Open access • Journal Article • DOI:10.1002/JAE.624

Rethinking an old empirical puzzle: econometric evidence on the forward discount anomaly — Source link

Alex Maynard, Peter C.B. Phillips

Institutions: University of Toronto, Cowles Foundation

Published on: 01 Nov 2001 - Journal of Applied Econometrics (John Wiley & Sons, Ltd.)

Topics: Forward premium anomaly and Forward rate

JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **16**: 671–708 (2001)
DOI: 10.1002/jae.624

RETHINKING AN OLD EMPIRICAL PUZZLE: ECONOMETRIC EVIDENCE ON THE FORWARD DISCOUNT ANOMALY

ALEX MAYNARD*[a] AND PETER C. B. PHILLIPS[b]

[a] *Department of Economics, University of Toronto, Canada*
[b] *Cowles Foundation for Research in Economics, Yale University, USA*

SUMMARY

Using both semiparametric and parametric estimation methods, this paper corroborates earlier findings of fractionally integrated behaviour in the forward premium. Two new explanations are also proposed to help reconcile earlier conflicting empirical evidence on the time series properties of the forward premium. Traditional regression approaches used to test the forward rate unbiasedness hypothesis are then evaluated, including regression in levels, in returns (Fama's, 1984, regression), and in error-correction format. Interesting statistical and/or interpretive implications are found in all three cases. For example, the predictions of the appropriate nonstandard limit theory are consistent with many of the standard empirical results reported from Fama's regression, including the commonly occurring, yet puzzling negative correlations between spot returns and the forward premium. It is suggested that the principal failure of unbiasedness, may be due instead to the difference in persistence between these two series. Copyright © 2001 John Wiley & Sons, Ltd.

1. INTRODUCTION

For nearly two decades the failure of forward exchange rates to forecast future spot rates has posed one of the central puzzles in international finance. Most surprisingly, negative estimates from the regression of the spot return on the forward premium suggest that the forward rate predicts changes in the spot rate with the wrong sign. For example, if the forward rate is above the current spot rate, it seems that one should actually expect the spot rate to fall. This conclusion is not only intuitively unappealing, but has also proved extremely difficult to explain using the traditional models of international finance.[1]

This interpretation of these regression results relies to a large extent on the presumed short-memory stationarity of both the spot return and the forward premium. It has been well established that the spot rate itself follows a unit root process, implying a stationary spot return. However, although the short-memory stationarity of the forward premium was once taken for granted, the current evidence from unit root and cointegration tests appears to lead to conflicting conclusions.

Using a semiparametric approach to estimating the fractional differencing parameter and some recent asymptotic theory of Phillips (1999a), and Kim and Phillips (1999a,b), we join Baillie and Bollerslev (1994) in finding evidence of fractionally integrated behaviour in the forward premium. This finding is further confirmed using parametric maximum likelihood estimation on parsimonious

* Correspondence to: Professor Alex Maynard, Department of Economics, 150, St. George Street, Toronto, Ontario, Canada M55 3G7. E-mail: amaynard@chass.utoronto.ca

Contract/grant sponsor: Sloan Foundation; Contract/grant number: SBR 94-22922.
Contract/grant sponsor: NSF; Contract/grant number: SBR 97-30295.

[1] Previous attempts to understand this puzzle have focused on time-varying risk premia, learning, peso problems, and central bank feedback rules. Surveys of the field are provided by Engel (1996), Lewis (**1994**) and Hodrick (1987).

Received 25 September 1998
Revised 30 March 2001

ARFIMA specifications. As Baillie and Bollerslev (1994) also argue, such long-memory behaviour may help to explain why traditional tests have had difficulty in distinguishing between the more extreme alternatives of a unit root and short-memory stationarity. Complementing this explanation, we also identify two potentially strong sources of finite sample bias that may have contributed to the conflicting unit root and cointegration test results mentioned above.

Our estimates of the fractional differencing parameter suggest that traditional statistical theory may not be applicable to many of the regressions commonly used to test forward rate unbiasedness. New limit theories are therefore developed, based on stochastic integrals of fractional Brownian motion, and they are shown to be compatible with the principal empirical regularities from the three most widely employed specifications.

The difference in persistence between the short-memory spot return and long-memory forward premium does not admit a valid regression relation in returns and the slope coefficient in the Fama (1984) regression is found to converge to zero. Given a symmetric distribution this would appear incompatible with the preponderance of negative estimates reported in the literature. However, an interesting feature of the limit theory for the slope coefficient in this regression is that it entails a highly biased and skewed distribution, similar to those found in unit root asymptotics. In particular, the estimated long-run parameters imply a long left tail. These results would seem to cast some doubt on the meaningfulness of the empirical estimates and suggest that the above-mentioned mismatch in persistence may instead bear primary responsibility for the rejection of unbiasedness.[2]

A brief informal discussion of some possible of explanations for the persistent or long-memory behaviour in the forward premium (most of which are not new to this paper) is also provided. Our econometric results suggest that this line inquiry may be worth further investigation.

The paper is divided into seven sections. Section 2 contains a background discussion. Empirical estimates of the fractional differencing parameter are provided in Section 3. Section 4 presents evidence of finite sample biases that may affect unit root and cointegration tests. In Section 5 we derive the econometric implications of the long-memory forward premium for three regression specifications commonly used to test forward rate unbiasedness. In Section 6, we then discuss these results in the context of the relevant empirical and economic literature. Section 7 concludes. Technical lemmas and proofs are contained in the Appendix.

2. BACKGROUND DISCUSSION

Tests of forward rate unbiasedness provided the original motivation for much of the empirical work behind the forward discount anomaly. Written as $E_t s_{t+k} = f_{t,k}$, where s_t and $f_{t,k}$ denote the (log) spot and forward rates and k is the length of the forward contract, the hypothesis states that the forward rate should act as an unbiased forecaster for the future spot rate. It is generally interpreted as a joint test of market efficiency, rational expectations, and risk neutrality. Unbiasedness is also equivalent to Uncovered Interest Parity (UIP),[3] a condition featured prominently in many international macroeconomic models.

[2] Baillie and Bollerslev (2000) and Kuersteiner (1996) contain some similar ideas, although they focus on only one of the three regressions considered here, and give no analytic results. Both authors kindly provided us with copies of their papers during our research on this topic.

[3] This follows by covered interest parity.

Some of the earliest tests of unbiasedness (e.g. Frenkel, 1976) were conducted in exchange rate levels by regressing the future spot rate on the forward rate:

$$s_{t+k} = \alpha_1 + \beta_1 f_{t,k} + \varepsilon_{1,t+k}. \tag{1}$$

The results from these regressions seemed promising at first, since estimates of β_1 were generally 'close' to one. However, due to the non-stationarity in the spot and forward rates, the regression in levels (1) has more recently been interpreted as a cointegrating regression. In this case, as Barnhart and Szakmary (1991) argue, estimates of β_1 close to one indicate only a tendency for the spot and forward rates to move together in the 'long run'. This is a necessary but not a sufficient condition for forward rate unbiasedness.

Following Fama (1984), perhaps the most influential tests have been conducted using a regression in returns,[4] in which the spot return $(s_{t+k} - s_t)$ is regressed on the forward premium $(f_{t,k} - s_t)$:

$$s_{t+k} - s_t = \alpha + \beta(f_{t,k} - s_t) + \varepsilon_{t+k}. \tag{2}$$

Under the null hypothesis one would expect an estimate of β close to one.[5] However, quite surprisingly, empirical estimates of β from (2) are typically negative, and often significantly so. These negative estimates form the central stylized fact underlying the forward discount anomaly. They imply such a strong rejection of unbiasedness that forward rates consistently mispredict even the direction in which the spot rate will move and the UIP condition must be reversed in sign. Furthermore, although such a rejection could in principle be explained by the existence of a risk premium, estimates of this kind also lead to discouraging implications for capital asset pricing models. In particular, Fama (1984) shows that a negative value of β in (2) requires the variance of the risk premium $f_{t,k} - E_t s_{t+k}$ to exceed that of the expected spot return $E_t s_{t+k} - s_t$. Although theoretically possible, such models have difficulty reproducing this inequality under reasonable levels of risk aversion.[6]

The final regressions we consider are the approximate error-correction (ECM) models such as:

$$s_{t+1} - s_t = \alpha_2 + \beta_2(f_{t,1} - s_t) + \gamma' Q_t + \varepsilon_{2,t+1} \tag{3}$$

where Q_t includes lagged differences of the spot and forward rate.[7] This specification adds to the simpler regression in returns by allowing for extra dynamics in the spot and forward rates through the inclusion of Q_t. This regression also yields negative estimates of β_2.

In all three regressions, inference procedures have typically relied upon the presumption of certain short-memory stationarity properties of the data. For example, inference in the returns regression has been conducted under the assumption of a short memory forward premium. Likewise, inference on the cointegrating vector requires the assumption of full I(1) cointegration between the spot and forward rates.[8]

[4] This regression is often referred to as a return/premium or uncovered interest parity regression.
[5] The null hypothesis of unbiasedness in this regression is $\alpha = 0$, $\beta = 1$, and $E_t s_{t+k} = 0$.
[6] See recent surveys by Engel (1996) and Lewis (1994) for further discussion.
[7] We refer to this equation as approximate, because it assumes a cointegrating vector equal to $(-1, 1)$. There is also a second variation of this specification, discussed later.
[8] Short-memory stationarity in the forward premium requires that the cointegration between the spot and forward rates be fully I(1), as well as 1 : 1.

The recent empirical literature provides less than full support for these assumptions. In fact, evidence from both unit root tests and cointegrating regressions have yielded surprisingly conflicting results. Crowder (1994, 1995) and Kuersteiner (1996) fail to reject unit roots in several forward premia series and, using the KPSS test, Crowder is also able to reject stationarity in both of his data sets. However, this conclusion appears to contrast sharply with evidence from similar tests conducted by Hai *et al*. (1997), Horvath and Watson (1994), and Barnhart and Szakmary (1991). Likewise, as Engel (1996) discusses, conflicting evidence has also been reported from tests of one for one (1 : 1) cointegration between the spot and forward rates (i.e. that $\beta_1 = 1$ in 1).[9]

The conflicting nature of the unit root tests suggests that perhaps neither the usual short-memory assumptions nor the unit root model are entirely appropriate for the data at hand. In fact, Baillie and Bollerslev (1994) find that a fractionally integrated model appears to fit the data quite well. They argue that this intermediate result helps to explain the conflicting unit root tests results. After confirming their empirical results, we take the argument further by showing that long memory in the cointegrating errors can also lead to spurious rejection of 1 : 1 cointegration between the spot and forward rates. In addition, we argue that severe data contamination, as well as large differences in magnitude between the spot return and forward premium, may lead to substantial biases in the unit root and cointegration tests. In the remainder of the paper we then consider the implications of this long memory process for the three regressions mentioned above, as well as for the forward discount anomaly more generally.

3. EVIDENCE OF FRACTIONAL INTEGRATION

Testing for a unit root forces a choice between what can be regarded as extreme alternatives. Thus, in the simple autoregression

$$x_t = \alpha + \rho x_{t-1} + \varepsilon_t \tag{4}$$

the long-run behaviour of x_t is very different at $\rho = 1$ and, say, $\rho = 0.99$, having innovations ε_t which are permanent in the first case, yet disappear geometrically in the second. This knife-edge property of the model is sometimes seen as being too rigid to adequately capture the long-run behaviour of certain data. The conflicting results reported from previous unit root tests suggest that this may be relevant in the case of the forward premium, indicating that the fractionally integrated model

$$(1 - L)^d (x_t - \mu) = u_t \tag{5}$$

where u_t is a general short-memory time series, may be better suited to the data.

This second approach allows for a much smoother transition between the two extremes in (4). The fractional parameter d determines the degree of long range persistence in x_t. A value of $d = 0$ implies short-memory stationarity, corresponding to $\rho < 1$ in the autoregressive model,

[9] At an abstract level, the assumption of a unit root has untenable implications for the forward premium. In particular, it implies that with probability one the forward premium will eventually wander off, becoming arbitrarily large, behaviour that might be ruled out *a priori*. Nevertheless, the local wandering behaviour of the forward premium could well be explained by a unit root process, giving rise to inference problems in the regressions mentioned above. It is possible, as Maynard (1998) discusses, that the forward premium could follow a regulated unit root process within a certain distance of the origin. Then, if the finite sample of data did not contain realizations outside the range it would be impossible to identify the nature of the regulation of the process. The fact that the overall scale of the forward premium is tiny in comparison with that of the spot return makes this possibility more compelling, because the time to arrival of the forward premium at a distant boundary is much longer.

while $d = 1$ corresponds to a unit root ($\rho = 1$). The intermediate values of d can then be split into two ranges. For $0 < d < 1/2$ the process remains stationary (with mean μ), but is said to contain long memory, having innovations that disappear hyperbolically rather than geometrically. On the other hand, for $1/2 < d < 1$, x_t is non-stationary (with initialization parameter μ) and is recurrent. Shocks are non-permanent and the unconditional variance grows at a slower rate than in the case of a unit root.

Using daily data on six currencies against the dollar from November 1986 to March 1998 we employ both a parametric and a semiparametric frequency domain approach to estimation of the fractional differencing parameter. The advantages of the semiparametric estimator we use are its robustness to non-stationarity, its capacity to work well over the region $\frac{1}{2} < d < 2$, which is relevant to much economic time series data, and its semiparametric nature, requiring only weak assumptions on the short-memory process u_t. However, as with other semiparametric methods there is the proviso of possible finite sample bias in the estimation that may arise from strongly autoregressive short-memory (see, for example, Agiakloglou *et al.*, 1993). While less robust in large sample, the accompanying parametric ARFIMA estimates, which explicitly allow for autoregressive and moving average terms in u_t, are less prone to finite sample bias. It is reassuring to note that the two sets of estimates match fairly closely.

3.1. Modified Log Periodogram Estimation

Consider the fractionally integrated process x_t, given by (5) where u_t is a short-memory, mean zero time series satisfying some rather weak regularity conditions such as those given in Phillips (1999a). A new semiparametric approach to the estimation of d that applies over stationary and non-stationary domains of d is suggested in Phillips (1999a). The approach uses the exact form of the discrete Fourier transform of a fractionally integrated process and non-linear versions of the log periodogram (Geweke and Porter-Hudak, 1983; Robinson, 1995a) estimator and local Whittle (Künsch, 1987; Robinson, 1995b) estimator are suggested. A simpler version of the log periodogram (LP) procedure involves a straightforward modification to the periodogram ordinates, is easy to apply in practical work and has been justified by Kim and Phillips (1999b) to apply over the range $0 < d < 2$, so it seems to be well suited to economic applications. We will use that procedure in our empirical work.

LP regression involves the least squares regression (over $s = 1, \ldots, m$)

$$\ln(I_x(\lambda_s)) = \widehat{c}_{\mathrm{LP}} - \hat{d}_{\mathrm{LP}} \ln \left|1 - \mathrm{e}^{i\lambda_s}\right|^2 + \text{ error}$$

where $I_x(\lambda_s)$ is the periodogram of x_t at the fundamental frequencies $\lambda_s = 2\pi s/\mathrm{n}$ for $s = 1, \ldots, m$. Modified LP regression involves the similar linear regression

$$\ln(I_v(\lambda_s)) = \widehat{c}_{\mathrm{MLP}} - \hat{d}_{\mathrm{MLP}} \ln \left|1 - \mathrm{e}^{i\lambda_s}\right|^2 + \text{ error} \tag{6}$$

in which the periodogram ordinates, $I_x(\lambda_s)$, are replaced by the modified periodogram ordinates $I_v(\lambda_s) = v_x(\lambda_s)v_x(\lambda_s)^*$ where

$$v_x(\lambda_s) = w_x(\lambda_s) + \frac{\mathrm{e}^{i\lambda_s}}{1 - \mathrm{e}^{i\lambda_s}} \frac{x_n}{\sqrt{2\pi n}}. \tag{7}$$

Since (see Phillips, 1999a),

$$w_{\Delta x}(\lambda_s) = (1 - e^{i\lambda_s}) w_x(\lambda_s) + \frac{e^{i\lambda_s}}{\sqrt{2\pi n}}(x_n - x_0) \tag{8}$$

we have

$$\begin{aligned} I_v(\lambda_s) &= \frac{1}{\left|1 - e^{i\lambda_s}\right|^2} \left| w_{\Delta x}(\lambda_s) + e^{i\lambda_s} \frac{x_0}{\sqrt{2\pi n}} \right|^2 \\ &= \frac{1}{\left|1 - e^{i\lambda_s}\right|^2} \left[I_{\Delta x}(\lambda_s) + 2\frac{x_0}{\sqrt{2\pi n}} \text{ Re } (w_{\Delta x}(\lambda_s) e^{-i\lambda_s}) + \frac{x_0^2}{2\pi n} \right] \end{aligned}$$

or

$$I_v(\lambda_s) = \frac{I_{\Delta x}(\lambda_s)}{\left|1 - e^{i\lambda_s}\right|^2} \tag{9}$$

when $x_0 = 0$. Thus, modified LP regression is equivalent to LP regression with differenced data when $x_0 = 0$, and essentially equivalent to it when $x_0 = O_p(1)$. Kim and Phillips (1999b) show that the modified LP estimator $\widetilde{d}$ is consistent for all $d \in (0, 2)$ and has the following limit theory:

$$\sqrt{m}(\hat{d}_{\text{MLP}} - d) \xrightarrow{d} N\left(0, \frac{\pi^2}{24}\right), \tag{10}$$

for $d \in (\frac{1}{2}, 2)$. Thus, the limit theory for $\hat{d}_{\text{MLP}}$ is the same as that of the conventional LP estimator $\hat{d}_{\text{LP}}$ in the stationary case (Robinson, 1999a; Hurvich, Deo and Brodsky, 1998). By contrast, the usual log periodogram estimator $\widehat{d}$ has a mixed normal limit theory when $d = 1$, as shown in Phillips (1999b) and is inconsistent when $d > 1$ (Kim and Phillips, 1999a). Simulations in Kim and Phillips (1999b) also show $\hat{d}_{\text{MLP}}$ to compare favourably to $\hat{d}_{\text{LP}}$ for $0.5 < d < 1$. Thus, the modified regression (6) is especially useful in the non-stationary case when $d > \frac{1}{2}$.

3.2. Empirical Estimates

Estimates of the fractional differencing parameter are shown in Table I.[10] We include both $\hat{d}_{\text{MLP}}$ and estimates from two simple ARFIMA(p,d,q) specifications, using exact maximum likelihood (ML)

Table I. Semiparametric and parametric estimates for d

	$\hat{d}_{\text{MLP}}$	$\hat{d}_{\text{ML}}$ $p=1$	$\hat{d}_{\text{ML}}$ $q=1$		$\hat{d}_{\text{MLP}}$	$\hat{d}_{\text{ML}}$ $p=1$	$\hat{d}_{\text{ML}}$ $q=1$
Aus$	0.957	0.948	0.965	DM	0.888	0.929	0.957
	(0.032)	(0.021)	(0.027)		(0.033)	(0.017)	(0.022)
Can$	0.937	0.956	0.957	Yen	0.882	0.92924	0.934
	(0.033)	(0.023)	(0.027)		(0.032)	(0.020)	(0.022)
FR	0.885	0.858	0.8691	UK	0.993	0.9583	0.992
	(0.032)	(0.021)	(0.025)		(0.032)	(0.020)	(0.028)

$m = n^{3/4}$.

[10] Due to data errors discussed in the following section, one-month London nominal interest rate differentials are used in place of the forward premium. By covered interest arbitrage these two series are equivalent.

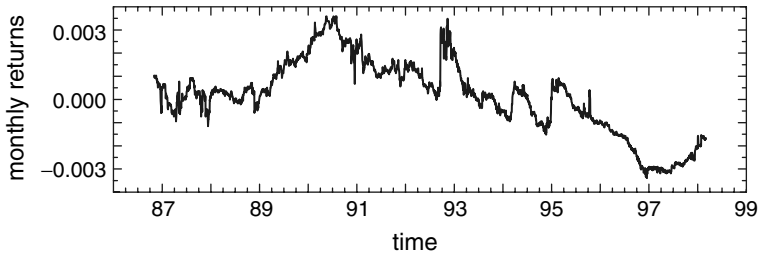

Figure 1. Nominal interest differential (Can$/US$)

estimation.[11] The estimates we obtain are somewhat larger than those first reported by Baillie and Bollerslev (1994), yet still fall into the same general range.[12] Estimating a low-order ARFIMA model on the monthly forward premium from 1974 to 1991, Baillie and Bollerslev (1994) report point estimates for d of 0.45, 0.77, and 0.55 for Canada, Germany, and the UK respectively. In both cases the estimates suggest that the forward premium probably displays less persistence than a unit root process ($d < 1$) yet still too much persistence to satisfy stationarity ($d > 0.5$). It therefore seems reasonable to model the forward premium as a non-stationary fractionally integrated process with $0.5 < d < 1$. (Figure 1 shows the nominal interest differential for the Canadian Dollar.)

4. FINITE SAMPLE BIAS AND CONFLICTING UNIT ROOT AND COINTEGRATION TEST RESULTS

As Baillie and Bollerslev (1994) argue, the fractional integration of the forward premium may help to explain the conflicting unit root tests found in the previous literature[13] and we show later, in Section 5, that similar arguments can be made for the conflicting evidence on tests of 1 : 1 cointegration between the spot and forward rates. However, as we show below, there appear also to be two rather strong sources of finite sample bias, which may complement this explanation.

4.1. Evidence of Data Contamination

Matched daily spot and one-month forward rates from five sources are compared with each other and with the matched daily one-month London Eurocurrency nominal interest differential. The spot and forward rates come from Reuters Instrument Code (RIC), Midlands Bank (MB), Bankers Trust (BT), National Westminster Bank (NW), and Barclays Bank (BB).[14]

[11] Using the ARFIMA package on Ox (see Doornik and Ooms, 1999) the ML estimator was applied to the first differenced series and the resulting estimate of d was then increased by one.

[12] Ongoing work has led us to suspect that this difference primarily reflects the differing sampling periods. The more modern sampling period, which our data covers, appears to display particularly strong long-memory characteristics. Different estimation methods may also play a role. Based on the ARFIMA estimates in Table I, finite sample bias doesn't seem to have played a role and sampling our data on a monthly basis actually resulted in higher (though more variable) estimates (available upon request).

[13] For example, many of the commonly employed unit root tests, including both the ADF and Z_t tests, are based on t statistics from models similar to (4). Sowell (1990) shows that the t statistic in (4) diverges to negative infinity when x_t is fractionally integrated with $d < 1$, suggesting an eventual rejection for large enough sample sizes. On the other hand, unit root tests are also known to have low finite sample power against fractional alternatives.

[14] The data from RIC comes from Data Resources (DRI), the four bank series are from Data Stream International (DS), and the interest rates are from the Financial Times, obtained from DS.

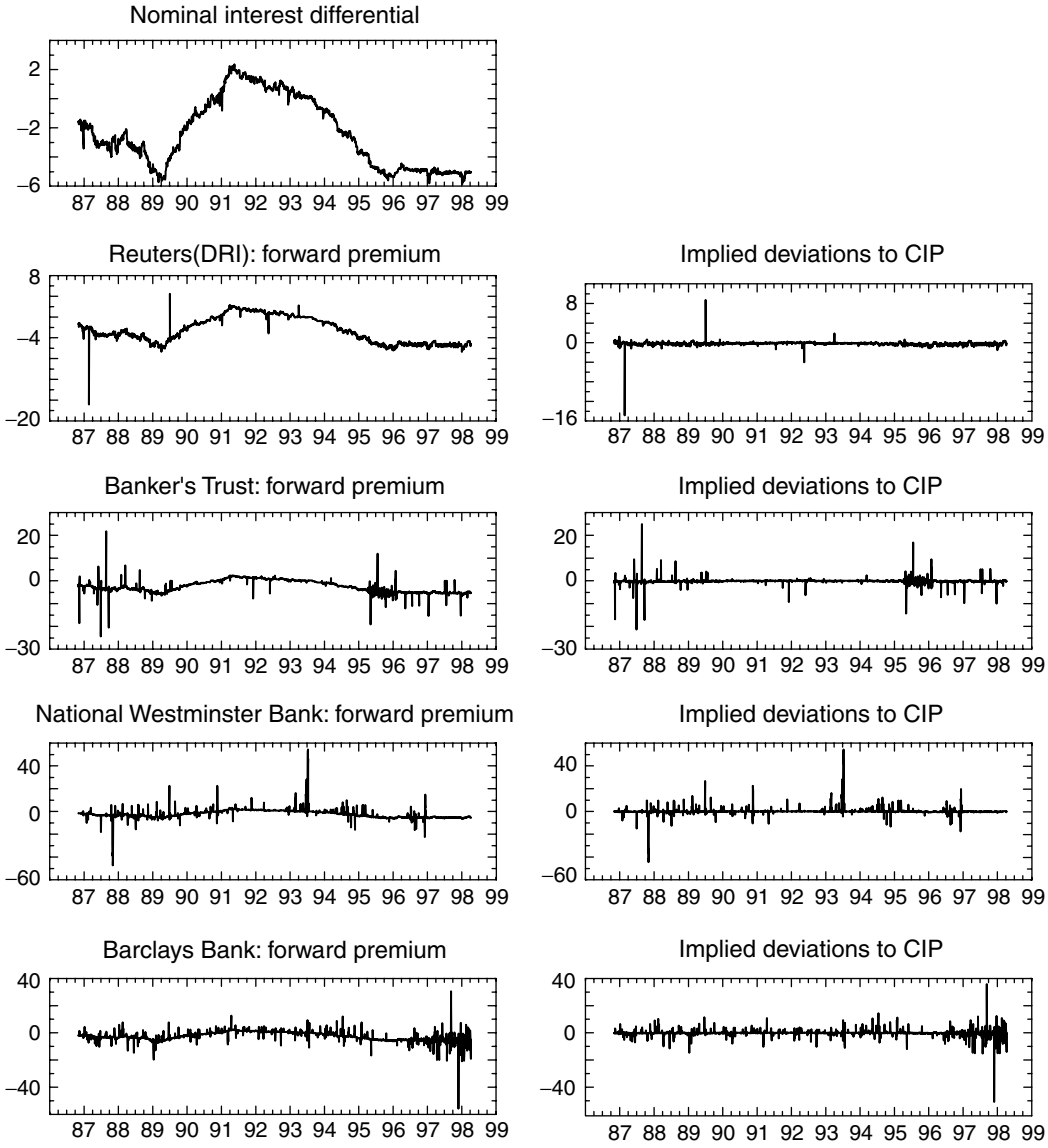

Figure 2. Evidence of data contamination: Matched daily Yen/Dollar rates in annual percentage return

Figure 2 shows the interest differential for Japan in the top-left panel, followed by each of the forward premia in the five panels below.[15] To varying degrees all five forward premia show large one-day fluctuations not found in the interest differential.[16] This is somewhat surprising, since

[15] Midlands Bank is excluded from Figure 2 for considerations of space.

[16] The relative quality of the data from Reuters (RIC) is due to an automatic filter designed to eliminate inaccurate bank quotes.

by covered interest arbitrage the forward premium and nominal interest differential should be identical.[17] The right-hand panels of Figure 2 show the implied deviations from covered interest parity converted into annual percentage returns. Even the more modestly sized spikes imply rather enormous deviations from CIP, which are far in excess of those commonly reported in tests of CIP (see for example, Taylor, 1987, 1989).[18] Furthermore, if there were any underlying economic motivations for the spikes, we would expect them to appear on the same dates regardless of the data source. However, Figure 2 in fact shows that the one-day spikes appear on different days for different sources and are more frequent in some sources than in others.

To understand how errors of this kind could go unnoticed observe that in the top two panels of Plate 1 all five spot and forward rates appear identical.[19] It is only in the bottom panel, which shows the forward premium for each series, that the differences become apparent. This is because the errors are only 'large' relative to the size of the forward premium. They are, in fact, minuscule in comparison with the spot and forward rates themselves. Furthermore, in monthly data it becomes more difficult to distinguish one day spikes from one month movements, suggesting that all but the very largest errors might be difficult to detect, even in the forward premium itself.

An unfortunate consequence of this data contamination is that it tends to obfuscate the true time series properties of the forward premium, creating a clear (finite sample) bias in favor of stationarity. While the interest differential shown in Figure 2 displays strong signs of a stochastic trend, this same trend is far less apparent in the forward premia shown below. The one-day spikes have effectively masked the underlying persistence in the true forward premium.

Table II shows the effect of the data contamination on the estimates of d, as well as on the estimate of the autoregressive parameter. The data sources are roughly ordered from the cleanest to the most contaminated. The results generally confirm the visual evidence shown in Figure 2, with the least contaminated series showing the highest levels of persistence.[20] In an earlier version of the paper (available by request) we showed that the measurement error had similar effects on unit root tests. Not surprisingly, the standard tests tended to provide greater evidence against a unit root the greater the degree of data contamination. To minimize the impact of data errors throughout the

Table II. Effects of measurement error on measures of persistence (Yen/US$)

Data source	AR coeff.	$\widehat{d}_{\mathrm{MLP}}$	$\widehat{d}_{\mathrm{Ml}}$ $p=1$	$\widehat{d}_{\mathrm{Ml}}$ $q=1$	Data source	AR coeff.	$\widehat{d}_{\mathrm{MLP}}$	$\widehat{d}_{\mathrm{ML}}$ $p=1$	$\widehat{d}_{\mathrm{Ml}}$ $q=1$
$i-i*$	0.999	0.88	0.93	0.94	BT	0.818	0.31	0.43	1.05
RIC	0.978	0.60	0.64	1.15	NWB	0.788	0.30	−0.36	0.47
MB	0.870	0.34	0.48	1.05	BB	0.629	0.24	0.34	0.36

[17] The exact formula for CIP is $\ln(1+i_{t,k})-\ln(1+i^*_{t,k})=f_{t,k}-s_t$. However, replacing this by the approximation $i_{t,k}-i^*_{t,k}=f_{t,k}-s_t$ did not produce any noticeable differences.

[18] Since the data fails to satisfy the criteria set out by Taylor (1987) that it be recorded 'at the same instant in time in which a trader could have dealt', CIP need not hold exactly in this data set. Transactions costs have also been ignored, but these are estimated by Clinton (1988) as no more than 0.1% in annual returns. The implied deviations shown in Figure 2 seem far too large to be accounted for by either of these reasons. Japan did experience some large deviations to CIP in the 1970s due to capital controls. However, according to Ito (1992, p. 323), CIP has 'held without exception' since these controls were abolished in 1980.

[19] Bekaert and Hodrick (1993) also report finding errors in the forward premium, but do not provide further detail.

[20] For $q=1$ we were surprised to find that a few of the contaminated estimates of d turned up larger than one, technically implying explosive behaviour. However, these estimates were accompanied by estimates of the moving average parameter close to negative one, suggesting that this more likely reflects a problem of identification.

rest of the paper spot rates are taken from DRI and both the forward rate and forward premium are calculated implicitly by CIP, using the nominal interest differential.

4.2. Scale Differences and Evidence from Tests of Cointegration

Together with the standard assumption of a unit root in the spot rate, a fractionally integrated forward premium $(f_t - s_t)$ implies 1 : 1 fractional cointegration between the spot and forward rates. Trivially,

$$s_t = f_{t,k} + \varepsilon^*_{1,t}$$

where the fractionally integrated residual $\varepsilon^*_{1,t}$ is simply the (negative) forward premium. Likewise, regression (1) may be rewritten as

$$s_{t+k} = f_{t,k} + \varepsilon_{1,t+k}$$

where $(\alpha_1, \beta_1) = (0, 1)$ and the residual

$$\varepsilon_{1,t+k} = s_{t+k} - f_{t,k} = (s_{t+k} - s_t) - (f_{t,k} - s_t)$$

is also integrated of order d.

As argued in Section 5, this suggests that the long-memory property of the forward premium may bear some responsibility for the conflicting tests of 1 : 1 cointegration between s_t and $f_{t,k}$ reported in the literature (see Engel, 1996). However, it also appears that cointegration tests may be further biased when based on $(s_{t+k}, f_{t,k})$, as in (1), rather than on the contemporaneous rates $(s_t, f_{t,k})$. Given the focus on prediction, tests on based on $(s_{t+k}, f_{t,k})$ naturally have intuitive appeal. Nevertheless, cointegration of $(s_t, f_{t,k})$ is equivalent to cointegration of $(s_{t+k}, f_{t,k})$ and it turns out that use of the latter adds considerable noise into the cointegrating residual. This extra noise may substantially impact the finite sample behaviour of the test statistic. An informal discussion of this problem is given below. A more rigorous treatment would require either an extensive simulation or the use of small sigma asymptotics.

Cointegration tests on $(s_t, f_{t,k})$ resemble unit root tests on the forward premium itself, while cointegration tests on $(s_{t+k}, f_{t,k})$ resemble unit root tests on $\varepsilon_{1,t+k}$, the sum of the spot return and the (negative) forward premium. In either case, given the stationary I(0) behaviour of $s_{t+k} - s_t$, the order of integration of the regression residual is determined solely by $f_{t,k} - s_t$. The inclusion of $s_{t+k} - s_t$ in $\varepsilon_{1,t+k}$ simply adds additional noise that could bias the results in finite sample. For example, when non-overlapping data is used ($k = 1$), the nearly independent behaviour of $s_{t+k} - s_t$ (see, for example, Baillie and Bollerslev, 2000) would tend to make $\varepsilon_{1,t+k}$ appear less persistent, possibly inducing a bias in favor of cointegration. With overlapping data ($k > 1$) $s_{t+k} - s_t$ is not only noisy, but also follows an MA(k) (see for example, Baillie and Osterberg, 1997). In this case, there is also a possible finite sample bias against cointegration due to the short-run persistence in $s_{t+k} - s_t$.

One might normally expect such potential biases to be of minor importance in reasonable sample sizes and any such biases of course disappear as $n \to \infty$. What makes them more problematic in the current context is the fact that the extra noise term $s_{t+k} - s_t$ dwarfs the term of interest $f_{t,k} - s_t$ in sheer magnitude (see Plate 2 and Table III). Consequently, as shown in Plate 4, the estimated cointegrating residual $(\hat{\varepsilon}_{1,t+k})$ using $(s_{t+k}, f_{t,k})$, is almost indistinguishable from the spot return itself. It therefore seems unlikely that we can learn much about the nature of the persistence

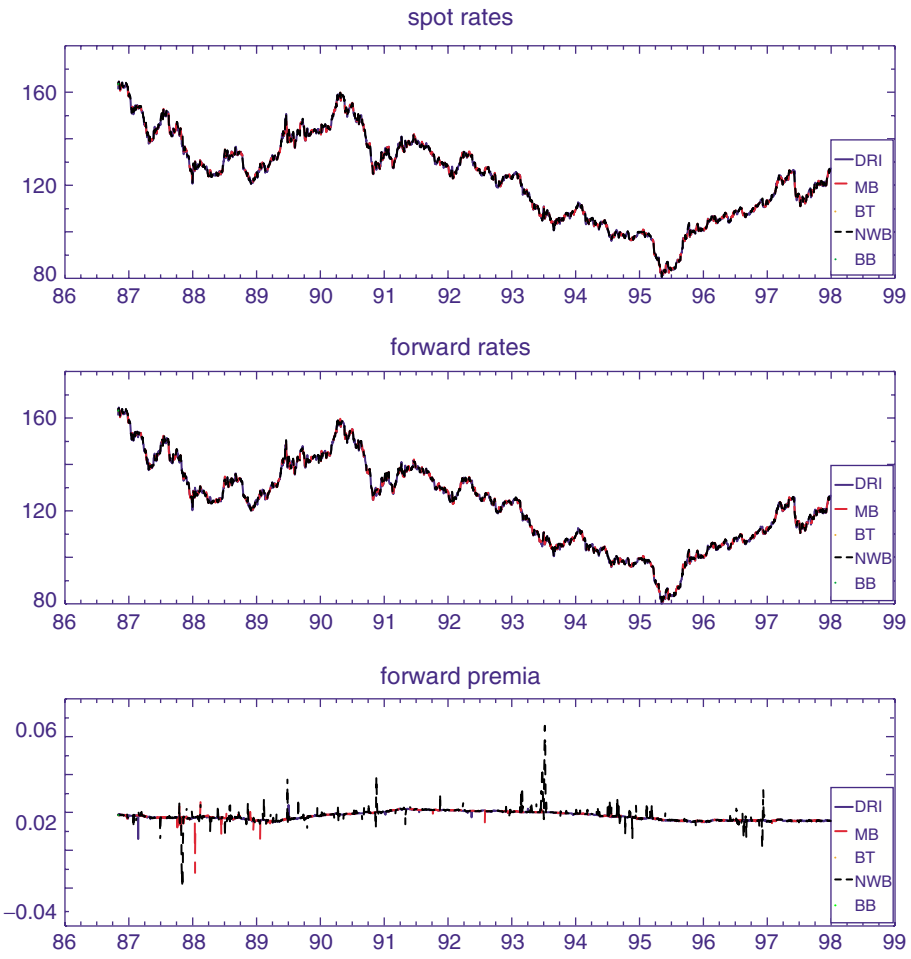

Plate 1. Comparing data from several sources: matched Yen/Dollar rates

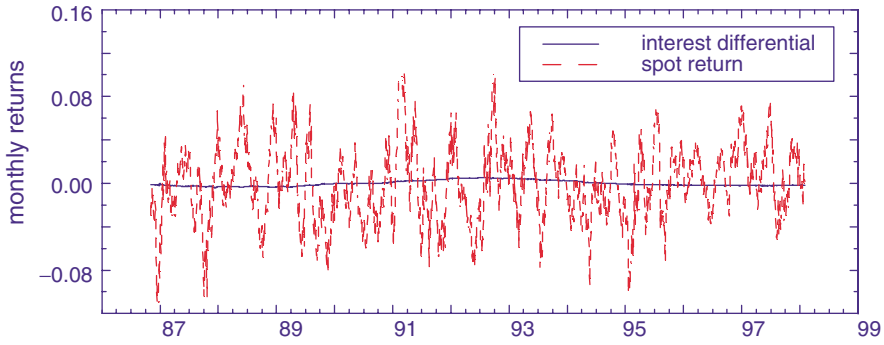

Plate 2. Spot return plotted with nominal interest differential (DM/US$)

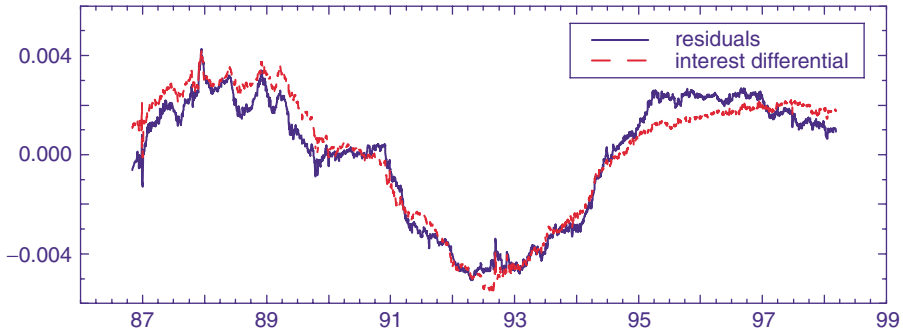

Plate 3. Residuals from regression of s_t on $f_{t,k}$ plotted with the (mirror image) of the nominal interest differential (DM/US$)

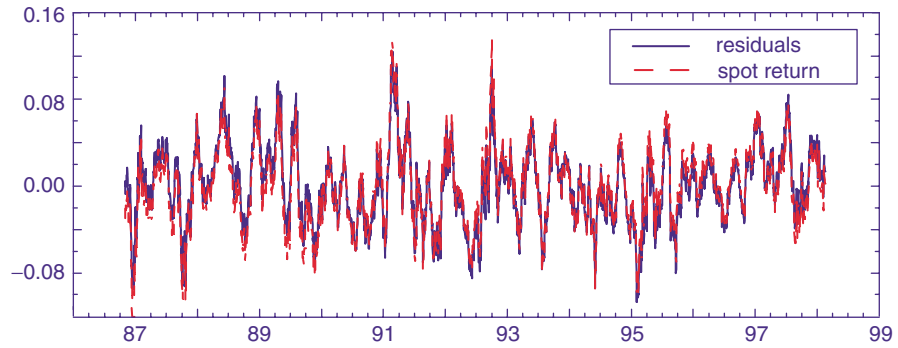

Plate 4. Residuals from regression of s_{t+k} on $f_{t,k}$ plotted with the spot return: $s_{t+k} - s_t$ (DM/US$)

Table III. Sample standard deviations (in annual %)

	$s_{t+k} - s_t$	$i_{t,\cdot} - i^*_{t,\cdot}$	$(1-L)^d (i_t - i^*_t)$		$s_{t+k} - s_t$	$i_t - i^*_t$	$(1-L)^d (i_t - i^*_t)$
Aus$	0.91	0.081	0.0078	DM	1.24	0.089	0.0035
Can$	0.48	0.055	0.0047	Yen	1.30	0.068	0.0034
FR	1.17	0.083	0.0083	UK	1.24	0.067	0.0051

$k = 25$.

Table IV. Semiparametric estimates of d: spot return and forecast error

	$\hat{d}_{\mathrm{LP}}$			$\hat{d}_{\mathrm{LP}}$	
	$s_{t+1} - s_t$	$s_{t+1} - f_t$		$s_{t+1} - s_t$	$s_{t+1} - f_t$
Aus$	0.12	0.18	DM	−0.06	0.00
Can$	−0.01	0.03	Yen	0.11	0.17
FR	0.00	0.06	UK	−0.04	0.00

$m = n^{3/4}$. The standard error is 0.10 for the MLP estimator.

in the forward premium from $\varepsilon_{1,t+k} = s_{t+k} - f_{t,k}$. By contrast, as shown in Plate 3, the estimated residual using $(s_t, f_{t,k})$ quite closely matches the mirror image of the forward premium.[21]

It appears that this extra noise term can strongly influence estimation and test results. For example, estimates of the fractional parameter for $\varepsilon_{1,t+k} = s_{t+k} - f_{t,k}$ in Table IV[22] nearly equal estimates for $s_{t+k} - s_t$, but look nothing like the estimates for $f_{t,k} - s_t$ reported in Table I. Likewise, residual-based cointegration tests shown in Table V strongly reject no cointegration using $(s_{t+k}, f_{t,k})$, yet fail to find as compelling evidence for full I(1) cointegration using $(s_t, f_{t,k})$.

5. ECONOMETRIC IMPLICATIONS FOR COMMON REGRESSION SPECIFICATIONS

This section derives the econometric implications of the long-memory properties of the forward premium for the three common regression specifications mentioned above: regressions in returns, in levels, and in error-correction format. The technical results are presented here and further discussion of the empirical and economic implications follows in the next section.

[21] A similar point is also made in recent work by Newbold *et al*. (1998). Working in an ARIMA context, they identify a large negative moving average term in $s_{t+1} - f_{t,1}$ and argue that this would help to conceal the apparent unit root found in the forward premium.

[22] These estimates were computed using data sampled at every $k = 22$ business days. Estimates using daily data appeared to suffer from an upward bias, probably as a consequence of the large-order moving average process in the spot return that results from the use of overlapping data. In particular, estimates of d for the k-period spot return using daily data were far larger than those for the one period spot return using either daily or monthly data (available upon request). By contrast, the overlapping nature of the data does not in any way appear to be responsible for the finding of long memory in the forward premium. In fact, using monthly data, we obtained even larger estimates than those reported in Table I (available upon request).

Table V. Residual-based cointegration tests

	s_{t+k} and $f_{t,k}$		s_t and $f_{t,k}$			s_{t+k} and $f_{t,k}$		s_t and $f_{t,k}$	
	z_α	z_t	z_{alpha}	z_t		z_α	z_t	z_{alpha}	z_t
Aus$	-128.5^{c}	-8.07^{c}	-8.2	-2.70	DM	-101.3^{c}	-7.17^{c}	-2.0	-0.98
Can$	-120.0^{c}	-7.78^{c}	-22.9^{b}	-3.58^{b}	Yen	-95.0^{c}	-6.94^{c}	-2.3	-0.91
FR	-105.1^{c}	-7.31^{c}	-9.5	-2.12	UK	-93.5^{c}	-6.88^{c}	-9.7	-2.47

[a]Rejection at 10%, [b]at 5%, [c] at 1%. The critical values are -27.9, -20.5, and -16.9 for z_α and -3.88, -3.36, and -3.04 for z_t.

5.1. Assumptions and Notation

A simple econometric model is used to re-evaluate regressions (1), (2), and (3) in light of the non-stationary long-memory properties of the forward premium. Employing covered interest arbitrage as the only underlying structural relationship in Assumption I, we treat all three regressions simultaneously, without imposing any controversial economic or econometric specifications.[23]

Assumption II characterizes the nominal interest rate differential as a nonstationary fractionally integrated process. This follows the empirical evidence presented both here and in Baillie and Bollerslev (1994). The spot rate is modelled as a unit root process in (III), an assumption which is by now non-controversial, and the forward rate is defined implicitly by CIP in (I).

Using the approach of Phillips and Solo (1992) we then represent the innovations to the spot rate and interest differential jointly, in terms of a linear process in (IV). However, in order to ensure weak convergence to fractional Brownian motion we are forced to use stronger moment and homogeneity assumptions in (V) than are required by Phillips and Solo (1992).[24] The very large difference in scale between the spot rate innovations v_t and the innovations to the interest differential ξ_t shown in Table III also plays a central role in all three regressions, but is not modelled formally.

Assumptions

(I) $(CIP) f_{t,k} - s_t = i^{+}_{t,k}$, *where* $i^{+}_{t,k} = i_{t,k} - i^{*}_{t,k}$.
(II) $i^{+}_{t,k} = (1-L)^{-d}\xi_t$ *for* $\frac{1}{2} < d < 1$.
(III) $\Delta s_t = v_t$ *and* $s_0 = v_0 = 0$.
(IV) $(v_t, \xi_t)' = C(L)\varepsilon_t$, *where* $C(L) = \sum_{j=0}^{\infty} c_j L^j$ *and* $\sum_{j=1}^{\infty} j\,|c_j| < \infty$.
(V) $\varepsilon_t' = (\varepsilon_{1,t}, \varepsilon_{2,t}) \sim i.i.d.(0, I)$ *and* $E\,||\varepsilon_t||^g < \infty$ *for*

$$g = \max\left(4, \frac{4(1-d)}{d-\frac{1}{2}}\right)$$

[23] By assuming that CIP holds exactly we abstract away from the data contamination discussed earlier. So long as the data contamination process has at least two plus epsilon moments and is uncorrelated with the other variables this should not affect our results. Otherwise, the presence of data contamination can cause additional complications, not captured by our analytic results.

[24] See Sowell (1990). The moment condition in V is more restrictive as $d \searrow \frac{1}{2}$. For $d = \frac{3}{4}$, fourth moments are required, whereas for $d = \frac{5}{8}$, twelfth moments are needed. The latter is strong for exchange rate data. However, it is not yet clear whether moment conditions as strong as these are actually necessary.

Notation

The linear process representation helps to simplify notation. Let $W(r)$ denote a two dimensional vector of independent standard Brownian motions, and define $B(r) = (B_s(r), B_i(r))' = C(I)W(r) = BM(\Omega)$, where

$$\Omega = \begin{pmatrix} \Omega_{\nu\nu} & \Omega_{\nu\xi} \\ & \Omega_{\xi\xi} \end{pmatrix} = C(I)C(I)'$$

is the long-run covariance matrix of $(\nu_t, \xi_t)'$. By standard functional limit theory (e.g. Phillips and Solo, 1992)

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}\varepsilon_t \Rightarrow W(r),\ \frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}\nu_t \Rightarrow B_s(r),\ \text{ and } \frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}\xi_t \Rightarrow B_i(r)$$

where $\Rightarrow$ signifies convergence in distribution. Standardizing B_s and B_i as $W_s = \Omega_{\nu\nu}^{-1/2}B_s$ and $W_i = \Omega_{\xi\xi}^{-1/2}B_i$ and employing the conditioning formula given in Phillips (1989), B_s can then be decomposed as

$$B_s = \Omega_{\xi\nu}\Omega_{\xi\xi}^{-1/2}W_i + \sigma_{s|i}W_{s\cdot i} \tag{11}$$

where $W_{s\cdot i}$ is independent of W_i and $\sigma_{s|i}^2 = \Omega_{\nu\nu} - \Omega_{\xi\nu}^2\Omega_{\xi\xi}$ is the variance of B_s conditional on B_i.

Two-dimensional standard fractional Brownian motion is then defined as

$$W_{(q)}(r) = \frac{1}{\Gamma(q+1)}\int_0^r (r-x)^q dW(x)$$

for $-0.5 < q < 0.5$ and extended to fractional Brownian motions with more general covariance structures by

$$\begin{aligned} B_{(q)}(r) = C(I)W_{(q)}(r) &= \frac{1}{\Gamma(q+1)}\int_0^r (r-x)^q C(I)dW(x) \\ &= \frac{1}{\Gamma(q+1)}\int_0^r (r-x)^q dB(x) \end{aligned}$$

for $-\frac{1}{2} < q < \frac{1}{2}$. We also define $W_{i(q)}(r) = \Omega_{\xi\xi}^{-1/2}B_{i(q)}(r)$.

The following notational conventions are also used throughout the remainder of the paper. A bar under a random variable represents deviations from the mean in the discrete case and deviations from the Lebesque integral from 0 to 1 in the continuous case. For example, $\underline{i}_{t,k}^+ = i_{t,k}^+ - \frac{1}{n}\sum_{r=1}^n i_{r,k}^+$ and $\underline{B}(r) = B(r) - \int_0^1 B(r)\,dr$. Weak and almost sure convergence are represented by $\Rightarrow$ and $\rightarrow$ respectively. The dependence of discrete variables on k and of continuous random variables, such as $B(r)$, on r, as well as the integrand dr and the limits of integration when integrating over (0, 1), are often suppressed for notational convenience. For example, $i_{t,k}^+$ is written as i_t^+, and $\int_0^1 B(r)dr$ as $\int B$.

5.2. Implications for the Regression in Returns

The simple model described in Section 5.1 leads to some interesting implications for the regression in returns (2). Define $\widehat{\beta}$ as the OLS estimate of β in (2) and the modified statistic

$$\widehat{\beta}_m = \widehat{\beta} - \frac{n(\Lambda_{\Delta i,sr})}{\sum_{t=1}^{n}(\underline{f}_{t,k} - \underline{s}_t)^2}$$

where $\Lambda_{\Delta i,sr} = \sum_{j=0}^{\infty} E\left[\Delta i_0^+ (s_{j+k} - s_j)\right] = \sum_{j=0}^{\infty} E\left[\Delta(f_{0,k} - s_0)(s_{j+k} - s_j)\right]$ is the one-sided long-run covariance of the spot return and the first differenced interest differential (or forward premium). $\widehat{\beta}_m$ is modified to remove second-order endogeneity effects and allows us to capture an important part of limit theory, which does not show up in the asymptotic distribution of $\hat{\beta}$. In the result below note that $d > 2d - 1$ for $d \in (\frac{1}{2}, 1)$, so $\widehat{\beta}_m$ has a faster convergence rate than $\hat{\beta}$.

Theorem 1 *Under Assumptions I to V* $\hat{\beta} \to 0$,

$$n^{(2d-1)}\widehat{\beta} \Rightarrow \frac{\Lambda_{\Delta i,sr}}{\int \underline{B}_{i(d-1)}^2} = \left(\frac{\Lambda_{\Delta i,sr}}{\Omega_{\xi\xi}}\right) \frac{1}{\int \underline{W}_{i(d-1)}^2}$$

and

$$n^d\widehat{\beta}_m \Rightarrow \frac{k \int \underline{B}_{i(d-1)} dB_s}{\int \underline{B}_{i(d-1)}^2}$$
$$= \left(\frac{k\Omega_{\xi v}}{\Omega_{\xi\xi}}\right) \frac{\int \underline{W}_{i(d-1)} dW_i}{\int \underline{W} i_{(d-1)}^2} + \left(\frac{k\sigma_{s|i}}{\Omega_{\xi\xi}^{1/2}}\right) \frac{\int \underline{W}_{i(d-1)} dW_{s\cdot i}}{\int \underline{W}_{i(d-1)}^2}$$

The limiting behaviour of the coefficient in the returns regression is best understood as a hybrid of the asymptotics for the cointegrating regression given in Phillips and Hansen (1990) and the fractional unit root distribution derived by Sowell (1990). Since the regressor in (2) is non-stationary, but the dependent variable has short memory, the resulting residuals can only be stationary for a true value of β equal to zero. Otherwise, they must inherit the stochastic trend in the regressor. In this way (2) is similar to a cointegrating regression with cointegrating vector (1, 0) and the convergence of $\widehat{\beta}$ to zero is therefore quite natural. Furthermore, the three components that appear in the limit theory given in Theorem 1, i.e.

$$\left(\frac{\Lambda_{\Delta i,sr}}{\Omega_{\xi\xi}}\right) \frac{1}{\int \underline{W}_{i(d-1)}^2}, \quad \left(\frac{k\Omega_{\xi v}}{\Omega_{\xi\xi}}\right) \frac{\int \underline{W}_{i(d-1)} dW_i}{\int \underline{W}_{i(d-1)}^2} \quad \text{and} \quad \left(\frac{k\sigma_{s|i}}{\Omega_{\xi\xi}^{1/2}}\right) \frac{\int \underline{W}_{i(d-1)} dW_{s\cdot i}}{\int \underline{W}_{i(d-1)}^2}$$

also correspond loosely to the endogeneity, serial correlation, and mixed normal terms, discussed by Phillips and Hansen (1990). (Examples of the first two terms are shown in Figures 3 and 4.)

By contrast, the difference in the convergence rates for $\widehat{\beta}$ and $\widehat{\beta}_m$ bears a closer resemblance to the asymptotic behaviour of the fractional unit root discussed by Sowell (1990). For $d < 1$, the endogeneity term dominates the other two terms asymptotically and the limiting distribution therefore has its entire weight to one side of zero. However, because this distribution is only valid

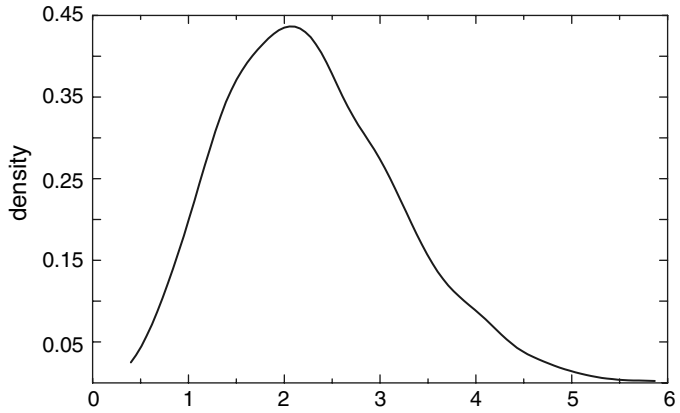

Figure 3. Endogeneity term

$$\frac{\Lambda_{\Delta i,sr}}{\Omega_{\xi\xi}} \frac{1}{\int \underline{W}_{i(d-1)}^2} \text{ for } \frac{\Lambda_{\Delta i,sr}}{\Omega_{\xi\xi}} = 1$$

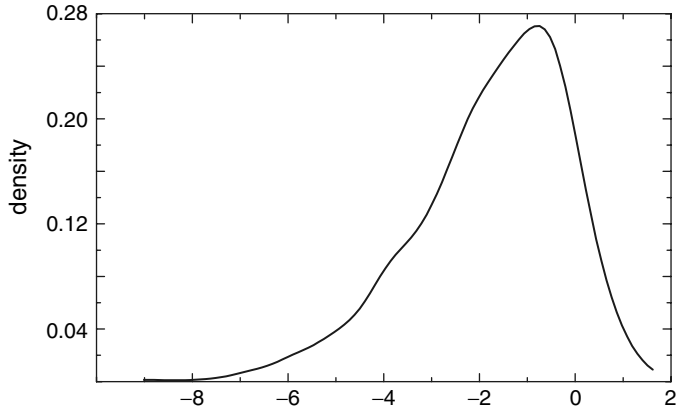

Figure 4. Serial correlation term

$$\left(\frac{k\Omega_{\xi v}}{\Omega_{\xi\xi}}\right) \frac{\int \underline{W}_{i(d-1)} dW_i}{\int \underline{W}_{i(d-1)}^2} \text{ for } \frac{k\Omega_{\xi v}}{\Omega_{\xi\xi}} = 1$$

up to an error of order $O_p(n^{-(1-d)})$, these other terms are likely to play an important role in finite sample, especially as d approaches one. The rate of convergence also depends on d. For $d = 1$, rate n convergence is achieved. However, for values of $d < 0.75$ the convergence rate is slower than $n^{1/2}$, becoming logarithmic as d approaches one half.

Asymptotic behaviour for the test statistics is equally non-standard. Let t denote the standard t statistic, t^* the test statistic for the hypothesis that $\beta = 1$, R^2 the coefficient of determination, and $\sigma_{sr} = \text{var}(s_k - s_0) = \sum_{j=1}^{k} \sum_{l=1}^{k} E(v_k v_l)$ the variance of the spot return.

Corollary 1.1 *Under Assumptions I to V,*

$$\frac{t}{n^{(1-d)}} \Rightarrow \frac{\Lambda_{\Delta i,sr}}{\sigma_{sr}\left(\int \underline{B}^2_{i(d-1)}\right)^{1/2}} \text{ for } \Lambda_{\Delta i,sr} \neq 0$$

$$t \Rightarrow \frac{k\int \underline{B}_{i(d-1)}dB_s}{\sigma_{sr}\left(\int \underline{B}^2_{i(d-1)}\right)^{1/2}} \text{ for } \Lambda_{\Delta i,sr} = 0, \text{ and}$$

$$\frac{1}{n^{2d}}t^* \Rightarrow -\frac{1}{\sigma_{sr}}\left(\int \underline{B}^2_{i(d-1)}\right)^{1/2}$$

Corollary 1.2 *Under Assumptions I to V* $R^2 \to 0$

$$n^{(2d-1)}R^2 \Rightarrow \frac{1}{\sigma^2_{sr}}\frac{\Lambda^2_{\Delta i,sr}}{\int \underline{B}^2_{i(d-1)}} \text{ for } \Lambda_{\Delta i,sr} \neq 0$$

and

$$nR^2 \Rightarrow \frac{k^2\left(\int \underline{B}_{i(d-1)}dB_s\right)^2}{\sigma^2_{sr}\int \underline{B}^2_{i(d-1)}} \text{ for } \Lambda_{\Delta i,sr} = 0$$

Just as in the case of a regression of a random walk on a linear trend, the standard t statistic can lead to the inference that the coefficient is significant, even though the estimator itself converges to the 'correct' value of zero. Furthermore, in close resemblance to the test statistic for the fractional unit root, it diverges for $d < 1$ and $\Lambda_{\Delta i,sr} \neq 0$, going towards $-\infty$ for $\Lambda_{\Delta i,sr} < 0$ and $+\infty$ for $\Lambda_{\Delta i,sr} > 0$. The divergence of the t statistic when $\Lambda_{\Delta i,sr} \neq 0$ is explained by the fact that there is long run covariation between the spot return and interest differential (forward premium) and this covariation is being picked up by the regression, albeit at a second-order level via $n^{2d-1}\hat{\beta}$. The coefficient of determination converges to zero and unbiasedness is rejected in the limit, since t^* diverges.

5.3. Implications for Regression in Levels

The long-run relationship between the spot and forward rate appears remarkably strong in levels. Figure 5, for example, shows the general good fit of the regression equation (1) to the data. In fact, as shown in Figure 6, using the current spot rate as the independent variable, as in the contemporaneous levels regression

$$s_t = \alpha_1 + \beta_1 f_{t,k} + \varepsilon^*_{1,t} \tag{12}$$

produces an even tighter fit. However, the test results in Table V suggest that the spot and forward rates may not be fully cointegrated in the traditional sense. In fact, the fractional integration of the interest differential implies a similar fractional cointegration between the spot and forward rates.[25]

[25] Perhaps this is best understood from the perspective of the CIP: $s_t = f_{t,k} - (i_{t,k} - i^*_{t,k}) - u_t$, which allows us to rewrite the levels specifications (1) and (12) as $s_t = f_{t,k} + \varepsilon^*_t$ for $\varepsilon^*_t = -(i_{t,k} - i^*_{t,k}) - u_t$ and $s_{t+k} = f_{t,k} + \varepsilon_{1,t+k}$ for $\varepsilon_{1,t+k} = (s_{t+k} - s_t) - (i_{t,k} - i^*_{t,k}) - u_t$, respectively. This formulation also suggests that the regression in levels may simply be reproducing the CIP relation, albeit with the interest differential hidden in the residual.

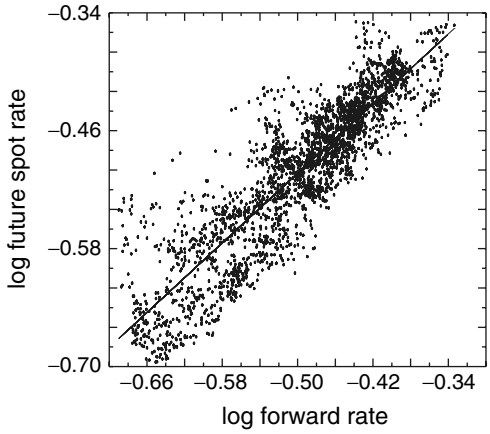

Figure 5. Plot for regression of s_{t+k} on $f_{t,k}$ (UK£/US$)

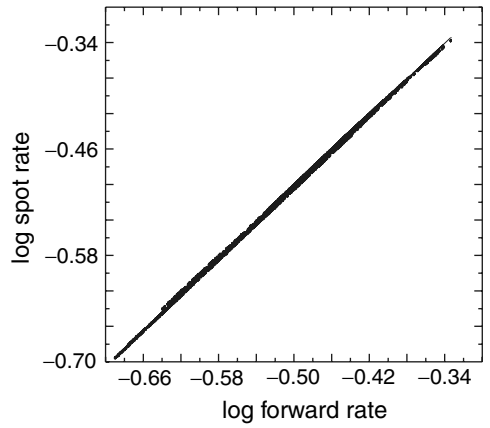

Figure 6. Plot for regression of s_t on $f_{t,k}$ (UK£/US$)

For $0.5 < d < 1$ the cointegrating residuals are not only fractionally integrated, but also lie in the non-stationary range. As Theorem 2 below shows, the asymptotic behaviour for the regressions in levels therefore falls between the case of the standard cointegrating regression and the spurious regression (see Phillips, 1986, 1998). Let $\hat{\beta}_{1,p}$, $t^*_{1,p}$, and $R^2_{1,p}$ denote the slope coefficient, the t statistic for the test of $\beta_{1,p} = 1$, and the coefficient of determination respectively in the regression

$$s_{t+p} = \hat{\alpha} + \hat{\beta}_{1,p} f_{t,k} + \widehat{\varepsilon}_{1,t+p} \quad \text{for} \quad p \geq 0$$

The contemporaneous regression of s_t on $f_{t,k}$ (12) is covered by $p = 0$ and the regression of s_{t+k} on $f_{t,k}$ as in (1) corresponds to $p = k$.

Theorem 2 *Under Assumptions I to V* $\hat{\beta}_{1,p} \to 1$,

$$n^{1-d}(\hat{\beta}_{1,p} - 1) \Rightarrow \frac{-\int \underline{B}_s B_{i(d-1)}}{\int \underline{B}_s^2} = -\left(\frac{\Omega_{\xi\xi}}{\Omega_{vv}}\right)^{1/2} \frac{\int \underline{W}_s W_{i(d-1)}}{\int \underline{W}_s^2}$$

$$\frac{t^*_{1,p}}{n^{\frac{1}{2}}} \Rightarrow \frac{-\int \underline{B}_s B_{i(d-1)}}{\left\{\int \underline{B}_s^2 \int \underline{B}_{i(d-1)}^2 - (\int \underline{B}_s B_{i(d-1)})^2\right\}^{1/2}}, \text{ and } R^2_{1,p} \to 1$$

Since the spot and forward rates are fully I(1), while the residual is only integrated of order $d < 1$, the signal of the regressor grows faster than the noise of the residual. Therefore, the estimator remains consistent despite the non-stationarity in the residual and the R^2 statistic also tends to unity. However, the rate of convergence is considerably slower than normal and the test statistic for the unbiasedness hypothesis $t^*_{1,p}$ diverges. This divergence is particularly interesting since it may well explain why cointegration-based tests of $\beta_1 = 1$ often lead to rejection (e.g. Phillips and McFarland, 1997; Goodhart *et al*., 1997; and Evans and Lewis, 1995) even though the estimates themselves are close to unity.[26] Results from regressions (1) and (12) are shown in Table VI, and seem generally to match the implications from Theorem 2. In particular, the slope coefficient and regression R^2's are close to one, while many of the t statistics are quite large.

To understand the behaviour of these regressions one must also appreciate the enormous difference in scale between the innovations to the spot rate and interest differential shown in Plate 2 and Table III. First, as argued in Section 3, this difference tends to obscure the long memory component in the residuals from the regression of the future spot rate on the current forward rate as in (1), causing it to appear as if it were fully cointegrated. Second, because this residual is so small compared to the regressor, the spot and forward rates appear to lie virtually on top of one another, generating the very tight fit in Figures 5 and 6 . This explains why the long-run relationship between the spot and forward rate appears so strong, despite the fractional non-stationarity of the residual. Finally, this scale difference also helps to explain why estimates of β_1 are so close to one in practice. The small size of the residual innovations produces a high finite sample signal to noise ratio and a very tight asymptotic distribution for $\hat{\beta}_{1,p}$ as witnessed by the very tiny ratio of $\Omega_{\xi\xi}$ to Ω_{vv} estimated in Table V.

Table VI. Regression in levels: estimates and underlying parameters

Currency	$\hat{\beta}_{1,0}$	$t^*_{1,0}$	$R_{1,0}$	$\hat{\beta}_{1,k}$	$t^*_{1,k}$	$R_{1,k}$	$\frac{\hat{\Omega}_{\xi\xi}}{\hat{\Omega}_{vv}}$
Aus$/US$	1.004	5.91	0.999	0.901	−12.84	0.827	0.00036
Can$/US$	1.016	57.82	1.000	1.002	0.625	0.961	0.00116
FF/US$	1.004	7.44	0.999	0.907	−12.38	0.837	0.00004
DM/US$	1.007	13.75	0.999	0.909	−12.31	0.843	0.00007
Yen/US$	0.996	−18.07	1.000	0.948	−11.21	0.936	1.3×10^{-8}
UK/US$	1.011	26.85	1.000	0.885	−13.37	0.791	0.00069

$k = 25$. Using daily data creates overlapping observations for $p = k$ but not $p = 0$, thereby inflating $t^*_{1,k}$ but not $t^*_{1,0}$.

[26] Such tests are based on modified versions of $t^*_{1,p}$ (see, for example, Phillips and Hansen, 1990).

5.4. Implications for Error-correction Models

Many of the more recent tests of unbiasedness have been conducted using approximate error-correction specifications.[27] Based on (12), the exact error-correction model (ECM) is given by

$$s_{t+1} - s_t = \alpha_2 + \beta_2(\beta_1 f_{t,1} - s_t) + \gamma_2' Q_t + \varepsilon_{2,t+1} \tag{13}$$

where Q_t contains lagged differences of the spot and forward rates.[28] An alternative specification, based on (1), is given by

$$s_{t+1} - s_t = \alpha_2 + \beta_2(\beta_1 f_{t-1,1} - s_t) + \gamma_2' Q_t + \varepsilon_{2,t+1}. \tag{14}$$

In their commonly employed two-step estimation procedure, Engle and Granger (1987) suggest estimating β_1 first, using regression in levels. Then, relying on the super consistency of the cointegrating regression, (13) or (14) may be estimated by linear regression with β_1 replaced by $\hat{\beta}_1$. However, due to the large number of estimates for β_1 near one, the common practice is instead simply to assume $\beta_1 = 1$, as in (3) or

$$s_{t+1} - s_t = \alpha_2 + \beta_2(f_{t-1,1} - s_t) + \gamma_2' Q_t + \varepsilon_{2,t+1} \tag{15}$$

We therefore refer to these two models as approximate error-correction specifications.

Because they attempt to estimate both the short- and long-run dynamics, error-correction models do not benefit from the same robustness properties as the (semiparametric) regression in levels. In fact, the nonstationary long memory of the forward premium causes the procedure of Engle and Granger (1987) to break down in two distinct ways. To begin with, the first-stage estimate of β_1 is no longer superconsistent. In fact, as shown in Theorem 2, it converges at a rate slower than $n^{1/2}$. This undermines our ability to treat the estimate from the first-stage regression as if it were known when conducting inference in the second stage (i.e. we have a generated regimes problem). A second, and more serious problem, occurs because even if we were to substitute $\beta_1 = 1$ (which is the true value in this case) for $\hat{\beta}_1$ as in (3) or (15), the error-correction terms: $f_{t-1,1} - s_t = (f_{t,1} - s_{t-1}) - \Delta s_t$ and $s_t - f_{t,1}$ contain the forward premium and are therefore fractionally non-stationary.

Consider the approximate ECM

$$s_{t+1} - s_t = \hat{\alpha}_{2,p} + \hat{\beta}_{2,p}(f_{t-p,1} - s_t) + \hat{\gamma}_{2,p}' Q_t + \hat{\varepsilon}_{2,t+1} \tag{16}$$

where $p = 0$ and $p = 1$ correspond to specifications (3) and (15), respectively. Defining $\hat{e}_{t,p}$ as the residual from the regression of the error-correction term on Q_t:

$$\underline{f}_{t-p,1} - \underline{s}_t = \hat{\delta}_{1,p}' \underline{Q}_t + \hat{e}_{t,p} \tag{17}$$

[27] These include Hakkio and Rush (1989), Baillie (1989), Barnhart and Szakmary (1991), Bekaert (1995), and Clarida and Taylor (1993).

[28] Non-overlapping data is generally employed so that k is normally equal to one.

$\hat{\beta}_{2,p}$ can be written compactly as the estimated coefficient in the regression of $s_{t+1} - s_t$ on $\hat{e}_{t,p}$:

$$\hat{\beta}_{2,p} = \frac{\sum_{t=1}^{n} \hat{e}_{t,p}(s_{t+1} - s_t)}{\sum_{t=1}^{n} \hat{e}_{t,p}^2}.$$

This, then, allows us to define the modified statistic

$$\hat{\beta}_{2,m} = \hat{\beta}_{2,p} - \frac{n(\Lambda_p - \Lambda_p^* - \Sigma_p)}{\sum_{t=1}^{n} \hat{e}_{t,p}^2}$$

where $\Lambda_p = \sum_{j=0}^{\infty} E(\Delta i_{-p}^{+} \Delta s_{j+1})$, $\Lambda_p^* = \gamma_{sq}' \left\{ \sum_{s=0}^{\infty} E(Q_s \Delta i_{-p}^{+}) - EQ_0 \left[s_0 - s_{-p} \right] \right\}$, $\Sigma_p = E \left[\Delta s_1 (s_0 - s_{-p}) \right]$, and $\gamma_{sq}' = E(\Delta s_1 Q_0') \left[E(Q_0 Q_0') \right]^{-1}$ reflect the various sources of endogeneity between the spot return and the error-correction term.[29] As in Theorem 1, such corrections allow us to account for the serial correlation and mixed normal terms, whose strong finite sample influence fades only slightly faster than that of the endogeneity terms, which dominate the asymptotic distribution. Let also,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} Q_t \Rightarrow B_Q(r)$$

define the Brownian motion B_Q.

Theorem 3 *Under Assumptions I to V*

$$n^{(2d-1)} \hat{\beta}_{2,p} \Rightarrow \frac{\Lambda_p - \Lambda_p^* - \Sigma_p}{\int \underline{B}_{i(d-1)}^2}, \quad n^d \hat{\beta}_{2,m} \Rightarrow \frac{\int \underline{B}_{i(d-1)} d(B_s - \gamma_{sq}' B_Q)}{\int \underline{B}_{i(d-1)}^2}$$

$$\frac{t_{2,p}}{n^{(1-d)}} \Rightarrow \frac{\Lambda_p - \Lambda_p^* - \Sigma_p}{\sigma_{sr} \left(\int \underline{B}_{i(d-1)}^2 \right)^{1/2}} \textit{ for } \Lambda_p - \Lambda_p^* - \Sigma_p = 0$$

and

$$t_{2,p} \Rightarrow \frac{\int \underline{B}_{i(d-1)} d(B_s - \gamma_{sq}' B_Q)}{\sigma_{sr} \left(\int \underline{B}_{i(d-1)}^2 \right)^{1/2}} \textit{ for } \Lambda_p - \Lambda_p^* - \Sigma_p \neq 0$$

This asymptotic theory is similar to the results of Theorem 1 and we therefore refer the reader to Section 5.2 for an in-depth discussion.

[29] For $p = 0$, $\Sigma_p = 0$ and $\Lambda_p^* = \Lambda_{\Delta i, sr}$.

6. DISCUSSION

6.1. A Comparison of the Asymptotic Theory and Empirical Regularities

Although the negative estimates of β in the returns regression (2) have played a central role in the forward discount anomaly, the non-stationary long memory of the forward premium suggests a statistical imbalance in the regression itself. One simply cannot explain a (short-memory) stationary variable, like the spot return, using a regressor with a stochastic trend. The 'true' value of β in (2) must in fact be zero, otherwise this trend would be transferred to the spot return as well. Since the variance of the regressor increases over time, Theorem 1 shows that $\hat{\beta}$ does indeed converge to zero. This is confirmed quite clearly by the recursively calculated slope estimates in Figure 8.

The absence of any simple linear relationship between the spot return and forward premium is also revealed by the nearly circular regression plot shown in Figure 7 and the very low R^2 statistics reported in Table VII.[30] Nevertheless, despite the poor fit of this regression, the previous literature

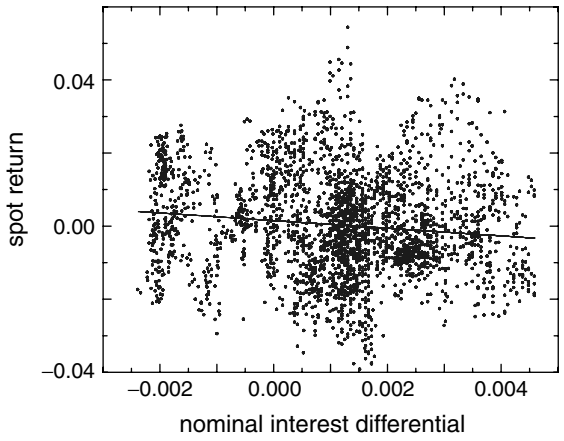

Figure 7. Plot for regression in returns: (CAN$/US$)

Table VII. Regression in returns: estimates and underlying parameters

Currency	$\hat{\beta}$	t $H_0: \beta = 0$	t^* $H_0: \beta = 1$	R^2	$\frac{\hat{\Lambda}_{\Delta i,sr}}{\hat{\Omega}_{\xi\xi}}$	$\frac{k\hat{\Omega}_{\xi v}}{\hat{\Omega}_{\xi\xi}}$
Aus$	−1.23	−1.29	−2.34*	0.01237	−69.8	73.8
Can$	−1.17	−1.67*	−3.11*	0.02064	−139.8	105.4
FR	−1.31	−0.41	−1.31	0.00127	−268.4	−217.8
DM	−0.19	−0.18	−1.11	0.00024	82.68	−217.23
Yen	−3.36	−2.29*	−2.98*	0.03818	−1126.7	6417.8
UK	0.07	0.05	−0.06	0.00002	−113.6	−16.4

*Significant at 5%. First four columns use month-end sampling to avoid inflation of t statistics due to overlapping data. (Daily results for $\hat{\beta}$ and R^2 are similar).

[30] In the traditional interpretation of this regression low R^2 statistics are explained by the extra noise added into the UIP equation when the expected spot return is replaced by its realization.

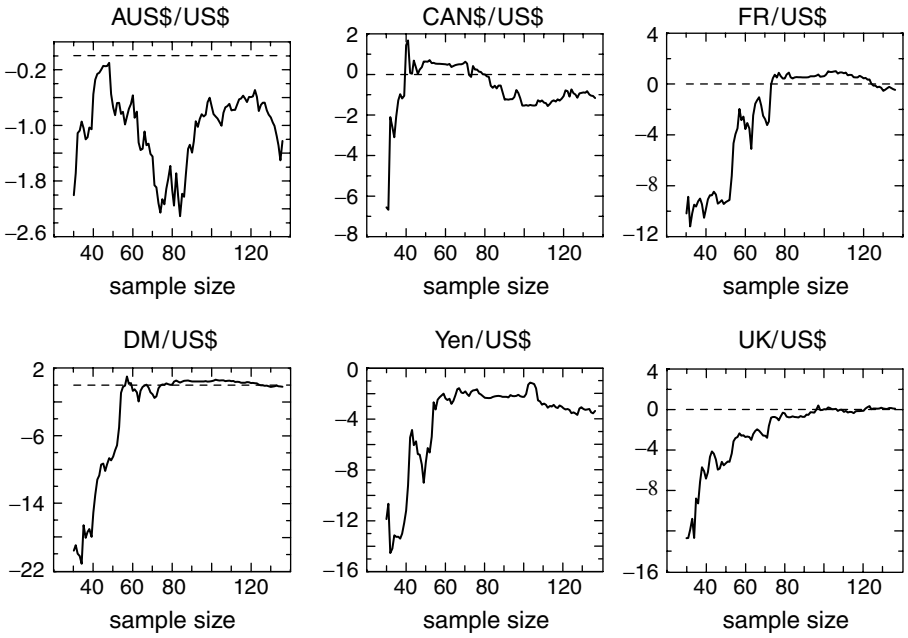

Figure 8. Recursive slope estimates for the regression in returns (sampled at month end)

has focussed a great deal of attention on these negative estimates. There are three principal reasons for this. The first is the preponderance of negative coefficients in empirical research, the second comes from their apparent magnitude, and the third from the often significantly negative t statistics. For example, in Table VII, five of the six estimates are negative, four of these are larger than one, and two are significant.

The results from Section 6.1 suggest an alternative way to account for these three empirical regularities. Consider first the preponderance of negative estimates. Given a symmetric limiting distribution, the fact that the vast majority of reported slope estimates are negative might seem at odds with the convergence of the estimator to zero. However, the non-standard limiting behaviour given in Theorem 1 is anything but symmetric and could quite easily give rise to such a skewed distribution of empirical estimates. In fact, the endogeneity and serial correlation terms are associated with long left-tailed distributions for values of $\Lambda_{\Delta i,sr} < 0$ and $\Omega_{\xi\nu} > 0$ respectively (see Figures 3 and 4). In this sense, it is possible for the vast majority of slope estimates to turn up negative, even through the estimator itself converges to zero.[31] This is somewhat analogous to

[31] Our theoretical results do not explain why the slope estimates on certain currencies have changed in recent subsamples, as reported by Baillie and Bollerslev (2000). However, neither are they inconsistent with this finding. The theory itself does not dictate a distribution skewed in any particular direction. At most it tells us that we should not count on a standard symmetrical distribution. Estimates of the underlying parameters do suggest a negative distribution for our particular sample, but these underlying parameters may have shifted over time, in which case the shape of the distribution may also have changed. Second, even holding these parameters constant the fact there are two terms with skewed limiting distributions and different rates of convergence suggests that the small sample distributions may look quite different than the distributions associated with larger samples. For this reason, estimates may show up positive in certain subsamples, even if they are generally negative across the entire sample.

the case of the unit root distribution, in which autoregressive estimates almost always turn up less than one, even when the true coefficient is itself equal to one.

Significantly negative t statistics can also be understood in terms of their non-standard limiting behaviour. For values of $\Lambda_{\Delta i,sr} < 0$ the t statistic diverges (possibly quite slowly) to negative infinity. (Figure 9 plots recursively estimated t statistics and this does appear to be the case for three of the six currencies.) Alternatively for $\Lambda_{\Delta i,sr} = 0$ and/or $d = 1$ the t statistic has a random but still highly non-standard limit, which may also lead to spurious inference.

Finally, at first sight, the apparent magnitude of the negative coefficients seems quite surprising as well. The Yen, for example, has a coefficient of -3.36. However, in light of the gigantic differences in scale between the spot return and forward premium discussed in Section 4.2 (and shown in Plate 2 and Table III), it hardly seems appropriate to refer to these estimates as 'large'. In fact, if the units of measurement were readjusted so as to make the forward premium and spot return of comparable size, the same estimates would actually look quite small.

The explanation given above is, for the most part, corroborated by the estimated coefficients on the endogeneity and serial correlation terms[32] $\left(\hat{\Lambda}_{\Delta i,sr}/\hat{\Omega}_{\xi\xi} \text{ and } k\hat{\Omega}_{\xi v}/\hat{\Omega}_{\xi\xi}\right)$ shown in the last two columns of Table VII.[33] $\hat{\Lambda}_{\Delta i,sr}$ is in fact negative for five of the six currencies and $\hat{\Omega}_{\xi v}$ is

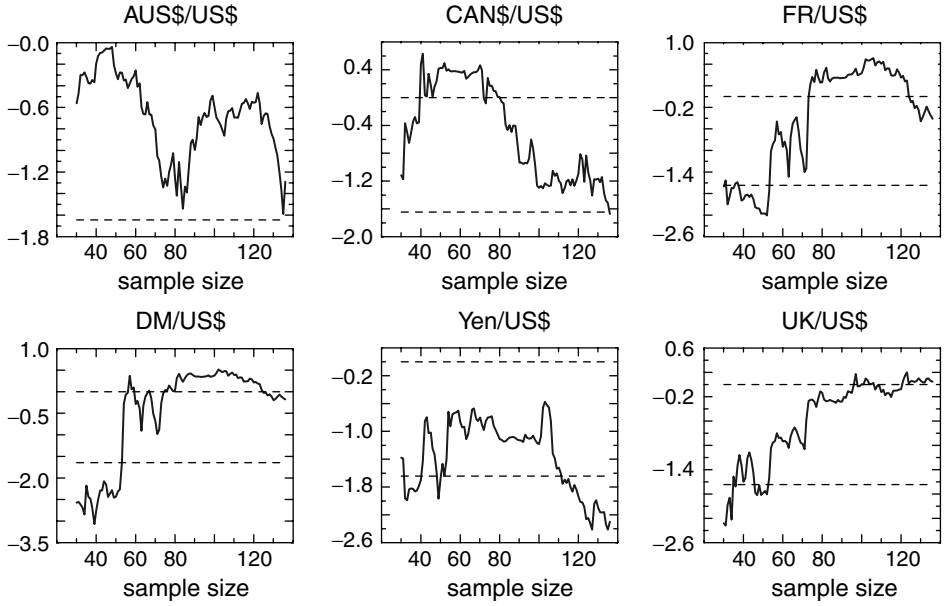

Figure 9. Recursive t statistics from the regression in returns (sampled at month end)

[32] Given $n \approx 2800$ and estimates of $\hat{d} \approx .9$ the finite sample influence of the endogeneity term, is roughly twice as large as that of serial correlation term.

[33] Using daily data and our estimates from Table I, the appropriate variables were either first or fractionally differenced, the first 100 observation were discarded to allow the fractional differencing to take effect, and the estimates were then computed using dspectra in COINT 2, which employs the approach of Andrews (1991) along with a Q-S kernel window.

positive for three. However, the estimates themselves appeared somewhat unstable, particularly for the Yen, and must be interpreted with some caution.[34]

While the convergence of $\hat{\beta}$ to zero is due to the time series imbalance between the spot return and forward premium, it is important to note that the non-standard behaviour of the regression estimate and t statistic is instead a direct consequence of the long memory properties of the forward premium itself. Another way of seeing this is to subtract the forward premium from both sides of (2) giving the alternative regression formulation[35]

$$s_{t+k} - f_{t,k} = \alpha + \delta(f_{t,k} - s_t) + \varepsilon_{t+k} \tag{18}$$

where $\delta = \beta - 1$. Under this formulation there is no apparent regression imbalance since both sides (the regressand and the regressor) contain fractionally integrated components. Nevertheless, this regression is essentially no different from (2) and hence is equally non-standard. Observe that under a null of $\beta = 1$ ($\delta = 0$) both the regressor and the regression error in (18) are fractionally integrated. On the other hand, if $\beta = 0$, then $\delta = -1$ and the regressor is fractionally integrated but the regression error is stationary.

6.2. Implications for Unbiasedness

Given the very standard assumption of a short-memory spot return, the long-memory behaviour of the forward premium already provides a rejection of forward rate unbiasedness. To make this point more intuitively, note that the unbiasedness hypothesis may be stated in terms of the orthogonality of the forward rate forecast errors $s_{t+k} - f_{t,k}$:

$$E_t(s_{t+k} - f_{t,k}) = 0.$$

If the forward premium $f_t - s_t$ is indeed a long-memory process and the spot return a short-memory process, then the forecast errors $s_{t+k} - f_{t,k} = (s_{t+k} - s_t) - (f_{t,k} - s_t)$ must themselves contain a long-memory component. By definition such a long-memory component is highly correlated with the past and thus partially forecastable.[36]

Nonetheless, our results may suggest some reorientation in the literature, away from attempts to explain the negative regression estimates themselves, and towards a fuller understanding of why the forward premium displays such strong long-memory characteristics. At best, one would hope to derive a realistic economic model that could deliver these long-memory characteristics as an implication. At the very least, one might strive for an intuitive understanding of just why the forward premium is so strongly persistent.

Several initial possibilities seem to present themselves. First, note that under a more general specification (such as an international CAPM) the error in (2) typically includes several 'missing

[34] Estimates varied substantially over subsamples, and were heavily influenced by initial observations. For example, leaving out just the first 50 observations for the Yen, $\hat{\Lambda}_{\Delta i,sr}/\ \hat{\Omega}_{\xi\xi}$ changed from -1126 to 79 and with only the first observation omitted it changed to 3845. The scale differences discussed above are reflected by the large size of these estimates, which imply a very wide asymptotic distribution for the estimator.

[35] We thank an anonymous referee for suggesting that we comment on this formulation.

[36] A pure long-memory process x_t with fractional parameter d can be forecasted using its AR(∞) representation $x_t = \Sigma_{j=0}^{\infty}\phi_j x_{t-j}$, where $\phi_j = \Gamma(j-d)/\{\Gamma(-d)(\Gamma(k+1))\}$. In the case of an ARFIMA(p,d,q) the value of the ϕ_j will depend, in part, on the ARMA coefficients for j small, but $\phi_j\{\Gamma(-d)\}^{-1}\ j^{-(1+d)}$ as $j \to \infty$.

variables', such as the conditional variance of the spot return[37] and a risk premium term. These missing variables clearly provide one possible explanation for the long memory behaviour. Baillie and Bollerslev (2000) for example, note that the conditional variance of the spot return displays long memory characteristics similar to those of the forward premium and suggest this as a possible explanation. Likewise, long memory in the forward premium could be due to long memory in the underlying sources of risk. [38]

A second possibility is suggested in Evans and Lewis (1995), who use a Peso problem argument to explain the evidence they find in favour of unit root behaviour in the forward premium.[39] Finally, its also quite possible that, contrary to the received wisdom, the expected spot return itself contains a long-memory component similar in size and persistence to the forward premium. If this were indeed the case then this predictable component would be nearly impossible to detect due to the large noise component in the spot return.[40]

6.3. Issues for Future Empirical Work

From a practical standpoint, one question which comes to mind is whether we can remedy these regression procedures. There are, in principle, two basic approaches that one might take. First, as suggested by Baillie and Bollerslev (2000), one could regress the spot return on the fractionally differenced forward premium[41]

$$s_{t+1} - s_t = \alpha^* + \beta^*(1 - L)^d(f_{t,k} - s_t) + \varepsilon^*_{t+1}.$$

Since the estimator for d converges at least as slowly as the regression coefficients, new formulas for the standard errors would need to be developed depending on the procedure used to estimate d. Doubtless this would increase the size of the standard errors. One might even contemplate a strategy of joint estimation of d and the regression parameters.

A more fundamental concern is that even after these adjustments are made, the modified regression no longer provides a proper framework for testing unbiasedness. In fact, given short memory in the sport return, a finding of long memory in the forward premium is already cause for rejection, so that no further tests are needed. Nevertheless, with properly modified standard errors, such a formulation might still yield interesting information with regard to the high-frequency dynamics between these two variables, even if there were no immediate economic interpretation to the results.

A second and perhaps more promising approach would be to attempt to impose balance by adding in appropriate missing variables rather than by differencing the data.[42] For example, one

[37] More generally, a JIT term, which simplifies to (a multiple of) the conditional variance of the spot return under the assumption of constant relative risk aversion. See Engel (1996) for more details.
[38] Kuersteiner (1996) also develops a cash in advance model relating the level of the exchange rate to the level of the interest rate.
[39] Although the data-generating mechanism they consider displays stationary characteristics over very long horizons, it mimics unit root type behaviour over more realistic time periods and might very well lead to econometric problems similar to those discussed here.
[40] Recall, for example, from Section 4 that the spot return dwarfs the forward premium in scale so that any long-memory component of comparable size would go virtually undetected.
[41] Newbold *et al*. (1998), pursue a similar strategy in the autoregressive/unit root context. They use the fitted values from an autoregression of the forecast error $s_{t+k} - f_{t,k}$ to produce a whitened version of the forward premium. They then run the regression in returns replacing the original forward premium by the whitened version. After this adjustment, they find that their estimates no longer turn up negative.
[42] We thank an anonymous referee for this suggestion.

could follow up on Baillie and Bollerslev's (2000) work by adding in an estimate of the Jensen's Inequality Term. In this case, one would look to see first whether such a term proved useful in balancing the equation (i.e. whether it was fractionally cointegrated with the forward premium) and if so, whether it led to more reasonable estimates of β. Clearly, there are also econometric complications involved here as well. Perhaps, most importantly, restoring balance to (2) does not, in itself, eliminate the non-standard nature of the limit theory, since this non-standard behaviour is primarily attributable to the long memory of the forward premium. One would be left with a fractionally cointegrated regression leading to regression asymptotics similar to those discussed earlier, albeit without the convergence of the estimator to zero. This might call for a modified estimator in the spirit of Phillips and Hansen (1990)[43]. Similar comments would apply to the error-correction regression.

The levels or cointegrating regression is by far the most robust of the three regressions, in that only the convergence rates and t statistics are affected by the long memory in the residuals. However, since this specification estimates only the long run relationship between the spot and forward rates, it is not clear whether any additional economic insight is gained from these improvements.

CONCLUSION

Using a new semiparametric estimator and some recent asymptotic theory of Kim and Phillips (1999a), we join Baillie and Bollerslev (1994) in finding evidence of non-stationary long-memory behaviour in the forward premium. This finding is further confirmed using parametric estimates based on parsimonious ARFIMA specifications. Two new explanations are also proposed for its conflicting empirical time series characterizations.

The non-stationary long memory of the forward premium implies an imbalance in the traditional regression of the spot return on the forward premium underlying much of the literature on the forward discount anomaly. Non-Standard limiting distributions for the estimator and test statistics are derived in terms of stochastic integrals of fractional Brownian motion. The slope and R^2 coefficients are shown to converge to zero, the t statistic is divergent, and the left-tailed limiting distribution appears consistent with puzzling negative estimates reported in the literature.

Interesting implications also follow for two related empirical applications traditionally used to test forward rate unbiasedness: regression in levels and in error-correction format. The regression in levels is found to be fractionally cointegrating, albeit with non-stationary residuals. The slope coefficient is still consistently estimated but the t statistic diverges, even when centred about the true coefficient. This may provide an explanation for results in several previous empirical studies, which have found the slope coefficient to be quite close to, yet still significantly different from, one. Estimation of the short run dynamics in the error-correction model is shown to be even more seriously affected.

There are two principal implications for future empirical work. First, the long memory of the forward premium provides in itself a rejection of forward rate unbiasedness, since the wandering behaviour predicted by the forward premium is not realized by the spot return. Given the presence of long memory it is therefore unnecessary to conduct further tests of unbiasedness. Second, in

[43] Kim and Phillips (**2001**) provide a fractional version of the FM estimator suited to the estimation of such fractionally cointegrated regressions.

order to correctly capture short-run dynamics between the spot and forward rates, the regressions in returns and in error-correction format must be adjusted to account for the stochastic trend in the forward premium. Work along these lines is currently underway.

Finally, our results may suggest a reorientation in the theoretical literature, away from attempts to explain the negative regression estimates themselves, and towards a fuller understanding of why the forward premium displays such strongly persistent behaviour. Evans and Lewis (1995) and Baillie and Bollerslev (2000) already contain some interesting suggestions along these lines.

APPENDIX

Technical Lemmas

The following two lemmas are used in the proofs of Theorems 1 and 3 respectively.

Lemma 1 *Let* $x_t = \sum_{j=0}^{\infty} \psi_{x,j}\varepsilon_{x,t-j}$ *and* $y_t = \sum_{j=0}^{\infty} \psi_{y,j}\varepsilon_{y,t-j}$ *satisfy the absolute summability conditions* $\sum_{j=0}^{\infty} |\psi_{x,j}|, \sum_{j=0}^{\infty} |\psi_{y,j}| < \infty$, *where* $(\varepsilon_{x,t}, \varepsilon_{y,t})$ *is i.i.d. with zero mean and finite fourth moments, and define their product as* $w_t = x_t y_t$. *Then the following hold: (i)* w_t *is weakly stationary, (ii)* $\sum_{h=0}^{\infty} |cov(w_0, w_h)| < \infty$, *(iii)* $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{w_t - E(w_t)\} = O_p(1)$, *and (iv)* $\frac{1}{n} \sum_{t=1}^{n} w_t \to E(w_0)$.

Remark 1.1 *This lemma applies in particular to fractionally integrated time series of order* $-1/2 < d \leq 0$.

Lemma 2 *Let* $x_n: \Omega \to R$ *be a sequence of random variables defined on the probability space* (Ω, R^p, P) *and suppose that* $x_n - x = o(n^{-a})$ *a.s. P, where x is a constant. If the (non-random) function* $g: R^p \to R^q$ *is continuously differentiable in a neighbourhood of x then* $g(x_n) - g(x) = o(n^{-a})$ *a.s. P.*

Remark 2.1 *This lemma also applies in the case in which x is a matrix.*

Proofs

Proof of Lemma 1 *(i)* $E(w_t) = E \lim_{n\to\infty} \sum_{j=0}^{n} \sum_{k=0}^{n} \psi_{x,j}\psi_{y,k}\varepsilon_{x,t-j}\varepsilon_{y,t-k} = E(\varepsilon_{x,0}\varepsilon_{y,0}) \sum_{j=0}^{\infty} \psi_{x,j}\psi_{y,j}$, *where dominated convergence follows from* $\sum_{j=0}^{\infty} \sum_{k=0}^{\infty} |\psi_{x,j}| |\psi_{y,k}| = \sum_{j=0}^{\infty} |\psi_{x,j}| \sum_{k=0}^{\infty} |\psi_{y,k}| < \infty$, *and*

$$\sum_{j=0}^{n} \sum_{k=0}^{n} \psi_{x,j}\psi_{y,k}\varepsilon_{x,t-j}\varepsilon_{y,t-k} \leq \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} |\psi_{x,j}| |\psi_{y,k}| (|\varepsilon_{x,t-j}\varepsilon_{y,t-k}| - E|\varepsilon_{x,t-j}\varepsilon_{y,t-k}|) + \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} |\psi_{x,j}| |\psi_{y,k}| E|\varepsilon_{x,t-j}\varepsilon_{y,t-k}|$$

$$= \sum_{j=0}^{\infty} |\psi_{x,j}| \, |\psi_{y,j}| \left(|\varepsilon_{x,t-j}\varepsilon_{y,t-j}| - E\,|\varepsilon_{x,0}\varepsilon_{y,0}| \right)$$

$$+ \sum_{j=0}^{\infty} \sum_{\substack{k=0 \\ k\neq j}}^{\infty} |\psi_{x,j}| \, |\psi_{y,k}| \left(|\varepsilon_{x,t-j}\varepsilon_{y,t-k}| - E\,|\varepsilon_{x,0}\varepsilon_{y,1}| \right)$$

$$+ E\,|\varepsilon_{x,0}\varepsilon_{y,0}| \sum_{j=0}^{\infty} |\psi_{x,j}| \, |\psi_{y,j}| + E\,|\varepsilon_{x,0}\varepsilon_{y,1}| \sum_{j=0}^{\infty} \sum_{\substack{k=0 \\ k\neq j}}^{\infty} |\psi_{x,j}| \, |\psi_{y,k}|$$

since $|\varepsilon_{x,t-j}\varepsilon_{y,t-j}| - E\,|\varepsilon_{x,0}\varepsilon_{y,0}|$ *is mean zero i.i.d. for* $j = 1, 2, 3, \ldots$ *and* $|\varepsilon_{x,t-j}\varepsilon_{y,t-k}| - E\,|\varepsilon_{x,0}\varepsilon_{y,1}|$ *is mean-zero i.d.d for* $j \neq k$. *Then, defining* $\underline{w}_t = w_t - E(w_t)$ *a similar dominated convergence argument implies that*

$$E(\underline{w}_0\underline{w}_h) = E(x_0 y_0 x_h y_h) - \left[E(x_0 y_0)\right]^2 = E(\varepsilon_{x,0}^2\varepsilon_{y,0}^2) \sum_{j=0}^{\infty} \psi_{x,j}\psi_{x,h+j}\psi_{y,j}\psi_{y,h+j}$$

$$+ E(\varepsilon_{x,0}^2)E(\varepsilon_{y,0}^2) \sum_{j=0}^{\infty}\sum_{r=0}^{\infty} \psi_{x,j}\psi_{x,h+j}\psi_{y,r}\psi_{y,h+r} + (E\left[\varepsilon_{x,0}\varepsilon_{y,0}\right])^2$$

$$\times \sum_{j=0}^{\infty}\sum_{r=0}^{\infty} \psi_{x,j}\psi_{x,h+r}\psi_{y,r}\psi_{y,h+j}$$

(ii) Note first that $\sum_{h=-\infty}^{\infty}\sum_{j=0}^{\infty} |\psi_{x,j}\psi_{x,j+h}|$, $\sum_{h=-\infty}^{\infty}\sum_{j=0}^{\infty} |\psi_{y,j}\psi_{y,j+h}|$, *and* $\sum_{h=-\infty}^{\infty}\sum_{j=0}^{\infty} |\psi_{x,j}\psi_{y,j+h}| < \infty$ *are implied by absolute summability. For example,*

$$\sum_{h=1}^{\infty}\sum_{j=0}^{\infty} |\psi_{x,j}\psi_{y,j+h}| \leq \sum_{j=0}^{\infty} \left(\sum_{h=1}^{\infty} |\psi_{y,j+h}| \right) |\psi_{x,j}|$$

$$\leq \left\{ \sum_{j=0}^{j^*} |\psi_{x,j}| \sum_{h=1}^{\infty} |\psi_{y,j+h}| + \sum_{j=j^*+1}^{\infty} |\psi_{x,j}| \right\} < \infty.$$

for j^* *such that* $\sum_{h=1}^{\infty} |\psi_{y,j+h}| < 1$ *for all* $j > j^*$.[44] *Then define* $a_j = \sum_{h=j}^{\infty} |\psi_{x,h}\psi_{y,h}|$, *and note that* $a_j \to 0$ *since* $\sum_{h=0}^{n} |\psi_{x,h}\psi_{y,h}| < \left(\sum_{h=0}^{n} \psi_{x,h}^2\right)^{\frac{1}{2}} \left(\sum_{h=0}^{n} \psi_{y,h}^2\right)^{\frac{1}{2}}$ *is increasing and convergent. Therefore, choosing* j^* *so that* $|a_j| < 1$ *for* $j > j^*$

$$\sum_{h=0}^{\infty} \left| \sum_{j=0}^{\infty} \psi_{x,j}\psi_{x,h+j}\psi_{y,j}\psi_{y,h+j} \right| \leq \sum_{j=0}^{\infty} a_j \, |\psi_{x,j}\psi_{y,j}| < \sum_{j=0}^{j^*} a_j \, |\psi_{x,j}\psi_{y,j}| + \sum_{j=j^*+1}^{\infty} |\psi_{x,j}\psi_{y,j}| < \infty.$$

[44] Note that since $\sum_{h=1}^{j} |\psi_{y,h}|$ is increasing and convergent, the tail sum $\sum_{h=1}^{\infty} |\psi_{y,j+h}| = \sum_{h=j+1}^{\infty} |\psi_{y,h}|$ must converge to zero.

Finally, we also have[45]

$$\sum_{h=0}^{\infty}\left|\sum_{j=0}^{\infty}\sum_{r=0}^{\infty}\psi_{x,j}\psi_{x,h+j}\psi_{y,r}\psi_{y,h+r}\right| \leq \left\{\sum_{h=0}^{\infty}\left(\sum_{j=0}^{\infty}\psi_{x,j}\psi_{x,h+j}\right)^2\right\}^{\frac{1}{2}}$$

$$\times\left\{\sum_{h=0}^{\infty}\left(\sum_{r=0}^{\infty}\psi_{y,r}\psi_{y,h+r}\right)^2\right\}^{\frac{1}{2}} < \infty$$

$$\sum_{h=0}^{\infty}\left|\sum_{j=0}^{\infty}\sum_{r=0}^{\infty}\psi_{x,j}\psi_{x,h+r}\psi_{y,r}\psi_{y,h+j}\right| \leq \left\{\sum_{h=0}^{\infty}\left(\sum_{j=0}^{\infty}\psi_{x,j}\psi_{y,h+j}\right)^2\right\}^{\frac{1}{2}}$$

$$\times\left\{\sum_{h=0}^{\infty}\left(\sum_{r=0}^{\infty}\psi_{x,h+r}\psi_{y,r}\right)^2\right\}^{\frac{1}{2}} < \infty$$

Therefore $\sum_{h=0}^{\infty}\left|E(\underline{w}_0\underline{w}_h)\right| < \infty$. *(iii) By the Toeplitz Lemma*

$$\frac{1}{n}E\left(\sum_{t=1}^{n}w_t\right)^2 = \sum_{h=-n+1}^{n-1}E(\underline{w}_0\underline{w}_h) - 2\frac{1}{n}\sum_{h=1}^{n-1}hE(\underline{w}_0\underline{w}_h) \to \sum_{h=-\infty}^{\infty}E(\underline{w}_0\underline{w}_h)$$

For any $\varepsilon > 0$, *choose* M_ε *such that* $M_\varepsilon^2 > \frac{1}{\varepsilon}\left\{\varepsilon + \sum_{h=-\infty}^{\infty}E(\underline{w}_0\underline{w}_h)\right\}$ *and* n_ε *such that* $\left|\frac{1}{n}E(\sum_{t=1}^{n}\underline{w}_t)^2 - \sum_{h=-\infty}^{\infty}E(\underline{w}_0\underline{w}_h)\right| < \varepsilon$ *for* $n > n_\varepsilon$. *Then, by Chebyshev's inequality*

$$\Pr\left\{\frac{1}{n^{1/2}}\sum_{t=1}^{n}\underline{w}_t > M_\varepsilon\right\} < \frac{1}{M_\varepsilon^2}\frac{1}{n}E\left(\sum_{t=1}^{n}\underline{w}_t\right)^2 < \frac{1}{M_\varepsilon^2}\left\{\varepsilon + \sum_{h=-\infty}^{\infty}E(\underline{w}_0\underline{w}_h)\right\} < \varepsilon$$

for all $n > n_\varepsilon$. *(iv) Therefore,* $\frac{1}{n}\sum_{t=1}^{n}\underline{w}_t = \frac{1}{n}\sum_{t=1}^{n}w_t - E(w_0) \to 0$, *implying that* $\frac{1}{n}\sum_{t=1}^{n}w_t \to E(w_0)$.

Proof of Lemma 2 *First, with probability one we draw an* $\omega\varepsilon\Omega$ *such that* $n^a(x_n - x) \to 0$. *For any* $\varepsilon > 0$ *choose* n_0 *and* $\delta_0 < \varepsilon$ *such that* g *is continuously differentiable over the closed region* $C = \{z : \|z - x\| < \delta_0\}$ *and that* $\forall n \geq n_0\ n^a\|x_n(\omega) - x\| \leq \delta_0$. *Next define*

$$M_{\delta_o} = \sup_{z\varepsilon C}\left\|\frac{dg(x)}{dx}\ |_{x=z}\right\| < \infty$$

and then choose

$$\delta_1 < \min\left(\frac{\varepsilon}{M_{\delta_0}}, \delta_0\right)$$

and $n \geq n_0$. *Then, by the mean-value theorem* $n^a\|g(x_n(\omega)) - g(x)\| \leq M_{\delta_o}n^a\|x_n(\omega) - x\| \leq \varepsilon$.

[45] These follow since $\sum_{h=0}^{\infty}\left|\sum_{j=0}^{\infty}\psi_{x,j}\psi_{x,h+j}\right|$, $\sum_{h=0}^{\infty}\left|\sum_{r=0}^{\infty}\psi_{y,r}\psi_{y,h+r}\right|$, $\sum_{h=0}^{\infty}\left|\sum_{j=0}^{\infty}\psi_{x,j}\psi_{y,h+j}\right|$, and $\sum_{h=0}^{\infty}\left|\sum_{r=0}^{\infty}\psi_{x,h+r}\psi_{y,r}\right| < \infty$ and absolute summability is stronger than square summability.

Proof of Theorem 1 *Define* $x_t = f_{t,k} - s_t = i_t^+$ *and* $y_{t+k} = s_{t+k} - s_t$, *and then write*

$$\hat{\beta} = \frac{\sum_{t=1}^{n} \underline{x}_t y_{t+k}}{\sum_{t=1}^{n} \underline{x}_t^2}$$

using the Frisch–Waugh–Lovell theorem. Using the Beveridge–Nelson (BN) decomposition we then express v_t *and* ξ_t *as the sum of a martingale difference sequence and a telescoping remainder:*

$$\begin{pmatrix} v_t \\ \xi_t \end{pmatrix} = C(L)\varepsilon_t = C(I)\varepsilon_t + \widetilde{\varepsilon}_{t-1} - \widetilde{\varepsilon}_t, \text{ for}$$

$$\widetilde{\varepsilon}_t = \widetilde{C}(L)\varepsilon_t = \sum_{l=0}^{\infty} \widetilde{c}_l \varepsilon_{t-l}, \widetilde{c}_l = \sum_{j=l+1}^{\infty} c_j,$$

$$C(I) = \sum_{j=0}^{\infty} c_j, \text{ and } \sum_{l=0}^{\infty} \left|\widetilde{c}_l\right| < \infty \tag{A1}$$

where summability of $\left|\widetilde{c}_l\right|$ *follows from Assumption V. We can then decompose* $i_t^+ = (1-L)^{-d}\xi_t$ *as*

$$i_t^+ = C_2(I)' i_{0,t}^+ + \widetilde{i}_{t-1}^+ - \widetilde{i}_t^+ = C_2(I)' i_{0,t}^+ - \widetilde{z}_t, \text{ where}$$
$$i_{0,t}^+ = (I-L)^{-d}\varepsilon_t, \widetilde{i}_t^+ = (1-L)^{-d}\widetilde{\varepsilon}_{2,t}, \quad \text{and} \quad \widetilde{z}_t = \widetilde{i}_t^+ - \widetilde{i}_{t-1}^+$$

By Assumption V and Theorem 2 of Sowell (1990)

$$\frac{1}{n^{d-1/2}} i_{0,[nr]}^+ = \frac{1}{n^{d-1/2}} \sum_{t=1}^{[nr]} (I-L)^{d-1}\varepsilon_t \Rightarrow W_{(d-1)}$$

and therefore

$$\frac{1}{n^{d-1/2}} x_{[nr]} \Rightarrow C_2(I)' W_{(d-1)} = B_{i(d-1)}, \frac{1}{n}\sum_{t=1}^{n} \frac{x_t}{n^{d-1/2}} \Rightarrow \int B_{i(d-1)}$$

$$\frac{1}{n^{d-1/2}} \underline{x}_{[nr]} \Rightarrow \underline{B}_{i(d-1)} = B_{i(d-1)} - \int B_{i(d-1)}, \text{ and } \frac{1}{n^{2d}}\sum_{t=1}^{n} \underline{x}_t^2 = \frac{1}{n}\sum_{t=1}^{n} \left(\frac{\underline{x}_t}{n^{d-1/2}}\right)^2$$
$$= \int_{1/n}^{(n+1)/n} \left(\frac{\underline{x}_{[nr]}}{n^{d-1/2}}\right)^2 dr \Rightarrow \int \underline{B}_{i(d-1)}^2$$

This gives the distribution for the denominator. By the BN decomposition (A1)

$$y_{t+k} = \sum_{p=1}^{k} v_{t+p} = C_1(I)' \sum_{p=1}^{k} \varepsilon_{t+p} + (\widetilde{\varepsilon}_{1,t} - \widetilde{\varepsilon}_{1,t+k}) \tag{A2}$$

allowing the numerator to be divided into two components:

$$\sum_{t=1}^{n} \underline{x}_t y_{t+k} = C_1(I)' \sum_{p=l}^{k} \sum_{t=1}^{n} \underline{i}_t^+ \varepsilon_{t+p} + \hat{\Lambda}_{\Delta i,sr} \text{ where } \hat{\Lambda}_{\Delta i,sr} = \sum_{t=1}^{n} \underline{i}_t^+ (\widetilde{\varepsilon}_{1,t} - \widetilde{\varepsilon}_{1,t+k})$$

Theorem 2.1 of Hansen (1992) then implies that

$$\frac{1}{n^d} \sum_{t=1}^{n} i_t^+ \varepsilon_{t+p} = \sum_{t=1}^{n} \left(\frac{i_t^+}{n^{d-1/2}} \right) \left(\frac{\varepsilon_{t+p}}{n^{1/2}} \right) \Rightarrow \int Bi_{(d-1)} dW$$

and therefore

$$\frac{1}{n^d} C_1(I)' \sum_{p=1}^{k} \sum_{t=1}^{n} \underline{i}_t^+ \varepsilon_{t+p} \Rightarrow k C_1(I)' \int \underline{B}_{i(d-1)} dW = k \int \underline{B}_{i(d-1)} dB_s. \tag{A3}$$

A telescoping sum can then be extracted from $\hat{\Lambda}_{\Delta i,sr}$*:*

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^{n} \underline{i}_t^+ \left(\widetilde{\varepsilon}_{1,t} - \widetilde{\varepsilon}_{1,t+k} \right) &= \frac{1}{n} \sum_{t=1}^{n} \left(\underline{i}_t^+ - \underline{i}_{t-k}^+ \right) \widetilde{\varepsilon}_{1,t} - \frac{1}{n} \sum_{t=1}^{n} \left(\underline{i}_t^+ \widetilde{\varepsilon}_{1,t+k} - \underline{i}_{t-k}^+ \widetilde{\varepsilon}_{1,t} \right) \\ &= \sum_{p=1}^{k} \frac{1}{n} \sum_{t=1}^{n} \left(\Delta i_{t-k+p}^+ \right) \widetilde{\varepsilon}_{1,t} + \sum_{p=1}^{k} \frac{1}{n} \underline{i}_{p-k}^+ \widetilde{\varepsilon}_{1,p} - \sum_{p=1}^{k} \frac{1}{n^{\frac{3}{2}-d}} \left(\frac{\underline{i}_{n+p-k}^+}{n^{d-\frac{1}{2}}} \right) \widetilde{\varepsilon}_{1,n+p} \end{aligned}$$

Since $\widetilde{\varepsilon}_{1,t}$ *and the fractionally overdifferenced variable* Δi_t^+ *satisfy the absolute summability condition in Lemma 1*

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{ \underline{i}_t^+ \left(\widetilde{\varepsilon}_{1,t} - \widetilde{\varepsilon}_{1,t+k} \right) - \sum_{p=1}^{k} E \left(\Delta i_{p-k}^+ \widetilde{\varepsilon}_{1,0} \right) \right\} = O_p(1)$$

$\sum_{p=1}^{k} E(\Delta i_{p-k}^+ \widetilde{\varepsilon}_{1,0}) = \sum_{p=0}^{k-1} E\left(\Delta i_0^+ \widetilde{\varepsilon}_{1,p}\right)$ *due to the weak stationarity of* Δi_t^+ *and* $\widetilde{\varepsilon}_{1,t}$*. Defining* $\varepsilon_t^* = |\varepsilon_t| - E|\varepsilon_0|$ *and noting that* $\sum_{l=1}^{\infty} |c_{1,l}| < \infty$ *by Assumption V and*

$$\sum_{l=1}^{n} c_{1,l} \varepsilon_{p+j-l} \leq \sum_{l=1}^{\infty} |c_{1,l}| \, |\varepsilon_{p+j-l}| = \sum_{l=1}^{\infty} |c_{1,l}| \, \varepsilon_{p+j-l}^* + E|\varepsilon_o| \sum_{l=1}^{\infty} |c_{1,l}|$$

we then apply dominated convergence to obtain:

$$\begin{aligned} \sum_{j=1}^{\infty} E_p \nu_{p+j} &= \sum_{j=1}^{\infty} \sum_{l=1}^{\infty} c_{1,l} E_p \varepsilon_{p+j-l} = \sum_{j=1}^{\infty} \sum_{l=j}^{\infty} c_{1,l} \varepsilon_{p+j-l} \\ &= \sum_{s=0}^{\infty} \left(\sum_{r=s+1}^{\infty} c_{1,r} \right) \varepsilon_{p-s} = \sum_{s=0}^{\infty} \widetilde{c}_{1,s} \varepsilon_{p-s} = \widetilde{\varepsilon}_{1,p} \end{aligned} \tag{A4}$$

Likewise,

$$\left|\sum_{j=1}^{n} E_p v_{p+j}\right| \leq \sum_{j=1}^{\infty} \left|E_p v_{p+j}\right| = \sum_{s=0}^{\infty} \left|\widetilde{c}_{1,s}\right| \left|\varepsilon_{p-s}\right|$$

$$= \sum_{s=0}^{\infty} \left|\widetilde{c}_{1,s}\right| \varepsilon^{*}_{p-s} + E\left|\varepsilon_o\right| \sum_{s=0}^{\infty} \left|\widetilde{c}_{1,s}\right| \tag{A5}$$

where $\sum_{s=1}^{\infty} \left|\widetilde{c}_{1,s}\right| < \infty$ *by Assumption V. Using (A5) to invoke dominated convergence it follows from (A4) that*

$$\sum_{p=0}^{k-1} E\left(\Delta i_0^{+} \widetilde{\varepsilon}_{1,p}\right) = \sum_{p=0}^{k-1} E\left(\Delta i_0^{+} \sum_{j=1}^{\infty} E_p v_{p+j}\right) = \sum_{j=1}^{\infty} E\left(\Delta i_0^{+} \sum_{p=0}^{k-1} v_{p+j}\right)$$

$$= \sum_{j=0}^{\infty} E\left[\Delta i_0^{+}\left(s_{j+k} - s_j\right)\right] = \Lambda_{\Delta i,sr}$$

Therefore, $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{\underline{x}_t y_{t+k} - \Lambda_{\Delta i,sr}\right\} = O_p(1)$, $\frac{1}{n} \sum_{t=1}^{n} \underline{x}_t y_{t+k} \to \Lambda_{\Delta i,sr}$, *and*

$$\frac{1}{n^d}\left(\sum_{t=1}^{n} \underline{x}_t y_{t+k} - n\Lambda_{\Delta i,sr}\right) = \frac{1}{n^d} C_1(I)' \sum_{p=l}^{k} \sum_{t=1}^{n} \underline{i}_t^{+} \varepsilon_{t+p} \Rightarrow k \int B_{i(d-1)} dB_s \tag{A6}$$

Applying the continuous mapping theorem and the decomposition of B_s *in (11) then gives the result in Theorem 1.*

Proof of Corollary 1.1 *Employing the Frisch–Waugh–Lovell theorem write*

$$\frac{t}{n^{(1-d)}} = \frac{n^{(2d-1)}\hat{\beta}}{\hat{\sigma}\left(\frac{1}{n^{2d}} \sum_{t=1}^{n} \underline{x}_t^2\right)^{-\frac{1}{2}}} \text{ for } \Lambda_{\Delta i,sr} \neq 0$$

$$t = \frac{n^{d}\hat{\beta}_m}{\hat{\sigma}\left(\frac{1}{n^{2d}} \sum_{t=1}^{n} \underline{x}_t^2\right)^{-\frac{1}{2}}} \text{ for } \Lambda_{\Delta i,sr} = 0$$

and

$$\frac{t^*}{n^d} = \frac{1}{\hat{\sigma}}\left(\frac{1}{n^{2d}} \sum_{t=1}^{n} \underline{x}_t^2\right)^{\frac{1}{2}} (\hat{\beta} - 1) = \hat{\beta}\frac{1}{\hat{\sigma}}\left(\frac{1}{n^{2d}} \sum_{t=1}^{n} \underline{x}_t^2\right)^{\frac{1}{2}} - \frac{1}{\hat{\sigma}}\left(\frac{1}{n^{2d}} \sum_{t=1}^{n} \underline{x}_t^2\right)^{\frac{1}{2}}$$

where

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{t=1}^{n} \left(\underline{y}_{t+k} - \hat{\beta}\underline{x}_t\right)^2 = \frac{1}{n} \sum_{t=1}^{n} \underline{y}_{t+k}^2 - 2\hat{\beta}\frac{1}{n} \sum_{t=1}^{n} \underline{x}_t y_{t+k} + \hat{\beta}\left(n^{(2d-1)}\hat{\beta}\right)\left(\frac{1}{n^{2d}} \sum_{t=1}^{n} \underline{x}_t^2\right)$$

From the proof of Theorem 1 $\frac{1}{n}\sum_{t=1}^{n}\underline{x}_t y_{t+k}$, $n^{(2d-1)}\hat{\beta}$, *and* $\frac{1}{n^{2d}}\sum_{t=1}^{n}\underline{x}_t^2$ *are all* $O_p(1)$, *while* $\hat{\beta}\to 0$ *a.s. Therefore* $\hat{\sigma}^2 \to E\left(\underline{y}_{t+k}^2\right)=\sigma_{sr}^2$. *These results and the limits given in Theorem 1 give the stated results.*

Proof of corollary 1.2 *Write the (standardized) coefficient of determination as*

$$n^{(2d-1)}R^2=\left(n^{(2d-1)}\hat{\beta}\right)^2\frac{\frac{1}{n^{2d}}\sum_{t=1}^{n}\underline{x}_t^2}{\frac{1}{n}\sum_{t=1}^{n}\underline{y}_{t+k}^2} \text{ for } \Lambda_{\Delta i,sr}\neq 0 \text{ and as}$$

$$nR_m^2=\left(n^{d}\hat{\beta}_m\right)^2\frac{\frac{1}{n^{2d}}\sum_{t=1}^{n}\underline{x}_t^2}{\frac{1}{n}\sum_{t=1}^{n}\underline{y}_{t+k}^2} \text{ for } \Lambda_{\Delta i,sr}=0$$

The rest follows from the application of the continuous mapping theorem to the results of Theorem 1.

Proof of Theorem 2 *By Assumption I* $f_{t,k}=s_t+x_t$. *Also define* $y_{p,t+p}=s_{t+p}-s_t$. *Then*

$$\hat{\beta}_{1,p}=\frac{\sum_{t=1}^{n}\underline{f}_{t,k}s_{t+p}}{\sum_{t=1}^{n}\underline{f}_{t,k}^2}=\frac{\sum_{t=1}^{n}\left(\underline{s}_t+\underline{x}_t\right)\left(s_t+y_{p,t+p}\right)}{\sum_{t=1}^{n}\left(\underline{s}_t+\underline{x}_t\right)^2}$$

$$=\frac{\sum_{t=1}^{n}\left(\underline{s}_t+\underline{x}_t\right)\left(\{s_t+x_t\}-\{x_t-y_{p,t+p}\}\right)}{\sum_{t=1}^{n}\left(\underline{s}_t+\underline{x}_t\right)^2}=1-\frac{\sum_{t=1}^{n}\left(\underline{s}_t+\underline{x}_t\right)\left(x_t-y_{p,t+p}\right)}{\sum_{t=1}^{n}\left(\underline{s}_t+\underline{x}_t\right)^2}$$

Recall that s_t *is I(1),* x_t *is I(d), and* $y_{p,t+p}$ *is I(0). Therefore* $\frac{1}{n^2}\sum_{t=1}^{n}\underline{s}_t^2\Rightarrow\int\underline{B}_s^2$ *and*

$$\frac{1}{n^{1+d}}\sum_{t=1}^{n}\underline{s}_t x_t=\frac{1}{n}\sum_{t=1}^{n}\left(\frac{\underline{s}_t}{n^{1/2}}\right)\left(\frac{x_t}{n^{d-1/2}}\right)=\int_{\frac{1}{n}}^{\frac{n+1}{n}}\left(\frac{\underline{s}_{[nr]}}{n^{1/2}}\right)\left(\frac{x_{[nr]}}{n^{d-1/2}}\right)dr\Rightarrow\int\underline{B}_s B_{i(d-1)}$$

The remaining terms are all of lower order: $\sum_{t=1}^{n}s_t y_{p,t+p}=O_p(n)$ *and from the proof of Theorem 1* $\sum_{t=1}^{n}x_t y_{p,t+p}=O_p(n)$ *and* $\sum_{t=1}^{n}x_t^2=O_p\left(n^{2d}\right)$. *Therefore* $\beta_{1,p}\to 1$ *and*

$$n^{1-d}\left(\hat{\beta}_{1,p}-1\right)=\frac{\frac{1}{n^{1+d}}\sum_{t=1}^{n}\left(-\underline{s}_t x_t-\underline{x}_t^2+\underline{s}_t y_{p,t+p}+\underline{x}_t y_{p,t+p}\right)}{\frac{1}{n^2}\sum_{t=1}^{n}\underline{s}_t^2+\frac{1}{n^2}\sum_{t=1}^{n}\left(2\underline{s}_t\underline{x}_t+\underline{x}_t^2\right)}\Rightarrow\frac{-\int\underline{B}_s B_{i(d-1)}}{\int\underline{B}_s^2}$$

Define

$$t^*_{1,p} = \frac{\left(\hat{\beta}_{1,p} - 1\right)}{\left\{\sum_{t=1}^{n} \underline{f}_t^2\right\}^{-1/2} \hat{\sigma}_1} \text{ where } \hat{\sigma}_1 = \left\{\frac{1}{n}\sum_{t=1}^{n}\left(\underline{s}_{t+p} - \hat{\beta}_{1,p}\underline{f}_{t,k}\right)^2\right\}^{\frac{1}{2}}$$

Then, substitute $\underline{s}_{t+p} - \hat{\beta}_{1,p}\underline{f}_{t,k} = \left(1 - \hat{\beta}_{1,p}\right)\underline{s}_t - \hat{\beta}_{1,p}\underline{x}_t + \underline{y}_{p,t+p}$ *so that*

$$\left(\underline{s}_{t+p} - \hat{\beta}_{1,p}\underline{f}_{t,k}\right)^2 = \left(\hat{\beta}_{1,p} - 1\right)^2 \underline{s}_t^2 + \hat{\beta}_{1,p}^2 \underline{x}_t^2 + 2\hat{\beta}_{1,p}\left(\hat{\beta}_{1,p} - 1\right)\underline{s}_t\underline{x}_t - 2\left(\hat{\beta}_{1,p} - 1\right)\underline{s}_t\underline{y}_{p,t+p} - 2\hat{\beta}_{1,p}\underline{x}_t\underline{y}_{p,t+p} + \underline{y}^2_{p,t+p}. \quad (A7)$$

The first three terms in (A7) are of order $O_p(n^{2d})$ *while the rest are of order* $O_p(n)$ *or lower.*[46] *Therefore*

$$\frac{1}{n^{2d-1}}\hat{\sigma}_1^2 = \frac{1}{n^{2d}}\sum_{t=1}^{n}\left(\underline{s}_{t+p} - \hat{\beta}_{1,p}\underline{f}_{t,k}\right)^2 \Rightarrow \frac{\int \underline{B}_s^2 \int \underline{B}_{i(d-1)}^2 - \left(\int \underline{B}_s B_{i(d-1)}\right)^2}{\int \underline{B}_s^2}$$

and

$$\frac{t^*_{1,p}}{n^{\frac{1}{2}}} = \frac{n^{(1-d)}\left(\hat{\beta}_{1,p} - 1\right)}{\left\{\frac{1}{n^2}\sum_{t=1}^{n}\underline{f}_t^2\right\}^{-1/2}\left\{\frac{1}{n^{2d-1}}\hat{\sigma}_1^2\right\}^{1/2}} \Rightarrow \frac{-\int \underline{B}_s B_{i(d-1)}}{\left\{\int \underline{B}_s^2 \int \underline{B}_{i(d-1)}^2 - \left(\int \underline{B}_s B_{i(d-1)}\right)^2\right\}^{1/2}}$$

[46] Recall that

$$n^{1-d}\left(\hat{\beta}_{1,p} - 1\right) \Rightarrow \frac{-\int \underline{B}_s B_{i(d-1)}}{\int \underline{B}_s^2}$$

$$\frac{1}{n^2}\sum_{t=1}^{n}\underline{s}_t^2 \Rightarrow \int B_s^2 \quad \frac{1}{n^{1+d}}\sum_{t=1}^{n}\underline{s}_t x_t \Rightarrow \int \underline{B}_s B_{i(d-1)}$$

$$\frac{1}{n^{2d}}\sum_{t=1}^{n}\underline{x}_t^2 \Rightarrow \int B_{i(d-1)}^2$$

$$\frac{1}{n}\sum_{t=1}^{n}\underline{s}_t\underline{y}_{p,t+p}, \frac{1}{n}\sum_{t=1}^{n}\underline{x}_t\underline{y}_{p,t+p}$$

and

$$\frac{1}{n}\sum_{t=1}^{n}\underline{y}^2_{p,t+p}$$

are all convergent.

For the R^2 statistic, we have

$$R^2_{1,p} = \frac{\hat{\beta}^2_{1,p}\sum_{t=1}^{n}\underline{f}^2_{t,k}}{\sum_{t=1}^{n}\underline{s}^2_{t+p}} = \frac{\hat{\beta}^2_{1,p}\sum_{t=1}^{n}\left(\underline{s}_t+\underline{x}_t\right)^2}{\sum_{t=1}^{n}\left(\underline{s}_t+\underline{y}_{p,t+p}\right)^2}$$

$$= \hat{\beta}^2_{1,p}\left(\frac{1}{1+\dfrac{\frac{1}{n^2}\sum_{t=1}^{n}\left(2\underline{s}_t\underline{y}_{p,t+p}+\underline{y}^2_{p,t+p}\right)}{\frac{1}{n^2}\sum_{t=1}^{n}\underline{s}^2_t}} + \frac{\frac{1}{n^2}\sum_{t=1}^{n}\left(2\underline{s}_t\underline{x}_t+\underline{x}^2_t\right)^2}{\frac{1}{n^2}\sum_{t=1}^{n}\left(\underline{s}_t+\underline{y}_{p,t+p}\right)^2}\right) \to 1$$

Proof of Theorem 3 *Define $\underline{w}_{p,t} = \underline{f}_{t-p} - \underline{s}_t = \underline{x}_{t-p} - \underline{y}_{p,t}$, where $x_t = f_t - s_t$ and $y_{t,p} = s_t - s_{t-p}$. The estimated coefficients in (17) can be written as $\hat{\delta}_1 = \left(\frac{1}{n}\sum_{t=1}^{n}\underline{Q}_t\underline{Q}'_t\right)^{-1}\frac{1}{n}\sum_{t=1}^{n}\underline{Q}_t\underline{w}_{p,t}$.*

By the ergodic theorem, $\frac{1}{n}\sum_{t=1}^{n}\underline{Q}_t\underline{Q}'_t \to E\left(Q_0Q'_0\right)$, as $E\left(Q_t\right)=0$ by assumption, and following the proof of Theorem 1, we have

$$\frac{1}{n^{d+1/2}}\sum_{t=1}^{n}\underline{w}_{p,t} = \frac{1}{n^{d+1/2}}\sum_{t=1}^{n}\left(\underline{x}_{t-p}-\underline{y}_{p,t}\right) \Rightarrow \int \underline{B}_{i(d-1)}$$

$$\frac{1}{n}\sum_{t=1}^{n}\underline{Q}_t\underline{w}_{p,t} = \frac{1}{n}\sum_{t=1}^{n}\underline{Q}_t\left(\underline{x}_{t-p}-\underline{y}_{p,t}\right) \to \left[\sum_{s=0}^{\infty}E\left(Q_s\Delta i^+_{-p}\right) - E\left(Q_0\left[s_0-s_{-p}\right]\right)\right]$$

$$\frac{1}{n^d}\sum_{t=1}^{n}\left\{\underline{Q}_t\underline{w}_{p,t} - \left[\sum_{s=0}^{\infty}E\left(Q_s\Delta i^+_{-p}\right) - E\left(Q_0\left[s_0-s_{-p}\right]\right)\right]\right\} \Rightarrow \int \underline{B}_{i(d-1)}dB_Q$$

$$\hat{\delta}_1 \to \widetilde{\Lambda}_p,\ n^{1-d}\left(\hat{\delta}_1-\widetilde{\Lambda}_p\right) \Rightarrow \left[E\left(Q_0Q'_0\right)\right]^{-1}\int \underline{B}_{i(d-1)}dB_Q$$

$$\frac{1}{n^{d-1/2}}\hat{e}_{p,[nr]} = \frac{1}{n^{d-1/2}}\left(\underline{x}_{[nr]-p}-\underline{y}_{p,[nr]}-\underline{Q}'_{[nr]}\hat{\delta}_1\right) \Rightarrow \underline{B}_{i(d-1)},\ \textit{and}$$

$$\frac{1}{n^{2d}}\sum_{t=1}^{n}\hat{e}^2_{p,t} \Rightarrow \int \underline{B}^2_{i(d-1)},\ \textit{where}\ \widetilde{\Lambda}_p = \left[E\left(Q_0Q'_0\right)\right]^{-1}\left[\sum_{s=0}^{\infty}E\left(Q_s\Delta i^+_{-p}\right) - E\left(Q_0\left[s_0-s_{-p}\right]\right)\right].$$

Substituting $\hat{e}_{p,t} = \underline{x}_{t-p} - \underline{y}_{p,t} - \hat{\delta}'_1\underline{Q}_t$ and $x_t = i^+_t$ and employing (A2) then write

$$\frac{1}{n}\sum_{t=1}^{n}\hat{e}_{p,t}\Delta s_{t+1} = C_1(I)\frac{1}{n}\sum_{t=1}^{n}\underline{i}^+_{t-p}\varepsilon_{t+1} - \frac{1}{n}\sum_{t=1}^{n}\Delta s_{t+1}\underline{Q}'_t\left(\hat{\delta}_1-\widetilde{\Lambda}_p\right) + \hat{\Lambda}_p - \hat{\Lambda}^*_p - \hat{\Sigma}_p$$

where $\hat{\Lambda}_p = \frac{1}{n}\sum_{t=1}^{n} \underline{i}_{t-p}^{+}(\widetilde{\varepsilon}_t - \widetilde{\varepsilon}_{t-1})$, $\hat{\Lambda}_p^* = \frac{1}{n}\sum_{t=1}^{n} \Delta s_{t+1}\underline{Q}_t^{'}\widetilde{\Lambda}_p$, *and* $\hat{\Sigma}_p = \frac{1}{n}\sum_{t=1}^{n}(s_t - s_{t-p})\Delta s_{t+1}$. *Following the proof of Theorem 1*

$$C_1(I)'\frac{1}{n^d}\sum_{t=1}^{n}\underline{i}_{t-p}^{+}\varepsilon_{t+1} \Rightarrow C_1(I)^{'}\int \underline{B}_{i(d-1)}dW$$

$$\frac{1}{n}\sum_{t=1}^{n}\Delta s_{t+1}\underline{Q}_t'(\hat{\delta}_1 - \widetilde{\Lambda}_p) \Rightarrow \gamma_{sq}^{'}\int \underline{B}_{i(d-1)}dB_Q$$

and $\hat{\Lambda}_p - \Lambda_p$, $\hat{\Lambda}_p^* - \Lambda_p^*$, *and* $\hat{\Sigma}_p - \Sigma_p$ *are all* $O_p(n^{-\frac{1}{2}})$. *Applying the continuous mapping theorem yields the results for the estimator. For the t statistic follow the proof of Corollary 1.1.*

ACKNOWLEDGEMENTS

Our thanks for helpful comments and suggestions on this work go to two referees, Richard Baillie, Giancarlo Corsetti, John Carlson, Ray Fair, Koichi Hamada, Oliver Linton, William Nordhaus, Chris Sims, Roger Moon, and participants at the Econometrics and Macro Lunch seminars at Yale University. Maynard gratefully acknowledges support from the Sloan Foundation and Phillips thanks the NSF for support under Grant Nos. SBR 94-22922 and SBR 97-30295.

REFERENCES

Agiakloglou C, Newbold P, Wohar M. 1993. Bias in an estimator of the fractional difference parameter. *Journal of Time Series Analysis* **14**(3): 235–246.

Andrews DWK. 1991. Heteroskedasticity and autocorrelation consistent covariance estimation. *Econometrica* **59**: 817–858.

Baillie RT. 1989. Tests of rational expectations and market efficiency. *Econometric Reviews* **8**: 151–168.

Baillie RT, Bollerslev T. 1994. The long memory of the forward premium. *Journal of International Money and Finance* **13**(5): 565–571.

Baillie RT, Bollerslev T. 2000. The forward premium anomaly is not as bad as you think. *Journal of International Money and Finance* **19**: 471–488.

Baillie RT, Osterberg WP. 1997. Central bank intervention and risk in the forward market. *Journal of International Economics* **43**: 483–497.

Barnhart SW, Szakmary AC. 1991. Testing the unbiased forward rate hypothesis: Evidence on unit roots, co-integration and stochastic coefficients. *Journal of Financial and Quantitative Analysis* **26**: 245–267.

Bekaert G. 1995. The time-variation of expected returns and volatility in foreign exchange markets. *Journal of Business and Economics Statistics* **13**(4): 397–408.

Bekaert G, Hodrick RJ. 1993. On biases in the measurement of foreign exchange risk premiums. *Journal of International Money and Finance* **12**: 115–138.

Clarida RH, Taylor MP. 1993. The term structure of forward exchange risk premia and the forecastability of spot exchange rates: Correcting the errors. Working paper no. 4442 (NBER: Cambridge, MA).

Clinton K. 1988. Transactions costs and covered interest arbitrage: Theory and evidence. *Journal of Political Economy* **96**(2): 358–370.

Crowder WJ. 1994. Foreign exchange market efficiency and common stochastic trends. *Journal of International Money and Finance* **13**(5): 551–564.

Crowder WJ. 1995. Covered interest parity and international capital market efficiency. *International Review of Economics and Finance* **4**(2): 115–132.
Doornik JA, Ooms M. 1999. A package for estimating, forecasting and simulating ARFIMA models: ARFIMA package 1.0 for Ox. Discussion paper, Nuffield College, Oxford.
Engel C. 1996. The forward discount anomaly and the risk premium: A survey of recent evidence. *Journal of Empirical Finance* **3**: 123–191.
Engle RF, Granger CWJ. 1987. Co-integration and error correction: Representation, estimation, and testing. *Econometrica* **55**(2): 251–276.
Evans M, Lewis KK. 1995. Do long term swings in the dollar affect estimates of the risk premia? *Review of Financial Studies* **8**(3): 709–742.
Fama E. 1984. Forward and spot exchange rates. *Journal of Monetary Economics* **14**: 319–338.
Frenkel JA. 1976. A monetary approach to the exchange rate: doctrinal aspects and empirical evidence. *Scandinavian Journal of Economics* **78**: 200–224.
Geweke J, Porter-Hudak S. 1983. The estimation and application of long memory time series models. *Journal of Time Series Analysis* **4**(4): 221–238.
Goodhart CAE, McMahon PC, Ngama YL. 1997. Why does the spot-forward discount fail to predict changes in future spot rates? *International Journal of Finance and Economics* **2**(2): 121–129.
Hai W, Nelson M, Yangru Wu. 1997. Understanding spot and forward exchange rate regressions. *Journal of Applied Econometrics* **12**: 715–734.
Hakkio C, Rush M. 1989. Market efficiency and cointegration: An application to the Sterling and Deutschemark exchange markets. *Journal of International Money and Finance* **8**: 75–88.
Hansen BE. 1992. Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory* **8**: 489–500.
Hodrick RJ. 1987. *The Empirical Evidence on the Efficiency of Forward and Futures Foreign Exchange markets*. Harwood: Chur.
Horvath MTK, Watson MW. 1994. Testing for cointegration when some of the cointegrating vectors are known. Technical working paper no. 171, National Bureau of Economic Research, Cambridge, MA.
Hurvich CM, Deo R, Brodsky J. 1998; The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time series. *Journal of Time Series Analysis* **19**: 19–46.
Ito T. 1992. *The Japanese Economy*. MIT Press: Cambridge, MA.
Kim CS, Phillips PCB. 1999a. Log periodogram regression: The nonstationary case. Yale University, mimeo.
Kim CS, Phillips PCB. 1999b. Modified log periodogram regression. Yale University, mimeo.
Kim CS, Phillips PCB. 2001. Fully modified estimation of fractional cointegration models. University of British Columbia and Yale University.
Kuersteiner GM. 1996. Net present value models for exchange rates. Dept of Economics, Yale University.
Künsch H. 1987; Statistical aspects of self-similar processes. In *Proceedings of the First World Congress of the Bernoulli Society*. Prokhorov Y, Sazanov VV (eds). VNU Science Press: Utrecht; 67–74.
Lewis K. 1994. Puzzles in international financial markets. In *The Handbook of International Economics*. North-Holland: Amsterdam.
Maynard A. 1998. Regression specification and tests of forward rate unbiasedness. Dept of Economics, Yale University.
Newbold P, Wohar ME, Rayner T, Kellard N, Ennew C. 1998. Two puzzles in the analysis of foreign exchange market efficiency. *International Review of Financial Analysis* **7**(2): 95–111.
Phillips PCB. 1986. Understanding spurious regressions in econometrics. *Journal of Econometrics* **33**: 311–140.
Phillips PCB. 1989. Partially identified econometric models. *Econometric Theory* **5**: 181–240.
Phillips PCB. 1998. New tools for understanding spurious regressions. *Econometrica*, **66**: 1299–1326.
Phillips PCB. 1999a. Discrete Fourier transforms of fractional processes. Yale University, working paper.
Phillips PCB. 1999b. Unit root log periodogram regression. Yale University, working paper.
Phillips PCB, Hansen BE. 1990. Statistical inference in instrumental variables regression with I(1) processes. *Review of Economic Studies* **57**: 99–125.
Phillips PCB, McFarland JW. 1997. Forward exchange market unbiasedness: the case of the Australian Dollar since 1984. *Journal of International Finance*, forthcoming.
Phillips PCB, Solo V. 1992. Asymptotics for linear processes. *Annals of Statistics* **20**(2): 971–1001.

Robinson PM. 1995a; Log periodogram regression of time series with long memory dependence. *Annals of Statistics* **23**: 1048–1072.
Robinson PM. 1995b. Gaussian semiparametric estimation of long range dependence. *Annals of Statistics* **23**: 1630–1661.
Sowell F. 1990. The fractional unit root distribution. *Econometrica* **58**(2): 495–505.
Taylor MP. 1987; Covered interest parity: A high frequency, high-quality data study. *Economica* **54**: 429–438.
Taylor MP. 1989. Covered interest arbitrage and market turbulence. *Economic Journal* **99**: 376–391.

