data (as opposed

[30] The tables containing the estimated structural break, order of integration of the demeaned series and residual from (12) under the one-break constraint are omitted to save space. However, they are available from the authors on request.

to daily) the unbiasedness hypothesis is rejected, often severely, by the bootstrap confidence intervals.

6. Conclusions

Recent literature has suggested that employing methodologies that allow for fractionally integrated behaviour in individual series render implied volatility an unbiased forecast of realized volatility. This paper extends this branch of literature in a number of ways. First, we conduct Monte Carlo experiments which reveal that uncommon structural breaks can spuriously cause the finding of fractional cointegration often found in this literature. Second, we test for uncommon multiple breaks in the mean levels of foreign exchange volatility series. Third, we explicitly examine the time series properties of break-free individual volatility series. Fourthly, we show via simulation, that confidence intervals for the bivariate realized-implied volatility regression become less reliable as the number and magnitude of breaks to be estimated rises. Consequently, we suggest a specific-to-general approach to estimating the break-free regression.

Using data on three currencies for the period 1991-2007, formal structural break procedures and spurious long memory tests both suggest that RV and IV contain structural breaks in mean. Interestingly, the breaks in the implied series never closely anticipate or co-break with those of the realized series, suggesting the market has no ability to forecast or mirror structural change. Moreover, occasionally implied volatility appears to break well before realized volatility. Amongst other reasons, one might reasonably attribute this is that traders perceive the risk of a possible jump in realized volatility at some point in the future and the current price of IV reflects a type of peso problem. It also suggests that implied volatility does not always adjust to realized volatility and, in fact, may not adjust for many time periods. Such results could not be seen if we forced breaks to be common.

When we allow for these structural breaks, this largely removes the persistence from both individual volatility series and the residuals, challenging the notion that realized volatility-implied volatility is a fractionally cointegrated relation, a result also suggested by our prior simulations. Furthermore, using the proposed specific-to-general approach within the bi-variate framework, the point estimate of the slope parameter falls away from unity and the null of unbiasedness is often rejected. In summary, allowing for uncommon structural breaks suggests the implied-realized volatility puzzle might not be solved after all and that implied volatility may not act like an efficient forecast.

References

Almeida, C., and J. Vicente. "Identifying volatility risk premia from fixed income Asian options." *Journal of Banking and Finance*, 33 (2009), 651-661.

Anderson, T.G., Bollerslev, T., Diebold, F., and P. Labys. "The distribution of realized exchange rate volatility." *Journal of American Statistical Association*, 96 (2001a), 42-55.

Anderson, T.G., Bollerslev, T., Diebold, F., and H. Ebens. "The distribution of realized stock return volatility." *Journal of Financial Economics*, 61 (2001b), 43-76.

Bai, J., and P. Perron. "Estimating and testing linear models with multiple structural changes." *Econometrica*, 66 (1998), 47-68.

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

Bai, J., and P. Perron. "Critical values in multiple structural change tests." *Econometrics Journal*, 6 (2003b), 72-78.

Bai, J., and P. Perron. 2004. “Multiple structural change models: A simulation study.” In: Corbea D, Durlauf S and Hansen B (Eds), Econometric essays, Cambridge University Press: Cambridge; England.

Bandi, F., and B. Perron. “Long memory and the relation between implied and realized volatility.” *Journal of Financial Econometrics*, 4 (2006), 636-670.

Bank of International Settlement. Statistical Annex. *BIS Quarterly Review*, December 2012.

Chambers, M. “Long Memory and Aggregation in Macroeconomic Time Series,” *International Economic Review*, 39 (1998), 1053–1072.

Chernov, M. “On the role of risk premia in volatility forecasting.” *Journal of Business and Economics Statistics*, 25 (2007), 411-426.

Choi K, and E. Zivot. “Long memory and structural changes in the forward discount: An empirical investigation.” *Journal of International Money and Finance*, 26 (2007), 342-363.

Choi, K., Yu, W. C., and E. Zivot. “Long memory versus structural breaks in modelling and forecasting realized volatility.” *Journal of International Money and Finance*, 29 (2010), 857-875.

Christensen, B., and C. Hansen. “New evidence on the implied-volatility relation.” *European Journal of Finance*, 8 (2002), 187-205.

Christensen, B., and P.S. de Magistris. “Level Shifts in Volatility and the Implied-Realized Volatility Relation.” *CREATES Research Paper*, 2010-60.

Christensen, B., and M. Nielson. “Asymptotic normality of narrow-band least squares in the stationary fractional cointegration model and volatility forecasting.” *Journal of Econometrics*, 133 (2006), 343-371.

Christensen B., Hansen, C., and N. Prabhala. “The telescoping overlap problem in options data.” (2001), University of Maryland.

Christensen, B., and N. Prabhala. “The relation between implied and realized volatility.” *Journal of Financial Economics*, 50 (1998), 125-150.

Coakley, J., Dollery. J., and N. Kellard. “Long Memory and Structural Breaks in Commodity Futures Markets”, *Journal of Futures Markets*, 31 (2011), 1076-1113.

Covrig, V. and Low, B.S. “The Quality of Volatility Traded on the Over-The-Counter Currency Market: A Multiple Horizons Study.” *The Journal of Futures Markets*. 23 (2003), 261–285.

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

Doornik, J. “Object-oriented Matrix Programming using Ox”. Timberlake Consultants Press, (2001), London.

Doran, J.S., and E.I. Ronn. “Computing the market price of volatility risk in the energy commodity markets”. *Journal of Banking and Finance*, 32 (2008), 2541–2552.

Dunis, C., and A. Keller. “Efficiency tests with overlapping data: An application to the currency options market.” *European Journal of Finance*, 4 (1995), 345-366.

Dunis, C., and X. Huang. “Forecasting and trading currency volatility: An application of recurrent neural regression and model combination.” *Journal of Forecasting*, 21 (2002), 317-354.

Geweke, J., and S. Porter-Hudak. “The estimation and application of long memory time series models.” *Journal of Time Series Analysis*, 4 (1983), 221-238.

Garvey, J.F., and L.A. Gallagher,. “The realised–implied volatility relationship: Recent empirical evidence from FTSE-100 stocks.” *Journal of Forecasting*, 31 (2012), 639-660.

Gerolimetto, M. “Frequency domain bootstrap for the fractional cointegration regression.” *Economics Letters* 91 (2006), 389-394.

Granger C, and N. Hyung. “Occasional structural breaks and long memory with an application to the S&P500 absolute stock returns.” *Journal of Empirical Finance*, 11 (2004), 399-421.

Kellard, N. “On the robustness of cointegration tests for market efficiency.” *Finance Research Letters*, 3 (2006), 57-64.

Kellard, N., and N. Sarantis, N. “Can exchange rate volatility explain persistence in the forward premium?” *Journal of Empirical Finance*, 15 (2008), 714-728.

Kellard, N., Dunis, C. and N. Sarantis. “Foreign Exchange, Fractional Cointegration and the Implied-Realized Volatility Relation,” *Journal of Banking and Finance*, 34 (2010), 882-891.

Li, Y. and P. Perron. “Modeling exchange rate volatility with random level shifts.” (2013), Boston University.

Marinucci, D., and P.M. Robinson. “Semiparametric fractional cointegration analysis.” *Journal of Econometrics*, 105 (2001), 225-247.

Maynard, A., and P.C.B. Phillips. "Rethinking an old empirical puzzle: Econometric evidence on the forward discount anomaly." *Journal of Applied Econometrics*, 16 (2001), 671-708.

Mikosch, T., and C. Starica. "Change of structure in financial time series, long range dependence and the GARCH Model." Working paper (1999).

Neely, C. "Forecasting foreign exchange volatility: Why is implied volatility biased and inefficient? And does it matter?" *Journal of International Financial Markets, Institutions and Money*, 19 (2009), 188-205.

Nielsen, M. "Local whittle analysis of stationary fractional cointegration and the implied-realized volatility relation". *Journal of Business and Economics Statistics*, 25 (2007), 427–446.

Ohanissian, A., Russell, J. and R. Tsay. "True or Spurious Long Memory? A New Test." *Journal of Business and Economic Statistics*, 26 (2008), 161-175.

Poteshman, A. "Forecasting future volatility from option prices." *Mimeo*, (2000), University of Illinois at Urbana-Champaign.

Qu, Z. "A test against spurious long memory." *Journal of Business and Economic Statistics*, 29 (2011), 423-438.

Robinson, P. M. "Log-periodogram regression of time series with long range dependence." *Annals of Statistics*, 23 (1995a), 1048-1072.

Robinson, P. M. "Gaussian semi-parametric estimation of long range dependence." *Annals of Statistics*, 23 (1995b), 1630-1661.

Robinson, P.M., Marinucci, D.,. Semiparametric frequency domain analysis of fractional cointegration. In: Robinson, P.M. (Ed.), Time Series with Long Memory. Oxford University Press: Oxford (2003).

Sarantis, N. "Testing uncovered interest rate parity using traded volatility, a time varying risk premium and heterogeneous expectations." *Journal of International, Money and Finance*, 25 (2006), 1168-1186.

Velasco, C. "Non-stationary log-periodogram regression." *Journal of Econometrics*, 91 (1999a), 325-371.

Velasco, C. "Gaussian semiparametric estimation of non-stationary time series." *Journal of Time Series Analysis*, 20 (1999b), 87-127.

Velasco, C. "Gaussian semiparametric estimation of fractional cointegration." *Journal of Time Series Analysis*, 24 (2003), 345-378.

Table 1: Monte Carlo Experiments

	$\bar{d}_{GSP}(y_t)$	$\bar{d}_{GSP}(x_t)$	$\bar{d}_{GSP}(\hat{u}_t)$
Common Breaks ($m = T^{0.5}$)	0.548	0.568	-0.046
Common Breaks ($m = T^{0.6}$)	0.507	0.535	-0.030
Common Breaks ($m = T^{0.7}$)	0.500	0.547	-0.022
Uncommon Breaks ($m = T^{0.5}$)	0.583	0.568	0.631
Uncommon Breaks ($m = T^{0.6}$)	0.515	0.535	0.491
Uncommon Breaks ($m = T^{0.7}$)	0.482	0.547	0.381

Table 2: Size of the $\hat{W}$ Tests

i	0.5			0.6			0.7		
M	4	5	6	4	5	6	4	5	6
5%	0.052	0.070	0.080	0.058	0.070	0.096	0.078	0.096	0.096
10%	0.106	0.126	0.152	0.124	0.144	0.156	0.130	0.164	0.168

Table 3: Power of the $\hat{W}$ Tests

i	0.5			0.6			0.7		
M	4	5	6	4	5	6	4	5	6
5%	0.077	0.057	0.046	0.512	0.461	0.411	0.803	0.788	0.777
10%	0.154	0.122	0.093	0.662	0.600	0.556	0.892	0.889	0.882

Table 4: Bootstrap Coverage Measure

(a) Common single break

$\delta_{j=k=1}$	-0.05	-0.15	-0.25
$T_{j=1}$	100	100	100
De-mean	0.946	0.946	0.938
Non de-mean	0.954	0.955	0.949

(b) Uncommon single break

$\delta_{j=k=1}$	-0.05		-0.15		-0.25	
$T_{j=1}$, $T_{k=1}$	100, 105	100, 150	100, 105	100, 150	100, 105	100, 150
De-mean	0.946	0.939	0.934	0.876	0.906	0.835
Non de-mean	0.958	0.949	0.947	0.827	0.924	0.531

(c) Uncommon double break

$\delta_{j=k=1}$, $\delta_{j=k=2}$	-0.05, 0.05		-0.15, 0.15		-0.25, 0.25	
$T_{j=1}$, $T_{k=1}$, $T_{j=2}$, $T_{k=2}$	50, 55 145, 150	50, 100 145, 150	50, 55 145, 150	50, 100 145, 150	50, 55 145, 150	50, 100 145, 150
De-mean	0.927	0.906	0.856	0.823	0.818	0.740
Non de-mean	0.959	0.937	0.927	0.763	0.878	0.442

Table 5: GSP Tests for the *d* of Individual Volatility Series

	$\hat{d}$	GSP $m = T^{0.5}$	GSP $m = T^{0.6}$	GSP $m = T^{0.7}$
UK£/US$	RV^d	0.467 (0.134)	0.424 (0.104)	0.542 (0.079)
	IV	0.638 (0.134)	0.583 (0.104)	0.625 (0.079)
	RV^h	0.582 (0.134)	0.526 (0.104)	0.564 (0.079)
US$/Yen	RV^d	0.282 (0.134)	0.421 (0.104)	0.281 (0.079)
	IV	0.542 (0.134)	0.545 (0.104)	0.539 (0.079)
	RV^h	0.501 (0.134)	0.493 (0.104)	0.496 (0.079)
SF/US$	RV^d	0.418 (0.134)	0.334 (0.104)	0.483 (0.079)
	IV	0.543 (0.134)	0.522 (0.104)	0.566 (0.079)
	RV^h	0.560 (0.134)	0.518 (0.104)	0.575 (0.079)

Note: numbers in parentheses beneath the estimates for d are the standard errors σ_d. RV^h and RV^d are the measures of realized volatility generated by high frequency or daily data respectively.

Table 6: NBLS estimates of (1)

		$\hat{\alpha}$	$\hat{\beta}$	95% CI for $\hat{\beta}$	90% CI for $\hat{\beta}$
US$/UK£	RV^d	-0.013	1.048	[0.827 -1.277]	[0.859 -1.247]
	RV^h	-0.654	0.702	[0.389 -1.035]	[0.434 -0.992]
US$/Yen	RV^d	-0.194	0.957	[0.792, 1.130]	[0.817, 1.107]
	RV^h	-0.079	0.932	[0.799, 1.069]	[0.815, 1.051]
SF/US$	RV^d	-0.106	0.987	[0.820, 1.151]	[0.848, 1.124]
	RV^h	-0.121	0.914	[0.684, 1.135]	[0.710, 1.108]

Note: The results in Table 3 above were originally shown in KDS and are reproduced for ease of comparison with later estimations in the current paper.

Table 7: GSP Tests for the Integration Order of the Residuals in (1)

		GSP $m = T^{0.5}$	GSP $m = T^{0.6}$	GSP $m = T^{0.7}$
US$/UK£	RV^d	0.322 (0.134)	0.193 (0.104)	0.210 (0.079)
	RV^h	0.667 (0.134)	0.594 (0.104)	0.453 (0.079)
US$/Yen	RV^d	-0.026 (0.134)	0.107 (0.104)	-0.035 (0.079)
	RV^h	0.270 (0.134)	0.282 (0.104)	0.216 (0.079)
SF/US$	RV^d	0.128 (0.134)	0.057 (0.104)	0.126 (0.079)
	RV^h	0.439 (0.134)	0.392 (0.104)	0.387 (0.079)

Table 8: Ohanissian *et al*. (2008) Test

		$\hat{W}$ $m = T^{0.5}$	$\hat{W}$ $m = T^{0.6}$	$\hat{W}$ $m = T^{0.7}$
UK£/US$	RV^d	0.578, 0.269, 0.300	**0.038, 0.075, 0.078**	**0.000, 0.000, 0.000**
	IV	0.805, **0.058, 0.017**	0.705, 0.837, 0.918	**0.004, 0.011, 0.021**
	RV^h	0.202, 0.159, **0.061**	**0.000, 0.000, 0.000**	**0.000, 0.000, 0.000**
US$/Yen	RV^d	0.200, 0.144, 0.183	0.113, **0.099**, 0.150	**0.000, 0.000, 0.000**
	IV	0.323, 0.383, 0.334	0.641, 0.700, 0.820	0.263, 0.408, 0.544
	RV^h	0.837, 0.910, 0.948	**0.000, 0.002, 0.003**	**0.000**, **0.000**, **0.000**
SF/US$	RV^d	0.907, 0.806, 0.895	**0.063, 0.022, 0.036**	**0.000**, **0.000**, **0.000**
	IV	0.229, 0.310, **0.065**	0.172, 0.113, 0.181	**0.020, 0.018, 0.028**
	RV^h	0.160, 0.262, 0.333	**0.001, 0.002, 0.003**	**0.000**, **0.000**, **0.000**

Note: the numbers are p-values. Significance at the 10% level or less are highlighted. Each cell contains three p-values with $M = 4$, $M = 5$ and $M = 6$ from equation (8) respectively.

Table 9: Bai and Perron statistics for tests of multiple structural breaks in monthly volatility series

Series	UDmax	WDmax (5%)	$F(1\|0)$	$F(2\|1)$	$F(3\|2)$	$F(4\|3)$	$F(5\|4)$
US$/UK£ IV	104.18***	104.18**	104.18***	3.90	1.31	0.44	0.00
US$/UK£ RV^d	52.16***	52.16**	52.16***	2.71	1.25	1.60	-
US$/UK£ RV^h	34.18***	45.63**	34.18***	5.87	3.77	5.64	1.00
US$/Yen IV	15.95***	22.28**	12.22**	10.48**	14.29**	1.84	0.00
US$/Yen RV^d	19.70***	23.42**	7.75*	16.30***	1.35	0.22	0.00
US$/Yen RV^h	25.23***	29.98**	19.69***	12.05**	5.13	5.13	0.00
SF/US$ IV	17.22***	26.93**	17.22***	15.73***	2.11	3.10	3.10
SF/US$ RV^d	20.37***	24.21**	20.19***	5.98	1.48	1.50	1.50
SF/US$ RV^h	27.87***	33.12**	18.45***	10.33**	3.85	1.29	0.22

Note: ***, **, * indicate significance at the 1, 5 and 10 percent levels respectively.

Table 10: Bai and Perron Regime Means and End Dates

Series	Regime 1	Regime 2	Regime 3	Regime 4
US$/UK£ IV	-2.045 (0.036)	-2.527 (0.031)	-	-
End date	(35)			
90% CI	[(28), (36)]			
95% CI	[(26), (37)]			
US$/UK£ RV^d	-2.134 (0.061)	-2.656 (0.038)	-	-
End date	(32)			
90% CI	[(23), (35)]			
95% CI	[(20), (37)]			
US$/UK£ RV^h	-2.201 (0.040)	-2.519 (0.036)	-	-
End date	(94)			
90% CI	[(82), (106)]			
95% CI	[(78), (111)]			
US$/Yen IV	-2.273 (0.048)	-1.946 (0.074)	-2.332 (0.020)	-2.493 (0.037)
End date	(76)	110	165	
90% CI	[(61),(89)]	[(108),(173)]	[(149),(171)]	
95% CI	[(55),(94)]	[(108),(173)]	[(143),(173)]	
US$/Yen RV^d	-2.391 (0.059)	-2.062 (0.056)	-2.471 (0.033)	
End date	(72)	109		
90% CI	[(64), (91)]	[(103), (115)]		
95% CI	[(60), (99)]	[(101), (117)]		

Note: The first number in each cell is the estimated mean for the regime; standard errors are reported in parentheses. The end date for each regime, in terms of observation number, is shown below the estimated mean.

Table 10: Bai and Perron Regime Means and End Dates Continued

Series	Regime 1	Regime 2	Regime 3	Regime 4
US$/Yen RV^h	-2.162 (0.047)	-1.885 (0.064)	-2.348 (0.027)	
End date	(72)	109		
90% CI	[(56), (89)]	[(106), (115)]		
95% CI	[(49), (96)]	[(105), (118)]		
SF/US$ IV	-2.043 (0.047)	-2.248 (0.021)	- 2.518 (0.074)	-
End date	(56)	169		
90% CI	[(47), (81)]	[(149), (173)]		
95% CI	[(43), (91)]	[(142), (176)]		
SF/US$ RV^d	-2.107 (0.050)	-2.384 (0.036)	-	-
End date	(55)			
90% CI	[(35), (69)]			
95% CI	[(27), (75)]			
SF/US$ RV^h	-2.044 (0.026)	-2.249 (0.0399)	-2.482 (0.059)	-
End date	(124)	(165)		
90% CI	[(111), (141)]	[(146), (174)]		
95% CI	[(106), (148)]	[(139), (178)]		

Note: The first number in each cell is the estimated mean for the regime; standard errors are reported in parentheses. The end date for each regime, in terms of observation number, is shown below the estimated mean.

Table 11: GSP Tests of *d* for Break-Free Individual Volatility Series

	$\hat{d}$	GSP $m = T^{0.5}$	GSP $m = T^{0.6}$	GSP $m = T^{0.7}$
UK£/US$	RV^d	0.052 (0.134)	0.160 (0.104)	0.260 (0.079)
	IV	0.023 (0.134)	0.101 (0.104)	0.270 (0.079)
	RV^h	0.299 (0.134)	0.311 (0.104)	0.393 (0.079)
US$/Yen	RV^d	0.036 (0.134)	0.213 (0.104)	0.130 (0.079)
	IV	-0.138 (0.134)	0.103 (0.104)	0.155 (0.079)
	RV^h	-0.037 (0.134)	0.105 (0.104)	0.188 (0.079)
SF/US$	RV^d	0.289 (0.134)	0.172 (0.104)	0.268 (0.079)
	IV	-0.095 (0.134)	-0.064 (0.104)	0.072 (0.079)
	RV^h	0.036 (0.134)	0.195 (0.104)	0.164 (0.079)

Table 12: GSP Tests for the Integration Order of the Residuals in Demeaned (12)

		GSP $m = T^{0.5}$	GSP $m = T^{0.6}$	GSP $m = T^{0.7}$
US$/UK£	RV^d	0.123 (0.134)	0.063 (0.104)	-0.035 (0.079)
	RV^h	0.254 (0.134)	0.408 (0.104)	0.169 (0.079)
US$/Yen	RV^d	0.150 (0.134)	-0.011 (0.104)	-0.023 (0.079)
	RV^h	0.032 (0.134)	-0.116 (0.104)	-0.048 (0.079)
SF/US$	RV^d	0.358 (0.134)	0.181 (0.104)	0.128 (0.079)
	RV^h	0.138 (0.134)	0.170 (0.104)	0.108 (0.079)

Table 13: NBLS Estimates of Demeaned (12) – Unrestricted Breaks

		$\hat{\alpha}$	$\hat{\beta}$	95% CI for $\hat{\beta}$	90% CI for $\hat{\beta}$
US$/UK£	RV^d	-0.002	0.884	[0.661, 1.102]	[0.691, 1.070]
	RV^h	0.008	0.633	[0.497, 0.768]	[0.524, 0.746]
US$/Yen	RV^d	0.001	0.711	[0.466, 0.963]	[0.498, 0.922]
	RV^h	-0.0002	0.609	[0.428, 0.789]	[0.450, 0.756]
SF/US$	RV^d	0.008	0.834	[0.625, 1.041]	[0.652, 1.007]
	RV^h	-0.017	0.293	[0.008, 0.583]	[0.043, 0.540]

Table 14: NBLS Estimates of Demeaned (12) - One Break Maximum

		$\hat{\alpha}$	$\hat{\beta}$	95% CI for $\hat{\beta}$	90% CI for $\hat{\beta}$
US$/UK£	RV^d	-0.002	0.884	[0.661, 1.102]	[0.691, 1.070]
	RV^h	0.008	0.633	[0.497, 0.768]	[0.524, 0.746]
US$/Yen	RV^d	-0.002	0.834	[0.592, 1.052]	[0.631, 1.025]
	RV^h	-0.003	0.646	[0.450, 0.866]	[0.472, 0.842]
SF/US$	RV^d	-0.007	0.412	[0.054, 0.773]	[0.102, 0.733]
	RV^h	-0.007	0.401	[0.237, 0.572]	[0.263, 0.546]

Table 15: NBLS Estimates of Demeaned (12) – Two Breaks Maximum

		$\hat{\alpha}$	$\hat{\beta}$	95% CI for $\hat{\beta}$	90% CI for $\hat{\beta}$
US$/UK£	RV^d	-0.002	0.884	[0.661, 1.102]	[0.691, 1.070]
	RV^h	0.008	0.633	[0.497, 0.768]	[0.524, 0.746]
US$/Yen	RV^d	-0.012	0.442	[0.170, 0.719]	[0.213, 0.682]
	RV^h	-0.013	0.336	[0.127, 0.558]	[0.163, 0.525]
SF/US$	RV^d	0.008	0.834	[0.625, 1.041]	[0.652, 1.007]
	RV^h	-0.017	0.293	[0.008, 0.583]	[0.043, 0.540]

Figure 1: UK£/US$ Implied Volatility

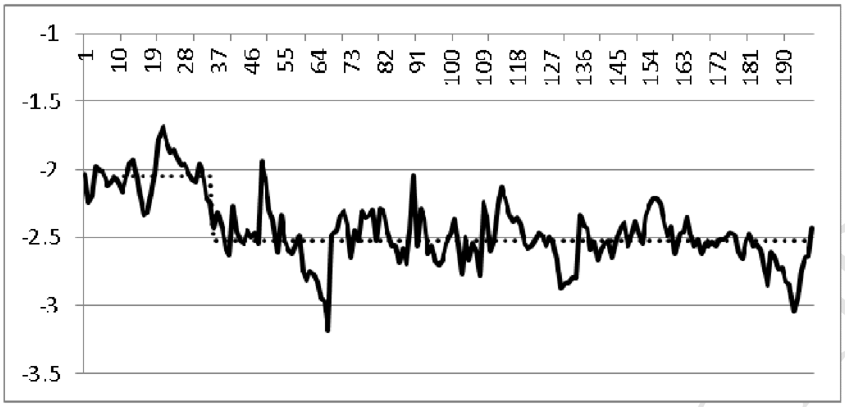

Figure 2: UK£/US$ Realized Volatility - Daily Data

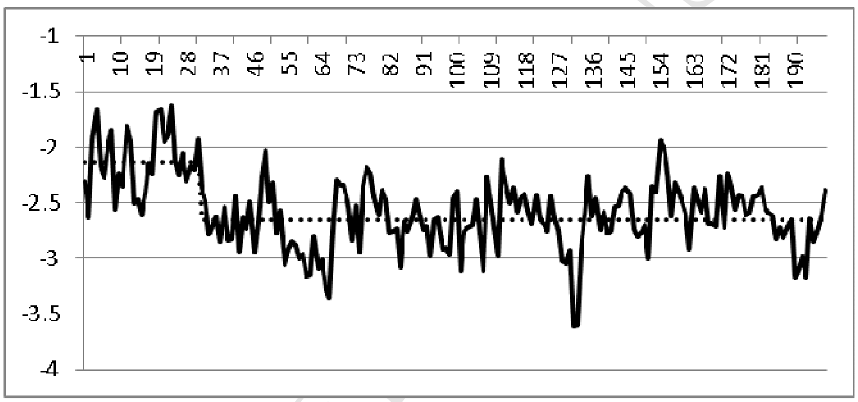

Figure 3: UK£/US$ Realized Volatility – High Frequency Data