

Testing forward rate unbiasedness allowing for persistent regressors

Wei Liu, Alex Maynard *

Department of Economics, University of Toronto, Canada

Accepted 27 May 2004

Available online 19 September 2005

Abstract

Standard spot return/forward premium regressions have long been known to provide a strong rejection of unbiasedness. However, due to the strong autocorrelation in the forward premium, which shows estimated autoregressive roots close and in some cases statistically indistinguishable from one, recent literature has cast doubt on the finite sample accuracy of these tests. In fact, finite sample size distortion has now come to be considered as one of several possible explanations behind the forward premium puzzle. In order to pursue this possibility further, we revisit the unbiasedness hypothesis using more appropriate inference procedures. In particular, rather than relying on standard stationarity-based asymptotics, we model the forward premium as a near-unit root process and then test unbiasedness using the bounds tests of Cavanagh et al. (1995) [Cavanagh, C.L., Elliott, G., Stock, J.H., 1995. Inference in models with nearly integrated regressors. *Econometric Theory* 11, 1131–1147.], which are explicitly designed to provide accurate size under near-unit root assumptions. To summarize our empirical findings, confidence intervals on the largest root confirm uncertainty regarding the stationarity/nonstationarity of the forward premium. However, estimates of the error correlation suggest only modest simultaneity bias. Consequently, we can still reject unbiasedness at the 5% level, even when using appropriately sized bounds tests. This evidence tends to suggest that the forward premium puzzle is more robust than previously imagined. It would be interesting in further work to explore to what extent such

* Corresponding author. 150 St. George St., Toronto, ON, Canada M5S 3G7. Tel.: +1 416 978 4358.

E-mail address: amaynard@chass.utoronto.ca (A. Maynard).

conclusions extend to alternative characterizations of the persistence in the forward premium, such as long-memory and structural break models.

© 2005 Elsevier B.V. All rights reserved.

JEL classification: F31; C22

Keywords: Forward premium; Local-to-unity; Predictability; Uncovered Interest Parity; Near-unit root

1. Introduction

The strong rejection of forward rate unbiasedness has constituted a long-standing empirical puzzle in international finance. Forward rate unbiasedness is the rather simple and intuitive hypothesis that the current log forward exchange rate f_t should provide an unbiased forecast of next period's log spot exchange rate s_{t+1} . Alternatively stated, under unbiasedness, the forward premium $f_t - s_t$ provides an unbiased prediction of the spot return $s_{t+1} - s_t$. With covered arbitrage holding the forward premium equal to the interest rate differential, this is in fact equivalent to Uncovered Interest Rate Parity (UIP), a condition incorporated in many open economy macroeconomic models. UIP is generally interpreted as a joint hypothesis of market efficiency, rational expectations, and risk neutrality. However, a constant risk premium simply adds an intercept into the UIP equation.

While there is a vast empirical literature on the topic, following [Fama \(1984\)](#), the most common tests involve a simple regression of the spot return on the forward premium

$$s_{t+1} - s_t = \beta_0 + \beta_1(f_t - s_t) + \varepsilon_{t+1}, \quad (1)$$

in which unbiasedness implies both $\beta_0 = 0$ and $\beta_1 = 1$. A joint test of this hypothesis may therefore be based on a standard F test. However, the literature has focused primarily on testing the restriction that $\beta_1 = 1$, one of the reasons being that it allows for the presence of a (constant) risk premium.¹

Results from such regressions not only provide a strong rejection of unbiasedness, but also generally yield negative estimates of β_1 . Far from being unbiased, the forward premium therefore appears if anything to be a perverse predictor of the spot return, predicting movements in the opposite direction from which they actually occur.

The empirical results thus contrast sharply with the theoretical predictions, giving rise to a puzzle. Possible explanations naturally fall into one of two categories. The problem lies either with the theoretical model, which could perhaps be overly simplified or unrealistic in some of its assumptions, or, alternatively, with the empirical estimates, which could, for various reasons, suffer from bias or size distortion. Possible explanations based on more realistic or elaborate models include a time-varying risk premium ([Hodrick and Srivastava, 1986](#); [Mark, 1985](#); [Hodrick, 1989](#); [Modjtahedi, 1991](#); [Kaminski and Peruga, 1990](#)), habit persistence ([Backus et al., 1993](#)), heterogeneous trading behavior ([Frankel and Froot, 1988](#)) and feedback from monetary policy

¹ The unbiasedness test is sometimes interpreted as a test for a time-varying risk premium (see [Engel, 1996](#)).

(McCallum, 1994; Chinn and Meredith, 2004). On the econometric side, bias or size distortion may arise, due either to peso effects (Evans and Lewis, 1995) or to the strong serial correlation in the regressor (see discussion below).

The focus of this paper is on the last of these possible explanations. Estimated autoregressive coefficients on the forward premium are generally large, to the extent that unit roots may sometimes be difficult to reject (Crowder, 1994, 1995; Evans and Lewis, 1995). Under unbiasedness, the forward premium is clearly pre-determined, in the sense that it cannot correlate with future exchange rate innovations, but not exogenous, as it may correlate with past exchange rate movements. Consequently, traditional inference procedures provide only an approximation based on stationary asymptotics. It is well known that when strict exogeneity is violated such approximations break down in finite sample for autoregressive roots close to unity.

This near-unit root problem has led recent work (Bekaert and Hodrick, 2001; Maynard, 2003b; Roll and Yan, 2000; Tauchen, 2001; Goodhart et al., 1997; Newbold et al., 1998) to question the validity of empirical inference procedures underlying the forward premium puzzle, suggesting potentially serious bias and/or size distortion. It seems important to note that this is a finite sample problem, the existence of which does not depend on either the presence of a unit root or the local-to-unity approximations discussed below. For a sample size of 200 (corresponding to roughly 16 years of monthly exchange rate data) simulations by Mankiw and Shapiro (1986, p. 142, Table 1) show that actual rejection rates for a nominal 5% test in regressions of the type given in (1) can run as high as 29% when the autoregressive root in the regressor is close to, but strictly less than, 1. When a trend is included, rejection rates can exceed 60%.²

It is natural therefore to seek more appropriate inference procedures in testing unbiasedness.³ One approach has been to change either the regression specification, for example by first-differencing or pre-filtering prior to estimation (Roll and Yan, 2000; Newbold et al., 1998), or the estimation technique, using robust methods such as sign tests (Maynard, 2003a). However, if one wants to stay within the framework of the original unbiasedness regression/test as in (1), it is perhaps more convenient to maintain the same estimator, but modify the critical values in order to adjust for the non-standard nature of the distribution. This is the approach taken by Bekaert and Hodrick (2001), who provide critical values by bootstrapping the residuals from a first-stage vector autoregression. They found this to provide weaker evidence against unbiasedness. However, while the bootstrap provides higher order accuracy for a fixed value of the

² See Mankiw and Shapiro (1986, p. 142, Table 2).

³ One possible interpretation (Maynard and Phillips, 2001; Maynard, 2003b) is that the persistent behavior in the forward premium, since not observed in the spot return, constitutes a regression imbalance. Consider for example an $I(0)/I(1)$ regression. In this case, inference on β itself is misleading, but size distortion per se is not an issue, as the regression imbalance also violates unbiasedness. However, so long as the forward is not truly nonstationary, it may be reasonably modelled as in (1) under the null hypothesis as the sum of a small but highly serially correlated component ($f_t - s_t$) and a large noise component (ε_{t+1}). As such, one can reasonably consider size distortions under the null hypothesis as in the simulations of Bekaert and Hodrick (2001) and Roll and Yan (2000).

autoregressive root below unity, it is not consistent under local-to-unity assumptions, and as such is not guaranteed to provide good small sample approximations for roots close to unity. In other words, the bootstrap becomes less reliable at the same point at which standard large sample inference breaks down. Therefore, while the bootstrap likely improves upon asymptotic inference, it may not fully correct the size distortion. Thus if the reason for the puzzle was solely the persistence of the forward premium, the bootstrap method could still fail to account for it.⁴

The goal of this paper is to test unbiasedness using inference procedures that remain appropriate in the presence of near-unit root behavior in the forward premium. As such, we explicitly model the forward premium as a local-to-unity process (Chan and Wei, 1987; Chan, 1988; Phillips, 1987). Not only does this conveniently nest both the stationary and unit root alternatives, but local-to-unity asymptotics have also been shown to provide accurate asymptotic approximations when roots are close but not equal to unity (Chan, 1988; Nabeya and Sørensen, 1994). This contrasts with both stationary and unit root inference procedures.

Under local-to-unity assumptions, the distribution of the regression t -statistic is nonstandard, so that conventional inference is not generally valid. However, the appropriate critical value depends on only two nuisance parameters, the local-to-unity coefficient (c) and a correlation coefficient measuring the departure from strict exogeneity (Cavanagh et al., 1995). Unfortunately, consistent first-stage estimation of c is not possible in a pure time series context. However, using the bounds procedures of Cavanagh et al. (1995), asymptotically valid inference on the regression coefficient is possible using either (i) a first-stage confidence interval on c (Bonferroni method), (ii) the entire range of plausible values for c (sup-bound), or (iii) a joint confidence interval on both c and the regression coefficient (Scheffe method). Simulations by Cavanagh et al. (1995) show these tests to work well with sample sizes as small as 100. We employ all three methods to provide tests of unbiasedness that remain robust to the near-unit root behavior in the forward premium.

Our empirical findings can be summarized as follows. First-stage confidence intervals for the largest root on the forward premium show a large degree of uncertainty, with confidence intervals often including both 1 and values substantially below 1. This confirms the questionable nature of the assumptions upon which traditional inference procedures rely, making this an appropriate application for the use of local-to-unity based procedures. On the other hand, employing the bounds tests we are still able to reject unbiasedness, often with p -values that are only somewhat larger than those from standard regression. This suggests that size distortions, while present in standard inference procedure, are in this case not large enough to substantially impact the overall qualitative nature of the results. This finding appears to have little to do with the degree of persistence in the data, but rather derives from the modest degree of observed simultaneity. While the forward premium correlates with past exchange innovations, these innovations are noisy enough to keep this correlation relatively small.

⁴ We thank an anonymous referee for encouraging us to clarify this point.

Two alternative approaches have been to model the forward premium either as a long-memory process (Baillie and Bollerslev, 1994; Maynard and Phillips, 2001) or as a process with infrequent structural breaks (Sakoulis and Zivot, 2002). In both cases, this has led to similar concerns regarding the validity of asymptotic inference procedures (Baillie and Bollerslev, 2000; Maynard and Phillips, 2001; Sakoulis and Zivot, 2002). Obtaining appropriate inference procedures on the regression coefficients may be somewhat more challenging under these assumptions⁵, but remains an interesting avenue of future research.

The remainder of the paper is organized as follows. Section 2 lays out the model assumptions, Section 3 describes the econometric methodology, Section 4 describes the data, Section 5 presents the empirical results, Section 6 provides a short discussion and Section 7 concludes. Tables are provided at the end of the paper.

2. Uncovered Interest Parity in a local-to-unity setting

UIP implies that the forward premium acts as an unbiased predictor for the spot return. The most common tests used to investigate unbiasedness employ the simple regression given in (1). To keep the notation simple, we will focus on the following compact form of the model setting for the analysis hereafter,

$$y_t = \beta_0 + \beta_1 x_{t-1} + \varepsilon_{1t}, \quad (2)$$

where y_t represents the spot return at time t and the lagged forward premium is given by x_{t-1} . Forward rate unbiasedness jointly implies $\beta_1 = 1$, $\beta_0 = 0$, and $E_{t-1}\varepsilon_{1t} = 0$. However, since a constant risk premium implies $\beta_0 \neq 0$, the focus is generally on testing $\beta_1 = 1$.

In principle, the null hypothesis imposes no restrictions whatsoever on the forward premium. However, given that little persistence is observed in the spot return, it has been argued that the persistent behavior in the forward premium amounts in itself to a violation of unbiasedness, as it implies a statistical imbalance in (1) (e.g. $I(0)$ return/ $I(1)$ premium). For a truly nonstationary forward premium, this seems an inescapable conclusion. However, more realistically, if its root is merely close to unity, then under the null hypothesis, the large noise component in the spot return could quite easily drown out the persistent component inherited from the forward premium. This is, for example, the implicit assumption in the simulations of Bekaert and Hodrick (2001) and Roll and Yan (2000). Similar issues also arise in stock return/dividend yield regressions (e.g. Torous et al., 2005), in which dividend yields are near-unit root persistent and stock returns are noisy but less strongly correlated. In order to provide a flexible parametric specification, we follow Cavanagh et al. (1995, Eq. 1.1) in modelling the forward premium as an AR(k) process

$$x_t = \mu_x + v_t \quad b(L)v_t = (1 - \alpha L)\tilde{b}(L)v_t = \varepsilon_{2t}, \quad (3)$$

with $\tilde{b}(L) = \sum_{i=0}^k \tilde{b}_i L^i$ and $\tilde{b}_0 = 1$, where we have factored out the largest root α of the polynomial lag operator $b(L)$ and the remaining roots are assumed to be less than 1 in

⁵ Maynard (2003a) provides a valid test of unbiasedness under long-memory assumptions, but uses a sign statistic rather than a regression coefficient.

absolute value. Alternatively, this may be written in the standard form used in unit root testing as⁶

$$\Delta x_t = \tilde{\mu}_x + (\alpha - 1) \tilde{b}(1)x_{t-1} + a(L)\Delta x_{t-1} + \varepsilon_{2t}. \quad (4)$$

Consider next the innovations, ε_{1t} and ε_{2t} . Although under unbiasedness the white noise assumption on the regression residual ε_{1t} implies that x_{t-1} is predetermined in (2), there is no presumption of strict exogeneity.⁷ In other words, unbiasedness imposes no restriction on the relation between past exchange rate movements and the current forward premium. We therefore allow ε_{1t} and ε_{2t} to covary with correlation coefficient δ and specify the innovations as a martingale difference sequence where the variance–covariance matrix Σ has the form

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \delta\sigma_1\sigma_2 \\ \delta\sigma_2\sigma_1 & \sigma_2^2 \end{pmatrix}. \quad (5)$$

It is well known that inference in (2) can critically depend on the value of α (Torous et al., 2005). As neither unit root test results, nor point estimates of α fully convey the uncertainty surrounding α , we focus instead on the local-to-unity setting in this paper,

$$\alpha = 1 + c/T, \quad (6)$$

where c is a fixed constant and T represents the sample size. It has been well established that the resulting local-to-unity asymptotic distributions provide good approximations to finite-sample distributions when the root is close to 1 (Chan, 1988; Nabeya and Sørensen, 1994).

3. Methodology

Previous empirical results come most commonly from the standard OLS testing procedure. However, it is well understood that when $\delta \neq 0$, this procedure may provide an inappropriate approximation if the explanatory variable is highly persistent or non-stationary. It can be shown that, under the null hypothesis ($\beta_1 = 1$), the limiting distribution of the t -statistic for regression (2) is given by

$$t_{\beta_1} \Rightarrow \delta\tau_c + (1 - \delta^2)^{1/2}z, \quad (7)$$

where τ_c is a functional of a diffusion process which depends on c ⁸ and z represents a random variable from a standard normal distribution which is independent of that

⁶ See Cavanagh et al. (1995, Eq. 1.3). Here $\tilde{\mu}_x = (1 - \alpha)\tilde{b}(1)\mu_x$, $a_j = -\sum_{i=j+1}^k \tilde{a}_i$, and $\tilde{a}(L) = L^{-1}(1 - (1 - \alpha L)\tilde{b}(L))$.

⁷ To be more precise, under the null, $\text{corr}(\varepsilon_{1t}, \varepsilon_{2t})$ is not restricted to be zero, though $\text{corr}(\varepsilon_{1t}, \varepsilon_{2t-1})$ has to be.

⁸ See, for example, Stock (1991), $\tau_c = (\int J_c^{\mu 2})^{-1/2} \int J_c^{\mu} dB$, where B is a Brownian motion, $J_c^{\mu}(s) = J_c(s) - \int_0^s J_c(w)dw$ with the diffusion process J_c being defined by $dJ_c = cJ_c(s)ds + dB(s)$. See Stock (1991) and Cavanagh et al. (1995) for additional regularity conditions involving moment conditions on the residuals used here and in the bounds tests described below.

diffusion process. Denote the corresponding probability measure by $\mathcal{P}_{c,\delta}$. Both c and δ play important roles in determining the location and shape of $\mathcal{P}_{c,\delta}$. Cavanagh et al. (1995) show that $\mathcal{P}_{c,\delta}$ has a significant shift at $c=0$ as compared with $c < -10$, while Torous et al. (2005) demonstrate the empirical relevance of these distortions for stock return predictability. Standard critical values are correct only for $c \ll 0$ and/or $\delta=0$. Otherwise, the distribution of $\mathcal{P}_{c,\delta}$ may be highly nonstandard. With regressors such as forward premium, there is reason to suspect a large, possibly unit, autoregressive root. Furthermore, strict exogeneity of x_t is not imposed under the null hypothesis. These call into question the applicability of conventional critical values.

Given a fixed value of α , one potential solution to this problem, the so-called two-step procedure, is to use a consistent sequence of pretests for a unit root, followed by either stationary regression or cointegration techniques, depending on the outcome of the first stage test (Elliott and Stock, 1994). Unfortunately, this approach is valid only pointwise in α but does not work in a local-to-unity context. Cavanagh et al. (1995) provided numerical evidence for the considerable size distortion of this approach when α is less than, but statistically indistinguishable, from 1.

3.1. Size conservative test methods

The fact that the asymptotic critical values depend on c implies that we should incorporate the information on this nuisance parameter into our test approach. Stock (1991) showed that the distribution of critical values for the Augmented Dickey–Fuller test (ADF) is almost a monotonic function of c , $f(c)$. Hence, although we are not able to consistently estimate c because of the variation of the sample size T , we can still conduct inference over a range of possible values of c inverted from the ADF statistic.

Combining the information on c and β_1 , Cavanagh et al. (1995) introduced the following 3 asymptotically valid methods for testing restrictions on β_1 .⁹ Throughout this paper, we denote η as the desired (or nominal) test size and restrict the range of c from -40 to 10 (Stock, 1991; Cavanagh et al., 1995).

3.1.1. Sup-bound method

For each possible value of c and δ , the distribution for the OLS test statistic, t_{β_1} , can be simulated from $\mathcal{P}_{c,\delta}$. Since δ can be consistently estimated by the sample correlation between $\hat{\varepsilon}_{1t}$ and $\hat{\varepsilon}_{2t}$ (estimated disturbances from Eqs. (2) and (3)), the main idea of the Sup-bound approach is to compare the OLS test statistics with the simulated distribution corresponding to each possible value of c . Define $(\underline{d}_{c,\eta/2}, \bar{d}_{c,1-\eta/2})$ as the $100\eta/2\%$ and $100(1-\eta/2)\%$ quantiles of $\mathcal{P}_{c,\delta}$ for a given value of c . Then,

$$\left(\underline{d}_{\eta/2}, \bar{d}_{1-\eta/2} \right) = \left(\min_c \underline{d}_{c,\eta/2}, \max_c \bar{d}_{c,1-\eta/2} \right) \quad (8)$$

⁹ Extension of these methods to multivariate tests, such as joint tests on β_0 and β_1 could provide an interesting topic for future research.

constitutes the sup-bound non-rejection region with asymptotic level at least $1 - \eta$. Note that for any given value of c , the asymptotic size of the sup-bound for testing Eq. (2) is

$$S_s(c, \eta) = \mathcal{P}_{c, \delta} \left[t_{\beta_1} \notin \left(\underline{d}_{\eta/2}, \bar{d}_{1-\eta/2} \right) \right] \leq \mathcal{P}_{c, \delta} \left[t_{\beta_1} \notin \left(\underline{d}_{c, \eta/2}, \bar{d}_{c, 1-\eta/2} \right) \right] = \eta.$$

Thus, the test is conservative in the sense that the asymptotic rejection rate never exceeds η .

3.1.2. Bonferroni method

A small modification of the Sup-bound method gives us a slightly more powerful test (Cavanagh et al., 1995). The only difference here is that, instead of allowing all possible values of c , we use only those lying in a $100(1 - \eta_1)\%$ first-stage confidence interval, say (c_l, c_u) .¹⁰ Then, defining

$$\left(\underline{d}^B(\eta_1, \eta_2), \bar{d}^B(\eta_1, \eta_2) \right) = \left(\min_{c_l \leq c \leq c_u} \underline{d}_{c, \eta_2/2}, \max_{c_l \leq c \leq c_u} \bar{d}_{c, 1-\eta_2/2} \right) \quad (9)$$

as the Bonferroni non-rejection region, the Bonferroni inequality implies that no matter what the true value of c is, the test will always reject with probability less than or equal to $\eta = \eta_1 + \eta_2$ in large sample.

3.1.3. Scheffe method

Rather than using a simple t -test, the Scheffe method is based on the joint hypothesis $H_0: \beta_1 = \beta_1^0, c = c_0$. The Wald statistic is a natural statistic to use to perform this test (Cavanagh et al., 1995). The strategy is to project a $100(1 - \eta)\%$ joint confidence set for (β_1, c) , say $C_{\beta_1, c}(\eta)$, which can be constructed by inverting a level- η Wald test, onto the β_1 axis, that is

$$C_{\beta_1}(\eta) = \{ \beta_1 : \exists c \text{ such that } (\beta_1, c) \in C_{\beta_1, c}(\eta) \}, \quad (10)$$

where $C_{\beta_1}(\eta)$ denotes the confidence interval for β_1 .¹¹ This set, by construction, will have asymptotic confidence level at least $100(1 - \eta)\%$. The Scheffe test rejects over a region defined in terms of β_1 and c , which under the null hypothesis has probability at most η in large sample.

3.1.4. Size adjustment

All three methods introduced above have size at most η . In most situations, the actual rejection rates are strictly less than the desired level as the test critical values are rarely picked from the distribution corresponding to the true value of α . Due to the rather conservative nature of these tests (Cavanagh et al., 1995), when applied to real data, a few adjustments to the critical values may be necessary in order to obtain relatively accurate

¹⁰ For each value of c , we first generate the simulated distribution of t_{ADF} and get the $100\theta\%$ quantile values of these distributions, $f_q(c, \theta)$. To be concrete, a 95% confidence interval of c , for a given value of ADF statistic (t_{ADF}), can be obtained by inverting these test statistics (Stock, 1991), giving $CI_{c, 95\%} = (f_q^{-1}(t_{ADF}, 2.5\%), f_q^{-1}(t_{ADF}, 97.5\%))$.

¹¹ The distribution of the Wald statistic is also nonstandard.

statistical inferences from these procedures. The size adjustment is a lengthy process and the reader is referred to Cavanagh et al. (1995) for the details.¹²

The advantage of Scheffe method is that the critical value is δ -independent because δ does not appear in the limiting distribution of the Wald statistics (Cavanagh et al., 1995). But the trade-off is that we have to give up size adjustment, since the test size does vary with δ .

In their work, Cavanagh et al. (1995) numerically assessed the validity of these testing procedures at a nominal level of 10% for sample sizes as small as 100. The results are fairly good. The Bonferroni and Sup-bound procedures have Monte Carlo sizes close to 10% in almost all cases. The Scheffe procedure is somewhat more conservative in finite sample, due to the fact that it is not size adjusted.

4. Data description

The empirical results reported below are based on month-end spot and 1-month forward rates for the USD price of the Australian dollar (AUD), Canadian dollar (CAD), French Franc (FFR), German Deutschemark (GDM), Japanese Yen (JPY), and British Pound (BRP), obtained from Data Resources International. With the exception of the Yen and Australian Dollar, for which the series begin in 79:10 and 86:11, respectively, the series run from July of 1973 (73:7) until March of 2000 (00:3). The forward rates are calculated implicitly using the LIBOR 1-month interest differential.¹³ Following Bauer (2001), in order to avoid measurement error due to the timing of maturity dates (see Cornell, 1989; Bekaert and Hodrick, 1993), the exact number of days between the last business day in each month is used to translate the annual rates of return to monthly rates.¹⁴ Tests are conducted both over the full sample and, where possible, over two half-samples, 73:7–86:10 (early sample) and 86:11–00:3 (late sample).

5. Empirical results

In order to look at the severity of near-unit root problem, we estimated the largest autoregressive roots, α , for the forward premia under all three sub-samples described in Section 4 above: full (73:7–00:3), early (73:7–86:10) and late (86:11–00:3). The confidence intervals are calculated by inverting the Augmented Dickey–Fuller test

¹² After a few experiments for the empirical rejection rate of the Bonferroni method, the optimal first stage size for getting a test size of 5% is 0.50 for $\delta=0$, 0.34 ($\delta=0.5$), 0.23 ($\delta=0.6$), 0.18 ($\delta=0.7$), 0.14 ($\delta=0.8$), and 0.11 ($\delta=0.9$). All experiments were carried out with 20,000 Monte Carlo simulations for $T=1000$. Our simulations also reconfirmed the optimal first stage sizes for $\eta=10\%$ suggested in Cavanagh et al. (1995).

¹³ Covered interest arbitrage holds the forward premium equal to the nominal interest differential and Maynard and Phillips (2001) report that the nominal interest differential provides a cleaner series than the forward premium. The only exception is the Canadian Dollar, for which the LIBOR series started much later than the forward rate series. In this case, use of the LIBOR series would have involved a substantial loss of data.

¹⁴ The actual number of days between the last business day of each month is divided by the total number of days in the year (365 for the British Pound, but 360 by convention for the other Eurocurrency rates—see Bauer, 2001).

statistics as introduced in Stock (1991). Estimation results are reported in Table 1. Prior to estimation, a criterion must be used to select the number of autoregressive lags (k) for the forward premium process. Since the standard Bayesian Information Criteria (BIC) tends to select too small a lag (k) in conjunction with unit root tests, we use the Modified Information Criteria (MIC) proposed by Ng and Perron (2001), which takes into account the fact that the bias in the sum of the autoregressive coefficients is highly dependent on k . At a 95% level, most of the independent variables have a confidence interval including one, with the exception of the BRP/USD. This uncertainty regarding degree of persistence in the explanatory variable suggests possible size distortion in standard tests of $\beta_1 = 1$. Monte Carlo evidences can be found in Cavanagh et al. (1995).

The findings above confirm previous concerns about potential over-rejection when utilizing standard OLS approach or its modified alternative, the two-step procedure. Instead, all three uniformly valid testing methods introduced previously are applied to our data and the results are presented in Table 2. The first column of Table 2 reports the estimated t -statistics ($t_{\hat{\beta}_1}$) for testing the null hypothesis $\beta_1 = 1$. The test rejects when $|t_{\hat{\beta}_1}|$ is greater than the absolute value of the size-adjusted lower and upper critical values from the Bonferroni and Sup-bound methods, B_l , B_u and S_l , S_u , respectively. Instead of providing the bounds for the critical values, we reported the test results of the Scheffé method using

Table 1
95% confidence level for the largest autoregressive root in the forward premium

	$\hat{\delta}$	t_{ADF}	α_l	α_u	k
<i>Full sample (7/1973–3/2000)^a</i>					
AUD/USD	0.029	−2.138	0.899	1.019	3.000
CAD/USD	0.110	−2.648	0.929	1.006	5.000
FFR/USD	−0.089	−3.088	0.909	1.001	6.000
GDM/USD	−0.045	−1.734	0.963	1.012	6.000
JPY/USD	0.131	−2.602	0.910	1.008	4.000
BRP/USD	0.041	−3.632	0.688	0.977	5.000
<i>Early sample (7/1973–10/1986)</i>					
CAD/USD	0.003	−2.609	0.861	1.012	5.000
FFR/USD	−0.055	−3.062	0.819	1.003	6.000
GDM/USD	−0.289	−2.691	0.854	1.011	6.000
BRP/USD	0.038	−3.374	0.788	0.972	1.000
<i>Late sample (11/1986–3/2000)</i>					
AUD/USD	0.029	−2.138	0.899	1.019	3.000
CAD/USD	0.267	−1.503	0.941	1.025	1.000
FFR/USD	−0.048	−1.103	0.962	1.028	6.000
GDM/USD	−0.044	−1.801	0.922	1.023	6.000
JPY/USD	0.048	−0.846	0.972	1.029	3.000
BRP/USD	−0.088	−1.529	0.939	1.025	6.000

^a For the AUS/USD and JPY/USD rates, the samples begin in 11/86 and 10/79, respectively. This table reports the lower (α_l) and upper (α_u) limit of the largest autoregressive roots for the forward premium. The full sample and two sub-samples are presented separately. $\hat{\delta}$ is estimated consistently from the two OLS residuals. k represents the optimal number of autoregressive lags selected by the MIC procedure suggested by Ng and Perron (2001).

Table 2
Test results for $H_0: \beta_1 = 1$

	OLS		Bonferroni			Sup-bound			Scheffe	
	t_{β_1}	p^z	B_l	B_u	p^b	S_l	S_u	p^s	R/F	p^{sch}
<i>Full sample (7/1973–3/2000)^a</i>										
AUD/USD	−2.855**	0.004	−2.017	1.949	0.007	−2.024	1.977	0.009	0.000	0.077
CAD/USD	−4.586***	0.000	−2.156	1.873	0.000	−2.079	1.908	0.000	1.000	0.000
FFR/USD	−3.022***	0.003	−1.968	1.960	0.004	−1.977	2.127	0.004	1.000	0.000
GDM/USD	−2.420**	0.016	−1.912	2.084	0.018	−1.977	2.053	0.018	0.000	0.053
JPY/USD	−3.717***	0.000	−2.189	1.853	0.001	−2.116	1.908	0.001	1.000	0.000
BRP/USD	−3.680***	0.000	−1.999	1.950	0.001	−2.045	1.977	0.001	1.000	0.000
<i>Early sample (7/1973–10/1986)</i>										
CAD/USD	−3.290***	0.001	−1.950	1.975	0.002	−1.980	1.977	0.002	1.000	0.007
FFR/USD	−2.537***	0.011	−1.940	2.007	0.013	−1.977	2.070	0.013	1.000	0.017
GDM/USD	−2.297***	0.022	−1.788	2.271	0.015	−1.840	2.269	0.017	1.000	0.035
BRP/USD	−3.700***	0.000	−1.998	1.946	0.000	−2.040	1.977	0.000	1.000	0.000
<i>Late sample (11/1986–3/2000)</i>										
AUD/USD	−2.855**	0.004	−2.017	1.949	0.007	−2.024	1.977	0.009	0.000	0.077
CAD/USD	−3.083**	0.002	−2.579	1.584	0.017	−2.236	1.840	0.010	0.000	0.059
FFR/USD	−1.856	0.063	−1.880	2.077	0.071	−1.977	2.057	0.071	0.000	0.300
GDM/USD	−1.467	0.142	−1.916	2.087	0.147	−1.977	2.051	0.147	0.000	0.638
JPY/USD	−2.324**	0.020	−2.067	1.881	0.038	−2.057	1.977	0.038	0.000	0.166
BRP/USD	−0.923	0.356	−1.962	1.976	0.359	−1.977	2.125	0.366	0.000	0.787

Column 2 shows the standard t -statistic (t_{β_1}) used in all four tests. Column 3 gives the OLS p -value (p^z). Columns 4–5 show lower and upper bounds (B_l , B_u) for the t -statistic in column 2 using the Bonferroni method. Column 6 gives the Bonferroni p -value (p^b). Columns 7–8 give lower and upper bounds (S_l , S_u) using the sup-bound. Column 9 gives the sup-bound p -value (p^s). A one in column 10 indicates a 5% rejection using the Scheffe bounds and column 11 gives the Scheffe p -value (p^{sch}).

^a For the AUS/USD and JPY/USD rates, the samples begin in 11/86 and 10/79, respectively.

** Denotes rejection at 5% significance level using both the Bonferroni and Sup-bound tests, but not for the Scheffe test.

*** Denotes rejection at 5% using all three tests.

an indicator variable, R/F , equal to one if the test rejects. An approximate p -value (p) for each method is also presented in Table 2. Defining p_c as the p -value for a given c , the reported p -value, p , is conservative in the sense that $p = \max_c p_c$ for all permissible c .¹⁵

Even after accounting for the possible near-unit root behavior of the forward premium, evidence against unbiasedness is found in all cases for both full samples and early sub-samples when using the Bonferroni and Sup-bound methods. However, it is only rejected in half of the samples in the late sub-period. The Scheffe method, which is a bit more conservative, also fails to reject for the AUD/USD, GDM/USD, JPY/USD and CAD/USD.

We next address a particularly puzzling aspect of the forward premium puzzle, namely the fact that the unbiasedness regression generally yields (often significantly) negative estimates of β_1 . Given the near-unit root problem, it is again not clear, from standard

¹⁵ Details on the p -value calculation are given in Appendix A.

regression, whether such negative estimates are actually significant or even whether the negative sign is really meaningful or simply the result of estimation bias. We revisit this issue by applying the bounds tests to the null hypothesis $H_0: \beta_1=0$. The results are presented in Table 3. The null is rejected by all techniques for JPY/USD in the full sample. Although we find some statistical evidence for $\beta_1<0$ for the CAD/USD and BRP/USD rates when employing the Bonferroni and Sup-bounds, this is not confirmed by the Scheffe method. No other slope coefficients are found to be significantly negative.

Despite the large uncertainty in the first-stage estimates of c , our overall findings tend to confirm the stylized facts underlying the forward premium puzzle. However, the Scheffe method does give a few different conclusions due to its conservative characteristic. In both the unbiasedness and negativity tests, we find cases in which the null is rejected by the standard OLS procedure and size adjusted bounds tests while it fails to be rejected by the Scheffe method.

Table 3

Test results for $H_0: \beta_1^0=0$

	OLS		Bonferroni			Sup-bound			Scheffe	
	$t_{\hat{\beta}_1}$	p^z	B_l	B_u	p^b	S_l	S_u	p^s	R/F	p^{sch}
<i>Full sample (7/1973–3/2000)^a</i>										
AUD/USD	–1.649	0.099	–2.017	1.949	0.156	–2.024	1.977	0.163	0.000	0.488
CAD/USD	–2.439**	0.015	–2.156	1.873	0.032	–2.079	1.908	0.040	0.000	0.063
FFR/USD	–1.136	0.256	–1.968	1.960	0.259	–1.977	2.127	0.266	0.000	0.109
GDM/USD	–1.014	0.311	–1.912	2.084	0.315	–1.977	2.053	0.321	0.000	0.579
JPY/USD	–2.663***	0.008	–2.189	1.853	0.018	–2.116	1.908	0.023	1.000	0.033
BRP/USD	–2.161**	0.031	–1.999	1.950	0.040	–2.045	1.977	0.045	0.000	0.113
<i>Early sample (7/1973–10/1986)</i>										
CAD/USD	–1.771	0.077	–1.950	1.975	0.107	–1.980	1.977	0.132	0.000	0.231
FFR/USD	–1.094	0.274	–1.940	2.007	0.278	–1.977	2.070	0.283	0.000	0.338
GDM/USD	–1.581	0.114	–1.788	2.271	0.080	–1.840	2.269	0.094	0.000	0.156
BRP/USD	–2.388**	0.017	–1.998	1.946	0.023	–2.040	1.977	0.033	0.000	0.070
<i>Late sample (11/1986–3/2000)</i>										
AUD/USD	–1.649	0.099	–2.017	1.949	0.156	–2.024	1.977	0.163	0.000	0.488
CAD/USD	–1.574	0.115	–2.579	1.584	0.391	–2.236	1.840	0.305	0.000	0.498
FFR/USD	–0.781	0.435	–1.880	2.077	0.438	–1.977	2.057	0.444	0.000	0.825
GDM/USD	–0.452	0.651	–1.916	2.087	0.658	–1.977	2.051	0.664	0.000	0.968
JPY/USD	–1.606	0.108	–2.067	1.881	0.174	–2.057	1.977	0.174	0.000	0.465
BRP/USD	–0.148	0.882	–1.962	1.976	0.888	–1.977	2.125	0.899	0.000	1.000

Column 2 shows the standard t -statistic ($t_{\hat{\beta}_1}$) used in all four tests. Column 3 gives the OLS p -value (p^z). Columns 4–5 show lower and upper bounds (B_l , B_u) for the t -statistic in column 2 using the Bonferroni method. Column 6 gives the Bonferroni p -value (p^b). Columns 7–8 give lower and upper bounds (S_l , S_u) using the sup-bound. Column 9 gives the sup-bound p -value (p^s). A one in column 10 indicates a 5% rejection using the Scheffe bounds and column 11 gives the Scheffe p -value (p^{sch}).

^a For the AUS/USD and JPY/USD rates, the samples begin in 11/86 and 10/79, respectively.

** Denotes rejection at 5% significance level using both the Bonferroni and Sup-bound tests, but not for the Scheffe test.

*** Denotes rejection at 5% using all three tests.

6. Discussion

These results tend to confirm the robustness of the puzzle, suggesting that size distortion may be somewhat less troublesome for this test than previously imagined. This contrasts somewhat with the related work on stock return/dividend yield regressions (Torous et al., 2005), in which the bounds tests lead to considerably weaker evidence in favor of return predictability. Some intuition for this difference comes from the estimated values of δ shown in Table 1. While fairly modest in our case, Torous et al. (2005) report values of δ as large as 0.9. For the other regressors they considered, such as term spreads and short-term interest rates, the value of δ was similar to ours, and as in our case, the qualitative conclusions from standard tests were reconfirmed by the bounds tests.

Some caveats are nevertheless in order. While the parametric setting employed here provides a flexible approach to modelling the dynamics of the forward premium, it does not incorporate some of the more complicated dynamics considered in the literature, such as long-memory and trend break models. Similar bias and size distortion have been shown to arise under these models (see Baillie and Bollerslev, 2000; Maynard and Phillips, 2001; Sakoulis and Zivot, 2002), suggesting that the general approach taken here might be promising in these cases too. However, the limiting distributions appear more complicated and the exact nature of the corrections to the critical values would likely be different under these assumptions.

The nonstandard component to the limit distribution is also simplified by the strict imposition of the null hypothesis, which requires a white noise assumption on the spot return innovation ε_{1t} . Given the strong serial correlation observed in the forward premium but not the spot return, it is not unreasonable to suspect serial correlation in the residual and this again leads to more complicated distributions (Maynard and Phillips, 2001; Maynard, 2003b), possibly requiring more elaborate bias adjustments of the type considered by Phillips and Hansen (1990). Such serial correlation constitutes a violation of unbiasedness, so that our tests would have correctly rejected unbiasedness. Yet they could still provide misleading inference on β . For example, in Baillie and Bollerslev's (2000) UIP-FIGARCH model $\beta_1 = 1$, but the residual (an omitted FIGARCH term) has long-memory. Although the long-memory of the residual in their model constitutes a violation of unbiasedness, it is still perfectly consistent with economic theory, whereas (a possibly erroneous) inference that $\beta_1 < 0$ might not be.

7. Conclusion

The highly autocorrelated or near-unit root behavior of the forward premium has called into question the validity of standard unbiasedness tests, thereby providing one potential explanation for the forward premium puzzle. Modelling the forward premium as a local-to-unity process, we apply tests that maintain good size properties for values of the largest autoregressive root close or equal to unity. We find that even after accounting for the near-unit root behavior of the forward premium, unbiasedness can still be strongly rejected. Some intuition for this finding comes from the finding that the estimated residual correlations governing the endogeneity bias are generally modest. While a number of

caveats apply, these results tend to suggest that size distortion may not provide the sole explanation behind this puzzle.

Acknowledgements

We thank two anonymous referees for useful comments and suggestions and Yoon-Jae Whang for helpful discussion. We are grateful to James Stock for providing us with Gauss programs for implementation of the bounds procedure. We also thank Serena NG and Pierre Perron for use of their publicly available Gauss code for the MIC selection procedure. Maynard thanks Connaught and SSHRC for research funding.

Appendix A. Details on p -value calculation

Consider the CAD/USD as an example. In this case, the range of c used in the first stage of the Bonferroni method was $(-30, -5)$. For each value of c in this range, we simulate the distribution of the test statistic and calculate the proportion of simulated test statistics below (above) the observed test statistic (below -4.586 for the CAD/USD). The maximum value of this proportion over the initial range of c is a conservative one-sided p -value. Since our tests are two-sided, we report twice this one-sided p -value in the tables. This takes a value less than α if, and only if, the test rejects. It constitutes an approximate two-sided p -value in the standard sense when the distribution is symmetric and can be interpreted as an equal-tailed two-sided p -value when the distribution is asymmetric.

The sup-bound size adjustment is accomplished via an adjustment to the conservative critical values (or bounds) given in (8). In order to obtain a logically consistent size-adjusted p -value, we need an equivalent size adjustment for the t -statistic itself. Assume that the unadjusted t -statistic $t_{\beta_1} > 0$. The unadjusted sup-bound critical value is then given by $\bar{d}_{1-\eta/2} = \bar{d}_{c^*, 1-\eta/2} = \max_{-40 \leq c \leq 10} \bar{d}_{c, 1-\eta/2}$, where $\bar{d}_{c, 1-\eta/2}$ is the critical value for given c , and c^* is the maximizing argument. Let $\bar{d}_{1-\eta/2}^a$ denote the adjusted sup-bound critical value, where the adjustment procedure is detailed in Cavanagh et al. (1995). A rejection occurs when the unadjusted t -statistic (t_{β_1}) exceeds the adjusted critical value ($\bar{d}_{1-\eta/2}^a$) (see Cavanagh et al., 1995). However, defining the adjusted t -statistic as

$$t_{\beta_1}^a = t_{\beta_1} + \bar{d}_{1-\eta/2} - \bar{d}_{1-\eta/2}^a. \quad (11)$$

this rejection occurs exactly when the adjusted t -statistic ($t_{\beta_1}^a$) exceeds the unadjusted critical value ($\bar{d}_{1-\eta/2}$). We therefore define the size-adjusted equal-tailed two-sided p -value as $2P_{c^*, \delta}[T > t_{\beta_1}^a]$. The analogous definition applies for $t_{\beta_1} < 0$.

References

- Backus, D., Gregory, A., Telmer, C., 1993. Accounting for forward rates in markets for foreign currency. *Journal of Finance* 48, 1887–1908.

- Baillie, R.T., Bollerslev, T., 1994. The long memory of the forward premium. *Journal of International Money and Finance* 13 (5), 565–571.
- Baillie, R.T., Bollerslev, T., 2000. The forward premium anomaly is not as bad as you think. *Journal of International Money and Finance* 19, 471–488.
- Bauer, G.H., 2001. The foreign exchange risk premium over the long run. Working Paper. William E. Simon Graduate School of Business Administration University of Rochester.
- Bekaert, G., Hodrick, R.J., 1993. On biases in the measurement of foreign exchange risk premiums. *Journal of International Money and Finance* 12, 115–138.
- Bekaert, G., Hodrick, R., 2001. Expectations hypothesis tests. *Journal of Finance* 56 (38), 1357–1394.
- Cavanagh, C.L., Elliott, G., Stock, J.H., 1995. Inference in models with nearly integrated regressors. *Econometric Theory* 11, 1131–1147.
- Chan, N., 1988. The parameter inference for nearly nonstationary time series. *Journal of the American Statistical Association* 83 (403), 857–862.
- Chan, N.H., Wei, C.Z., 1987. Asymptotic inference for nearly nonstationary AR(1) processes. *Annals of Statistics* 15 (3), 1050–1063.
- Chinn, M.D., Meredith, G., 2004. Monetary policy and long-horizon uncovered interest parity. *IMF Staff Papers* 51 (3).
- Cornell, B., 1989. The impact of data errors on measurement of the foreign exchange risk premium. *Journal of International Money and Finance* 8, 147–157.
- Crowder, W.J., 1994. Foreign exchange market efficiency and common stochastic trends. *Journal of International Money and Finance* 13 (5), 551–564.
- Crowder, W.J., 1995. Covered interest parity and international capital market efficiency. *International Review of Economics and Finance* 4 (2), 115–132.
- Elliott, G., Stock, J., 1994. Inference in time series regression when the order of integration of a regressor is unknown. *Econometric Theory* 10, 672–700.
- Engel, C., 1996. The forward discount anomaly and the risk premium: a survey of recent evidence. *Journal of Empirical Finance* 3, 123–191.
- Evans, M.D., Lewis, K.K., 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.
- Frankel, J., Froot, K.A., 1988. Explaining the demand for dollars: international rates of return and the expectations of chartists and fundamentalists. In: Chambers, R., Paarlberg, P. (Eds.), *Agriculture, Macroeconomics and the Exchange Rate*. Westview Press, Boulder, Co.
- Goodhart, C., McMahon, P., Ngama, Y., 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.
- Hodrick, R.J., 1989. U.S. international capital flows: perspectives from rational maximizing models. *Carnegie-Rochester Conference Series on Public Policy* 30, 231–288.
- Hodrick, R., Srivastava, S., 1986. The covariation of risk premiums and expected future spot rates. *Journal of International Money and Finance* 3, 5–30.
- Kaminski, G., Peruga, R., 1990. Can a time varying risk premium explain excess returns in the forward market for foreign exchange? *Journal of International Economics* 28, 47–70.
- Mankiw, N., Shapiro, M., 1986. Do we reject too often? Small sample properties of tests of rational expectations models. *Economics Letters* 20, 139–145.
- Mark, N., 1985. On time varying risk premia in the foreign exchange market: an econometric analysis. *Journal of Monetary Economics* 16, 3–18.
- Maynard, A., 2003a. The forward premium anomaly: statistical artifact or economic puzzle? New evidence from robust tests. Mimeo, University of Toronto.
- Maynard, A., 2003b. Testing for forward rate unbiasedness: on regression in levels and in returns. *Review of Economics and Statistics* 85 (2), 313–327.
- Maynard, A., Phillips, P.C.B., 2001. Rethinking an old empirical puzzle: econometric evidence on the forward discount anomaly. *Journal of Applied Econometrics* 16 (6), 671–708.
- McCallum, B.T., 1994. A reconsideration of the uncovered interest parity relationship. *Journal of Monetary Economics* 33, 105–132.

- Modjtahedi, B., 1991. Multiple maturities and time-varying risk premia in forward exchange markets: an econometric analysis. *Journal of International Economics* 30, 69–86.
- Nabeya, S., Sørensen, B., 1994. Asymptotic distributions of the least-squares estimators and test statistics in the near unit root model with non-zero initial value and local drift and trend. *Econometric Theory* 10, 937–966.
- Newbold, P., Wohar, M., Rayner, T., Kellard, N., Ennew, C., 1998. Two puzzles in the analysis of foreign exchange market inefficiency. *International Review of Financial Analysis* 7 (2), 95–111.
- Ng, S., Perron, P., 2001. Lag length selection and the construction of unit root tests with good size and power. *Econometrica* 69 (6), 1519–1554.
- Phillips, P.C.B., 1987. Towards a unified asymptotic theory for autoregression. *Biometrika* 74 (3), 535–547.
- Phillips, P.C.B., Hansen, B.E., 1990. Statistical inference in instrumental variables regression with $I(1)$ processes. *Review of Economic Studies* 57, 99–125.
- Roll, R., Yan, S., 2000. An explanation of the forward premium ‘puzzle’. *European Financial Management* 6 (2), 121–148.
- Sakoulis, G., Zivot, E., 2002. Time-Variation and Structural Change in the Forward Discount: Implications for the Forward Rate Unbiasedness Hypothesis. Department of Economics, University of Washington.
- Stock, J.H., 1991. Confidence intervals for the largest autoregressive root in U.S. macroeconomic time series. *Journal of Monetary Economics* 28, 435–439.
- Tauchen, G., 2001. The bias of tests for a risk premium in forward exchange rates. *Journal of Empirical Finance* 8, 695–704.
- Torous, W., Valkanov, R., Yan, S., 2005. On predicting stock returns with nearly integrated explanatory variables. *The Journal of Business* 77 (4), 937–966.