View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Research Papers in Economics

Department of Economics

Working Paper Series

A New Look at the Forward Premium Puzzle

08-009

Nikolay Gospodinov
Concordia University

Department of Economics, 1455 De Maisonneuve Blvd. West, Montreal, Quebec, Canada H3G 1M8
Tel 514–848–2424 # 3900 · Fax 514–848–4536 · econ@alcor.concordia.ca · alcor.concordia.ca/~econ/repec

A New Look at the Forward Premium Puzzle

Nikolay Gospodinov*†
Concordia University and CIREQ

December 2008

Abstract

This paper analyzes the sampling properties of the widely documented large negative slope estimates in regressions of future exchange returns on current forward premium. We argue that the abnormal behavior of the slope estimators in these regressions arises from the simultaneous presence of high persistence, low signal-to-noise ratio, strong endogeneity and an omitted variable problem. The paper develops the limiting theory for the slope parameter estimators in the levels and differenced forward premium regressions under some assumptions that match the empirical properties of the data. The asymptotic results derived in the paper help to reconcile the findings from the levels and difference specifications and provide important insights about the time series properties of the implied risk premium.

Keywords: Forward premium anomaly, high persistence, low signal-to-noise ratio, local-to-unity asymptotics.

JEL Classification: C13; C22; F31.

---

*Department of Economics, Concordia University, 1455 de Maisonneuve Blvd. West, Montreal, Quebec, H3G 1M8 Canada, tel: (514) 848-2424 ext.3935, fax: (514) 848-4536, nikolay.gospodinov@concordia.ca.

†I would like to thank the Editor (René Garcia), an Associate Editor and anonymous referees for very insightful comments and suggestions. Financial support from SSHRC (Canada) and FQRSC (Quebec) is gratefully acknowledged.

1 Introduction and Empirical Motivation

The anomalous results from regressing future exchange rate returns on the current forward premium constitutes one of the major puzzles in modern international finance. The voluminous empirical literature provides ample empirical evidence not only for the strong rejection of the hypothesis that the forward rate is an unbiased predictor of the future spot rate but also for the negative sign and the large magnitude of the estimated slope parameters in this model. The empirical evidence and possible explanations suggested in the literature are surveyed in Engel (1996) and Lewis (1995).

In this paper, we argue that the forward premium puzzle (the finding of large negative and highly unstable slope parameter estimates in the differenced forward premium regression) appears to arise from the simultaneous presence of a risk premium and several empirical characteristics of the spot and forward exchange rate data that render some standard estimators highly misleading. It is important to point out up front, however, that the paper is not directly concerned with testing the validity of the forward rate unbiasedness hypothesis and the uncovered interest parity. In fact, by allowing for a latent risk premium, we already deviate from the pure version of the expectations hypothesis. Instead, the main objective of the paper is to analyze the sampling properties of the estimators in forward premium regressions and reconcile the empirical findings in the literature with the predictions of economic theory.

To situate the discussion in its proper context, we first look at some of the empirical regularities that characterize the behavior of the major foreign exchange currencies. The data used in this paper are monthly observations for British pound (GBP), German mark (DM), Japanese yen (JPY), Canadian dollar (CAD) and Swiss franc (CHF) and cover the period January 1975 - May 2006 except for the JPY for which the sample period starts in August 1978. The data source and the construction of the series are described in Appendix A. Table 1 reports some summary statistics of the data.

The main stylized facts about the variables that enter the forward premium regression for exchange rates are the following: (i) the forward premium is a highly persistent process, (ii) the variability of the forward premium is only a small fraction of the variability of exchange rate returns, (iii) excess returns and spot returns exhibit very little persistence, (iv) spot and forward rates appear to be unit root processes with very similar descriptive statistics and (v) the errors that are driving the processes for next period spot and forward rates are almost perfectly correlated.

Most of these stylized facts have already been documented and discussed in the empirical literature. The implications of (i) on the forward premium puzzle has been studied and analyzed by Baillie and Bollerslev (1994, 2000), Liu and Maynard (2005), Maynard and Phillips (2001) and Sakoulis and Zivot (2002), among others. Maynard (2003) and Liu and Maynard (2005) parameterize the persistence as a near-integrated (local-to-unity) process; Baillie and Bollerslev (1994, 2000) and Maynard and Phillips (2001) model the forward premium as a long-memory process and Sakoulis and Zivot (2002) explore the possibility that the persistence in the forward premium arises from structural breaks. In this paper, we adopt the local-to-unity framework as a convenient device to capture the high persistence of the forward premium even though it may have been generated by some observationally equivalent but analytically less tractable representations such as long memory, structural break or nonlinear processes. Furthermore, we combine the local-to-unity parameterization of the forward premium with two other data characteristics that have been less exploited in the literature; more specifically, the small variability of the forward premium in (ii) and the almost perfectly correlated expectational errors in (v). The paper highlights the interaction of persistence with low variance and strong endogeneity and helps to resolve the seemingly conflicting results from the levels and differenced regressions reported in the literature.

A particular parameterization of the small variability of the regressor has been used recently by Torous and Valkanov (2000) and Moon, Rubia and Valkanov (2004) in a predictive regression of stock returns. We adopt a similar parameterization for the variability of the forward premium and show analytically that this appears to be the source of the highly disperse and unstable empirical estimates of the slope parameter in the differenced forward premium regression. This parameterization is also consistent with the results in Engel and West (2005) who argue that in rational expectations present-value models, the importance of the current information variables to the future exchange rate changes tends to zero as the discount factor approaches unity.

Regarding empirical regularity (v), Zivot (2000) discusses the high correlation between the spot and forward exchange rate errors but does not fully investigate its consequences for the forward premium regression in differences. The estimated error correlation in our sample is 0.997 for CAD and 0.998 for all other currencies implying an extreme endogeneity problem. We argue that accounting for this endogeneity plays an important role in determining the time series properties of the omitted risk premium.

Table 2 presents the regression results from the forward premium regression of $(s_{t+1} - s_t)$ on $(f_t - s_t)$, where s_t and f_t are logarithms of spot and one-period forward rates. Consistent with the findings reported in the literature, the standard forward premium regression delivers large negative (and statistically significant) estimates of the slope parameter. As we show in the paper, these results are driven by the high persistence and the low variability of the regressor (forward premium), reported in Table 1, as well as some severe endogeneity that arises from the presence of a latent risk premium.

The paper takes into account the salient features of the data and the forward premium model and derives the limiting behavior of the estimators in forward premium regressions that provides guidance to understanding some puzzling results reported in numerous empirical studies. The methodological contributions of the paper can be summarized as follows. Interestingly, while a linear combination of the spot and forward rates (the forward premium) is not a stationary but a near-unit root process, this does not give rise to a spurious regression problem because its innovation variance is orders of magnitude smaller and approaches zero with the sample size. As a result, the estimator in the levels regression is consistent although its rate of convergence is slower than the rate of convergence of the standard cointegrating estimator. Second, the commonly used OLS slope estimator in the differenced forward premium regression is asymptotically biased and plausible values of the high persistence, low signal-to-noise ratio and degree of endogeneity, that are attributed to different components of the bias term, can easily produce the magnitude of the negative estimates of the slope parameter reported in the literature. The limiting distribution of the slope estimator is inversely related to the localizing constant in the signal-to-noise ratio which explains the extreme instability of the empirically documented estimates when this constant is close to zero and provides warnings against the practical relevance of the results obtained from this model. These theoretical results are in agreement with some empirical findings that the forward premium puzzle is not a pervasive phenomenon (Bansal and Dahlquist, 2000) and can be linked to the variability of the forward premium.

Importantly, incorporating a risk premium in the model is instrumental in matching the magnitude of the bias in the empirical studies. While high persistence and low signal-to-noise ratio can produce downward biased slope estimates, the large negative values of the slope parameter cannot be replicated for realistic parameter configurations in the absence of a risk premium term. At the same time, accounting for a risk premium is not sufficient to explain the bias unless some specific time series properties are imposed on the risk premium process.

The contribution of the paper can be put more broadly in the context of unbalanced regressions in which the dependent and the explanatory variables have different order of integration. Maynard and Phillips (2001) and Maynard (2003) analyze the unbalanced nature of the forward premium regression and conclude that the only possible value for the slope in this type of regressions is zero which creates tension with the economic (no-arbitrage) theory that predicts a value of one. Another popular example of unbalanced regression is the predictability of stock returns with near nonstationary regressors as valuation ratios and interest rates which has been extensively studied in the financial economics literature. The inherently inconsistent structure of unbalanced regressions poses theoretical problems on how to model statistically the relationship between variables with different stationarity (or memory) properties. Marmer (2008) proposes nonlinear transformations of the nonstationary predictor that change its memory properties and strength of the signal. We use an alternative way to dampen the signal from the possibly nonstationary regressor by localizing its variance to zero. As a result, the dependent and explanatory variables are of the same order of magnitude which helps to reconcile the seemingly contradictory predictions from the statistical model and economic theory. Note that this balancing transformation is invariant to the way we model the persistence and memory properties of the regressor. As mentioned above, Torous and Valkanov (2000) and Moon, Rubia and Valkanov (2004) employ a similar variance localization to balance the predictive regression of stock returns.

The paper also sheds light on several sub-puzzles regarding the properties of the forward premium $(f_t - s_t)$ and excess returns $(s_{t+1} - f_t)$. While $(s_{t+1} - f_t)$ exhibit very little persistence (see stylized fact (iii) above) which implies that s_{t+1} and f_t are cointegrated with cointegrating vector $(1, -1)$, $(f_t - s_t)$ is highly persistent and the empirical evidence for cointegration is mixed despite the fact that two series trace each other very closely. Since the first cointegration relationship implies the latter, our model resolves this sub-puzzle by explicitly modeling the small variance of $(f_t - s_t)$ as local-to-zero which does not allow f_t and s_t to drift apart in the long run. This localization also helps to reconcile another puzzling feature of the data that both spot and excess returns are not persistent but the forward premium, which is the difference of the two $(f_t - s_t) = (s_{t+1} - s_t) - (s_{t+1} - f_t)$, is highly persistent.

The rest of the paper is structured as follows. Section 2 introduces the model structure and the assumptions. Section 3 studies the limiting behavior of the slope estimators in the levels and differenced forward premium regressions. Section 4 investigates the numerical small-sample properties of the estimators under various model parameterizations. Section 5 concludes.

2 Models and Assumptions

The forward rate unbiasedness hypothesis under rational expectations and risk neutrality postulates that $E(s_{t+1}|\mathcal{F}_t) = f_t$ or $s_{t+1} = f_t + \varepsilon_{t+1}$, where ε_{t+1} are rational expectations forecast errors that satisfy $E(\varepsilon_{t+1}|\mathcal{F}_t) = 0$ and $\mathcal{F}_t$ denotes an increasing sequence of sigma-fields $(...\mathcal{F}_{t-3} \subset \mathcal{F}_{t-2} \subset \mathcal{F}_{t-1})$ generated by the history of the series to date t. The unbiasedness hypothesis implies that the regression model $s_{t+1} = \alpha + \beta f_t + \varepsilon_{t+1}$ should satisfy the restrictions $\alpha = 0$, $\beta = 1$ and $E(\varepsilon_{t+1}|\mathcal{F}_t) = 0$.

If the agents in the economy are risk averse, Fama (1984) argues that

$$s_{t+1} = f_t - rp_t + \varepsilon_{t+1}, \tag{1}$$

where $rp_t = f_t - E(s_{t+1}|\mathcal{F}_t)$ is a time-varying rational expectations risk premium. The risk premium rp_t drives a wedge between $E(s_{t+1}|\mathcal{F}_t)$ and f_t and represents the conditional bias in the forward rate forecast of the spot rate (Engel, 1996).[1]

[1] More generally, Bekaert and Hodrick (1993), Engel (1996) and Baillie and Bollerslev (2000) demonstrate that the Euler equation for a risk averse investor implies that $f_t - E(s_{t+1}|\mathcal{F}_t) = 0.5Var(s_{t+1}|\mathcal{F}_t) - Cov(s_{t+1}p_{t+1}|\mathcal{F}_t) - Cov(s_{t+1}q_{t+1}|\mathcal{F}_t)$, where q_{t+1} is the logarithm of the intertemporal marginal rate of substitution, p_{t+1} is the

Now we specify a model that captures the salient features of the data and study the effect of omitting the unobserved risk premium on the slope estimators in the levels and difference specifications of the forward premium regression. Since the empirical evidence indicates that the spot and forward exchange rates follow a unit root process, then $E(f_{t+1}|\mathcal{F}_t) = f_t$ and the statistical model that describes the joint dynamics of spot and forward rates can be represented conveniently in the triangular form

$$\begin{aligned} s_{t+1} &= \alpha + \beta f_t - rp_t + \varepsilon_{1,t+1} \\ f_{t+1} &= f_t + \varepsilon_{2,t+1}. \end{aligned} \tag{2}$$

Assumption A. *Assume that in model (2),* $\alpha = 0$, $\beta = 1$, f_0 *is* $o_p(T^{1/2})$, $\varepsilon_{2,t} = C(L)\xi_t = \sum_{j=0}^{\infty} C_j \xi_{t-j}$ *with* $\sum_{j=0}^{\infty} j|C_j| < \infty$, $E[(\varepsilon_{1,t+1},\, \xi_{t+1})'|\mathcal{F}_t] = 0$, $\sup_t E(\varepsilon_{1,t+1}^4) < \infty$, $\sup_t E(\xi_{t+1}^4) < \infty$ *and* $E[(\varepsilon_{1,t+1},\, \xi_{t+1})'(\varepsilon_{1,t+1},\, \xi_{t+1})] = \begin{bmatrix} \sigma_\varepsilon^2 & \rho\sigma_\varepsilon\sigma_\xi \\ \rho\sigma_\varepsilon\sigma_\xi & \sigma_\xi^2 \end{bmatrix}$ *with* $\rho = Corr(\varepsilon_{1,t+1}, \xi_{t+1})$.

Throughout the paper, we assume that the true data generating process satisfy the restrictions imposed by the economic theory ($\alpha = 0$, $\beta = 1$) and investigate analytically if the limiting behavior of the estimators can mimic their sampling properties from the actual data even when the restrictions hold. The conditions in Assumption A require that $(\varepsilon_{1,t+1},\, \xi_{t+1})'$ is a two-dimensional martingale difference sequence which can exhibit conditional heteroskedasticity provided that $(\varepsilon_{1,t+1},\, \xi_{t+1})'$ is unconditionally homoskedastic with finite fourth moments. While model (2) is fairly general, it can be further extended to serially correlated forecast errors $\varepsilon_{1,t+1}$ that might arise from "peso effects" (Evans and Lewis, 1994) and more flexible deterministic components.

It is interesting to note that subtracting the first from the second equation in (2) yields $f_{t+1} - s_{t+1} = -\alpha + (1-\beta)f_t + rp_t + \varepsilon_{2,t+1} - \varepsilon_{1,t+1}$. If $\alpha = 0$, $\beta = 1$, the variances of $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ are of similar magnitude and their correlation is in the vicinity of one (as the data indicate), the unobserved risk premium and the forward premium are characterized by similar dynamics. The latter result is important because it implies that the highly persistent nature of the forward

logarithm of the domestic price level for the numeraire currency and $rp_t = -Cov(s_{t+1}q_{t+1}|\mathcal{F}_t)$ is the risk premium. The first two conditional second moment terms on the right-hand side (Jensen's inequality terms) are found to be empirically very small at monthly frequency and are typically omitted from the analysis (Bekaert and Hodrick, 1993). Alternatively, one could define $f_t - E(s_{t+1}|\mathcal{F}_t)$ as the 'true' rational expectations risk premium that includes the Jensen inequality terms (Engel, 1996).

premium and the low signal-to-noise ratio in the differenced forward premium regression are most likely due to a small but highly persistent risk premium component.

The typical differenced regression used in the literature is obtained by subtracting s_t from both sides of (1). Let $y_{t+1} = s_{t+1} - s_t$ and $x_t = f_t - s_t$ denote the spot returns and forward premium, respectively. Then, the standard differenced specification is given by

$$y_{t+1} = \alpha_2 + \beta_2 x_t + e_{t+1}, \tag{3}$$

where $e_{t+1} = -rp_t + \varepsilon_{1,t+1}$ if $\alpha = 0$ and $\beta = 1$. Equation (3) is a restricted error-correction representation of model (2). It is worth pointing out that (3) constrains both the long-run and the short-run behavior of the spot and forward rates (Zivot, 2000). More specifically, the model restricts the cointegrating and speed of adjustment parameters to be the same so that the adjustment to the long-run equilibrium takes place in one period.

The properties of the forward premium are specified in the following assumption.

Assumption B. *Assume that the forward premium is generated by*

$$x_{t+1} = \phi_T x_t + \tau_T v_{t+1}, \tag{4}$$

where x_0 *is* $o_p(T^{1/2})$, $v_t = D(L)\upsilon_t = \sum_{j=0}^{\infty} D_j \upsilon_{t-j}$ *with* $\sum_{j=0}^{\infty} j|D_j| < \infty$, $E(\upsilon_{t+1}|\mathcal{F}_t) = 0$, $E(\upsilon_{t+1}^2) = \sigma_\xi^2$, $\sup_t E(\upsilon_{t+1}^4) < \infty$, $\delta = Corr(\upsilon_{t+1}, \xi_{t+1})$, $\phi_T = 1 + c/T$ *for some fixed constant* $c \leq 0$ *and* $\tau_T = \lambda/\sqrt{T}$ *for some fixed constant* $\lambda > 0$ is *the signal-to-noise ratio.*

A few remarks regarding the assumed dynamic behavior of the forward premium are in order. In Assumption B, we reparameterize ϕ_T and τ_T as local-to-unity and local-to-zero sequences in order to mimic the high persistence of the forward premium x_t and the substantially lower variability of the errors driving the forward premium process compared to the variability of the noise component in the forward premium regression. The normalization factors T and $T^{1/2}$ for the local-to-unity and local-to-zero parameterizations are chosen to match the asymptotics of the estimators of ϕ_T and τ_T. The latter follows from the fact that x_{t+1} is observed so that τ_T can be estimated directly from the residuals of model (4) and inherit the same rate of convergence as the estimator of the standard deviation.[2] The local-to-zero parameterization

[2] Alternatively, the local-to-unity and local-to-zero parameterizations can be used for the latent risk premium

has been used recently in a predictive regression framework by Torous and Valkanov (2000) and Moon, Rubia and Valkanov (2004). In a different context, Ng and Perron (1997) adopt a similar parameterization to study the effect of low signal-to-noise ratio of the regressor on the sampling properties of cointegrating vector estimators.

The dual localization proves to be instrumental in producing a process that is stochastically bounded and hence consistent with both statistical and economic theory. Unlike regular near-unit root processes that are of order $O_p(T^{1/2})$, the local-to-zero variance localization dampens the stochastic trend behavior of x_{t+1} and keeps it stochastically bounded ($O_p(1)$).[3] The dual localization removes the economically unappealing possibility that the forward premium can wander off and preserves the cointegration between spot and forward rates. Also, the regression of future spot returns on forward premium is now balanced as both the dependent variable and the regressor are stochastically bounded.

The local-to-unity parameterization of ϕ_T as $1 + c/T$ requires further clarification. The near unit root specification of the forward premium has been used previously by Maynard (2003) and Liu and Maynard (2005). It should be emphasized that this parameterization is chosen as a convenient device to approximate the high persistence in the forward premium process. This high persistence can possibly arise from long memory (Baillie and Bollerslev, 1994, 2000; Maynard and Phillips, 2001), structural instability (Sakoulis and Zivot, 2002) or regime-switching nonlinearity (Bansal, 1997; Baillie and Kiliç, 2006). It is well known that the statistical differentiation between some of these models could be quite challenging in finite samples. Consider, for example, an ARFIMA $(1, d, 0)$ model $(1 - \phi L)(1 - L)^d x_t = v_t$ that can be rewritten as $x_t = (\phi + d)x_{t-1} + \left[\frac{d(1-d)}{2} - \phi d\right] x_{t-2} + ... + v_t$. Then, it is straightforward to see that the AR representations of the near-integrated model with ϕ close to one and d close to 0

process. While this approach specifies directly the dynamics of the more primitive structure of the model in levels, it complicates the form of the asymptotic distributions derived in the paper although the nature of the limiting results is qualitatively similar. Moreover, since the economic theory provides very little guidance regarding the dynamics of the risk premium, the assumptions on the latent risk premium process may appear somewhat arbitrary. Also, our specification of the forward premium can accommodate a wide range of possible parameterizations of the risk premium depending on how large the noise component $(\xi_{t+1} - \varepsilon_{t+1})$, that distorts the relationship between x_{t+1} and rp_t, is. Finally, our assumption on the forward premium can be consistent with both a small random-walk risk premium component and "peso" effects that could help to explain the forward premium puzzle (Evans and Lewis, 1994). We present some simulation results and further discussion for the risk premium specification in Section 4.

[3] More specifically, x_t converges weakly to an Ornstein-Uhlenbeck process without any normalization that depends on the sample size (see Preliminary Lemma in Appendix B).

and the long-memory model with d close to one and ϕ close to 0 are observationally equivalent in finite samples. We use the local-to-unity framework as a suggestive analytical tool that delivers better finite-sample approximations of persistent processes without claiming whether it reflects an underlying data generating process that is more realistic than any of the other persistent specifications. In fact, the main argument of this paper can be easily extended to fractionally integrated or regime-switching processes but at the cost of increased complexity that arises from the non-linear estimation of these models. In contrast, the local-to-unity framework provides a very convenient framework for analysis based on OLS estimation and well developed limiting theory.

To see if the data lend some support to the assumed specification of the forward premium, we estimated the largest AR root and the localizing constant c from the five exchange rate series. The median unbiased estimates and the 90% confidence intervals for these two parameters are reported in Table 3. The results confirm the near unit root behavior of the forward premium with median unbiased estimates of c in the range of -1 to -15. The interval estimates also reveal that, except for the British pound, the presence of a unit root (lack of cointegration between forward and spot rates) cannot be rejected at 5% significance level. Although this finding may be attributed to the small sample size and relatively low power of the test with the chosen long lag length, it is in line with the mixed evidence on cointegration between s_t and f_t (Baillie and Bollerslev, 1994; Zivot, 2000). Note, however, that our local-to-zero parameterization of τ_T suggests that this can only be a finite sample problem since the variance of the forward premium vanishes asymptotically and forward and spot rates are cointegrated as $T \to \infty$. Unlike the previous literature, the vanishing signal-to-noise ratio helps to reconcile the near unit root specification of the forward premium with the economic theory and common intuition.

Lastly, it is instructive to assess if the adopted statistical framework is in agreement with certain conditions derived by Fama (1984) that render the OLS estimate of β negative. Fama (1984) showed that a negative estimate of β requires that $Cov[E(s_{t+1}|\mathcal{F}_t) - s_t, rp_t] < 0$ and $Var(rp_t) > Var[E(s_{t+1}|\mathcal{F}_t) - s_t]$. Note again that if $\alpha = 0$ and $\beta = 1$, $E(s_{t+1}|\mathcal{F}_t) - s_t = x_t - rp_t$ and $x_{t+1} = rp_t + \varepsilon_{2,t+1} - \varepsilon_{1,t+1}$ or, after substituting for x_t, $E(s_{t+1}|\mathcal{F}_t) - s_t = -(rp_t - rp_{t-1}) + (\varepsilon_{2,t} - \varepsilon_{1,t})$. Then, it is straightforward to demonstrate that if, as the empirical evidence from Table 1 suggests, x_t (and hence rp_t) is a near unit root process with an asymptotically vanishing signal-to-noise ratio and $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ are highly correlated with similar variances (so that $Var(\varepsilon_{2,t} - \varepsilon_{1,t}) \approx 0$), then $Var[E(s_{t+1}|\mathcal{F}_t) - s_t]$ is very close to zero and is easily dominated by the variance of rp_t. Also, $Cov[E(s_{t+1}|\mathcal{F}_t) - s_t, rp_t] = Cov[x_t - rp_t, rp_t] < 0$ since $Corr(x_t, rp_t) < 1$.

Even though our model satisfies Fama's conditions, they do not seem to generate large negative values of the slope coefficient in the differenced specification. The reason for this is that the variability of the risk premium implied by (2) and (4) is very small. In Fama's framework, the empirically large negative estimates require a large risk premium that is typically generated by excessive levels of risk aversion (Engel, 1996). Instead, we maintain our assumption of a small slowly varying risk premium but augment the model by taking explicitly into account the fact that the forward premium is only predetermined and not strictly exogenous (Tauchen, 2001). Fama's analysis does not allow for possible feedback between the expectational errors and the forward premium. For example, the expectational errors can be correlated with future values of the forward premium (as in Assumption B) as the traders use the information in $\varepsilon_{1,t+1}$ to update the forward premium at time $t+1$ (Tauchen, 2001). Tauchen (2001) and Liu and Maynard (2005) discuss the possible feedback between the expectational errors and the forward premium and its implications on the distribution of the estimator and the test statistic. We demonstrate below that positive (and possibly large) values of this correlation, combined with the other properties mentioned above, can replicate the sampling properties of the slope estimates, documented in the empirical literature, without the need of large risk premium and excessive levels of risk aversion.

3 Theoretical Results and Discussion

3.1 Forward Premium Regression in Levels

A version of the triangular representation (2) has been considered previously by Phillips (1991) and more recently by Campbell and Yogo (2006) in the context of a predictive regression for stock returns. A convenient way to remove the endogeneity that arises from the correlation between $\varepsilon_{1,t+1}$ and $\varepsilon_{2,t+1}$ is based on the control variable approach. Let $\varepsilon_t = (\varepsilon_{1,t}, \varepsilon_{1,t})'$ denote the joint innovation process with covariance matrix $\Omega = [(\omega_{11}, \omega_{12}), (\omega_{21}, \omega_{22})]$, $\omega_{11} = \sigma_\varepsilon^2$, $\omega_{12} = \omega_{21} = C(1)\rho\sigma_\varepsilon\sigma_\xi$ and $\omega_{22} = C(1)^2\sigma_\xi^2$.

Now let $\varepsilon_{1,t+1} = \gamma\varepsilon_{2,t+1} + \eta_{t+1}$, where $\gamma = \rho\sigma_\varepsilon/[C(1)\sigma_\xi]$, so that the error term η_{t+1} is orthogonal to $\varepsilon_{2,t+1}$ by construction. Substituting for $\varepsilon_{1,t+1}$ into the first equation of (2) yields

$$s_{t+1} = \alpha + \beta f_t + \gamma \triangle f_{t+1} + u_{t+1}, \tag{5}$$

where $u_{t+1} = -rp_t + \eta_{t+1}$. If rp_t is assumed constant, Phillips (1991) shows that the OLS estimation of equation (5) is equivalent to the MLE of the joint system (2).

The next theorem characterizes the limiting behavior of the estimator of β in model (5) under Assumptions A and B.

Theorem 1. *Under Assumptions A and B, the limiting distribution of the estimator of β from (5), denoted by $\widehat{\beta}$, is*

$$\sqrt{T}\left(\widehat{\beta} - 1\right) \Rightarrow -\lambda\frac{\int_0^1 \overline{J}_c(s)\overline{B}_1(s)ds}{\int_0^1 \overline{B}_1(s)^2 ds},$$

where $\Rightarrow$ denotes weak convergence, $J_c(r)$ is an Ornstein-Uhlenbeck process generated by the stochastic differential equation $dJ_c(r) = cJ_c(r)dr + dB_2(r)$, $\{B_1(r), B_2(r) : r \in [0,1]\}$ is a bivariate Brownian motion with covariance matrix Σ and $\overline{J}_c(r) = J_c(r) - \int_0^1 J_c(s)ds$ and $\overline{B}_1(r) = B_1(r) - \int_0^1 B_1(s)ds$ denote the demeaned versions of $J_c(r)$ and $B_1(r)$.

Proof. See Appendix B. ■

Theorem 1 shows that under Assumptions A and B, the estimator of β is consistent with a rate of convergence $\sqrt{T}$. It is interesting to note that even though the error term in (5) contains a near-unit root process, this does not give rise to a spurious regression problem because the importance of this highly persistent component dissipates with the sample size. At the same time, the variance of the near nonstationary component does not shrink fast enough in order to achieve the super-consistency of the cointegration estimator. The estimator is unbiased if the correlation between $B_1(r)$ and $B_2(r)$ is equal to zero. For non-zero correlation, the estimator is biased but the magnitude of the bias is also controlled by λ and is small when λ is in the range of values that are relevant for the forward premium example.

The limiting distribution in Theorem 1 also possesses some interesting features. Due to the presence of a near-integrated component in the error term, the limiting distribution has a form typically found in spurious regressions. However, as pointed out above, the shrinking variance of this near random walk term does not affect the consistency of the estimator although it slows down its rate of convergence. The comparison between the asymptotic behavior of this estimator and the standard OLS estimator of s_{t+1} on f_t also deserves a few remarks. While the OLS estimator is biased and inefficient under Assumptions A and B, the omitted risk premium has only a second-order effect since its variance is dominated by the variance of the error term $\varepsilon_{1,t+1}$. As a result, the super-consistency of the OLS estimator is unaffected but its limiting distribution is much more dispersed due to the larger variance of the error term. In contrast, even though $\triangle f_{t+1}$ has no direct asymptotic effect on the estimator $\widehat{\beta}$, it alters the error term that is now dominated by the omitted risk premium whose implied properties (near unit root and small variance) determine the form of the limiting representation in Theorem 1.

Despite the consistency of the estimator $\widehat{\beta}$, the statistical inference on this parameter has to be conducted with caution since the conventional t-statistic of $H_0 : \beta = 1$ is diverging at rate $T^{1/2}$. This can be easily seen using standard errors that do not account for the serial correlation in the error term. In this case, from the limiting results in Appendix B it follows that the standard error is $O_p(T^{-1})$ while the numerator in the t-statistic is $O_p(T^{-1/2})$. Similar results emerge if the standard error is computed with the conventional HAC estimator with a bandwidth of order $o(T)$. In order to get a well-behaved t-statistic that can be used for inference, one needs to employ a HAC estimator with a bandwidth that is proportional to the sample size (Sun, 2004). This HAC estimator would ensure that the strong dependence in the regression errors is properly captured and the t-statistic is stochastically bounded with a non-degenerate limiting distribution (for details, see Sun, 2004).

The results from estimating model (5) are presented in Table 4. The OLS estimates for β are slightly downward biased[4] with larger standard errors. The estimates from model (5) are very close to 1 for all currencies and the increase in the precision of these estimates compared to the OLS is 7-10 fold. Similar findings that indicate the gains of the estimator in (5) over the OLS estimator and the empirical support for $\beta = 1$ from the levels regression have been reported in Hai, Mark and Wu (1997) and Zivot (2000), among others.

[4] The bias of the OLS estimator in this context is analyzed by Zivot (2000).

3.2 Asymptotic Analysis of the Forward Premium Bias in Differenced Regression

While the evidence from the levels regression of spot and forward rates typically favours the hypothesis that $\beta = 1$, the results from the differenced forward premium regression appear anomalous not only in terms of rejecting overwhelmingly this hypothesis but also in terms of the negative sign and the large magnitude of the slope coefficient. The differenced specification has been historically preferred over the levels regression for two reasons. First, the difference transformation is expected to render the variables stationary and standard statistical inference can be performed. Second, given the strength of the signal coming from the nonstationary variables, the orthogonality of the errors to $\mathcal{F}_t$ and the presence of a risk premium cannot be properly studied in the levels regression.

In what follows, we use the statistical model (equations (3) and (4) in Section 2)

$$
\begin{aligned}
y_{t+1} &= \alpha_2 + \beta_2 x_t + e_{t+1} \\
x_{t+1} &= \phi_T x_t + \tau_T v_{t+1}
\end{aligned}
\tag{6}
$$

to study analytically the sampling behavior of the estimators in the differenced regression.[5] The properties of the forward premium are specified in Assumption B.

In order to visualize the nature of the problem in the differenced regression, Figures 1 and 2 plot the rolling sampling estimates (with rolling window of 50 observations) of β_2, the correlation between e_{t+1} and v_{t+1} and the corresponding signal-to-noise ratio, multiplied by $T^{1/2}$, for GBP and DM.[6] Figures 1 and 2 reveal that the estimates of the slope coefficient is varying widely ranging from -15 to 15 for the GBP and -18 to 12 for the DM.[7] The correlation coefficient is also time-varying and there is a pronounced negative comovement between the slope estimates and the correlation between the residuals from the forward premium regression and the model

[5] Model (6) can also be rewritten in a predictive regression framework as $s_{t+1} - f_t = \alpha + (\beta - 1)x_t + e_{t+1}$. The predictability of excess returns $s_{t+1} - f_t$ and its implications for asset pricing models have been analyzed by Bekaert and Hodrick (1992), Bauer (2001), among others.

[6] The results for the other currencies are similar and are omitted to preserve space.

[7] The instability of the parameter estimates may be due to omitted nonlinearities in the expected returns-forward premium relationship that can be modeled using a regime-switching framework (Bansal, 1997; Baillie and Kiliç, 2006; among others). We do not pursue this possibility in this paper. Instead, it can be shown that some alternative specifications tend to remove almost completely any instability in the parameter estimates (for more details, see Gospodinov, 2008).

for x_{t+1}. The signal-to-noise ratio is initially small and stabilizes at even lower values from the middle of the sample onwards. These graphs suggest that the highly unstable estimates of the slope parameter may be caused by the simultaneous presence of low signal-to-noise ratio, near nonstationarity and endogeneity of the regressor.

The next theorem provides some theoretical guidance towards understanding the extremely volatile and economically unintuitive OLS estimates of the slope parameter from the differenced forward premium regression model.

THEOREM 2. *Under Assumptions A and B, the limiting distribution of the OLS estimator of β_2 from model (6), denoted by $\widehat{\beta}_2$, is given by*

$$\sqrt{T}\left(\widehat{\beta}_2 - 1 + \phi_T\right) \Rightarrow \frac{1}{\lambda}\frac{\int_0^1 \overline{J}_c(s)dB_1(s) + \Lambda_{21}}{\int_0^1 \overline{J}_c(s)^2 ds}, \tag{7}$$

where $\Lambda = E(U_t U_t') + \sum_{k=0}^{\infty} E(U_k U_0')$ and $U_t = (\varepsilon_{2,t}, \upsilon_t)'$.

PROOF. See Appendix B. ■

Several remarks regarding the limiting distribution in Theorem 2 are in order. First, the OLS estimator of β_2 from model (6) is inconsistent for any $\phi \neq 0$ and converges in probability to zero. Interestingly, this is the same result obtained by Maynard and Phillips (2001) but it arises for a completely different reason. While the bias towards zero in Maynard and Phillips (2001) is driven by the unbalanced nature of the regression, the asymptotic bias in Theorem 2 is due to the omitted risk premium. Furthermore, if the errors are serially and mutually uncorrelated, $\sqrt{T}\left(\widehat{\beta}_2 - 1 + \phi_T\right) \Rightarrow N\left(0, \lambda^{-2}\left(\int_0^1 J_c(s)^2 ds\right)^{-1}\right)$ and the estimator $\widehat{\beta}_2$ is normally distributed with mean zero and variance that depends inversely on the signal-to-noise ratio. In the more general case when the errors are correlated, the OLS estimator $\widehat{\beta}_2$ has an additional bias term that vanishes at rate $T^{1/2}$.

It is instructive to analyze the bias of $\widehat{\beta}_2$ in the absence of a risk premium and serial correlation. In this case, $\sqrt{T}\left(\widehat{\beta}_2 - 1\right) \Rightarrow \frac{1}{\lambda}\left[\frac{\delta}{2}\frac{\overline{J}_c(1)^2 - 2c\int_0^1 \overline{J}_c(s)^2 ds - 1}{\int_0^1 \overline{J}_c(s)^2 ds} + (1-\delta^2)^{1/2} z\right]$, where z is a standard normal random variable distributed independently of (B_1, J_c) (Stock, 1991) and $\delta = Corr(\upsilon_{t+1}, \xi_{t+1})$. Since $E\left[\frac{\overline{J}_c(1)^2 - 2c\int_0^1 \overline{J}_c(s)^2 ds - 1}{\int_0^1 \overline{J}_c(s)^2 ds}\right] < 0$, the bias is negative if $\delta > 0$ and positive if $\delta < 0$. While Liu and Maynard (2005) and Tauchen (2001) report small positive or even negative estimates of δ that further robustify the forward premium puzzle, these estimates are obtained from the composite error term that contains a risk premium. In the simulation section, we present evidence that the presence of a risk premium severely distorts the estimation of δ and large positive values of δ are empirically plausible. Furthermore, the fact that the limiting distribution is premultiplied by λ^{-1} and λ is small (in the range 0.05-0.4) has two important implications: (i) it amplifies the bias of $\widehat{\beta}_2$ and (ii) it increases the variability of $\widehat{\beta}_2$ which helps explain the highly volatile nature of the rolling estimates plotted in Figures 1 and 2. Despite this, the high persistence and low signal-to-noise ratio alone cannot generate the large negative slope coefficients from the empirical studies. The inclusion of a risk premium, however, further shifts the distribution of the slope parameter to the left and can help to reproduce numerically the forward premium anomaly. For example, our numerical results in Section 4 indicate that if $\lambda = 0.1$, $\delta = 0.8$, $c = -5$ and $T = 400$, the mean of the (simulated) exact distribution of $\widehat{\beta}_2$ is -1.92 which is in the range of values reported in the empirical literature.

Moon, Rubia and Valkanov (2004) adopt a different parameterization of the signal-to-noise ratio that explicitly links the persistence of the regressor and its local-to-zero variance. In our case, this parameterization can be written as $\tau_T = \lambda\sqrt{-c}/\sqrt{T}$ for $c < 0$ and λ in the limiting representations in Theorems 1 and 2 needs to be replaced by $\lambda\sqrt{-c}$. While the rate of convergence and the form of the asymptotic distributions in Theorems 1 and 2 do not change, there is an extra parameter that affects the dispersion of the estimator through the persistence of the regressor. We do not use this parameterization in our paper for the following reasons. First, the parameterization in Moon, Rubia and Valkanov (2004) rules out the case $c = 0$ and does not allow for the possibility of a small random walk risk premium component as in Evans and Lewis (1994). Also, we intentionally choose to separate the signal-to-noise and persistence parameters because we do not observe empirically that the reduction in the signal-to-noise ratios, plotted in Figures 1 and 2, is accompanied by an increase in the persistence of the forward premium.

Our results are also consistent with the evidence in Bansal and Dahlquist (2000) that the forward premium puzzle is not a pervasive phenomenon and does not seem to be present in some low-income countries with high inflation and inflation volatility. High inflation volatility, for example, is expected to lead to a more volatile froward premium, higher signal-to-noise ratio and, hence, less biased slope estimates. In the context of oil prices, Alquist and Kilian (2008) also report estimates from similar regressions that are close to the theoretically predicted values which appears to be due to the large time variation of the difference between futures and spot prices.

Finally, while the form of the limiting representation (7) in Theorem 2 is similar to that of a regression with near integrated regressors (Cavanagh *et al.*, 1995, for example), it has a slower (root-T) rate of convergence. Similarly to the limiting result in Theorem 1, the rate of convergence is affected by the presence of a near-integrated component in the error term whose variance vanishes at rate $\sqrt{T}$.

Even though this theoretical framework helps us to understand better the statistical behavior of the OLS estimates in the forward premium regression, the asymptotic results are of limited practical importance for conducting inference on the parameters of interest. First, the limiting distribution (7) depends on nuisance parameters which complicates the inference on the parameter β_2 especially because the localizing constant c is not consistently estimable. In this case, one could resort to asymptotically conservative procedures (Cavanagh, Elliott and Stock, 1995; Liu and Maynard, 2005) or subsampling (Politis, Romano and Wolf, 1999). More importantly, the results from Figures 1 and 2 show that the correlation coefficient δ is time-varying and highly unstable. This feature further limits the applicability of these asymptotic and subsampling techniques for inference.

An alternative differenced specification can be obtained by subtracting s_t from both sides of (5) and not from (1) as the standard forward premium regression does. This gives rise to a singe equation, conditional error-correction model of y_{t+1} on x_t and $\triangle f_{t+1}$ which removes a significant source of the sampling bias and variability of the slope estimator in (6). In particular, the improved sampling properties of the estimator in this conditional model arise primarily from its independence of λ which is due to the fact that the signal and the noise component are of similar magnitude. For more details and asymptotic results, see Gospodinov (2008).

4 Simulation Experiment

The data generating process in this Monte Carlo experiment is intended to replicate the main stylized facts in the exchange rate data. In particular, we generate sample paths from an empirical version of model (2), where the parameters are calibrated to match the empirical estimates and properties of the forward premium.

More specifically, the data generating process is

$$s_{t+1} = f_t - rp_t + \varepsilon_{1,t+1} \tag{8}$$

$$f_{t+1} = f_t + \varepsilon_{2,t+1}$$

$$x_{t+1} = (1 + c/T)x_t + (\lambda/\sqrt{T})v_{t+1}, \tag{9}$$

where $x_{t+1} = f_{t+1} - s_{t+1}$ and $(\varepsilon_{1,t+1}, \varepsilon_{2,t+1}, v_{t+1})' \sim iidN(0, \Psi)$ with $\Psi = \begin{bmatrix} 1 & \rho & \delta \\ \rho & 1 & \delta \\ \delta & \delta & 1 \end{bmatrix}$. The risk premium is determined implicitly using that $x_{t+1} = rp_t + \varepsilon_{2,t+1} - \varepsilon_{1,t+1}$.

We consider sample sizes $T = 50$ and 400 that correspond to the rolling sample and the whole sample in our empirical example. The correlation coefficient ρ is set to 0.9999. Since in practice, the estimation of this coefficient is distorted by the presence of a latent risk premium, the average estimated values of ρ in our simulation exercise are around 0.998-0.999 which are the values that are estimated from the actual data. We should note that the value of ρ does not affect our regression results and is useful only for backing out the properties of the risk premium. The values for the other correlation coefficient δ are set to 0, 0.4 and 0.8. Negative values of this correlation coefficient have a symmetric effect and are not reported to preserve space. Finally, the localizing constants c and λ are set to (-20, -5, 0) and (0.1, 0.3), respectively.

We present estimates from three regression specifications (M1, M2 and M3) obtained from 50,000 Monte Carlo replications. Model M1 regresses s_{t+1} on f_t; model M2 is a regression of s_{t+1} on f_t and $\triangle f_{t+1}$ and model M3 is the standard differenced regression of $\triangle s_{t+1}$ on $(f_t - s_t)$. All estimated models include an intercept.

Tables 5 and 6 report some summary statistics (mean, median, 10th and 90th percentiles) of the Monte Carlo estimators of the slope parameter from models M1, M2 and M3. The results show that our theoretical framework can approximate well the bias and the dispersion of the empirical estimates from the forward premium regressions and lend support to the theoretical predictions from Theorems 1 and 2. For example, for $\lambda = 0.1$, $c = 0$ and $\delta = 0.8$, the mean estimate for the OLS estimator from the usual differenced forward premium regression is -5.75 and -2.12 for $T = 50$ and 400, respectively. Our results are also qualitatively similar to the simulation findings in Baillie and Bollerslev (2000) obtained by allowing for a very persistent (long-memory) volatility and varying the importance of the permanent and transitory components in the daily spot rates.

Since the forward discount anomaly is not only that the slope parameters in the differenced forward premium regression are negative but also that they are statistically different than one (based, typically, on the asymptotic normal approximation), we also report the finite-sample behavior of the t-test of $H_0 : \beta_2 = 1$ from model (6). Table 7 presents the 0.025 and 0.05 quantiles of the Monte Carlo distribution of the t-statistic along with the empirical size of the two-sided and one sided t-tests at 5 % significance level using the asymptotically normal critical values. As our analytical results suggest, the distribution of the t-statistic is shifted to the left with critical values that well exceed the standard normal critical values. As a result, the t-test based on the standard normal approximation significantly overrejects especially for large values of δ and small values of c. The critical values and overrejections for $\lambda = 0.3$ are bigger than those for $\lambda = 0.1$ because the omitted term $\phi_T/s.e.(\widehat{\beta}_2)$ is larger for $\lambda = 0.3$ due to the smaller standard error of $\widehat{\beta}_2$ in this case. When the t-statistic is recentered with this term, the critical values and the rejection rates for $\lambda = 0.3$ and $\lambda = 0.1$ are practically identical.

Next, we modify our simulation experiment by replacing the specification for the forward premium (9) with the following process for the risk premium

$$rp_{t+1} = (1 + \widetilde{c}/T)rp_t + (\widetilde{\lambda}/\sqrt{T})\varsigma_{t+1}, \tag{10}$$

where $(\varepsilon_{1,t+1}, \varepsilon_{2,t+1}, \varsigma_{t+1})' \sim iidN(0, \Phi)$ with $\Phi = \begin{bmatrix} 1 & \rho & \widetilde{\delta} \\ \rho & 1 & \widetilde{\delta} \\ \widetilde{\delta} & \widetilde{\delta} & 1 \end{bmatrix}$ and $\rho = 0.9999$.[8] Although this case is not explicitly analyzed in the paper, it is still interesting to investigate numerically the effect of parameterizing the risk premium as a small, near unit root process on the parameters

[8] Another possibility is to parameterize the signal-to-noise ratio as $\widetilde{\lambda}/T$, where the normalization factor T is chosen to match the asymptotics of this parameter in an unobserved component model. More specifically, note that under Assumption A and assuming for simplicity that $c = 0$, model (8)-(10) has a time-varying parameter

of the forward premium regressions. Another reason to consider this model is the fact that our theoretical framework in Section 3 requires a relatively large, constant parameter value of δ in order to explain the large negative bias of the OLS estimator in the differenced regression. In practice, this parameter is highly time-varying (see Figures 1 and 2) with an average value for the whole sample around zero. As we will see below, this may be due to the fact that δ captures the correlation between the risk premium and the expectational errors but this correlation is contaminated by the noise component $\varepsilon_{2,t+1} - \varepsilon_{1,t+1}$ whose magnitude depends on the correlation coefficient ρ.

The Monte Carlo results from parameterization (10) with $\widetilde{\lambda} = 0.05$ are presented in Table 8. Overall, these results are numerically very similar to the results in Table 5. One interesting finding that emerges from this parameterization is that the parameters $\widetilde{c}$, $\widetilde{\lambda}$ and $\widetilde{\delta}$ cannot be estimated directly from the observable dynamics of the forward premium due to the component $\varepsilon_{2,t+1} - \varepsilon_{1,t+1}$ that distorts the relationship between the risk and forward premia. Table 9 reports that with $\rho = 0.9999$, λ is overestimated and is around 0.135 for $T = 50$ and 0.360 for $T = 400$ when the data are generated with $\widetilde{\lambda} = 0.05$ and $\widetilde{c} = 0$. Similarly, for $\widetilde{\delta} = 0.8$, $\widetilde{\lambda} = 0.05$ and $\widetilde{c} = 0$, the estimates for δ from the forward premium are 0.220 and 0.076 for $T = 50$ and 400, respectively. These simulation results provide evidence that may help to reconcile the low estimates of the error correlation and the large bias of the estimates from the usual differenced regression whose magnitude can be supported only by the presence of strong feedback between the expectational errors and the forward premium.

form

$$\begin{aligned} s_{t+1} - f_t &= rp_t + \varepsilon_{t+1} \\ rp_t &= rp_{t-1} + \tau\varsigma_t. \end{aligned}$$

Taking differences, we obtain $\Delta(s_{t+1} - f_t) = \tau\varsigma_t + \Delta\varepsilon_{t+1}$. It is straightforward to show that this model possesses the same autocorrelation structure as the MA(1) model $\Delta(s_{t+1} - f_t) = e_{t+1} - \theta e_t$ with the constraint that $0 \leq \theta \leq 1$. In fact, there exists a one-to-one mapping between the parameters of the two representations τ and θ; namely $\tau = \sqrt{\frac{(1-\theta)^2}{\theta}}$ and $\theta = 1 + \frac{\tau^2 - \sqrt{\tau^4 + 4\tau^2}}{2}$ which are monotonic in θ and τ, respectively. It can be seen that reparameterizing $\tau_T = \lambda/T$ with $\lambda \geq 0$ is equivalent to reparameterizing the MA root θ as local-to-unity $\theta_T = 1 + \phi/T + O_p(T^{-2})$ for $\phi \leq 0$. Therefore, this specification gives rise to the standardization factor T in the signal-to-noise ratio and was used by Gospodinov (2002) for estimating the risk premium from interest rate data. Some numerical and analytical results for this parameterization are available from the author upon request.

5 Concluding Remarks

This paper studies the sampling properties of the slope parameters in levels and differenced forward premium exchange rate regressions. The novelty of the paper is to highlight the interaction of the high persistence with the low variability and endogeneity of the regressor and reconcile some seemingly contradictory results in the forward premium literature. The analysis suggests that the large negative values and highly unstable behavior of the slope estimates in the usual differenced specification, reported in many empirical studies, appear to be due to the simultaneous presence of a risk premium and a number of data characteristics that have not been fully incorporated into the inference procedure.

Our methodological framework also helps to resolve several "sub-puzzles" pertaining to some properties of the exchange rate data. First, the adopted local-to-zero variance parameterization dampens the near unit root behavior of the forward premium and "balances" the forward premium regression as the variables that enter the regression share the same stochastic order of magnitude. Furthermore, the variance localization reconciles some seemingly contradictory evidence regarding the cointegration properties of spot and forward exchange rates.

Finally, while the main focus of the paper is to study the behavior of the estimators in forward premium regressions, the paper delivers other interesting findings that deserve closer attention. For example, our parameterizations and estimation results point to some evidence that the latent risk premium follows a near-integrated process with small variability. Slowly moving habit persistence (Verdelhan, 2008) and cross-country heterogeneity (Sarkissian, 2003) can provide important insights for understanding and mimicking the time series properties of the risk premium. Another important dimension for future research is to relate the findings in this paper to the growing literature on exchange rate predictability and returns from currency speculation (Bacchetta and van Wincoop, 2007; Burnside *et al.*, 2006; among others).

A Appendix A: Data Source and Description

Following Bauer (2001), we construct our data series from daily mid-market observations of spot exchange rates for GBP, DM, JPY, CAD and CHF and one-month Eurocurrency rates for US, UK, Germany, Japan, Canada and Switzerland obtained from Datastream. The spot exchange rates are recorded in British pounds and converted into US dollars using the GBP/US rate. The monthly spot rates are constructed by taking the observation on the last business day of each month. The variable s_t in the paper is a logarithmic transformation of these monthly series.

The monthly interest rate data is also end-of-the month and the annual rates are converted into monthly rates using continuous compounding based on the exact number of days between two consecutive end-of-the-month observations divided by 365 for the British pound and by 360 for all the other currencies (Bauer, 2001). The forward premium is calculated as the difference between the US and the corresponding Eurocurrency monthly rates. Finally, the logarithm of one-month forward rate is obtained from the covered interest parity.

B Appendix B: Mathematical Proofs

B.1 Preliminary Lemma

Let $\varepsilon_{2,t} = C(L)\xi_t$, $v_t = D(L)\upsilon_t$, $U_t = (\varepsilon_{2,t},\ v_t)'$ *and* $B(r) \equiv BM(\Sigma)$ *denote a bivariate Brownian motion with a covariance matrix* Σ. *Let also* $\Sigma = \Lambda + \Gamma'$ with $\Lambda = \Gamma_0 + \Gamma$, $\Gamma_0 = E(U_tU_t')$, $\Gamma = \sum_{k=0}^{\infty} E(U_kU_0')$, $f_t = f_{t-1} + \varepsilon_{2,t} = \sum_{j=1}^{t} \varepsilon_{2,j} + f_0$ *and* $x_t = (1 + c/T)x_{t-1} + \lambda T^{-1/2} v_t = \lambda T^{-1/2} \sum_{i=1}^{t} (1 + c/T)^{t-i} v_i + x_0$. *Then, under Assumptions A and B,*

(i) $T^{-1/2} \sum_{i=1}^{[Tr]} U_i \Rightarrow B(r)$

(ii) $T^{-1} \sum_{t=1}^{T} f_t\varepsilon_{2,t+1} \Rightarrow \int_0^1 B_1(s)dB_1 + \Gamma_{11}$

(iii) $T^{-2} \sum_{t=1}^{T} f_t^2 \Rightarrow \int_0^1 B_1(s)^2 ds$

(iv) $x_t \Rightarrow \lambda J_c(r)$

(v) $T^{-1} \sum_{t=1}^{T} x_t v_{t+1} \Rightarrow \int_0^1 J_c(s)dB_1(s) + \Gamma_{22}$

(vi) $T^{-2} \sum_{t=1}^{T} x_t^2 \Rightarrow \int_0^1 J_c(s)^2 ds$

(vii) $T^{-1} \sum_{t=1}^{T} x_t\varepsilon_{2,t+1} \Rightarrow \int_0^1 J_c(s)dB_1(s) + \Lambda_{21}$

(viii) $T^{-2} \sum_{t=1}^{T} x_{t+1}f_t \Rightarrow \int_0^1 J_c(s)B_1(s)ds,$

where $J_c(r) = \exp(cr) \int_0^r \exp(-cs) dB_2(s)$. ■

B.2 Proof of Theorem 1

Let bars denote the demeaned variables, $\widetilde{f}_t = \overline{f}_t - \left[\sum_{j=1}^{T-1} \overline{f}_j \triangle \overline{f}_{j+1}\right] \left[\sum_{j=1}^{T-1} (\triangle \overline{f}_{j+1})^2\right]^{-1} \triangle \overline{f}_{t+1}$ and note that $-rp_t + \eta_{t+1} = (1-\gamma)\varepsilon_{2,t+1} - \overline{x}_{t+1}$. Then, under $\alpha = 0$ and $\beta = 1$, the estimator $\widehat{\beta}$ has the form

$$\sqrt{T}\left(\widehat{\beta} - 1\right) = \frac{T^{-3/2}(1-\gamma)\sum_{t=1}^{T-1} \widetilde{f}_t \varepsilon_{2,t+1}}{T^{-2}\sum_{t=1}^{T-1} \widetilde{f}_t^2} - \frac{T^{-3/2}\sum_{t=1}^{T-1} \widetilde{f}_t \overline{x}_{t+1}}{T^{-2}\sum_{t=1}^{T-1} \widetilde{f}_t^2}.$$

From results (ii), (iii) and $(viii)$ in Preliminary Lemma, $T^{-2}\sum_{t=1}^{T-1} \widetilde{f}_t^2 = T^{-2}\sum_{t=1}^{T-1} \overline{f}_t^2 + o_p(1)$, $T^{-1}\sum_{t=1}^{T-1} \widetilde{f}_t \varepsilon_{2,t+1} = T^{-1}\sum_{t=1}^{T-1} \overline{f}_t \varepsilon_{2,t+1} + o_p(1)$ and $T^{-3/2}\sum_{t=1}^{T-1} \widetilde{f}_t \overline{x}_{t+1} = T^{-3/2}\sum_{t=1}^{T-1} \overline{f}_t \overline{x}_{t+1} + o_p(1)$. Thus,

$$\sqrt{T}\left(\widehat{\beta} - 1\right) = -\frac{\lambda T^{-2}\sum_{t=1}^{T-1} \overline{f}_t \overline{x}_{t+1}}{T^{-2}\sum_{t=1}^{T-1} \overline{f}_t^2} + o_p(1).$$

Substituting for $x_{t+1} = \lambda T^{-1/2}\sum_{i=1}^{t+1}(1 + c/T)^{t-i} v_i + x_0$ and using the limiting results (iii) and $(viii)$ in Preliminary Lemma, we obtain

$$\sqrt{T}\left(\widehat{\beta} - 1\right) \Rightarrow -\lambda \frac{\int_0^1 \overline{J}_c(s) \overline{B}_1(s) ds}{\int_0^1 \overline{B}_1(s)^2 ds},$$

where $\overline{J}_c(r) = J_c(r) - \int_0^1 J_c(s) ds$ and $\overline{B}_1(r) = B_1(r) - \int_0^1 B_1(s) ds$ denote the demeaned versions of $J_c(r)$ and $B_1(r)$. ■

B.3 Proof of Theorem 2

Using that $\overline{x}_{t+1} = rp_t + \varepsilon_{2,t+1} - \varepsilon_{1,t+1}$, the OLS estimator of β_2 in (6) is given by

$$\widehat{\beta}_2 - 1 = \frac{\sum_{t=2}^{T-1} \overline{x}_t \left(-\overline{x}_{t+1} + \varepsilon_{2,t+1}\right)}{\sum_{t=2}^{T} \overline{x}_t^2}.$$

or equivalently

$$\widehat{\beta}_2 - 1 + \phi_T = -\frac{\frac{\lambda}{\sqrt{T}}\sum_{t=1}^{T-1} \overline{x}_t v_{t+1}}{\sum_{t=2}^{T} \overline{x}_t^2} + \frac{\sum_{t=1}^{T-1} \overline{x}_t \varepsilon_{2,t+1}}{\sum_{t=2}^{T} \overline{x}_t^2} + o_p(1). \tag{11}$$

Substituting for $x_{t+1} = \lambda T^{-1/2}\sum_{i=1}^{t+1}(1 + c/T)^{t-i} v_i + x_0$ and applying results (v), (vi) and (vii) from Preliminary Lemma, we get $\frac{\frac{\lambda}{T}\sum_{t=1}^{T-1} \overline{x}_t v_{t+1}}{\frac{1}{T}\sum_{t=2}^{T} \overline{x}_t^2} = o_p(1)$ and

$$\sqrt{T}\left(\widehat{\beta}_2 - 1 + \phi_T\right) \Rightarrow \frac{1}{\lambda} \frac{\int_0^1 \overline{J}_c(s) dB_1(s) + \Lambda_{21}}{\int_0^1 \overline{J}_c(s)^2 ds}. \tag{12}$$

■

References

[1] Alquist, R., and L. Kilian. (2008). "What do we learn from the price of crude oil futures?" *Journal of Applied Econometrics*, forthcoming..

[2] Andrews, D.W.K. (1991). "Heteroskedasticity and autocorrelation consistent covariance matrix estimation." *Econometrica* 59, 817-858.

[3] Bacchetta, P., and E. van Wincoop. (2007). "Random walk expectations and the forward discount puzzle." NBER Working Paper 13205.

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

[5] Baillie, R.T., and T. Bollerslev. (2000). "The forward premium anomaly is not as bad as you think." *Journal of International Money and Finance* 19, 471-488.

[6] Baillie, R.T., and R. Kiliç. (2006). "Do asymmetric and nonlinear adjustments explain the forward premium anomaly." *Journal of International Money and Finance* 25, 22-47.

[7] Bansal, R. (1997). "An exploration of the forward premium puzzle in currency markets." *Review of Financial Studies* 10, 369-403.

[8] Bansal, R., and M. Dahlquist. (2000). "The forward premium puzzle: different tales from developed and emerging economies." *Journal of International Economics* 51, 115-144.

[9] Bauer, G. (2001). "The foreign exchange risk premium over the long run." Unpublished manuscript, University of Rochester.

[10] Bekaert, G., and R.J. Hodrick. (1992). "Characterizing predictable components in excess returns on equity and foreign exchange markets." *Journal of Finance* 47, 467-509.

[11] Bekaert, G., and R.J. Hodrick. (1993). "On biases in the measurement of foreign exchange risk premiums." *Journal of International Money and Finance* 12, 115-138.

[12] Burnside, C., M. Eichenbaum, I. Kleshchelski, and S. Rebelo. (2006). "The returns to currency speculation." NBER Working Paper 12489.

[13] Campbell, J.Y., and M. Yogo. (2006). "Efficient tests of stock return predictability." *Journal of Financial Economics* 81, 27–60.

[14] Cavanagh, C.L., G. Elliott, and J.H. Stock. (1995). "Inference in models with nearly nonstationary regressors." *Econometric Theory* 11, 1131-1147.

[15] Elliott, G., T.J. Rothenberg, and J.H. Stock. (1996). "Efficient tests for an autoregressive unit root." *Econometrica* 64, 813-836.

[16] Engel, C. (1996). "The forward discount anomaly and the risk premium: A survey of recent evidence." *Journal of Empirical Finance* 3, 123-192.

[17] Engel, C., and K.D. West. (2005). "Exchange rates and fundamentals." *Journal of Political Economy* 113, 485-517.

[18] Evans, M.D.D., and K.K. Lewis. (1994). "Do risk premia explain it all? Evidence from the term structure." *Journal of Monetary Economics* 33, 285-318.

[19] Fama, E. (1984). "Forward and spot exchange rates." *Journal of Monetary Economics* 14, 319–338.

[20] Gospodinov, N. (2002). "Bootstrap-based inference in models with a nearly noninvertible moving average component." *Journal of Business and Economic Statistics* 20, 254-268.

[21] Gospodinov, N. (2008). "A new look at an old puzzle: Understanding the sampling behavior of slope estimators in forward premium regressions." Working paper, Concordia University.

[22] Hai W., N. Mark, and Y. Wu. (1997). "Understanding spot and forward exchange rate regressions." *Journal of Applied Econometrics* 12, 715-734.

[23] Lewis, K.K. (1995). "Puzzles in international financial markets." In G. Grossman and K. Rogoff (eds.), *Handbook of International Economics*, Amsterdam: North-Holland.

[24] Liu, W., and A. Maynard. (2005). "Testing forward rate unbiasedness allowing for persistent regressors." *Journal of Empirical Finance* 12, 613-628.

[25] Marmer, V. (2008). "Nonlinearity, nonstationarity, and spurious forecasts." *Journal of Econometrics* 142, 1-27.

[26] Maynard, A. (2003). "Testing for forward-rate unbiasedness: On regression in levels and returns." *Review of Economics and Statistics* 85, 313-327.

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

[28] Moon, R., A. Rubia, and R. Valkanov. (2004). "Long-horizon regressions when the predictor is slowly varying." Unpublished manuscript, University of Southern California.

[29] Ng, S., and P. Perron. (1997). "Estimation and inference in nearly unbalanced nearly cointegrated systems." *Journal of Econometrics* 79, 53-81.

[30] Phillips, P.C.B. (1991). "Optimal inference in cointegrated systems." *Econometrica* 59, 283-306.

[31] Politis, D.N., J.P. Romano, and M. Wolf. (1999). *Subsampling*, New York: Springer-Verlag.

[32] Sakoulis, G., and E. Zivot. (2002). "Time variation and structural change in the forward discount: Implications for the forward rate unbiasedness hypothesis." Unpublished manuscript, University of Washington.

[33] Sarkissian, S. (2003). "Incomplete consumption risk sharing and currency risk premiums." *Review of Financial Studies* 16, 983-1005.

[34] Stock, J.H. (1991). "Confidence intervals for the largest autoregressive root in U.S. macroeconomic time series." *Journal of Monetary Economics* 28, 435-459.

[35] Sun, Y. (2004). "A convergent t-statistic in spurious regressions." *Econometric Theory* 20, 943-962.

[36] Tauchen, G. (2001). "The bias of tests for a risk premium in forward exchange rates." *Journal of Empirical Finance* 8, 695-704.

[37] Torous, W., and R. Valkanov. (2000). "Boundaries of predictability: Noisy predictive regressions." Unpublished manuscript, UCLA.

[38] Verdelhan, A. (2006). "A habit-based explanation of the exchange rate risk premium." *Journal of Finance*, forthcoming.

[39] Zivot, E. (2000). "Cointegration and forward and spot exchange rate regressions." *Journal of International Money and Finance* 19, 785-812.

Table 1. Summary statistics of exchange rate data.

	s_t	f_t	$f_t - s_t$	$s_{t+1} - f_t$	$s_{t+1} - s_t$
GBP					
mean	0.5229	0.5211	-0.0017	0.0011	-0.0006
std. dev.	0.1408	0.1407	0.0021	0.0307	0.0304
AC(1)	0.968	0.968	0.917	0.087	0.071
DM					
mean	-0.6570	-0.6558	0.0012	0.0000	0.0011
std. dev.	0.1968	0.1959	0.0023	0.0321	0.0318
AC(1)	0.984	0.984	0.958	0.034	0.019
JPY					
mean	-4.9681	-4.9654	0.0027	-0.0011	0.0016
std. dev.	0.3111	0.3108	0.0022	0.0355	0.0350
AC(1)	0.991	0.991	0.937	0.029	0.004
CAD					
mean	-0.2418	-0.2425	-0.0007	0.0004	-0.0003
std. dev.	0.1171	0.1166	0.0013	0.0155	0.0153
AC(1)	0.984	0.983	0.884	0.014	-0.003
CHF					
mean	-0.4962	-0.4935	0.0026	-0.0007	0.0019
std. dev.	0.2284	0.2275	0.0027	0.0357	0.0353
AC(1)	0.981	0.981	0.963	0.055	0.038

Notes: AC(1) denotes the first-order autocorrelation coefficient and s_t and f_t are logarithms of spot and one-period forward rates.

Table 2. Estimation results from the regression of $(s_{t+1} - s_t)$ on $(f_t - s_t)$.

	GBP	DM	JPY	CAD	CHF
$(s_{t+1} - s_t) = \alpha + \beta(f_t - s_t) + error$					
estimate of α	-0.004	0.002	0.009	-0.001	0.006
	(0.002)	(0.002)	(0.003)	(0.001)	(0.003)
estimate of β	-1.736	-0.984	-2.935	-1.135	-1.431
	(0.967)	(0.837)	(0.792)	(0.503)	(0.767)

Notes: Newey-West standard errors with automatic bandwidth.

Table 3. Median unbiased and interval estimates of the largest AR root and the localizing constant c.

	largest AR root		c	
	MUE	90% CI	MUE	90% CI
GBP	0.960	[0.931, 0.993]	-15.03	[-25.87, -2.72]
DM	0.996	[0.973, 1.008]	-1.48	[-10.22, 2.98]
JPY	0.986	[0.960, 1.007]	-4.69	[-13.32, 2.49]
CAD	0.978	[0.954, 1.005]	-8.15	[-17.43, 1.73]
CHF	0.996	[0.973, 1.008]	-1.36	[-10.11, 2.99]

Notes: MUE denotes median unbiased estimate. The median unbiased and interval estimates are obtained by inverting the DF-GLS test for a unit root (Elliott, Rothenberg and Stock, 1996) from a model with 12 lags and a constant (for details, see Stock, 1991).

Table 4. Estimation results from model (2).

	GBP		DM		JPY		CAD		CHF	
	OLS	ML	OLS	ML	OLS	ML	OLS	ML	OLS	ML
α	0.017 (0.008)	0.001 (0.001)	−0.005 (0.007)	0.002 (0.001)	−0.030 (0.037)	0.001 (0.006)	−0.002 (0.002)	0.002 (0.001)	−0.006 (0.005)	−0.001 (0.001)
β	0.970 (0.015)	1.001 (0.002)	0.992 (0.010)	1.005 (0.002)	0.994 (0.007)	1.001 (0.001)	0.991 (0.007)	1.004 (0.002)	0.990 (0.009)	1.004 (0.002)
ρ/γ	0.998	1.008 (0.006)	0.998	1.009 (0.003)	0.998	1.012 (0.005)	0.997	1.009 (0.005)	0.998	1.010 (0.004)

Notes: Newey-West standard errors in parentheses with a bandwidth equal to $0.2T$. OLS denotes the estimates from a regression of s_{t+1} on f_t and ML denotes the estimates from a regression of s_{t+1} on f_t and $\triangle f_{t+1}$. The last row for each currency reports the estimates of ρ (correlation between the errors) and γ, respectively.

Table 5. Monte Carlo simulation results for model (8)-(9) with $\lambda = 0.1$.

	$c=-20$				$c=-5$				$c=0$			
	mean	med	Q10	Q90	mean	med	Q10	Q90	mean	med	Q10	Q90
T=50												
$\delta=0$												
M1	0.90	0.92	0.79	0.98	0.90	0.92	0.79	0.98	0.90	0.92	0.79	0.99
M2	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.01	1.00	1.00	0.99	1.01
M3	0.57	0.56	-10.6	11.8	0.20	0.18	-7.54	7.97	0.09	0.05	-5.51	5.75
$\delta=0.4$												
M1	0.90	0.92	0.79	0.98	0.90	0.91	0.79	0.98	0.89	0.91	0.78	0.98
M2	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	0.99	0.99	0.98	1.00
M3	-1.21	-0.87	-12.5	9.66	-2.30	-1.77	-10.6	5.21	-2.84	-2.21	-9.13	2.54
$\delta=0.8$												
M1	0.90	0.92	0.79	0.98	0.89	0.91	0.78	0.98	0.89	0.91	0.78	0.97
M2	1.00	1.00	1.00	1.00	0.99	0.99	0.99	1.00	0.99	0.99	0.98	1.00
M3	-2.82	-2.22	-14.3	7.77	-4.83	-3.80	-13.6	2.60	-5.75	-4.68	-12.9	-0.04
T=400												
$\delta=0$												
M1	0.99	0.99	0.97	1.00	0.99	0.99	0.97	1.00	0.99	0.99	0.97	1.00
M2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
M3	0.05	0.02	-4.27	4.39	0.04	0.03	-2.77	2.82	0.02	0.02	-1.94	1.95
$\delta=0.4$												
M1	0.99	0.99	0.97	1.00	0.99	0.99	0.97	1.00	0.98	0.99	0.97	1.00
M2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00
M3	-0.77	-0.56	-5.31	3.50	-0.93	-0.69	-3.95	1.81	-1.06	-0.81	-3.27	0.82
$\delta=0.8$												
M1	0.99	0.99	0.97	1.00	0.98	0.99	0.97	1.00	0.98	0.99	0.97	0.99
M2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00
M3	-1.57	-1.19	-6.34	2.68	-1.92	-1.48	-5.26	0.85	-2.13	-1.71	-4.71	-0.07

Notes: The table presents the mean, median, 10th and 90th percentiles of the Monte Carlo distributions of the different estimators.

Table 6. Monte Carlo simulation results for model (8)-(9) with $\lambda = 0.3$.

	$c = -20$				$c = -5$				$c = 0$			
	mean	med	Q10	Q90	mean	med	Q10	Q90	mean	med	Q10	Q90
T=50												
$\delta = 0$												
M1	0.90	0.92	0.79	0.99	0.90	0.92	0.79	0.99	0.90	0.92	0.78	0.99
M2	1.00	1.00	0.99	1.01	1.00	1.00	0.98	1.02	1.00	1.00	0.97	1.03
M3	0.47	0.45	-3.25	4.19	0.20	0.18	-2.39	2.81	0.11	0.09	-1.73	2.00
$\delta = 0.4$												
M1	0.90	0.92	0.79	0.98	0.89	0.91	0.78	0.98	0.88	0.90	0.77	0.97
M2	1.00	1.00	0.99	1.00	0.99	0.99	0.97	1.01	0.98	0.98	0.95	1.01
M3	-0.06	0.03	-3.80	3.53	-0.63	-0.46	-3.30	1.82	-0.88	-0.68	-2.93	0.89
$\delta = 0.8$												
M1	0.90	0.92	0.79	0.98	0.89	0.90	0.78	0.97	0.87	0.88	0.76	0.95
M2	1.00	1.00	0.99	1.00	0.98	0.98	0.97	1.00	0.97	0.97	0.95	0.99
M3	-0.65	-0.43	-4.36	2.76	-1.47	-1.14	-4.32	0.93	-1.84	-1.49	-4.17	0.01
T=400												
$\delta = 0$												
M1	0.99	0.99	0.97	1.00	0.99	0.99	0.97	1.00	0.99	0.99	0.97	1.00
M2	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.01	1.00	1.00	0.99	1.01
M3	0.07	0.06	-1.40	1.54	0.03	0.02	-0.91	0.97	0.01	0.01	-0.63	0.66
$\delta = 0.4$												
M1	0.99	0.99	0.97	1.00	0.98	0.99	0.97	1.00	0.98	0.98	0.96	1.00
M2	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	0.99	0.99	0.98	1.00
M3	-0.22	-0.16	-1.73	1.20	-0.30	-0.23	-1.31	0.60	-0.34	-0.26	-1.07	0.28
$\delta = 0.8$												
M1	0.98	0.99	0.97	1.00	0.98	0.98	0.96	0.99	0.97	0.98	0.96	0.99
M2	1.00	1.00	0.99	1.00	0.99	0.99	0.99	1.00	0.99	0.99	0.98	0.99
M3	-0.49	-0.36	-2.09	0.93	-0.63	-0.48	-1.74	0.29	-0.70	-0.56	-1.57	-0.01

Notes: The table presents the mean, median, 10th and 90th percentiles of the Monte Carlo distributions of the different estimators.

Table 7. Monte Carlo critical values and rejection probabilities of t-test of $H_0 : \beta_2 = 1$.

	$c = -20$				$c = -5$				$c = 0$			
	CV1	CV2	size1	size2	CV1	CV2	size1	size2	CV1	CV2	size1	size2
$\lambda = 0.1$												
$T = 50$												
$\delta = 0$	-2.13	-1.79	0.06	0.07	-2.24	-1.89	0.07	0.08	-2.35	-2.01	0.07	0.09
$\delta = 0.4$	-2.29	-1.94	0.07	0.09	-2.54	-2.21	0.09	0.13	-2.93	-2.58	0.15	0.22
$\delta = 0.8$	-2.41	-2.09	0.08	0.11	-2.83	-2.49	0.13	0.21	-3.36	-3.05	0.31	0.44
$T = 400$												
$\delta = 0$	-2.26	-1.93	0.06	0.09	-2.45	-2.14	0.08	0.12	-2.85	-2.51	0.13	0.20
$\delta = 0.4$	-2.40	-2.08	0.07	0.12	-2.69	-2.39	0.12	0.20	-3.28	-2.97	0.27	0.38
$\delta = 0.8$	-2.54	-2.23	0.09	0.15	-2.92	-2.63	0.19	0.30	-3.60	-3.33	0.51	0.67
$\lambda = 0.3$												
$T = 50$												
$\delta = 0$	-2.30	-1.97	0.07	0.09	-2.56	-2.21	0.09	0.13	-3.00	-2.61	0.14	0.21
$\delta = 0.4$	-2.41	-2.08	0.08	0.11	-2.86	-2.52	0.13	0.21	-3.50	-3.14	0.29	0.40
$\delta = 0.8$	-2.54	-2.21	0.09	0.14	-3.07	-2.75	0.20	0.31	-3.92	-3.59	0.55	0.69
$T = 400$												
$\delta = 0$	-2.87	-2.53	0.14	0.22	-3.58	-3.21	0.31	0.41	-5.18	-4.63	0.56	0.66
$\delta = 0.4$	-2.93	-2.63	0.17	0.27	-3.67	-3.37	0.43	0.56	-5.53	-5.01	0.78	0.85
$\delta = 0.8$	-3.00	-2.72	0.21	0.32	-3.64	-3.39	0.61	0.76	-5.74	-5.23	0.97	0.99

Notes: CV1 and CV2 denote the 0.025 and 0.05 Monte Carlo quantiles (critical values) of the t-test of $\beta_2 = 1$; size1 and size2 denote the empirical rejection probabilities of the two-sided and one-sided t-tests of $\beta_2 = 1$ based on the standard normal critical values at 5% significance level.

Table 8. Monte Carlo simulation results for model (8)-(10) with $\widetilde{\lambda} = 0.05$.

	$\widetilde{c} = -20$				$\widetilde{c} = -5$				$\widetilde{c} = 0$			
	mean	med	Q10	Q90	mean	med	Q10	Q90	mean	med	Q10	Q90
T=50												
$\widetilde{\delta} = 0$												
M1	0.90	0.92	0.79	0.98	0.90	0.92	0.79	0.98	0.90	0.92	0.79	0.98
M2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.01
M3	0.86	0.91	-10.6	12.2	0.66	0.65	-9.24	10.5	0.44	0.35	-7.72	8.67
$\widetilde{\delta} = 0.4$												
M1	0.90	0.92	0.79	0.98	0.90	0.91	0.79	0.98	0.90	0.91	0.79	0.98
M2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00
M3	0.30	0.33	-11.0	11.6	-0.91	-0.79	-10.8	8.72	-2.36	-2.16	-10.6	5.47
$\widetilde{\delta} = 0.8$												
M1	0.90	0.92	0.79	0.98	0.90	0.91	0.79	0.98	0.89	0.91	0.78	0.98
M2	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	0.99	0.99	0.99	1.00
M3	-0.36	-0.29	-11.6	10.8	-2.56	-2.23	-12.2	6.59	-5.21	-4.81	-13.2	2.19
T=400												
$\widetilde{\delta} = 0$												
M1	0.99	0.99	0.97	1.00	0.99	0.99	0.97	1.00	0.99	0.99	0.97	1.00
M2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
M3	0.80	0.78	-3.24	4.83	0.60	0.58	-2.82	4.07	0.41	0.38	-2.38	3.25
$\widetilde{\delta} = 0.4$												
M1	0.99	0.99	0.97	1.00	0.99	0.99	0.97	1.00	0.99	0.99	0.97	1.00
M2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
M3	0.57	0.59	-3.44	4.56	0.03	0.08	-3.42	3.38	-0.60	-0.55	-3.41	2.11
$\widetilde{\delta} = 0.8$												
M1	0.99	0.99	0.97	1.00	0.99	0.99	0.97	1.00	0.98	0.99	0.97	1.00
M2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
M3	0.35	0.39	-3.61	4.27	-0.57	-0.44	-3.94	2.63	-1.60	-1.47	-4.34	0.91

Notes: The table presents the mean, median, 10th and 90th percentiles of the Monte Carlo distributions of the different estimators.

Table 9. Estimates of λ and δ when the data are generated from model (8)-(10) with $\widetilde{\lambda} = 0.05$.

	$\widetilde{c} = -5$			$\widetilde{c} = 0$		
	$\widetilde{\delta} = 0$	$\widetilde{\delta} = 0.4$	$\widetilde{\delta} = 0.8$	$\widetilde{\delta} = 0$	$\widetilde{\delta} = 0.4$	$\widetilde{\delta} = 0.8$
$T = 50$						
estimate of λ	0.126	0.127	0.126	0.134	0.135	0.136
estimate of δ	-0.011	0.129	0.273	-0.013	0.103	0.220
$T = 400$						
estimate of λ	0.338	0.338	0.339	0.359	0.360	0.360
estimate of δ	-0.011	0.041	0.094	-0.013	0.032	0.076

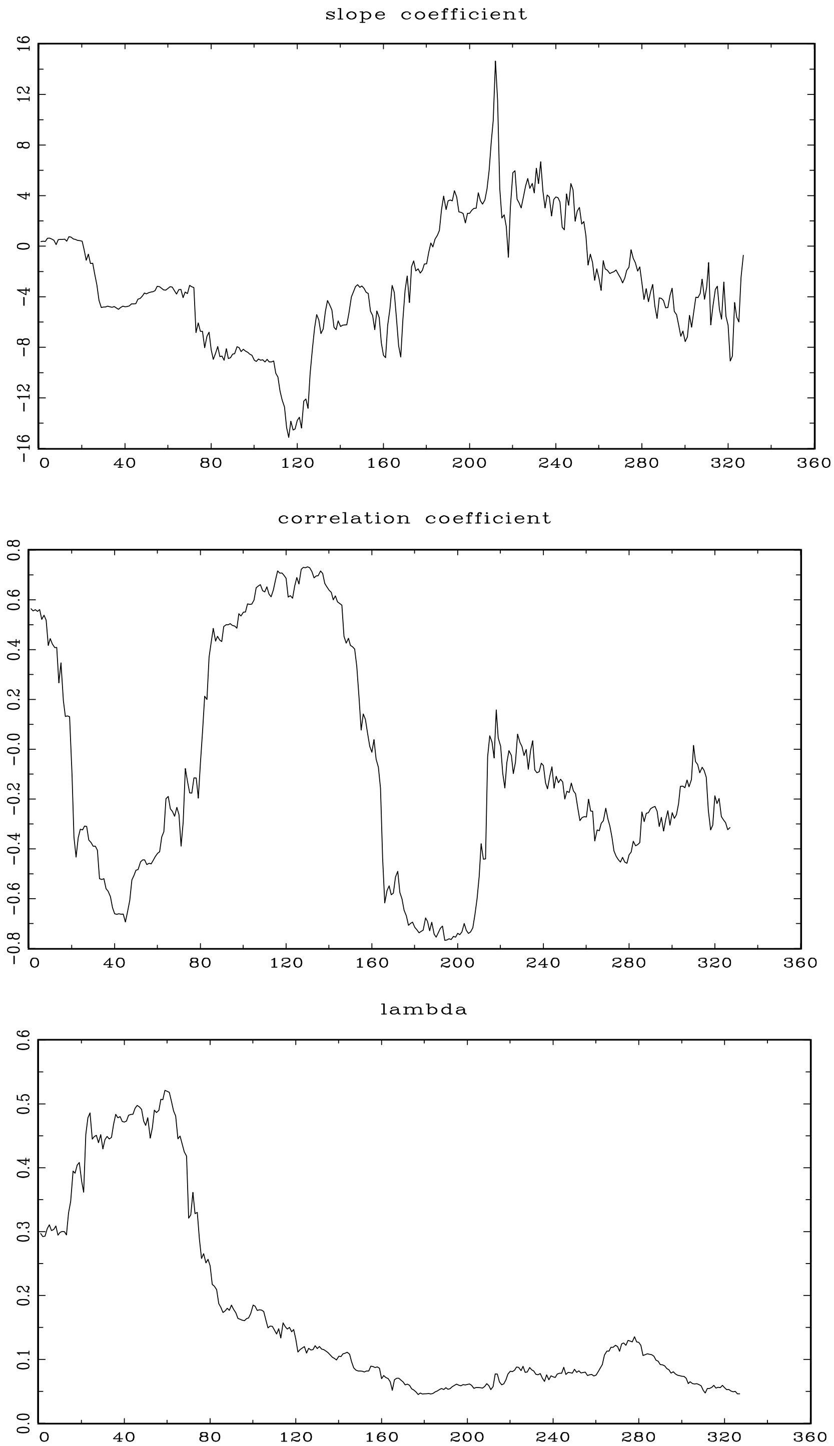

FIGURE 1. Rolling sample estimates of β, δ and λ from the standard differenced regression (6) for British pound. The size of the rolling window is 50 observations. The estimates of δ and λ are obtained from the long-run covariance matrix computed using a Quadratic spectral kernel with an automatic bandwidth (Andrews, 1991).

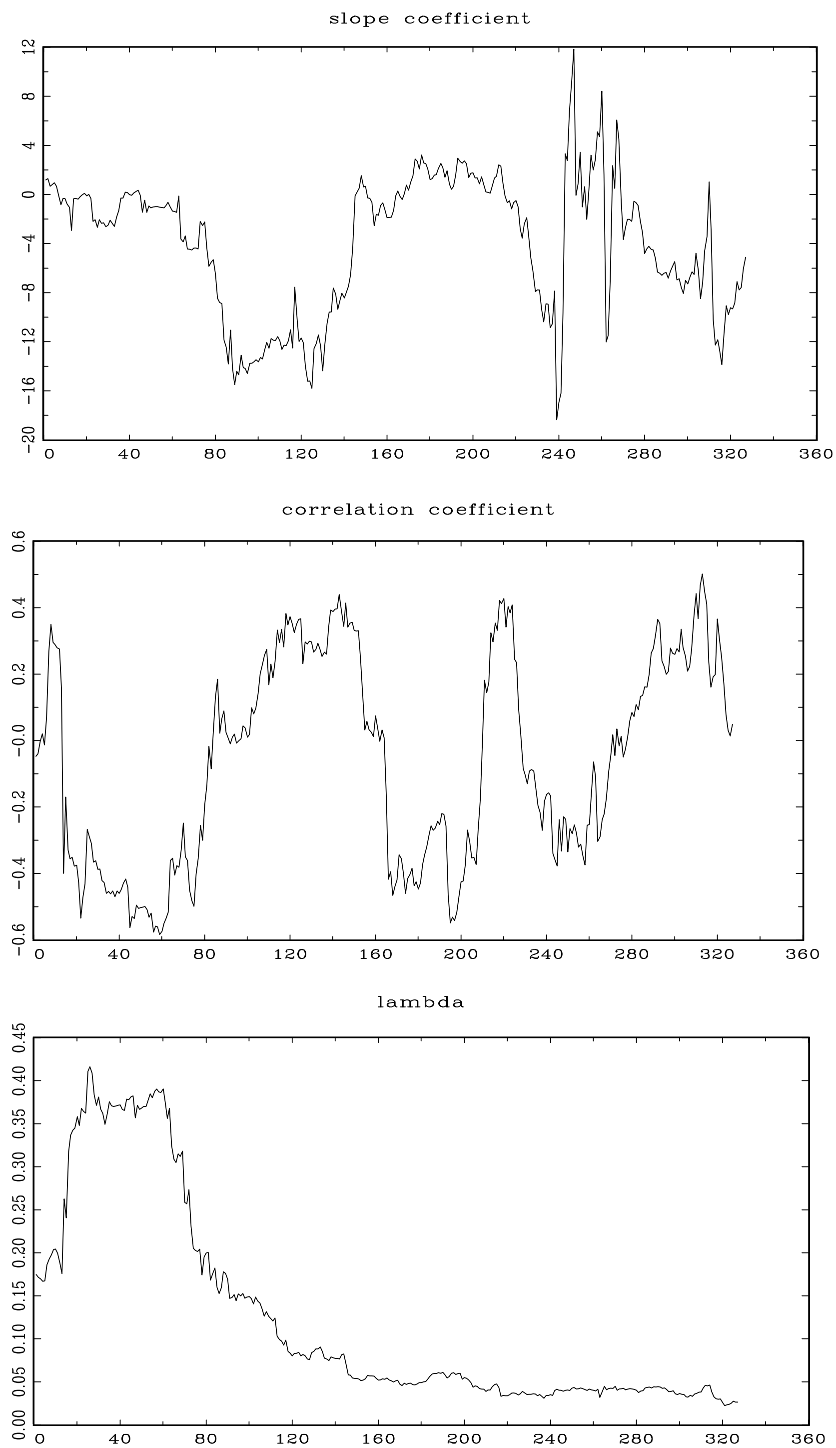

FIGURE 2. Rolling sample estimates of β, δ and λ from the standard differenced regression (6) for German mark. The size of the rolling window is 50 observations. The estimates of δ and λ are obtained from the long-run covariance matrix computed using a Quadratic spectral kernel with an automatic bandwidth (Andrews, 1991).