Time Variation in the Standard Forward Premium Regression: Some new Models and Tests*

Richard T. Baillie[a,b,c,†] Dooyeon Cho[d,‡]

[a]Departments of Economics and Finance, Michigan State University, USA

[b]School of Economics and Finance, Queen Mary, University of London, UK

[c]Rimini Center for Economic Analysis, Italy

[d]Department of Economics, Kookmin University, Seoul, Korea

*The authors are very grateful for helpful comments from two anonymous referees, and from the editor Menelaos Karanasos, and also seminar participants at the Bank of Korea, Hanyang University, KAIST, Michigan State University, University of Liverpool, and QMUL. Any remaining errors are our responsibility.

†Department of Economics and Department of Finance, Michigan State University, East Lansing, MI 48824-1024, USA; Tel: +1 517 355 1864, Fax: +1 517 432 1068, E-mail: baillie@msu.edu

‡Corresponding author. Department of Economics, Kookmin University, Seoul 136-702, Korea; Tel: +82 2 910 5617, Fax: +82 2 910 4519, Email: dcho@kookmin.ac.kr

Time Variation in the Standard Forward Premium Regression: Some new Models and Tests

Abstract

This paper makes two contributions to trying to understand the forward premium anomaly and the apparent breakdowns of Uncovered Interest Rate Parity (*UIP*). First, investigation of the time series properties of the forward premium reveals either four or five breaks in the last twenty three years and evidence of long memory within each sub period. In fact the forward premium is highly nonlinear and appears to defy classification as a process with a constant order of integration. The second aspect of the paper is concerned with the time varying nature of the estimate of the slope parameter when spot returns are regressed on the lagged forward premium. We compare rolling type regression estimates, with Bayesian estimation and also a new Time Varying Parameter (*TVP*) method that is motivated by the *TVP* autoregression of Giraitis et al. (2014). The procedure is a form of kernel weighted regression and delivers relatively tight standard errors on the parameter estimates. We find the existence of the forward premium anomaly with large negative beta coefficients in the 1980s and 1990s. For some currencies there is also evidence of large positive coefficients and a reversal of the forward premium anomaly after the financial crisis of 2008.

JEL Classification: C22; F31; G01; G15

Keywords: Forward premium anomaly; Time-varying parameter regression

1 Introduction

A major issue in international finance has been the apparent failure of the theory of Uncovered Interest Rate Parity (UIP). The usual method for testing UIP is to estimate the slope coefficient in a regression of spot returns on the lagged forward premium, or equivalently, the lagged interest rate differential. While the slope coefficient should be unity under UIP, most studies have found statistically significant rejections of the UIP hypothesis, with the slope coefficient estimate invariably being quite large and negative. This has become known as the forward premium anomaly; and research to account for the anomaly has focused on (i) the presence of time dependent risk premium, (ii) irrational agents in segmented markets, (iii) peso problems, and (iv) econometric issues with the testing of UIP. The dominant approach has been to explain the phenomenon by modeling a time dependent risk premium, and overall this approach has not been particularly successful empirically.[1] Other research has simply considering the empirical relationship between currency returns and lagged interest rate differentials. For example, see Bansal (1997), Bansal and Dahlquist (2000) and Lothian and Wu (2011).

The first contribution of this paper is to consider the evidence on the extent of nonlinearity in the forward premium series for eight different currencies in the post Bretton Woods era. We present a range of tests for structural breaks and use robust Local Whittle estimates of the long memory parameter in the various sub periods. The evidence clearly indicates the futility of trying to classify forward premium series as being stable, time invariant processes; such as fractionally integrated processes with constant, time invariant, orders of integration. Rather, the evidence strongly suggests that the order of fractional integration substantially changes over the sub periods.

The second contribution of this paper is in terms of estimating the relationship between spot returns and the lagged forward premium. Previous work by Bansal (1997) and Baillie and Kiliç (2006) have allowed for the slope parameter, beta, in the forward premium regression to be time varying and to move between two regimes. While, Baillie and Bollerslev (2000), Lothian and Wu (2011) use rolling regressions with the beta continuously being updated by including a new observation and discarding the earliest observation. The rolling regression method

[1] In the carry trade literature, FX market participants actively speculate against UIP and attempt to exploit the forward premium anomaly. Hence investors buy currencies associated with relatively high interest rate countries and exploit the forward premium anomaly in order to make a short term profit. This is apparently a high risk strategy before the inevitable bursting of the bubble.

is inevitably quite arbitrary; and this paper implements an alternative, new method for time varying parameter (TVP) in a kernel weighted regression. The method is a direct extension of the TVP autoregressive approach due to Giraitis et al. (2014). The method compares favorably with rolling regression, and also Bayesian estimation where the slope coefficient follows a random walk. The new TVP estimation procedure delivers reduced confidence intervals compared with the rolling regressions. We find evidence of large negative beta coefficients in the 1980s and 1990s; and for many currencies, there is also evidence of large positive coefficients and a reversal of the forward premium anomaly for some of the sample, particularly after the financial crisis of 2008. Hence the widely cited meta study of Froot and Thaler (1990), who find the average value of the slope coefficient to be -0.88 is not representative across the whole realization in the last twenty three years. These findings cast doubt on the approach of assessing the validity of calibrated, or simulated models, by checking to see if they generate a negative correlation between spot returns and lagged interest rate differentials.

The plan of the rest of the paper is as follows: the next section notes some issues concerning the specification of the forward premium regressions, while section 3 presents econometric evidence on discontinuities in forward premium time series and how these series defy being classified as being generated by processes with constant orders of integration. This issue is particularly relevant for the estimation of the forward premium regressions in sub regimes. Section 4 then considers various approaches at time varying parameter (TVP) methods for estimating the parameters in the forward premium regression. The rolling regression and Bayesian approach assuming a random walk for the coefficient are used; but the primary focus is on the TVP kernel weighted regression. This method has some distinct advantages in terms of producing smooth estimates of the slope coefficient and also in terms of generating relatively tight confidence intervals around the point estimates. The paper then ends with a brief section of concluding remarks.

2 Model set up

There have been numerous studies that have regressed currency returns on lagged interest rate differentials, or equivalently the forward premium. Following Fama (1984) it has become standard to test the theory from the regression equation

$$\Delta s_{t+1} = \alpha + \beta(f_t - s_t) + u_{t+1}, \tag{1}$$

where the theory of UIP implies $\alpha = 0$, $\beta = 1$ and u_{t+1} being serially uncorrelated. The forward premium anomaly of a negative slope coefficient has generally occurred for most freely floating currencies and appears robust to the choice of numeraire currency. This appeared to motivate the analysis of Fama (1984). In the widely cited study by Froot and Thaler (1990), it is found that the average value of the estimated β across 75 studies was -0.88.

The usual economic theory behind the UIP hypothesis is based on the standard discrete time, consumption based asset pricing model of Lucas (1978), which uses a risk adjusted version of the real returns from going long or short in the currency market. Then, for the representative investor,

$$E_t\left(\frac{S_{t+1} - F_t}{P_{t+1}}\right)\left(\frac{U'(C_{t+1})}{U'(C_t)}\right) = 0, \tag{2}$$

where S_{t+1} is the spot exchange rate at time $t+1$ which is defined as number of units of foreign currency per one US dollar, and F_t represents the corresponding one period ahead forward rate, P_{t+1} is the domestic price level and E_t is the conditional expectation based on a sigma field of information available at time t. Furthermore, $U'(C_t)$ is the marginal utility of consumption in period t. Then,

$$E_t(\Delta s_{t+1}) = (f_t - s_t) - \left(\frac{1}{2}\right) Var_t(\Delta s_{t+1}) + Cov_t\left(\Delta s_{t+1} p_{t+1}\right) + \rho_t, \tag{3}$$

where ρ_t is the conditional covariance at time t of the spot returns with the natural logarithm of the intertemporal marginal rate of substitution and is generally called the "risk premium". The above theory dates back at least to Hansen and Hodrick (1983), and further explanations are available in Hodrick (1987) and Engel (1996). Hence expressing the theory of UIP under rational expectations and a constant risk premium as

$$E_t(\Delta s_{t+1}) = (f_t - s_t) = (i_t^* - i_t) \tag{4}$$

is always an approximation which neglects the Jensen inequality terms, and possible time dependent risk premium. While possible peso problems, segmented markets, and heterogenous trading behavior have all been suggested as resolutions of the anomaly, the presence of time-dependent risk premia has generally seemed the most persuasive explanation. In particular, see Hodrick (1987, 1989) and Mark and Wu (1997). It is worth noting that other formulations of UIP, such as deriving it from $\frac{E_t S_{t+1}}{S_t} = \frac{(1+i_t^*)}{(1+i_t)}$ can also be used to derive formulations such as $E_t \Delta \ln(S_{t+1}) \approx (i_t^* - i_t)$. While, Chang (2013) has an alternative formulation which assumes that agents diversify in stocks and bonds, and finds that the slope coefficient depends on equity differential across countries.

3 Discontinuities in the Forward Premium

One set of explanations for the existence of the anomaly focuses on econometric issues. Estimation of Equation (1) involves a classic problem of regressing the very volatile, virtually uncorrelated spot returns on the very persistent, highly autocorrelated forward premium. The variance of spot returns is often twenty times that of the forward premium. The recognition of this stylized fact dates back to Cornell (1977) and is mentioned by many subsequent authors. The properties of the forward premium have been widely discussed by Hai et al. (1997), Baillie and Bollerslev (1994), Maynard and Phillips (2001), Choi and Zivot (2007) and Sakoulis et al. (2010). Several articles find the forward premium to be a long memory $I(d)$ process where $0 < d < 1$. The implication of this finding is that the Fama regression is unbalanced and this avenue is pursued by both Baillie and Bollerslev (2000) and Maynard and Phillips (2001).

The finding of long memory in the forward premium suggests the omission of another variable (possibly associated with a time dependent risk premium), which also has long memory and which is fractionally cointegrated with the forward premium, and hence balances the regression equation. The findings in this paper suggest the situation is not as simple as this and that the relationship is extremely nonlinear. In fact the time series properties and order of integration of the forward premium appear to change within the sample. As discussed later in this paper, there is additional strong evidence that the slope parameter in the basic Fama regression, Equation (7) is also highly time dependent. This leads to different interpretations on when and why the theory of UIP is valid.

This article uses data on the eight freely floating currencies of the Australian dollar (AUD), Canadian dollar (CAD), Swiss franc (CHF), Danish krone (DKK), Japanese yen (JPY), British pound (GBP), Norwegian krone (NOK), and New Zealand dollar (NZD). The data is monthly for spot exchange rates and one month forward exchange rates and is from December 1988 through September 2011; which is a total of 273 observations. All the exchange rate data have the US dollar as the numeraire currency.

While Baillie and Bollerslev (1994) and Maynard and Phillips (2001) found evidence of long memory, or fractional integration, in the whole realization of the forward premium, $(f_t - s_t)$, Choi and Zivot (2007) found evidence for long memory in separate sub regimes of the forward premium. In the following, we use the multiple mean break methodology of Bai and Perron (1998, 2003) to detect possible structural changes in the forward premium. The m break model and the $m+1$ regime model are defined as

$$(f_t - s_t) = c_j + u_t, \;\; t = T_{j-1}+1, T_{j-1}+2, ..., T_j \tag{5}$$

where $j = 1, 2, ..., m+1$, and $T_0 = 0$, $T_{m+1} = T$, and c_j is the mean of the forward premium for each regime. For each one of the m partitions, the OLS estimate of the parameter c_j is obtained by minimizing the quantity,

$$S_T = \sum_{j=1}^{m+1} \sum_{t=T_j-1}^{T_j} (y_t - c_j)^2. \tag{6}$$

The application of Bai and Perron tests reported in Table 1 indicates that most of the forward premium series have four or five break points. Table 2 reports the estimated break points with 95% confidence intervals for these estimated break points. Some of the dates for the break points have natural interpretations in terms of the *ERM* crises in the Fall of 1992 and 1993, and the financial crisis in the Fall of 2008. Others are not necessarily interpretable in terms of known economic and political events. It should be noted that the above testing for breaks is predicated on the process between breaks being weakly stationary. However, identification of these break points is a useful starting place for investigating the time variation within the forward premium processes.

Table 3 then presents the Local Whittle estimates of the long memory parameter for each sub regime. It can be seen that there is substantial variation of the estimates of the long memory parameter across regimes. Many of the sub periods have estimates of the long memory parameter that are statistically different from zero and many also have estimates in excess of 0.5 which is indicative of non stationary long memory. In general the results are consistent with the view of strong persistence of the forward premium series across regimes. These properties strongly suggest that it cannot be simply classified as a process with a certain order of integration, or fractional integration. Hence the issue of balanced and unbalanced regressions becomes more complicated in this context with the lagged forward premium used as an explanatory variable. Table 4 then presents $SURE$ estimation results for a model that has different betas in the forward premium regression for each currency and for all of its different sub periods. Although some of the sample sizes are relatively small and have beta parameter estimates with correspondingly wide standard errors, the results are nevertheless strongly indicative of variation of beta across regimes. Wald tests for the stability of the betas across the six regimes are rejected at the .05 level for four of the eight currencies.

4 TVP Estimation of β_t

There are many possible approaches at estimating β_t, the following forward premium regression, when the slope coefficient is allowed to be time dependent:

$$\Delta s_{t+1} = \alpha_t + \beta_t(f_t - s_t) + u_{t+1}. \tag{7}$$

Previous studies by Wu and Zhang (1996), Zhou (2002), Bansal (1997) and Bansal and Dahlquist (2000) all noted the asymmetry in the forward premium anomaly, which tended to be more extreme with larger negative slope coefficient estimates, when US interest rates were below foreign equivalents. These authors typically allowed the β_t to switch from one level to another throughout the sample. A more general approach is taken by Baillie and Kiliç (2006) who allowed the slope coefficient to vary so that it followed a smooth transition regression function of the form $\beta_t = \beta_1(1 - G(z_t;\gamma,c)) + \beta_2 G(z_t;\gamma,c)$, where $G(\cdot)$ was a logistic smooth transition function $G(z_t;\gamma,c) = (1 + \exp(-\gamma(z_t - c)/\sigma_{z_t}))^{-1}$, $\gamma > 0$, in the range of $(0,1)$. Hence z_t is the transition variable, and Baillie and Kiliç (2006) used interest rate differential, money growth differential, income differentials, and the volatility of US money growth rate. Lothian and Wu (2011), also used the formulation

$$\Delta s_{t+1} = \alpha + \beta_1(i_t - i_t^* - \mu) + \beta_2(1 - e^{-\lambda(i_t - i_t^* - \mu)^2})(i_t - i_t^* - \mu) + u_{t+1} \tag{8}$$

where μ is the long run mean of the interest rate differential, and the transition function $G = (1 - e^{-\lambda(i_t - i_t^* - \mu)^2})$ is between 0 and 1.

Previous work on using random coefficient models in the forward premium regression is in Wolff (1987) who used a Kalman filtering model and a stationary $AR(1)$ to represent the time variation in the β_t. In this paper three different methods were considered; the simple rolling regression, a Bayesian approach where it is assumed the regression parameters follow a random walk and a new TVP kernel weighted regression.

Figure 1 plots the generated β_t from rolling regressions with a window length of 60 monthly observations. The real difficulty of the approach is that the attempt to capture "local variation" by having short intervals of data, is incompatible with the desire of having tight standard errors and hence tight confidence intervals on the estimated β_t parameters. Experimentation with having differing window lengths was not very successful and is not reported here.

An alternative approach is to use the Bayesian approach developed by Koop et al. (2009), where the slope coefficient is assumed to follow the Random Walk model,

$$\beta_t = \beta_{t-1} + v_t \tag{9}$$

and $v_t \sim NID(0, \Sigma)$, $\beta_0 = 0$ and $v_0 \sim NID(0, \Sigma_0)$ and α follows the inverse Wishart distribution and Σ_0 follows the inverse Gamma distribution. The model was estimated from methodology described by Nakajima (2011). The results of estimating this model can be seen in Figure 2 which reports the posterior estimates of β_t for each point of time. The results were found to be particularly sensitive to the specification of the prior distribution's variance, and still imply fairly wide confidence intervals on the estimated β_t so that it is hard to make definitive tests of the UIP hypothesis.

The third and new method for estimating the TVP of β_t is by a kernel weighted regression and is a modification of the work on autoregressions by Giraitis et al. (2014).[2] The TVP $AR(1)$ model is

$$y_{t+1} = \phi_t y_t + u_{t+1} \tag{10}$$

[2]The motivation for TVP approach can also be justified from Granger (2008) who showed that any nonlinear stationary time series process can always be expressed as an $AR(1)$ process with TVP. Although this may potentially be a very inefficient approach to modeling a TVP, it is nevertheless a basic benchmark. For example it is trivially possible to express the $AR(2)$ process, $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t$ as $y_t = \phi_t y_{t-1} + u_t$, where $\phi_t = (\phi_1 + \phi_2 \frac{y_{t-2}}{y_{t-1}})$. Although the forecast variance would likely be poor compared with using the original $AR(2)$ process.

where u_{t+1} is $i.i.d.(0, \sigma_u^2)$ and there is some initialization of y_0. The stability of the model depends on the TVP nature of the AR parameters satisfying various smoothness classes. Giraitis et al. (2014) have modeled the TVP denoted by ϕ_t for an $AR(1)$ as a rescaled random walk, and where $\{a_t\}$ is a non stationary process which defines the random drift, and $-1 < \phi < 1$. However, in this context ϕ_t is a standardized version of a_t so that

$$\phi_t = \phi \frac{a_t}{\max_{0 \leq k \leq t} |a_k|} \tag{11}$$

It is assumed that $a_t = a_{t-1} + w_t$ which is a driftless random walk, where w_t is a stationary process with zero mean. If w_t is white noise then the process is identical to that of Cogley and Sargent (2005). Following Giraitis et al. (2014), it can be assumed that it is a general stationary process with possible slow hyperbolic decay in its autocorrelation function. They show that the coefficient process $\{\phi_t; t = 1, ..., T\}$ converges in distribution as T increases to the limit $\{\phi_t; 0 \leq \tau \leq 1\} \rightarrow_D \{\phi \tilde{W}_\tau; 0 \leq \tau \leq 1\}$, where

$$\tilde{W}_\tau := \frac{W_\tau}{\sup_{0 \leq s \leq 1} |W_s|} \tag{12}$$

An important point is that W_τ can be either regular Brownian Motion or fractional Brownian Motion. Also the parameter ϕ_t evolves around a mean of zero and can take any value in the interval $[-|\phi|, |\phi|]$. Giraitis et al. (2014) also find some bounds on the second moment properties of the process such that

$$Cov(y_{t+k}, y_t) \leq \frac{\phi^{|k|}}{1 - \phi^2} \left\{\sigma_u^2 + E(y_0^2)\right\} \tag{13}$$

and find convergence to its limiting distribution. The approach for estimating the TVP is to take

$$\hat{\beta}_t = \frac{\sum_{t=1}^{H} K(\frac{t-k}{H}) y_t y_{t-1}}{\sum_{t=1}^{H} K(\frac{t-k}{H}) y_{t-1}^2} \tag{14}$$

where $K(\frac{t-k}{H})$ is a kernel and is a continuously bounded function. For example the Epanechnikov kernel with finite support and the Gaussian kernel with infinite support are potential candidates. Then,

$$H^{1/2}(1 - \hat{\beta}_t^2)^{-1/2}(\hat{\beta}_t - \beta_t) \sim N(0, 1) \tag{15}$$

In the context of estimating the TVP regression in equation (7), the general TVP regression can be expressed as

$$y_t = x_t'\beta_t + u_t \tag{16}$$

with β_t being a bounded random walk, and in the context of equation (7), $x_t' = [1, (f_t - s_t)]$. In general the kernel weighted regression estimator for β_t is

$$\hat{\beta}_t = \left(\sum_{j=1} w_{jt} x_j x_j'\right)^{-1} \left(\sum_{j=1} w_{jt} x_j y_j\right). \tag{17}$$

Then

$$\hat{\beta}_t - \beta_t = \left(\sum_{j=1} w_{jt} x_j x_j'\right)^{-1} \sum_{j=1} w_{jt} x_j y_j - \beta_t = \left(\sum_{j=1} w_{jt} x_j x_j'\right)^{-1} \sum_{j=1} w_{jt} x_j \left(x_t'\beta_t + u_t\right) - \beta_t \tag{18}$$

$$= \left(\sum_{j=1} w_{jt} x_j x_j'\right)^{-1} \sum_{j=1} w_{jt} x_j x_j'\beta_j + \left(\sum_{j=1} w_{jt} x_j x_j'\right)^{-1} \sum_{j=1} w_{jt} x_j u_t - \beta_t$$

$$= \left(\sum_{j=1} w_{jt} x_j x_j'\right)^{-1} \sum_{j=1} w_{jt} x_j u_t + \left(\sum_{j=1} w_{jt} x_j x_j'\right)^{-1} \sum_{j=1} w_{jt} x_j x_j' \left(\beta_j - \beta_t\right)$$

If the bandwidth is $o_p\left(T^{1/2}\right)$, then the term $\left(\sum_{j=1} w_{jt} x_j x_j'\right)^{-1} \sum_{j=1} w_{jt} x_j x_j' \left(\beta_j - \beta_t\right)$ is negligible. Hence it is satisfactory to focus on the term $\left(\sum_{j=1} w_{jt} x_j x_j'\right)^{-1} \sum_{j=1} w_{jt} x_j u_t$. One expression for the estimator of the variance of the TVP if u_t is homoskedastic is given by

$$Var\left(\hat{\beta}_t\right) = \hat{\sigma}_u^2 \left(\sum_{j=1} w_{jt} x_j x_j'\right)^{-1} \sum_{j=1} w_{jt}^2 x_j x_j' \left(\sum_{j=1} w_{jt} x_j x_j'\right)^{-1} \tag{19}$$

where $\hat{\sigma}_u^2 = \frac{1}{T}\sum_{i=1}^{T}\left(y_t - x_t'\beta_t\right)^2$. If u_t is heteroskedastic then the covariance matrix of the TVP parameter estimates is given by

$$Var\left(\hat{\beta}_t\right) = \left(\sum_{j=1} w_{jt} x_j x_j'\right)^{-1} \left(\sum_{j=1} w_{jt}^2 x_j x_j' \hat{u}_t^2\right) \left(\sum_{j=1} w_{jt} x_j x_j'\right)^{-1} \tag{20}$$

Similarly, an extremum estimator of the form

$$\hat{\theta}_t = \arg\min_{\theta} \sum_{j=1}^{T} w_{jt} l_j\left(y_j; \theta\right) \tag{21}$$

will have a sandwich type estimator for the covariance matrix of the TVP parameter estimates given by

$$Var\left(\hat{\beta}_t\right) = \left(\sum_{j=1}^{T} w_{jt} \frac{\partial^2 l_j\left(y_j;\hat{\theta}_t\right)}{\partial\theta\partial\theta'}\right)^{-1} \left(\sum_{j=1}^{T} w_{jt}^2 \left(\frac{\partial l_j\left(y_j;\hat{\theta}_t\right)}{\partial\theta}\right)\left(\frac{\partial l_j\left(y_j;\hat{\theta}_t\right)}{\partial\theta}\right)'\right) \left(\sum_{j=1}^{T} w_{jt} \frac{\partial^2 l_j\left(y_j;\hat{\theta}_t\right)}{\partial\theta\partial\theta'}\right)^{-1} \tag{22}$$

The results of estimating β_t by the TVP by the new kernel weighted regression β_t are presented in Figure 3. The 95% confidence intervals of the estimated β_t are also shown in the same diagram and are obtained from equation (19). It is clear that these confidence intervals are considerably smaller than those generated from the rolling regressions, or by the Bayesian procedure with coefficients following random walks. The method can be seen to have the apparent advantage of producing quite smooth estimates of the slope coefficient and correspondingly relatively tight confidence intervals around the point estimates of the slope coefficient.

In general it can be seen that there are many periods where the β_t are significantly negative and other periods particularly towards the end of the sample, where its value becomes relatively high and sometimes exceeds unity. For example, the New Zealand dollar has a β_t that is consistently negative from 2001 through 2008 and is as small as -10 in some periods. However, after 2008 it is large and positive and $+10$ with two sided 95% confidence intervals easily exceeding $+3$. In general the financial crisis of the Fall of 2008 with lower nominal interest rates has coincided with the AUD, CHF, NOK and NZD all having large and positive β_t coefficients from the rolling regression. Another interesting aspect is that the shape of the estimated β_t graphed over time has distinct similarities between currencies. There also seem to be significant "over reactions" to UIP that occur for the Australian dollar, the Swiss franc and Norwegian krone, as with the pound being the numeraire. This positive version of the forward premium anomaly is in contrast to the conventional or classic forward premium anomaly evident in other time periods. These stylized facts appear to be generally true regardless of which is the numeraire currency. Overall, there is strong evidence of predictable and consistent evidence of the standard forward premium anomaly with $\beta < 0$ at specific periods, and other periods, particularly at the beginning and end of the sample when $\beta > 0$, and there is an apparent over reaction of UIP with greater rates of depreciation that expected for the higher interest rate country.

Clearly this work raises issues concerning the reasons and causes of the breakdown of UIP

at specific time points. One other explanation of the instability of the slope parameter estimate would be in terms of omitted variables to represent a risk premium. In this light of this, and in view of the models developed by Hodrick (1989), we also estimated several models with the volatility of money growth included in a TVP setup where the volatility of money growth was defined as the conditional variance generated by estimating an $ARFIMA - GARCH(1,1)$ model. These models appear to have some explanatory power for post 2008 and could perhaps be combined with models involving cross sectional variation between the returns of high and low interest rate currencies as in Lustig and Verdelhan (2007). However a full TVP analysis of these models is quite challenging and is a subject for future research.

A final corollary concerns the use of the standard forward premium anomaly to be used as a model checking device to assess the adequacy of calibrated *DSGE* models as in Backus et al. (2001), Lustig and Verdelhan (2007) and Verdelhan (2010), among others. These papers have routinely partly assessed the validity of their models by checking to see if they generate a negative correlation between spot returns and lagged interest rate differentials. Our results cast doubt on whether this is a meaningful approach given the huge variation in the slope coefficient estimate in the forward premium regression. Contrary to the perceived wisdom in the profession there appears ample evidence of over reaction to UIP with countries with relatively high interest rates often having a currency that depreciates significantly more than expected from UIP.

Finally, it should be noted that all the results in the figures and tables are for the US dollar being the numeraire currency. Similar analysis was conducted for the numeraire currency being the British pound and the Japanese yen. These did not qualitatively change the results and the details are suppressed in the interests of saving space, but are available from the authors on request.

5 Conclusions

This paper has provided more evidence on the time series properties of the univariate forward premium series for eight countries and has found four or five breaks in the series. The forward premium is highly nonlinear and appears to defy classification as a process with a constant order of integration. The second aspect of the paper is concerned with the time varying nature of the estimate of the slope parameter when spot returns are regressed on the lagged forward premium, or equivalently the lagged interest rate differential. We compare rolling type regression estimates, with Bayesian estimation and also a new Time Varying Parameter (TVP) method that is motivated by the TVP autoregression of Giraitis et al. (2014). The procedure is a form of kernel weighted regression and delivers relatively tight standard errors on the parameter estimates. We find the existence of the forward premium anomaly with large negative beta coefficients in the 1980s and 1990s. For some currencies there is also evidence of large positive coefficients and a reversal of the forward premium anomaly after the financial crisis of 2008.

Contrary to the perceived wisdom in the profession there appears ample evidence of relatively frequent over reaction to UIP with countries with relatively high interest rates often having a currency that depreciates significantly more than expected from UIP. The extent and persistence of the estimated TVP beta coefficients appear more extreme than previously recognized. To understand the economic and financial reasons for this behavior is an important subject for future research.

References

[1] Bacchetta, P., van Wincoop, E., 2007. Random walk expectations and the forward discount puzzle. American Economic Review 97, 346–350.

[2] Backus, D.K., Foresi, S., Telmer, C.I., 2001. Affine term structure models and the forward premium anomaly. Journal of Finance 56, 279–304.

[3] Bai, J., Perron, P., 1998. Estimating and testing linear models with multiple structural changes. Econometrica 66, 47–78.

[4] Bai, J., Perron, P., 2003. Computation and analysis of multiple structural change models. Journal of Applied Econometrics 18, 1–22.

[5] Baillie, R.T., Bollerslev, T., 1994. The long memory of the forward premium. Journal of International Money and Finance 13, 565–571.

[6] Baillie, R.T., Bollerslev, T., 2000. The forward premium anomaly is not as bad as you think. Journal of International Money and Finance 19, 471–488.

[7] Baillie, R.T., Kiliç, R., 2006. Do asymmetric and nonlinear adjustments explain the forward premium anomaly? Journal of International Money and Finance 25, 22–47.

[8] Bansal, R., 1997. An exploration of the forward premium puzzle in currency markets. Review of Financial Studies 10, 369–403.

[9] Bansal, R., Dahlquist, M., 2000. The forward premium puzzle: different tales from developed and emerging economies. Journal of International Economics 51, 115–144.

[10] Chang, S.S., 2013. Can cross-country portfolio rebalancing give rise to forward bias in FX markets? Journal of International Money and Finance 32, 1079–1096.

[11] Choi, K., Zivot, E., 2007. Long memory and structural breaks in the forward discount: an empirical investigation. Journal of International Money and Finance 26, 342–363.

[12] Cogley, T., Sargent, T.J., 2005. Drifts and volatilities: monetary policies and outcomes in the post WWII US. Review of Economic Dynamics 8, 262–302.

[13] Cornell, B., 1977. Spot rates, forward rates and exchange market efficiency. Journal of Financial Economics 5, 56–65.

[14] Engel, C., 1996. The Forward discount anomaly and the risk premium: a survey of recent evidence. Journal of Empirical Finance 3, 123–192.

[15] Fama, E.F., 1984. Forward and spot exchange rates. Journal of Monetary Economics 14, 319–338.

[16] Froot, R.A., Thaler, R.H., 1990. Anomalies: foreign exchange. Journal of Economic Perspectives 4, 179–192.

[17] Giraitis, L., Kapetanios, G., Yates, T., 2014. Inference on stochastic time-varying coefficient models. Journal of Econometrics 179, 46–65.

[18] Granger, C.W.J., 2008. Non-Linear models: where do we go next - time varying parameter models. Studies in Nonlinear Dynamics and Econometrics 12, 1–9.

[19] Hai, W., Mark, N.C., Wu, Y., 1997. Understanding spot and forward exchange rate regressions. Journal of Applied Econometrics 12, 715–734.

[20] Hansen, L.P., Hodrick, R.J., 1983. Risk averse speculation in forward exchange markets: an econometric analysis, in: Frenkel, J.A. (Eds.), Exchange Rates and International Macroeconomics. University of Chicago Press.

[21] Hodrick, R.J., 1987. The empirical evidence on the efficiency of forward and futures foreign exchange markets. Harwood, London.

[22] Hodrick, R.J., 1989. Risk, uncertainty and exchange rates. Journal of Monetary Economics 23, 433–459.

[23] Koop, G., Leon-Gonzalez, R., Strachan, R., 2009. On the evolution of the monetary policy transmission mechanism. Journal of Economic Dynamics and Control 33, 997–1017.

[24] Lothian, J.R., Wu, L., 2011. Uncovered interest-rate parity over the past two centuries. Journal of International Money and Finance 30, 448–473.

[25] Lucas, R.E., 1978. Asset prices in an exchange economy. Econometrica 46, 1429–1445.

[26] Lustig, H., Verdelhan, A., 2007. The cross section of foreign currency risk premia and consumption growth risk. American Economic Review 97, 89–117.

[27] Mark, N.C., Wu, Y., 1997. Rethinking deviations from uncovered interest rate parity: the role of covariance risk and noise. Economic Journal 108, 1686–1706.

[28] Maynard, A., Phillips, P.C.B., 2001. Rethinking an old empirical puzzle: econometric evidence on the forward discount anomaly. Journal of Applied Econometrics 16, 671–708.

[29] Nakajima, J., 2011. Time-varying parameter VAR model with stochastic volatility: an overview of methodology and empirical applications. Monetary and Economic Studies 29, 107–142.

[30] Sakoulis, G., Zivot, E., Choi, K., 2010. Time variation and structural change in the forward discount: implications for the forward rate unbiasedness hypothesis. Journal of Empirical Finance 17, 957–966.

[31] Verdelhan, A., 2010. A habit-based explanation of the exchange rate risk premium. Journal of Finance 65, 123–146.

[32] Wolff, C.C., 1987. Forward foreign exchange rates, expected spot rates, and premia: a signal-extraction approach. Journal of Finance 42, 395–406.

[33] Wu, Y., Zhang, H., 1996. Asymmetry in forward exchange rate bias: a puzzling result. Economic Letters 50, 407–411.

[34] Zhou, S., 2002. The forward premium anomaly and the trend behavior of the exchange rates. Economic Letters 76, 273–279.

TABLE 1. MULTIPLE STRUCTURAL CHANGE TESTS

Statistic	Currency (vis-à-vis the US Dollar)							
	AUD	CAD	CHF	DKK	GBP	JPY	NOK	NZD
Tests								
$\sup F_T(1)$	52.992***	65.219***	22.859***	16.485***	56.069***	3.843	8.625*	8.419*
$\sup F_T(2)$	32.484***	40.579***	45.589***	16.406***	51.004***	14.382***	7.249	33.271***
$\sup F_T(3)$	39.829***	51.483***	13.833***	21.325***	32.082***	8.901**	9.887***	19.623***
$\sup F_T(4)$	48.950***	40.376***	14.374***	18.804***	40.035***	150.782***	15.107***	10.017***
$\sup F_T(5)$	38.387***	43.220***	56.830***	17.605***	62.802***	100.124***	8.324***	16.071***
$\text{UD}_{\max}$	52.992***	65.219***	56.830***	21.325***	62.802***	150.782***	15.107***	33.271***
$\text{WD}_{\max}$	78.204***	74.768***	98.312***	30.906***	108.643***	240.896***	24.136***	41.262***
$\sup F_T(2\|1)$	10.884**	6.260	36.668***	11.255**	2.560	17.765***	5.738	6.062
$\sup F_T(3\|2)$	20.392***	5.059	8.873	7.771	1.472	3.309	3.636	18.653***
$\sup F_T(4\|3)$	10.312	6.186	16.742***	7.771	14.249**	6.792	22.018***	7.416
$\sup F_T(5\|4)$	3.861	38.527***	9.168	15.446**	4.032	13.510**	11.137*	5.703
Number of breaks selected								
Sequential	1	1	3	2	1	0	0	0
LWZ	5	5	5	5	5	5	5	5
BIC	5	5	5	5	5	5	5	5

Note. The test results for multiple structural changes in the forward premium as in Bai and Perron (1998, 2003) are reported. *, **, *** indicate 10%, 5%, 1% significance, respectively.

TABLE 2. ESTIMATES FOR MULTIPLE STRUCTURAL BREAK MODEL

AUD				CAD			
$\hat{c}_1$	0.603 (0.017)	$\hat{T}_1$	91:03 [90:12, 91:08]	$\hat{c}_1$	0.276 (0.010)	$\hat{T}_1$	93:02 [92.02, 93:08]
$\hat{c}_2$	0.165 (0.011)	$\hat{T}_2$	96:11 [95:08, 98:01]	$\hat{c}_2$	0.071 (0.012)	$\hat{T}_2$	96:03 [95:10, 97:01]
$\hat{c}_3$	-0.019 (0.012)	$\hat{T}_3$	01:08 [01:05, 02:02]	$\hat{c}_3$	-0.161 (0.015)	$\hat{T}_3$	97:12 [97:07, 99:11]
$\hat{c}_4$	0.278 (0.014)	$\hat{T}_4$	05:05 [05:01, 06:01]	$\hat{c}_4$	-0.046 (0.011)	$\hat{T}_4$	01:03 [97:09, 01:04]*
$\hat{c}_5$	0.098 (0.016)	$\hat{T}_5$	07:12 [07:07, 08:03]	$\hat{c}_5$	0.087 (0.011)	$\hat{T}_5$	04:11 [04:05, 09:07]*
$\hat{c}_6$	0.308 (0.016)			$\hat{c}_6$	-0.029 (0.008)		
CHF				**DKK**			
$\hat{c}_1$	-0.166 (0.027)	$\hat{T}_1$	89:12 [88:12, 90:01]	$\hat{c}_1$	0.064 (0.031)	$\hat{T}_1$	91:03 [90:10, 92:01]
$\hat{c}_2$	0.166 (0.013)	$\hat{T}_2$	94:08 [94:08, 08:10]*	$\hat{c}_2$	0.429 (0.043)	$\hat{T}_2$	92:06 [90:07, 92:09]
$\hat{c}_3$	-0.310 (0.011)	$\hat{T}_3$	00:12 [00:03, 01:02]	$\hat{c}_3$	0.863 (0.043)	$\hat{T}_3$	93:09 [93:01, 94:07]
$\hat{c}_4$	-0.076 (0.014)	$\hat{T}_4$	05:01 [04:10, 06,01]	$\hat{c}_4$	0.207 (0.046)	$\hat{T}_4$	94:10 [94:03, 97:07]
$\hat{c}_5$	-0.270 (0.017)	$\hat{T}_5$	07:12 [07:11, 08:03]	$\hat{c}_5$	-0.123 (0.019)	$\hat{T}_5$	00:12 [97:08, 04:08]
$\hat{c}_6$	-0.050 (0.017)			$\hat{c}_6$	0.021 (0.015)		
GBP				**JPY**			
$\hat{c}_1$	0.390 (0.019)	$\hat{T}_1$	89:12 [89:08, 95:09]*	$\hat{c}_1$	-0.315 (0.023)	$\hat{T}_1$	89:12 [88:12, 91:02]
$\hat{c}_2$	0.519 (0.012)	$\hat{T}_2$	92:10 [92:09, 93:01]	$\hat{c}_2$	0.010 (0.011)	$\hat{T}_2$	94:08 [94:08, 00:10]*
$\hat{c}_3$	0.244 (0.018)	$\hat{T}_3$	94:01 [91:06, 94:02]	$\hat{c}_3$	-0.435 (0.009)	$\hat{T}_3$	01:08 [01:06, 01:11]
$\hat{c}_4$	0.054 (0.007)	$\hat{T}_4$	01:08 [00:12, 05:07]	$\hat{c}_4$	-0.138 (0.013)	$\hat{T}_4$	05:01 [02:10, 05:03]
$\hat{c}_5$	0.207 (0.010)	$\hat{T}_5$	05:05 [04:11, 06:03]	$\hat{c}_5$	-0.347 (0.012)	$\hat{T}_5$	08:09 [08:08, 09:10]
$\hat{c}_6$	0.039 (0.009)			$\hat{c}_6$	-0.038 (0.017)		
NOK				**NZD**			
$\hat{c}_1$	0.224 (0.035)	$\hat{T}_1$	91:05 [88:12, 91:07]*	$\hat{c}_1$	0.385 (0.013)	$\hat{T}_1$	91:07 [90:10, 92:11]
$\hat{c}_2$	0.723 (0.039)	$\hat{T}_2$	93:05 [93:03, 95:01]	$\hat{c}_2$	0.219 (0.008)	$\hat{T}_2$	98:07 [97:07, 98:11]
$\hat{c}_3$	-0.008 (0.020)	$\hat{T}_3$	01:02 [97:09, 01:10]	$\hat{c}_3$	-0.027 (0.012)	$\hat{T}_3$	01:04 [00:09, 01:11]
$\hat{c}_4$	0.340 (0.033)	$\hat{T}_4$	03:11 [03:03, 06:03]	$\hat{c}_4$	0.292 (0.008)	$\hat{T}_4$	07:11 [06:12, 09:05]*
$\hat{c}_5$	-0.079 (0.028)	$\hat{T}_5$	07:11 [07:02, 10:02]	$\hat{c}_5$	0.458 (0.020)	$\hat{T}_5$	08:12 [08:11, 09:08]
$\hat{c}_6$	0.173 (0.032)			$\hat{c}_6$	0.215 (0.015)		

Note. $\hat{c}_i$ is the estimated intercept parameter and $\hat{T}_i$ is the estimated break date. Asymptotic standard errors are in parenthesis next to the corresponding parameter estimates. The 95% confidence intervals are reported in brackets. * indicates the 90% confidence interval.

TABLE 3. LOCAL WHITTLE (*LW*) ESTIMATES OF THE LONG MEMORY PARAMETER

	Bandwidth (m)	AUD	CAD	CHF	DKK	GBP	JPY	NOK	NZD
$\widehat{d}$	$T^{0.5}$	0.713	0.692	1.237	0.670	0.867	0.889	0.665	0.764
		(0.123)	(0.123)	(0.123)	(0.123)	(0.123)	(0.123)	(0.123)	(0.123)
	$T^{0.8}$	0.942	0.767	0.973	0.741	1.064	1.078	0.607	0.980
		(0.053)	(0.053)	(0.053)	(0.053)	(0.053)	(0.053)	(0.053)	(0.053)
$\widehat{d}_1$	$T^{0.5}$	0.726	0.769	0.521	0.527	0.411	2.488	0.045	0.848
		(0.217)	(0.187)	(0.263)	(0.217)	(0.263)	(0.263)	(0.214)	(0.210)
	$T^{0.8}$	0.553	0.496	0.284	0.151	0.427	1.154	0.301	0.771
		(0.132)	(0.104)	(0.179)	(0.132)	(0.179)	(0.179)	(0.128)	(0.125)
$\widehat{d}_2$	$T^{0.5}$	0.901	0.289	1.202	0.458	1.097	0.927	0.964	0.743
		(0.174)	(0.203)	(0.183)	(0.254)	(0.207)	(0.183)	(0.226)	(0.165)
	$T^{0.8}$	0.783	0.928	1.021	-0.088	1.079	0.866	0.585	0.935
		(0.092)	(0.118)	(0.100)	(0.169)	(0.122)	(0.100)	(0.140)	(0.085)
$\widehat{d}_3$	$T^{0.5}$	0.754	1.343	1.251	0.527	0.266	1.660	0.600	0.356
		(0.182)	(0.234)	(0.169)	(0.254)	(0.254)	(0.165)	(0.161)	(0.209)
	$T^{0.8}$	0.853	1.163	1.030	0.407	0.075	1.008	0.526	0.740
		(0.099)	(0.148)	(0.088)	(0.169)	(0.169)	(0.085)	(0.082)	(0.123)
$\widehat{d}_4$	$T^{0.5}$	1.436	0.767	0.261	0.811	0.683	0.911	1.818	1.111
		(0.193)	(0.200)	(0.189)	(0.263)	(0.162)	(0.198)	(0.209)	(0.168)
	$T^{0.8}$	1.194	0.766	0.955	0.801	0.894	0.896	1.219	0.989
		(0.109)	(0.115)	(0.105)	(0.179)	(0.082)	(0.113)	(0.123)	(0.087)
$\widehat{d}_5$	$T^{0.5}$	0.618	1.353	1.330	1.130	1.443	1.235	1.410	0.900
		(0.212)	(0.194)	(0.206)	(0.170)	(0.193)	(0.194)	(0.190)	(0.263)
	$T^{0.8}$	0.327	1.144	1.082	1.003	1.099	0.768	1.014	0.732
		(0.127)	(0.110)	(0.121)	(0.089)	(0.109)	(0.110)	(0.106)	(0.179)
$\widehat{d}_6$	$T^{0.5}$	0.963	1.040	0.428	1.239	0.508	0.527	1.134	1.152
		(0.193)	(0.166)	(0.193)	(0.148)	(0.169)	(0.204)	(0.192)	(0.209)
	$T^{0.8}$	0.975	0.735	0.494	0.785	0.653	0.489	1.230	0.930
		(0.109)	(0.086)	(0.109)	(0.072)	(0.088)	(0.119)	(0.108)	(0.123)

Note. The Local Whittle (*LW*) estimates of the long memory parameters are reported for the bandwidths of $T^{0.5}$ and $T^{0.8}$. Standard errors are also reported below the corrsponding estimates in parentheses. $\hat{d}$ denotes the *LW* estimate for the entire sample period, and $\hat{d}_i$, for each sub period divided by structural break points, where $i = 1, 2, ..., 6$.

TABLE 4. BETA ESTIMATES FOR EACH SUB PERIOD BY STRUCTURAL BREAKS IN AN *SURE* SYSTEM

	AUD	CAD	CHF	DKK
$\hat{\beta}_1$	0.384 (0.868)	-0.285 (1.001)	-1.054 (1.446)	0.213 (0.807)
$\hat{\beta}_2$	-0.389 (1.833)	-2.373 (3.141)	-0.219 (1.544)	-0.714 (0.937)
$\hat{\beta}_3$	2.247 (5.873)	1.036 (2.331)	-0.935 (1.008)	-0.126 (0.430)
$\hat{\beta}_4$	-0.966 (1.678)	6.102 (5.256)	-2.569 (3.246)	-0.598 (1.474)
$\hat{\beta}_5$	0.515 (4.279)	-3.718 (3.076)	-0.884 (1.466)	-1.021 (1.475)
$\hat{\beta}_6$	-0.138 (1.414)	2.886 (3.243)	9.896 (3.478)	-0.495 (1.290)
Wald	2.02	4.43	20.20	12.75
	GBP	JPY	NOK	NZD
$\hat{\beta}_1$	0.942 (1.464)	-4.659 (2.466)	0.449 (1.041)	-0.646 (1.372)
$\hat{\beta}_2$	0.509 (0.797)	-4.081 (3.526)	0.351 (0.493)	-2.642 (1.654)
$\hat{\beta}_3$	-1.511 (2.329)	-1.686 (1.026)	-1.075 (1.061)	5.012 (5.686)
$\hat{\beta}_4$	-1.490 (2.976)	-4.447 (3.802)	-0.379 (0.952)	-2.505 (1.448)
$\hat{\beta}_5$	-0.570 (1.746)	-0.711 (1.496)	-0.192 (2.255)	2.286 (1.468)
$\hat{\beta}_6$	3.248 (2.863)	8.872 (11.871)	0.705 (1.602)	-4.210 (2.474)
Wald	3.27	15.53	7.62	14.29

Note. The estimates for beta coefficients from the Fama regressions are reported for each sub period divided by structural breaks. Standard errors are also reported in parenthesis next to the corresponding parameter estimates. *Wald* is the robust *Wald* statistic for testing H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 1$; it is asymptotically χ^2 distributed with six degrees of freedom.

TABLE 5. THE NUMBER OF TIME PERIODS THAT *UIP* DOES NOT HOLD IN EITHER DIRECTION FOR THE NEW *TVP* KERNEL WEIGHTED REGRESSION MODEL

	AUD	CAD	CHF	DKK	GBP	JPY
Number of months	53	0	47	41	0	58
that $\beta^{ub} < 1$	(19.41%)	(0%)	(17.22%)	(15.02%)	(0%)	(21.25%)
Number of months	0	0	0	0	0	0
that $\beta^{lb} > 1$	(0%)	(0%)	(0%)	(0%)	(0%)	(0%)
	NOK	NZD				
Number of months	70	60				
that $\beta^{ub} < 1$	(25.64%)	(21.98%)				
Number of months	26	57				
that $\beta^{lb} > 1$	(9.52%)	(20.88%)				

Note. β^{ub} denotes the upper bound of the 95% confidence intervals, and β^{lb} denotes the lower bound of the 95% confidence intervals.

Figure 1. Slope coefficient estimates from Rolling *UIP* Regressions using 5-year subsamples. The dashed lines represent 95 percent confidence bands.

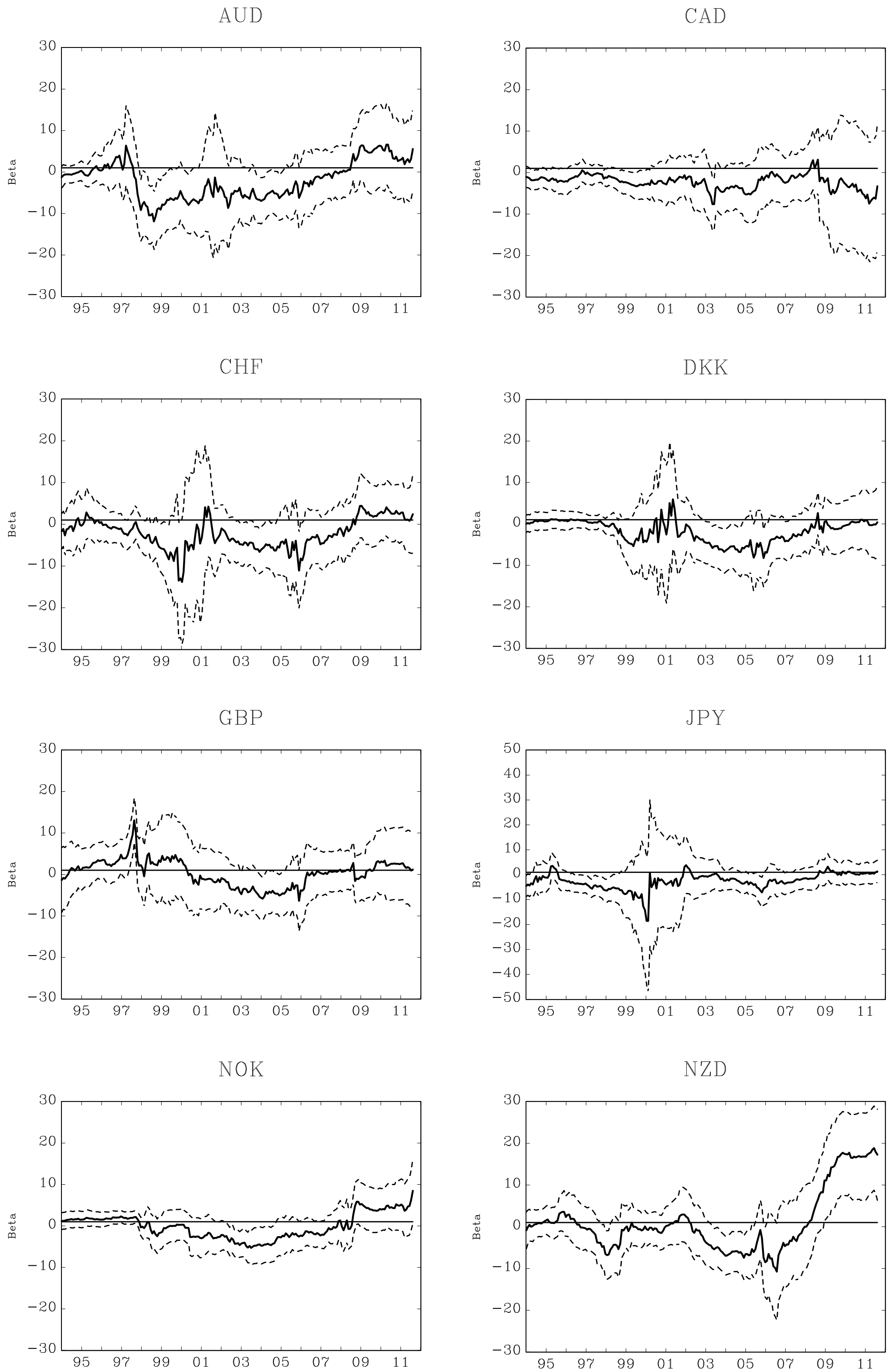

FIGURE 2. POSTERIOR ESTIMATES OF β_t FROM BAYESIAN *TVP* REGRESSIONS. THE SOLID LINE DENOTES POSTERIOR MEANS, AND THE DASHED LINES, TWO STANDARD ERROR CONFIDENCE BANDS. THE STRAIGHT LINE REPRESENTS THE VALUE OF UNITY.

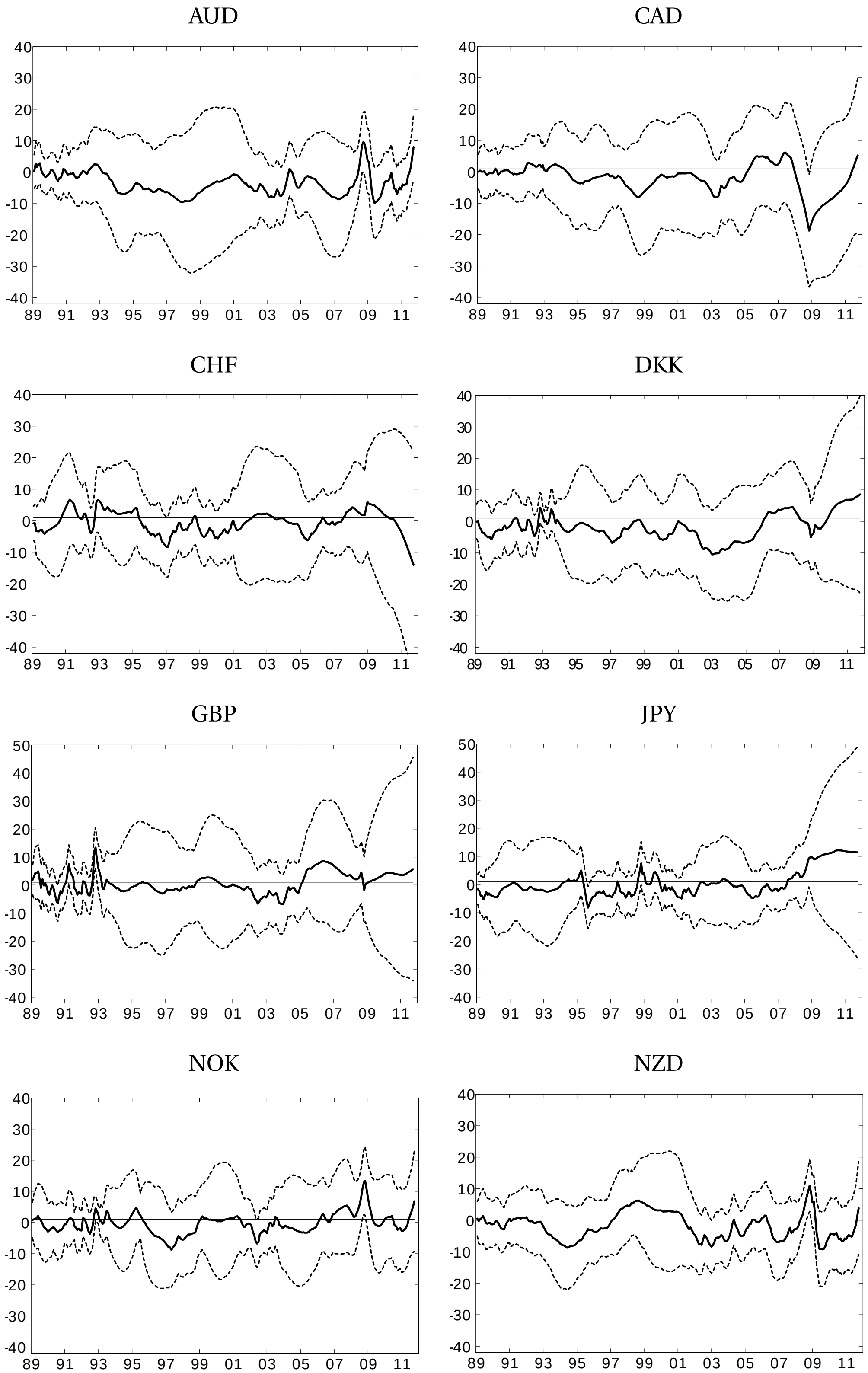

Figure 3. Slope parameter estimates from new *TVP* kernel weighted regressions. The dashed lines represent 95 percent confidence bands. The straight line represents the value of unity.