10-25-2012

Long Memory Regressors and Predictive Regressions: A two-stage rebalancing approach

Alex Maynard

Aaron Smallwood

Mark E. Wohar
University of Nebraska at Omaha

Long Memory Regressors and Predictive Regressions: A two-stage rebalancing approach

Alex Maynard, Aaron Smallwood,* and Mark E. Wohar

Department of Economics, University of Guelph, McKinnan Building Guelph, ON, Canada, N1G 2W1.

Department of Economics, University of Texas-Arlington, Arlington, TX 76051

Department of Economics, University of Nebraska at Omaha, RH 512K, Omaha, NE 68182-0286

December 3, 2008

Abstract

Predictability tests with long memory regressors entail both size distortion and potential regression imbalance. Addressing both problems simultaneously, this paper proposes a two-step procedure that rebalances the predictive regression by fractionally differencing the predictor based on a first-stage estimation of the memory parameter. A full set of asymptotic results are provided. The second-stage t-statistic used to test predictability has a standard normal limiting distribution. Extensive simulations indicate that our procedure has good size, is robust to estimation error in the first stage, and can yield improved power over cases in which an integer order is assumed for the regressor. We use our procedure to provide a valid test of forward rate unbiasedness that allows for a long memory forward premium.

JEL classifications: C22, C12, F31

Key words: Predictive regressions, long memory, forward rate unbiasedness

*Corresponding author. *E-mail:* smallwood@uta.edu, *Fax:* 817-272-3145, *Phone:* 817-272-3062. This paper has substantially benefited from comments by two referees and an associate editor. We thank Arthur Lewbel and Katsumi Shimotsu for useful discussion. We also thank seminar participants at the University of Oklahoma, the University of Guelph (FARE), the 2005 International Conference in Economics and Finance in Washington D.C., and the Conference on International Economics and Economics Research held in 2008 at the University of Saskatchewan. Maynard thanks the SSHRC for research funding. Any remaining errors are our own.

1 Introduction

A common aspect of many predictive regressions is the highly persistent behavior of the regressor. Examples include stock return predictability tests using dividend-price ratios, earning-price ratios or interest rates as regressors, tests of the permanent income hypothesis, and tests of forward rate unbiasedness. It has been understood since Mankiw and Shapiro (1986) that this persistence may lead to size distortion. The extant literature has also focused on the potential for regression imbalance (i.e. stationary dependent variable, near nonstationary regressor) in excess returns regressions, where returns typically exhibit little or no persistence.

The problem of size distortion has led to a large empirical literature and the development of techniques designed to address these issues, with much of the literature concentrated in the context of local-to-unity models (Cavanagh *et al.* 1995, Jansson and Moreira 2006). However, persistence can also manifest itself in the form of long memory, which has been documented in many predictive regressors including the forward premium (Baillie and Bollerslev 1994, Maynard and Phillips 2001), volatility (Baillie and Bollerslev 2000) and dividend yields (Koustas and Serletis 2005).

With a few exceptions (Campbell and Dufour 1997), the econometric literature on predictive regressions has focused on the case of near unit root regressors. The most common approach has been to maintain the same regression specification, but to adjust the critical values in order to preserve correct test size. This may be attractive in some applications when economic theory suggests this form of the alternative. Moreover, if the largest root of the regressor is merely close, but not equal to unity, then the original regression specification may still be compatible with a stationary return series for the dependent variable. Thus, while size distortion is of central importance in predictive regressions with near unit roots, it has sometimes been argued that problems of regression imbalance may be avoided. This is no longer true when predictive regressors have long memory, since imbalance may exist when a short memory variable is regressed on a long memory regressor.

In this paper we propose a simple, intuitive two-stage rebalancing procedure that addresses both the regression imbalance and size distortion discussed above, while allowing for (without imposing) long memory behavior in the predictive regressor. In the first stage, either a semi-parametric or parametric estimator may be used to estimate the degree of long memory in the regressor. Then, in the second stage, the predictive regression is rebalanced by fractionally differencing the regressor. This rebalances the alternative hypothesis, while leaving the null hypothesis unchanged and thus allowing for a valid test of predictability. By fractionally differencing the regressor, we also remove the source of size distortion, yielding a t-statistic in the second-stage regression with correct size.

We derive the large sample theory for our proposed technique and demonstrate its applicability by a detailed Monte Carlo study. The simulation study confirms the potential for size distortion in the absence of rebalancing (or other size adjustment), while showing that our two-stage procedure works well. We also find that estimation and inference in the second stage are robust to estimation error or even modest misspecification in the first stage. We see this as an important practical benefit, since the memory parameter can be difficult to estimate in small samples (see Nielsen and Frederiksen (2005), for a survey).

As an empirical application of our two-step method, we consider tests of the forward rate unbiasedness hypothesis (FRUH). This hypothesis may be re-written as a test of the predictability of excess foreign exchange rate returns using the information in the lagged forward premium. While excess returns are arguably stationary, beginning with Baillie and Bollerslev (1994), several studies have documented long memory in the forward premium and have underlined the potential importance of the resulting imbalance for understanding the strong and rather paradoxical rejection of the FRUH resulting from these regressions (Baillie and Bollerslev 2000, Maynard and Phillips 2001).

Our method provides a reliable test of FRUH in the presence of the presence of a long memory forward premium. This contrasts with standard tests, which may either overstate the evidence against FRUH due to size distortions or understate this evidence due to the power reductions inherent in an imbalanced alternative model. We fail to reject unbiasedness for two of five currencies, providing some support to the contention that the evidence against unbiasedness may be overstated due to the long-memory characteristics of the data. Nonetheless, our tests reconfirm the validity of earlier rejections of FRUH for the remaining currencies.

The rest of the paper is organized as follows. In Section 2, we provide background on the FRUH, highlighting the relevant econometric issues underlying our analysis. Section 3 outlines the proposed two-step predictability test using long memory regressors, provides its large sample properties, and discusses fist-stage estimation of the long memory parameter. Extensive simulation evidence is provided in Section 4. Section 5 contains the results of our empirical investigation of the FRUH, and Section 6 provides a summary of our results. An appendix contains the proofs affiliated with the asymptotic properties of our two-step procedure.

2 Background

While the methodology we propose applies to any predictability test with long memory regressors, we motivate our procedure with a discussion of the forward rate unbiasedness hypothesis (FRUH). The empirical results from tests of the FRUH have provided the stylized facts underpinning what is often referred to as the forward discount anomaly. The FRUH states that the current (log) forward exchange rate (f_t) should provide an unbiased forecast of next period's (log) spot exchange rate (s_t), i.e. $E_t s_{t+1} = f_t$. This implies the orthogonality or non-predictability condition

$$E_t[s_{t+1} - f_t] = 0, \tag{1}$$

in which next period's forecast error ($s_{t+1} - f_t$) is unpredictable using any information available at time t. Thus, the FRUH can be thought of as a test of excess return predictability.

The classic predictability regression

$$s_{t+1} - f_t = c_1 + b_1(f_t - s_t) + e_{1t+1} \tag{2}$$

provides a simple specification in which to formulate the alternative hypothesis, along with the testable restriction $b_1 = 0$. This regression is equivalent to a spot return/forward premium regression

$$s_{t+1} - s_t = c_2 + b_2(f_t - s_t) + e_{2t+1}, \tag{3}$$

where $b_2 = b_1 + 1 = 1$ under the FRUH. While these two equivalent regressions are the most common in the literature, it will be important to our analysis below to note that only the form of the null hypothesis in (1) is implied by the FRUH. Theory does not dictate the exact form of the alternative hypothesis and the regressions given above are simply convenient specifications.

The empirical results from the predictability regressions in (2) and (3) are quite puzzling. Not only is unbiasedness strongly rejected (i.e. $b_1 \neq 0$; $b_2 \neq 1$), but the estimates of b_2 are invariably negative. In other words, the forward premium is not only found to be a biased predictor, it is also a perverse predictor, even mispredicting the direction of change in exchange rates.

Recently it has been recognized that the forward premium has long memory characteristics (Baillie and Bollerslev 1994) and that this calls into question the statistical validity of standard tests of FRUH (Baillie and Bollerslev 2000, Maynard and Phillips 2001). Nevertheless, to date few studies have specifically attempted to design predictability tests that allow for long memory regressors.[1] In this paper, we provide a valid test for predictability in the context of long memory. A second issue that arises with long memory regressors in predictive regressions, such as (2), is a possible statistical imbalance, since the return variables on the LHS are generally short memory. For example, under the FRUH, the forecast error $(s_{t+1} - f_t)$ must not only have short memory, but must also be serially uncorrelated in order to meet the restriction in (1). Empirically, its short memory characteristics are apparent in the data. For example, a plot of the log of excess returns for Canada from June 1973 to March 2000 is depicted in Figure 1. By contrast, a time series plot of the forward premium for Canada for the same time period, in Figure 2, exhibits very different and much more persistent behavior. The autocorrelations for these two series are depicted in Figure 3. These figures clearly indicate that the forward premium has much stronger memory characteristics than the excess returns, which show very little autocorrelation.

[FIGURES 1-3 ABOUT HERE]

Although it may cause size distortion, the apparent imbalance between the components in (2) is not inconsistent with the FRUH, which implies $b_1 = 0$, in which case $s_{t+1} - f_t$ and $f_t - s_t$ are free to exhibit different orders of integrations. If test size were the only issue, corrections to the critical values could feasibly be derived. However, the apparent imbalance in (2) can cause fundamental problems under the alternative hypothesis as

[1] Proposed corrections have generally been undertaken employing an autoregressive or near unit root model. For corrections in the stock return predictability literature see Stambaugh (1999), Rapach and Wohar (2006), Torous *et al.* (2005), and the references within.

characterized by this regression specification. In fact, if the order of integration of the RHS variable exceeds 0.5, such that it is non-stationary, then the regression attempts to relate a stationary dependent variable to a non-stationary regressor. Since the RHS variable has a tendency to wander off, whereas the LHS variable does not, b_2 =0 is the only possible parameter value consistent with this statistical imbalance and show that the OLS estimate of b_2 converges to zero with a nonstandard distribution.[2]

Note that our ultimate interest lies in testing the non-predictability of the forward rate forecast error in (1) and not simply the parameter restriction in the convenient but rather simple regression specification given by (2). In other words, the parameter restriction b_1 =0 is only necessary, but not sufficient for the FRUH. From this perspective, the imbalance in (2) (short- memory excess returns, long memory forward premium) does not necessarily imply that the null hypothesis in (1) holds but rather indicates that (2) does not provide a meaningful parametric specification in which to couch the alternative. In particular, it does not allow for a rejection of unbiasedness due to the presence of a stationary short-memory risk premium. This imbalance thus calls for a test that not only maintains correct size but also allows alternative specifications in which the dependent and independent variable are both integrated of the same order.

Although there has been much previous interest on this question, our application is the first we know of to provide an asymptotically justified regression-based test of the FRUH that is valid for long-memory regressors. Baillie and Bollerslev (2000), and Maynard and Phillips (2001) provide simulation evidence and asymptotic theory that demonstrate the problems inherent in traditional tests but do not offer or employ any empirical tests. Liu and Maynard (2005) and Rossi (2005) , test the FRUH using local-to-unity based procedures, which do not allow for long-memory regressors. Departing from the standard regression approach, Maynard (2006) tests for sign predictability using the nonparametric procedure of Campbell and Dufour (1997), which remains valid under long memory assumptions. However, additional symmetry conditions are required to equate a lack of sign predictability with unbiasedness, and it is not clear that the sign test specifically addresses the question of regression imbalance considered here.[3]

Thus, as discussed in the previous literature, long memory regressors, such as the forward premium pose substantial difficulties for predictive regression tests. While the previous literature discussed above has clearly delineated these obstacles, few solutions to this testing problem have been proposed. We contribute to this literature by providing a simple intuitive two-step predictability test in the presence of long memory regressors

[2]Maynard and Phillips (2001) make a very similar argument, except that they focus on the imbalance in (3), whereas we focus on (2). Only an imbalance in (2) is compatible with the FRUH. In other words, since it is impossible to impose the null in their setting, the FRUH cannot be tested in the framework of Maynard and Phillips (2001).

[3]Simulations in Maynard (2006, Table 6) show that the procedure of Campbell and Dufour (1997) has good power against standard regressions but somewhat lower power against the type of rebalanced alternatives considered here. The covariance test of Maynard and Shimotsu (forthcoming) does have power against rebalanced alternatives of this type but has only been developed for use with I(0), I(1), and local-to-unity regressors. Likewise, the tests of Jansson and Moreira (2006) exclude long-memory regressors and do not rebalance.

that remains valid under the null hypothesis and well balanced under the alternative hypothesis.

3 Econometric Methodology

Our two-step procedure is intended to rebalance predictive regressions that may have long memory regressors with dependent variables that are short memory. In this section, we consider the implications of not knowing the true integration order, d, of the regressor. In the first stage, the value of d is estimated, and the regressor, x_t, is fractionally differenced with the estimated value. In the second stage, the regression is run with this fractionally differenced variable. In Section 3.1, we validate our procedure, showing that the estimate of the slope coefficient from the rebalanced regression is consistent. Further, in instances where the null hypothesis of predictability can be re-written as a zero restriction on the slope coefficient, we show that the t-statistic from the rebalanced regression achieves a standard normal asymptotic distribution. In Section 3.3, we discuss the estimation alternatives for d that are available in the first stage.

3.1 Two-Stage Test Procedure

We model y_{t+1} as a linear function of the fractionally differenced predictor x_t, where[4]

$$y_{t+1} = \beta_0 + \beta_1 (1-L)^d x_t + \varepsilon_{1,t+1}. \tag{4}$$

Further, the process x_t is modeled as a type II fractionally integrated process (see Tanaka (1999) and Marinucci and Robinson (1999)):

$$x_t = (1-L)^{-d} \left(u_{2,t} 1_{\{t>0\}}\right) \tag{5}$$

where $1_{\{t>0\}}$ is an indicator function and in which, following Phillips and Solo (1992), $u_{2,t}$ is modeled as a general linear process of the form[5]

$$u_{2,t} = C_2(L)\,\varepsilon_t = \sum_{j=0}^{\infty} C_{2j}\varepsilon_{t-j} \text{ where} \tag{6}$$

$$\varepsilon_t = \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \sim \text{i.i.d}\,(0,\Sigma) \text{ and } \sum_{j=0}^{\infty} j^{\frac{1}{2}} \left\|C_{2j}\right\| < \infty. \tag{7}$$

Note that $C_2(L)$ is a 2-dimensional row vector, and thus for generality we allow u_{2t} to be linearly related to both ε_{1t} and ε_{2t}. An ARFIMA(p,d,q) model for x_t results when $C_2(L) = \left[\begin{array}{cc} 0 & \theta(L)/\phi(L) \end{array}\right]$, where $\theta(L)$ and

[4] Note that β_1 in (4) and b_1 are parameters on two different regressors (one in levels, one in fractional differences) and are thus only equal under the null hypothesis when both are zero.

[5] The moment conditions of Maynard and Phillips (2001, assumption V, p. 682) are stronger than those used here. Their conditions were needed to establish weak convergence to fractional Brownian motion. The limiting behavior of the rebalanced regression, is fundamentally unlike the unbalanced regression studied by Maynard and Phillips (2001) and does not rely on convergence to fractional Brownian motion.

$\phi(L)$ are moving average and autoregressive polynomials in the lag operator L, such that all roots to $\phi(L)$ and $\theta(L) = 0$ lie outside the unit circle.

Our main interest lies in tests of the hypothesis $H_0{:}\beta_1 = 0$. The first stage of our two-step method consists of obtaining a consistent estimate $(\hat{d})$ for (d), with convergence rate T^{α}, for $\alpha > 1/4$. Using $\hat{d}$, we then regress y_{t+1} on the fractional difference of x_t in the second-stage regression:

$$y_{t+1} = \beta_0 + \beta_1 (1-L)^{\hat{d}} x_t + \varepsilon_{1,t+1}. \tag{8}$$

This provides a feasible version of (4) with which to rebalance the relation between y_{t+1} and x_t. The standard t-test is then used to test the hypothesis that $\beta_1 = 0$. The following theorem establishes the large sample properties of our proposed two-step procedure, where, for any discrete random variable x_t, we define $\underline{x}_t = x_t - \bar{x}$.

Theorem 1 *Assuming (4), (5), (6), and (7), where $\hat{d}$ is a T^{α} consistent first-stage estimator of d for $\alpha > 1/4$, the regression coefficient in (8) satisfies:*

$$\sqrt{T}\left(\hat{\beta}_1 - \beta_1\right) - \beta_1 B_T \to_d N\left(0, (var[u_{2,t}])^{-1}\Sigma_{11}\right) \text{ where} \tag{9}$$

$$\begin{aligned} B_T &= \left(T^{-1}\textstyle\sum_{t=1}^{T-1}\hat{\underline{u}}_{2,t}^2\right)^{-1} T^{-1/2}\textstyle\sum_{t=1}^{T-1}\left(\underline{u}_{2t} - \hat{\underline{u}}_{2,t}\right)\hat{\underline{u}}_{2,t} = O_p\left(T^{1/2-\alpha}\right), \\ \underline{u}_{2,s} &= u_{2,s} - T^{-1}\textstyle\sum_{t=1}^{T-1} u_{2,t} \text{ and } \hat{u}_{2,t} = (1-L)^{\hat{d}} x_t. \end{aligned}$$

The theorem shows that $\hat{\beta}_1$ is consistent for β_1 with a convergence rate given by

$$\hat{\beta}_1 - \beta_1 = \begin{cases} O_p\left(T^{-1/2}\right), \ldots if\ \beta_1 = 0 & \text{(standard limiting behavior)} \\ O_p\left(T^{-\alpha}\right)\ , \ldots if \beta_1 \neq 0 & \text{(contamination from first stage estimation)}. \end{cases} \tag{10}$$

In general, the limit distribution in the second stage is contaminated by the estimation error in the first stage, leading to the additional term (B_T) of order $O_p(T^{1/2-\alpha})$. However, this contamination disappears in the special case when $\beta_1 = 0$, and thus no predictive relation exists for any value of d. In this case, the second-stage limit distribution obtained by estimating d in the first stage is the same as the distribution that would be obtained if d were known.

The asymptotic properties of the second-stage t-statistic for a predictability test $(\beta_1 = 0)$ are established in Corollary 3, which shows that the test statistic is standard normal under the null $(\beta_1 = 0)$ and diverges at rate $T^{1/2}$ under the alternative $(\beta_1 \neq 0)$. It thus provides a solid basis for predictability testing with long memory regressors. First, Corollary 2 shows that the residual variance is estimated consistently.

Corollary 2 $\hat{\sigma}^2 = T^{-1}\sum_{t=1}^{T-1}\hat{\varepsilon}_{1t+1}^2 \to_p \Sigma_{11}$, *where* $\hat{\varepsilon}_{1t+1} = y_{t+1} - \hat{\beta}_0 - \hat{\beta}_1\hat{u}_{2,t}$.

Corollary 3 *(a) If the null hypothesis $\beta_1 = 0$ holds, then $t \to_p N(0,1)$. (b) If the alternative $\beta_1 \neq 0$ holds then* $T^{-1/2}t \to_p \Sigma_{11}^{-1/2} var[u_{2,t}]^{1/2}\beta_1$, *with* $t = \hat{\sigma}^{-1}\left(T^{-1}\sum_{t=1}^{T-1}\hat{\underline{u}}_{2,t}^2\right)^{1/2} T^{1/2}\hat{\beta}_1$.

The results of the theorem show that the two-stage procedure results in a consistent estimate of β_1 for all values of this parameter. Further, if the test can be formulated in terms of a zero restriction on β_1, then the asymptotic distribution of the usual t-test is standard normal, and thus achieves the same asymptotic distribution that would have been obtained had d been known. This occurs because the effects of the first stage estimation of d are asymptotically negligible under the null hypothesis, so that in this case the two-step estimator $\hat{\beta}_1$ is asymptotically equivalent to the infeasible estimator that would result from the regression in (4) with known d.

Under the alternative hypothesis when $\beta_1 \neq 0$, the equivalence between the feasible and infeasible estimator no longer holds. The effect of the first-stage is captured by the term $\beta_1 B_T$ in (9) and depends on the specifics of the first-stage estimator for d, particularly its rate of convergence (α). This implies an efficiency and power loss relative to the infeasible estimator, which would be asymptotically equivalent to maximum likelihood when d is known. An alternative feasible procedure would be to estimate both β_1 and d jointly in (4) by maximum likelihood. However, in this case the value of d, would be unidentified under the null hypothesis that $\beta_1 = 0$.

Results matching ours are not directly available in the literature. As noted above, Maynard and Phillips (2001) derive asymptotic properties for a unbalanced regression in which the dependent variable is stationary, while the regressor is a non-stationary fractional process. In that case, there is no rebalancing, and thus no first stage estimation of the differencing parameter, resulting in a limiting distribution that is non-standard. Our results are perhaps closest to those of Dolado *et al.* (2002), who use a two step procedure to estimate the differencing parameter in the context of a fractional Dickey Fuller test for unit roots. For example, for a time series process, x_t, under the null hypothesis that d =1, their procedure regresses Δx_t on $(1-L)^{d_a} x_{t-1}$ and lags of Δx_t, where d_a is the estimated value of d obtained under the alternative. Note that their results are not entirely analogous, as the order of integration of the dependent variable differs from the regressor of interest, $(1-L)^{d_a} x_{t-1}$, under the null. Nonetheless, there results are comparable, since they set up a null hypothesis of a zero restriction on the slope coefficient on $(1-L)^{d_a} x_{t-1}$ using a two step procedure. Further, similar to the asymptotic theory developed here, Dolado *et al.* (2002) show that the t-statistic of the slope coefficient on $(1-L)^{d_a} x_{t-1}$ is standard normal by virtue of the consistency of their estimate of d.

In Section 4, we demonstrate the robust small sample properties of the two-stage estimator advocated here. Before doing so, we provide a generalization to the case in which the dependent variable also has long memory and then briefly discuss the choice of the estimator for d.

3.2 Allowing long-memory in y_t

We next generalize our model to allow y_t and x_t to exhibit different orders of integration under both the null and alternative and to allow fractional integration in y_t. More specifically, we maintain the same model (5) for

x_t but model its relation with y_t as

$$(1-L)^{d_y} y_{t+1} \quad = \quad \begin{cases} \beta_0 + \beta_1 (1-L)^{d_x} x_t + \varepsilon_{1t+1}, & \text{for} \quad t > 0 \\ 0 & \text{for} \quad t \leq 0 \end{cases} \tag{11}$$

The role of the initialization becomes apparent in solving for y_t:

$$y_{t+1} \quad = \quad \beta_1 (1-L)^{d_x - d_y} x_t + (1-L)^{-d_y} \left([\beta_0 + \varepsilon_{1t+1}] \, 1_{\{t>0\}} \right) \tag{12}$$

When $\beta_1 = 0$ and $0 < d_y < 1$ this implies y_t is the sum a type II fractionally integrated process and a non-linear time trend. In fact, using the MA(∞) representation $(1-L)^{-dy} = \sum_{j=0}^{\infty} \psi_j$, where $\psi_j = \Gamma(j + d_y)/\left[\Gamma(d_y)\Gamma(j+1)\right]$ and Γ denotes the Gamma function, we may re-express (12) as

$$y_t \quad = \quad \beta_1 (1-L)^{d_x - d_y} x_t + \beta_0 g(t-1) + \sum_{j=0}^{t-1} \psi_j \varepsilon_{t+1-j}$$

where $g(t-1) = \sum_{j=0}^{t-1} \psi_j$ is a non-linear time trend for $0 < d_y < 1$ and $\sum_{j=0}^{t-1} \psi_j \varepsilon_{t+1-j}$ is a type II fractionally integrated process.

The specification in (11) also suggests a feasible rebalanced regression specification of the form

$$(1-L)^{\hat{d}_y} y_{t+1} = \hat{\beta}_0 + \hat{\beta}_1 (1-L)^{\hat{d}_x} x_t + \hat{\varepsilon}_{1t+1} \tag{13}$$

with the regression coefficient given by

$$\hat{\beta}_1 = \frac{\sum_{t=1}^{T-1} \underline{(1-L)^{\hat{d}_x} x_t} (1-L)^{\hat{d}_y} y_{t+1}}{\sum_{t=1}^{T-1} w \left(\underline{(1-L)^{\hat{d}} x_t} \right)^2} = \frac{\sum_{t=1}^{T-1} \underline{\hat{u}}_{2,t} (1-L)^{\hat{d}_y} y_{t+1}}{\sum_{t=1}^{T-1} \underline{\hat{u}}_{2,t}^2} \tag{14}$$

To implement this, we require first-stage estimates of both fractional parameters, d_x and d_y. We make the same assumptions as before on $\hat{d}_x$ and now let $\hat{d}_y$ denote an α_y consistent estimator for d_y for $1/4 < \alpha_y \leq 1/2$ with limit distribution $G_{y\hat{d}}$:

$$T^{\alpha_y} \left(\hat{d}_y - d_y \right) \to_d G_{y,\hat{d}} \quad \text{for} \quad 1/4 < \alpha_y \leq 1/2 \tag{15}$$

Thus, we extend our earlier two step procedures by estimating the degree of long-memory in both x_t and y_t in the first step and then regressing the fractionally differenced y_t on fractionally differenced x_t.

The properties of such an estimator have not, to our knowledge, been previously investigated. The limit theory under the null hypothesis that $\beta_1 = 0$ is derived in the theorem below. To simplify notation we assume $\beta_0 = 0$.

Theorem 4 ***[Preliminary]*** *Assuming $\beta_0 = \beta_1 = 0$, (5), (6), (7), and (11), where $\hat{d}_x$ is a T^{α_x} consistent first-stage estimator of d_x for $\alpha_x > 1/4$ and $\hat{d}_y$ is a T^{α_y} consistent estimator satisfying (15). Then the limiting* *behavior of regression coefficient in (14) satisfies:*

$$T^{\alpha_y}\left(\hat{\beta}_1-\beta_1\right) \to_d var\left[u_{2,t}\right]^{-1} A\xi_{\delta^y}+1_{\{\alpha_y=1/2\}}\xi_\beta \text{ for } |A|<\infty \text{ where} \tag{16}$$

$$\xi_{\delta^y} \sim G_{y,\hat{d}}, \tag{17}$$

$$\xi_\beta \sim N\left(0,\left(var\left[u_{2,t}\right]\right)^{-1}\Sigma_{11}\right), \text{ and} \tag{18}$$

$$A = lim_{T\to\infty}T^{-1}\sum_{t=1}^{T-1}E\left[u_{2,t}\tilde{\varepsilon}_{1,t+1}\right]=-lim_{T\to\infty}T^{-1}\sum_{t=1}^{T-1}\sum_{j=0}^{t-1}\frac{1}{j+1}C_{2,j}\Sigma_{\cdot,1} \tag{19}$$

When $A=0$ and $\alpha_y=1/2$ then the distribution of $\hat{\beta}_1$ above specializes to $T^{1/2}\left(\hat{\beta}-\beta\right)\to_d N\left(0,\left(\mathrm{var}\left[u_{2,t}\right]\right)^{-1}\Sigma_{11}\right)$, the same distribution found in Theorem 1. This is also the distribution of the infeasible estimator when both d_y and d_x are known. In this special case, critical values may again be calculated in the usual way, ignoring the effect of the first stage estimation of d_x and d_y.

Unfortunately, this equivalence does not hold more generally. When $A\neq 0$ and a root-T estimator is employed for d_y (i.e. $\alpha_y=1/2$) the first-stage estimation of d_y complicates the limit distribution via the term $A\xi_{\delta^y}$. If $\hat{d}_y$ is asymptotically unbiased then this term will have mean-zero, but will contribute to (or detract from) the variance of the estimator, invalidating the normal standard errors which do not account for it. The exact distribution of $A\xi_{\delta^y}$ and its covariance with ξ_β depend on the estimator employed for $\hat{d}_y$. Depending on this covariance, the standard errors may either overestimate or underestimate the variance in $\hat{\beta}_1$.

When $A\neq 0$ and $\alpha_y<1/2$ the distribution of $\hat{\beta}_1$ is dominated by its first term, $T^{\alpha_y}\left(\hat{\beta}_1-\beta_1\right)\to_d A\xi_{\delta^y}$. This implies a reduced rate of convergence for the estimator, due to the additional noise resulting from the estimation of d_y. Since the normal standard errors are based on the assumption of $\sqrt{T}$ convergence to the second term ξ_β, they fail to account for this reduced convergence rate and are likely to vastly under-estimate the variance of estimator. In a testing context this suggests large size distortion.

As the above discussion makes clear, the critical values for the distribution in (16) depends on the method of estimation for d_y. The vast majority of proposed estimators satisfy:

$$T^{\alpha}\left(\hat{d}_y-d_y\right)\to_d \xi_{\delta^y}\sim N\left(0,V_{\delta,\delta}\right) \quad \text{for} \quad 1/4<\alpha_y\leq 1/2 \tag{20}$$

and have available a consistent variance estimator $\hat{V}_{\delta,\delta}$ for $V_{\delta,\delta}$ satisfying $\hat{V}_{\delta,\delta}\to_p V_{\delta,\delta}$.

On a case by case basis, it may be further established for many estimators that ξ_{δ^y} and ξ_β are jointly asymptotically normally distributed with covariance $V_{\delta,\beta}$, where the form of $V_{\delta,\beta}$ depends on the estimator employed for d_y.[6] In this case, setting $V_{\beta,\beta}=\mathrm{var}\left[u_{2,t}\right]^{-1}\Sigma_{11}$ we have:

$$\begin{pmatrix}\xi_{\delta^y}\\ \xi_\beta\end{pmatrix}\sim N\left(0,V\right) \quad \text{for} \quad V=\begin{bmatrix}V_{\delta,\delta} & V_{\delta,\beta}\\ V_{\delta,\beta} & V_{\beta,\beta}\end{bmatrix}$$

and the distribution in (16) simplifies to

$$T^{\alpha_y}\left(\hat{\beta}_1-\beta_1\right)\to_d N\left(0,q'Vq\right) \quad \text{for} \quad q'=\begin{bmatrix}\mathrm{var}\left[u_{2,t}\right]^{-1}A, & 1_{\{\alpha_y=1/2\}}\end{bmatrix}.$$

[6]Note that the separate normality assumptions in (18) and (20) are necessary, but not sufficient, to establish joint normality.

Then, employing a consistent estimators $\hat{q}$ for q and $\hat{V}$ for V we can define t= $T^{\alpha_y}\hat{\beta}_1/(\hat{q}'\hat{V}\hat{q})^{1/2}$ and, under the null hypothesis that $\beta_1 = 0$, we have $t \rightarrow_d N(0,1)$.

This approximation should work reasonably for $\alpha_y = 1/2$. When α_y is slightly less than one-half, it ignores the second term in (16), even though this term is only of slightly lower order. A finite sample adjustment that reflects this lower order term is given by $\tilde{t} = T^{\alpha_y}\hat{\beta}_1/(\tilde{q}_T'\hat{V}\tilde{q}_T)^{1/2}$, where $\tilde{q}_T' = (\ [\widehat{\text{var}}(u_{2,t}]^{-1}\hat{A}, \quad T^{\alpha_y - 1/2}\ )$. Since $T^{\alpha_y-1/2} = 1$ for $\alpha_y = 1/2$ and converges to zero for $\alpha_y < 1/2$ it is apparent that $\tilde{t} \rightarrow_d N(0,1)$ under the null hypothesis $H_0 : \beta_1 = 0$.

3.3 First-Stage Estimation of d

To utilize our two-step procedure a consistent estimate of d must be obtained for a long memory model, which for our analysis is the ARFIMA model. Fortunately, a plethora of techniques exist for the first-stage estimation of d, which range from parametric MLE, both in the time domain and frequency domain (e.g. Sowell (1992) and Fox and Taqqu (1986)) to semi-parametric, and wavelet based estimators. To highlight the feasibility of our approach, we use two estimators, one being a parametric estimator in the time domain (the constrained sum of squares, CSS, estimator) and one being a semi-parametric estimator in the frequency domain (the bias reduced log periodogram regression, BRLPR, based estimator).

In the time domain, the CSS estimator has become popular because of its relative simplicity and robustness to non-stationarity. For a sample of size T, the CSS estimates are the set of parameters that maximize the approximate maximum likelihood function, ψ, which is given by

$$\begin{aligned} &\psi(\mu, \phi', \theta', d, \sigma^2) = -\tfrac{T}{2}\log(2\pi) - \tfrac{T}{2}\log\sigma^2 - \tfrac{1}{2\sigma^2}\textstyle\sum_{t=2}^{T} a_t^2, \\ &a_t = \tfrac{\phi(L)}{\theta(L)}(1-L)^d(x_t - \mu), \end{aligned} \tag{21}$$

where a_t is a martingale difference sequence and $\phi(L)$ and $\theta(L)$ are autoregressive and moving average polynomials with all roots to $\phi(L) = 0$ and $\theta(L) = 0$ lying outside the unit circle.[7] It is necessary to initialize pre-sample values, which is usually accomplished by setting them equal to 0. The properties of the CSS estimator have been established by Beran (1995), who shows that the estimator of d converges at rate $T^{1/2}$ and is asymptotically normal for $d >$ -1/2. Although, in the current context, any of the time domain based estimators will likely work well, we choose to utilize the CSS estimator given its relative simplicity and robustness to non-stationary processes.

While the CSS estimator has good small sample properties when the ARFIMA model is correctly specified (Chung and Baillie (1993), and Nielsen and Frederiksen (2005)), it is well known that the estimator is inconsistent when the number of autoregressive and/or moving average parameters are incorrectly chosen (Robinson 1995). A number of semi-parametric estimators that avoid the concerns of misspecification have

[7]The CSS estimator also has the advantage that it can be modified to accommodate other distributions such as the t-distribution (see Baillie *et al.* (1996)). Further, heteroskedastic effects can also be considered by replacing σ^2 in (21) with σ_t^2.

been developed. These estimators include log periodogram regression (LPR) based estimators (Geweke and Porter-Hudak 1983, Robinson 1995, Andrews and Guggenberger 2003), and Whittle type estimators, including the local Whittle estimator (Robinson 1995), and the exact Whittle estimator (Phillips 1999, Shimotsu and Phillips 2005). The exact Whittle estimator is an attractive alternative, as (Shimotsu and Phillips 2005) show that it is asymptotically normal for any value of d.

The appeal of the LPR based estimators lie in their incredible simplicity. These estimators are based on the properties of the log of the spectral density function of a long memory fractional process (including an ARFIMA process), which satisfies,

$$\log f(\omega) \sim \log[g(\omega)] - 2d\log(\omega) \tag{22}$$

where "$\sim$ " denotes asymptotic equivalence as $\omega \to 0$, and $g\ (\omega\ )$ is an even function that is continuous at zero and finite. The original LPR based estimator of Geweke and Porter-Hudak (1983) replaces $\log[g(\omega)]$ with a constant and regresses the log periodogram at the first m frequencies on a constant and $-2log(\omega_j)$, $j = 1, \ldots, m$. Here, m is the user selected bandwidth, and ω_j denotes the Fourier frequencies given by, $\omega_j = 2\pi j/T$, $j = 1, \ldots, m$. The approximation of $\log[g(\omega)]$ with a constant may not be innocuous and can in fact lead to a sizeable small sample bias as shown by Agiakloglou *et al.* (1993). Andrews and Guggenberger (2003) have suggested that a decrease in the small sample bias can be obtained by approximating the term first term in (22) with a constant and the polynomial $\sum_{r=0}^{R} \omega_j^{2r}, j = 1, ..., m$. Recently, Nielsen and Frederiksen (2005) have shown that the use of the biased reduced LPR (BRLPR) estimator of Andrews and Guggenberger (2003) does substantially mitigate the small sample bias relative to other frequency based estimators. Based on this bias reduction, coupled with its simplicity, we chose to utilize the BRLPR estimator in our analysis below. Following Nielsen and Frederiksen (2005) we also set R =1 throughout.[8]

4 Monte Carlo Evidence

The simulation experiments in this section serve several purposes. First, we wish to demonstrate the potential pitfalls that exist when long memory regressors are used in predictive regressions. To this end, we allow the regressors to follow long memory ARFIMA(0,d,0) and ARFIMA(1,d,0) processes, while allowing the dependent variables to be white noise. We show that without rebalancing, estimates of the slope coefficient are substantially biased with t-statistics that are large in absolute value yielding an empirical test that is oversized. Second, we wish to evaluate the effectiveness of our proposed solution to this problem, and thus we report extensive simulation results based on our two-step estimation procedure using both a time domain and frequency based

[8]When $d \in (-1/2, 1/2)$, Andrews and Guggenberger (2003) show that their estimator is consistent and asymptotically normal. Although it appears likely that it remains consistent for $d < 1$, given the potential for non-stationarity, we follow others using LPR based estimators (e.g. Sun and Phillips (2003)), and apply the linear filter $(1 - L)^{0.50}$ prior to using the BRLPR estimation technique. The final estimate of d results by adding 0.50 to this value.

estimator to filter the long memory regressor prior to running the predictive regressions. We offer further evidence of the robustness of our approach by demonstrating the applicability of our procedure using the CSS estimator when the model is misspecified. Finally, we close with a brief power experiment to further highlight the validity of our two-step procedure.

Our simulations are based on the following model

$$\begin{aligned} y_{t+1} &= \beta_0 + \beta_1(1-L)^d x_t + \varepsilon_{1t+1} \\ x_t &= (1-L)^{-d} u_{2t},\ u_{2t} = (1-\phi L)^{-1}\varepsilon_{2t}\ , \\ \varepsilon_t &= \begin{pmatrix} \varepsilon_{1t}, & \varepsilon_{2t} \end{pmatrix}' \sim i.i.d.\ N(0,\Sigma) \end{aligned} \tag{23}$$

where $|\phi| < 1$, and Σ is a positive definite matrix with potentially non-zero off diagonal elements. To generate data, we first draw the residual vector ε_t, whose elements are correlated with correlation coefficient equal to $\rho = \Sigma_{12}/\sqrt{\Sigma_{11}\Sigma_{22}}$. To be consistent with our theoretical construct, x_t is created as a type II fractional process using the recursive structure of the operator (1-L)$^{-d}$. For example, following Tanaka (1999), we have

$$(1-L)^{-d}u_{2t} = \sum_{k=0}^{t-1} \psi_k u_{2t-k}, \tag{24}$$

where ψ_0 =1, ψ_k =($k + d$ -1)ψ_{k-1} /k with $k \geq 1$. Typically, β_1 is set equal to zero, although in our discussion of the power of our test, we allow β_1 to take on values between 0 and 1. Based on the empirical example in our paper, we chose a sample size of 350 and perform 3000 simulations. Finally, we allow the correlation coefficient across the residuals to vary from -0.95 to 0.95.[9]

Tables 1-2 motivate the problem by demonstrating the size distortion that results when long memory regressors are included in predictive regressions. The tables show simulation results under the null hypothesis (b_1 =0) for a standard predictability regression, where our two-step procedure is not applied. In our empirical application, this would correspond to the traditional tests of the FRUH when the forward premium displays long memory. Our objective here is to observe the consequences of not adequately accounting for long memory.

Table 1 contains our results when the regressor follows an ARFIMA(0,d,0) specification. The values in the first column of Table 1 give the integration order (d) of the regressor, while the correlation coefficients between the simulated residuals (ρ) are reported across the top of the table. Table 1a shows rejection rates under the null hypothesis for a predictability test when the regressor is I(d). The test becomes oversized in the presence of residual correlation when the value of d exceeds 0.50. The size distortion increases with both the absolute value of the correlation coefficient and the persistence of the regressor, with rejection rates as high as 29% in a nominal 5% test. These results are similar to those of Mankiw and Shapiro (1986), who analyze size distortion in predictive regressions with near unit root regressors. Tables 1b and 1c contain the simulated biases and

[9] We considered a couple of cases for the standard deviations of ε_{1t} and $\varepsilon_{2,t}$, including setting them equal to unity. Many predictive regressions, including the FRUH regressions, are characterized by a volatile dependent variable relative to the regressor, and thus we also allowed the standard deviations to differ based on our empirical example below. To conserve space, we report only the results based on the case where the standard deviations differ. See the notes to Table 1 for additional details.

variances of $\hat{b}_1$. The estimator is negatively (positively) biased when the correlation between the residuals is positive (negative) and this bias can be substantial. Finally, it is interesting to note that the variance of $\hat{b}_1$ declines as the regressor becomes more persistent, which is expected given the non-stationarity, and thus divergent variances, for most of the processes considered.

We next consider the effects of adding short run dynamics to the system in Table 2, where we allow x_t to follow various ARFIMA(1,0.80,0) specifications. The results are similar to those in Table 1, although it is interesting to note that the inclusion of short term dynamics can influence the tests. For example, rejection rates increase as the process becomes more persistent through the stable short run components. Note, rejection rates of the true null of no predictability can be as high as 26% with a nominal 5% test, whereas the highest rejection rate we encounter in Table 1 for an ARFIMA(0,0.80,0) process is 19%. This implies that short memory dynamics exacerbate the rejection rates associated with regression imbalance as documented in Table 1. Finally, we reach the same conclusion in Table 2 as we did in Table 1 regarding the mean bias and variance of $\hat{b}_1$.

The results of Tables 1-2 demonstrate the potential pitfall of using long memory regressors in predictive regressions. Tables 3-7 demonstrate the applicability of our suggested two-step approach. We consider both time and frequency domain estimators, while allowing for the possibility of misspecification using our time domain estimator. Tables 3-4 contain our results using the correctly specified CSS estimator in the first-step estimation of d, when x_t follows both a fractional noise process (Table 3) and an ARFIMA (1,0.80,0) process (Table 4).[10] The results for our two-step procedure are quite promising and contrast quite nicely to those found above when long memory regressors are employed in standard regressions without differencing. In each case, the empirical size of the test is approximately equal to the nominal size, with only one exception. In Table 4, when x_t is an ARFIMA(1,0.80,0) process with $\phi = 0.99$, we see that $\beta = 0$ is rejected too frequently when the residual correlation differs from 0. This is to be expected as the correctly differenced process is a near unit root variable and is thus persistent even though it is not long memory. The results of Table 3b also indicate that the bias in the first table is dramatically reduced by rebalancing. There are still some cases in Table 4 where the mean estimate of β_1 is not centered precisely at 0. Nonetheless, the resulting biases are usually smaller than those reported in Table 2.

Table 5 presents our results using the semi-parametric BRLPR estimator for first-stage calculation of d.[11] Here, we only analyze the case where x_t follows an ARFIMA(1,d,0) process to conserve space. Throughout, we allow the value of ϕ to vary from -0.99 to 0.99, but fix d to be equal to 0.80. The last panel of the table documents the exceptional performance of the BRLPR estimator. The bias here is consistent with previous

[10] Starting with Table 4, we present the bias in estimating d with each of the estimators we employ. The CSS estimator of d is remarkably accurate when there are no ARMA components, and thus, for brevity, we omit the bias of the estimated value of d from Table 3. These results are available upon request.

[11] As discussed above, a bandwidth parameter must be selected. We considered several bandwidths, ranging from m=$T^{0.55}$ to $m = T^{0.85}$. Here, for brevity, we report the results for $m = T^{0.75}$, which is the same bandwidth Maynard and Phillips (2001) use with their modified LPR results. Results are unchanged with other values of m.

results associated with this estimator including Nielsen and Frederiksen (2005) and Andrews and Guggenberger (2003). Unless a very large and positive autoregressive parameter is present, the estimated value of d is very near the true value. When strong autoregressive dynamics are present, the spectral density function near the origin is contaminated with both short and long memory components. The result is a substantial positive bias in the differencing parameter, which interestingly wipes out all of the activity near the origin, resulting in a correctly sized second-stage t-test. In other words, when $\phi \geq 0.80$, d is over-estimated resulting in x_t being slightly over-differenced. The result of this over-differencing is a mitigation of the over-all persistence of the process due to both autoregressive and long memory components, and thus a correctly sized second-stage t-test for all values of ϕ. Finally, there is a substantial bias reduction relative to the results in Table 2.

Tables 6 and 7 demonstrate that our second-stage test performs well even when the model in the first stage is misspecified or over-parameterized. Table 6 considers the case where x_t follows an ARFIMA(1,d,0) process but an ARFIMA(0,d,0) process is estimated using the CSS estimator. We fix $d = 0.80$ and allow ϕ to vary from -0.99 to 0.99. The value of d is not estimated well under the misspecification, a fact familiar to practitioners using parametric long memory estimators. For large negative values of ϕ, a substantial negative bias results for d, while d is dramatically over-estimated for large positive values of ϕ. In this case, d is burdened with the role of accounting for both short and long memory components. Nonetheless, the second-stage test has the correct size throughout, and the mean estimate of β_1 is very near 0. Table 7 considers the opposite scenario, in which the true process is fractional noise, but an ARFIMA(1,d,0) model is estimated. It is clear, from the last panel of the table, that d is frequently underestimated, as the algorithm will routinely select large autoregressive parameters rather than the correct value of d. However, the bias is reasonable, resulting in an accurate second-stage test, with an empirical size of about 5%.

As discussed above, the Monte Carlo results were obtained here based on the truncated type II fractional process to remain consistent with our theoretical results. We also considered results that are available upon request based on a type I fractional process, where the truncation is not applied, but where the data are generated using the autocovariances of the process. The conclusions are consistently the same as above, where empirical sizes are large without rebalancing with excessive biases, while our rebalancing procedure corrects the distortion and mitigates the biases. It should be noted, however, that the size distortion without rebalancing and the associated biases tend to be even larger when the data are generated without the truncation assumption.

We close this section by commenting on power. Under the null that $\beta_1 = 0$, moderately imprecise estimation of d does not result in a large size distortion. This does not suggest, however, that over-differencing is appropriate. Indeed, our approach does not force the researcher to take any a-priori stand on the order of integration of the regressor, be it I(0), I(d), I(1) or even I(2), for example. In addition, our procedure yields a consistent second-stage estimator for any value of β_1, an important property for test power. To highlight the performance of our two-step procedure under the alternative, we ran a brief power study. Based on equation (23), we allowed the true value of β_1 to vary from 0 to 1 with a step size of 0.10, and tested the hypothesis that $\beta_1 = 0$. For brevity, we chose a value of $\rho = 0.80$ and set the standard deviations of both disturbance sequences equal to unity. We allowed the regressor, x_t, to be a fractional noise process and also allowed d to vary from 0 to 1, with a step size of 0.20. The results clearly show that substantial power loss will generally occur unless the two-step procedure is used relative to cases in which no differencing is employed or over-differencing is utilized.[12] As an example, consider Figure 4, which depicts the power related to the use of our two-step procedure, application of a simple first difference, and the use of no differencing when x_t is a fractional variable with d =0.40 and y_{t+1} is related to x_t with the value of β_1 ranging from 0 to 0.30 depicted along the x-axis. As above, the sample size is set equal to 350, and we employ 3000 simulations. When $\beta_1 = 0$, the statistic displayed corresponds to the size of the test. The power is always greatest for our two-step procedure. Substantial power loss occurs for the case when nothing is done to rebalance the equation, even though the processes considered here are stationary. Over-differencing results in higher power relative to no differencing, but is clearly dominated by the application of our two-step procedure for all values of β_1.[13]

The results of our simulation section show that care must be taken in regressions involving short memory dependent variables and long memory regressors. In particular, the t-statistics are too large in absolute value and can result in substantial over-rejection. Our simulation results indicate that our two-step procedure results in a rebalanced regression whose t-statistic has the correct size. It is also robust, both with respect to the selected estimator and the potential for misspecification. Further, substantial power gains result when d is first estimated relative to the cases in which no differencing occurs or a simple first difference is used. We now apply our two-step procedure in the context of the FRUH.

5 Application to the FRUH

As discussed above, the FRUH is typically tested by the regression depicted in equation (3), where the change in the spot rate is regressed on the forward premium. Constructing a test based on equation (3) that accounts for the long memory behavior of the forward premium is difficult. In particular, in its present form, if the change in the spot rate is I(0), the finding of a non-stationary long memory forward premium implies an automatic rejection of the FRUH. A more natural way to test the FRUH, while allowing for long memory in the forward premium, is based on the matching regression depicted in (2). In particular, we base our test on the following regression:

$$s_{t+1} - f_t = \beta_0 + \beta_1 (1-L)^d (f_t - s_t) + \varepsilon_{1t+1}. \tag{25}$$

[12] Extremely small power gains were detected when the true value of d was 0 and unity with no differencing and the application of simple first difference, respectively, relative to our two-step procedure.

[13] The remaining power results are available upon request. To summarize, for $d < 0.40$, the power gain from differencing with the estimated d is even greater relative to the case where a simple first difference is used but decreases relative to the case with no differencing. The opposite occurs as the value of d rises, with the power generally remaining highest for the case in which our two-step procedure is employed.

If excess returns are I(0), as both intuition and empirical evidence suggest, then the regression in (25) contains components that are all integrated of the same order. The test for unbiasedness is then given by a simple t-test of the hypothesis $\beta_1 = 0$.

We consider exchange rate data for Canada, France, Germany, Japan, and the UK vis-à-vis the US from July 1973 to March 2000. The data are obtained from Data Resources International. We use the one month forward and spot US dollar price of the foreign currency, where the data are recorded on the last day of each month. See Liu and Maynard (2005) for precise details.[14] As a benchmark, Table 8 yields regression results for the standard FRUH equations shown in (2) and (3). We also report the probability values associated with the Ljung-Box Q statistics for the residuals and squared residuals. Under unbiasedness, we expect b_1 =0 and b_2 =1. Using the probability values for these hypotheses, we encounter a strong rejection of the unbiasedness hypothesis. In *every* case the estimated coefficient is negative, and when the change in the spot rate appears as the dependent variable, we reject the hypothesis of a unity slope coefficient at the 1% level for 3 of the 5 countries, while we are able to reject this hypothesis at the 5% level for every country in our sample. Precisely the same finding regarding unbiasedness emerges when we use excess returns. The results from the Ljung-Box Q-statistics show that the returns are free of serial correlation and generally homoscedastic. In particular, the squared residuals from the excess returns equation for the UK indicate the likely presence of heteroskedastic effects, while there is some evidence of higher order GARCH effects for Japan. The remaining countries do not appear to exhibit substantial volatility clustering.

The standard test results presented above cannot be fully relied on given the long-memory characteristics of the regressor, which can give rise to problems with both size and power. In principle, they could either exaggerate evidence against FRUH due to size distortion or alternatively they could understate the rejection on account of power loss arising from an imbalanced alternative. By employing our two-stage rebalancing procedure we can provide a more reliable test of FRUH. Table 9 presents our results using the CSS estimator. It is interesting to note that our findings are very much in line with previous research in that we find significant evidence of long memory dynamics in the forward premium. Using the numerical standard errors as our guide, we are able to reject the hypothesis that d is either 0 or 1 at the 5% level for every country in our sample, except Germany, where we fail to reject a unit root in the forward premium. After filtering the forward premium using the estimated value of d in the first stage, we run the regression associated with equation (25). First we note

[14] While daily data may contain more information than monthly data, its use would complicate the analysis in several respects. First there is no hard rule for the exact number of business days in each month. Yet returns must still be calculated on a monthly basis when using forward rates with a one month maturity. Secondly, sampling the monthly returns on a daily basis induces a large moving average process in the residuals. Under the traditional assumption that the horizon length is fixed or small relative to the sample size this can be handled via the use of robust standard errors. However, (Richardson and Stock 1989), and more recently (Valkanov 2003), show that the normal asymptotic distribution based on the assumption of fixed horizon lengths can provide poor approximations to finite sample behavior.

that there is one case where the sign switches from negative to positive (for Japan). Secondly, the probability values associated with the hypothesis of unbiasedness always exceed the same values in Table 8. We continue to reject the hypothesis that $\beta_1 = 0$ at the 1% level in three cases (Canada, France, and the UK). Now, however, we fail to reject the hypothesis at the 10% level for Germany and Japan. Thus, rebalancing makes a difference for two countries in our sample, and we conclude that when our two-step procedure is implemented, less evidence against unbiasedness is uncovered. As a robustness check, we also utilized the BRLPR estimator with several bandwidths. The results are quantitatively identical to those reported here. Further, the estimates of d are in ranges consistent with those reported in Table 9.

Finally, in terms of serial correlation and heteroskedasticity in the residuals, we reach the same conclusions as above where there is no evidence of serial correlation for any country, strong evidence of heteroskedasticity for the UK and mild evidence of potentially higher order GARCH effects for Japan. Interestingly, Baillie and Bollerslev (2000) analyze the effects of long memory in the conditional variance of spot returns in daily data. Using monthly data, we encounter no evidence supporting potential long memory in the conditional variance of the residuals of the excess returns series. However, we did consider heteroskedastic effects by estimating a GARCH model based on equation (25) with the inclusion of a GARCH-in-mean term. In no case, were the resulted altered. The inclusion of long memory GARCH effects in daily data possibly presents an interesting extension to our analysis.

6 Summary and Conclusion

A substantial literature exists on predictive regressions with near unit root regressors, but far less attention has been paid to a second empirically relevant case in which predictive regressors display long memory behavior. In both cases, size distortion can be problematic. However, the remedies employed in the context of near unit roots do not necessarily carry over to the long memory case. Moreover, while problems of regression imbalance are arguably of concern in the near unit root case (Maynard and Shimotsu forthcoming), they become unavoidable when regressors are fractionally integrated, particularly if returns are stationary, but the predictive regressors are integrated of order $d > 0.5$, as in tests of the forward rate unbiasedness hypothesis (FRUH).

In this paper we have suggested a two-stage predictive regression test in which the dependent variable is stationary, but which allows for, without imposing, long memory behavior in the predictor. The first stage involves obtaining a consistent estimate of the long memory parameter. Then in the second stage, the predictive regression is rebalanced by fractionally differencing the regressor with the estimated value of d from the first stage. A full set of asymptotic results are provided. The t-statistic in the second-stage predictability test has a standard normal limiting distribution. Likewise, extensive simulations suggest that the two-step procedure works remarkably well in practice. It has good size, is highly robust to estimation error in the first stage, and can yield improved power over cases in which either no differencing or over-differencing is employed. As an empirical application, we consider the puzzle affiliated with the forward rate unbiasedness hypothesis. We find that the forward premium is typically subject to long memory, while the standard regressands in the forward rate unbiasedness hypothesis regressions appear to be I(0), making it well suited to our two-stage test. We reverse a strong rejection of unbiasedness for two of the five currencies in our sample, while reconfirming the validity of previous rejections in the other three cases.

A Appendix

1.A Lemmas

Lemma 5 . *Define* $\bar{\delta} > 0$, *and let* $|\delta_T^*| < \bar{\delta}$. *Defining* $\tilde{u}_{2,t} = ln(1-L)u_{2,t}1_{\{t>0\}}$, $\tilde{\tilde{u}}_{2,t} = ln(l-L)\tilde{u}_{2,t}$, *and* $u^*_{2,t,T} = (1-L)^{\delta_T^*}\tilde{\tilde{u}}_{2,t}$, *we have*

$$max_{t\leq T} E\tilde{u}^2_{2,t} < \|\Sigma\| \left(\sum_{k=1}^{\infty} \|c_{2k}\|\right)^2 \left(\sum_{v=1}^{\infty} \frac{1}{v^2}\right) < \infty, \quad (A.1)$$

$$max_{t\leq T} E\tilde{\tilde{u}}^2_{2,t} = O(ln(T)^2), \quad (A.2)$$

$$max_{t\leq T} E(u^*_{2,t,T})^2 = O\left(\left(ln(T)T^{\bar{\delta}}\right)^2\right). \quad (A.3)$$

Lemma 6 *Using the same definitions in the statement of Theorem 1 and Lemma 5, the following convergence rates apply*

$$\textbf{\textit{a)}}\ T^{-1/2}\sum_{t=1}^{T-1} \underline{\tilde{u}}_{2,t}\varepsilon_{1,t+1} = O_p(1) \quad (A.4)$$

$$\textbf{\textit{b)}}T^{-1}\sum_{t=1}^{T-1} \underline{\tilde{u}}_{2,t}\underline{u}_{2,t} = O_p(1) \quad (A.5)$$

$$\textbf{\textit{c)}}T^{-1}\sum_{t=1}^{T-1} \underline{\tilde{u}}^2_{2,t} = O_p(1), \quad (A.6)$$

$$\textbf{\textit{d)}}\ T^{-1}\sum_{t=1}^{T-1} \underline{u}^*_{2,t,T}\varepsilon_{1,t+1} = O_p\left(ln(T)T^{\bar{\delta}}\right) \quad (A.7)$$

$$\textbf{\textit{e)}}\ T^{-1}\sum_{t=1}^{T-1} \underline{u}^*_{2,t,T}\underline{u}_{2,t} = O_p\left(ln(T)T^{\bar{\delta}}\right) \quad (A.8)$$

$$\textbf{\textit{f)}}\ T^{-1}\sum_{t=1}^{T-1} \underline{u}^*_{2,t,T}\underline{\tilde{u}}_{2,t} = O_p\left(ln(T)T^{\bar{\delta}}\right) \quad (A.9)$$

$$\textbf{\textit{g)}}\ T^{-1}\sum_{t=1}^{T-1} \left(\underline{u}^*_{2,t,T}\right)^2 = O_p\left(ln(T)^2T^{2\bar{\delta}}\right). \quad (A.10)$$

Lemma 7 *Define* $\bar{\delta}^y > 0$, *and let* $|\delta_T^{y*}| < \bar{\delta}^y$.

Defining $\tilde{\varepsilon}_{1,t+1} = ln(1-L)\left(\varepsilon_{1,t+1}1_{\{t>0\}}\right)$, $\tilde{\tilde{\varepsilon}}_{1,t+1} = ln(l-L)\tilde{\varepsilon}_{1,t+1}$, *and* $\varepsilon^*_{1,t+1,T} = (1-L)^{\hat{\delta}^*_T}\,\tilde{\tilde{\varepsilon}}_{1,t+1}$,

$$max_{t\le T}E\tilde{\varepsilon}^2_{1,t+1} \quad < \quad \Sigma_{1,1}\sum_{v=1}^{\infty}\frac{1}{v^2} < \infty, \tag{A.11}$$

$$max_{t\le T}E\tilde{\tilde{\varepsilon}}^2_{1,t+1} \quad = \quad O\left(ln(T)^2\right), \qquad and \tag{A.12}$$

$$max_{t\le T}E\left(\varepsilon^*_{1,t+1,T}\right)^2 \quad = \quad O\left(\left(ln(T)T^{\bar{\delta}^y}\right)^2\right). \tag{A.13}$$

The proof of Lemma 7 is omitted because it follows very closely the proof of Lemma 5.

Lemma 8 *The following convergence rates apply, where A is given by (19),*

$$\textbf{\textit{a)}}\ T^{-1}\sum_{t=1}^{T-1}\underline{u}_{2,t}\tilde{\varepsilon}_{1,t+1} \quad \to_p \quad A \tag{A.14}$$

$$\textbf{\textit{b)}}\ T^{-1}\sum_{t=1}^{T-1}\tilde{\varepsilon}_{1,t+1}\tilde{u}_{2,t} \quad = \quad O_p(1), \tag{A.15}$$

$$\textbf{\textit{c)}}\ T^{-1}\sum_{t=1}^{T-1}\tilde{\varepsilon}_{1,t+1}\underline{u}^*_{2,t,T} \quad = \quad O_p\left(ln(T)T^{\bar{\delta}^x}\right), \tag{A.16}$$

$$\textbf{\textit{d)}}\ T^{-1}\sum_{t=1}^{T-1}\varepsilon^*_{1,t+1,T}\underline{u}_{2,t} \quad = \quad O_p\left(ln(T)T^{\bar{\delta}^y}\right) \tag{A.17}$$

$$\textbf{\textit{e)}}\ T^{-1}\sum_{t=1}^{T-1}\varepsilon^*_{1,t+1,T}\underline{\tilde{u}}_{2,t} \quad = \quad O_p\left(ln(T)T^{\bar{\delta}^y}\right) \tag{A.18}$$

$$\textbf{\textit{f)}}\ T^{-1}\sum_{t=1}^{T-1}\varepsilon^*_{1,t+1,T}\underline{u}^*_{2,t,T} \quad = \quad O_p\left(ln(T)^2T^{\left(\bar{\delta}^y+\bar{\delta}^x\right)}\right) \tag{A.19}$$

1.B Proofs

Proof of Lemma 5

(A.1) follows by (7) and the series expansion $\ln(x) = \sum_{j=1}^{\infty}(-1)^{j-1}\frac{(x-1)^j}{j}$:

$$\begin{aligned}
\tilde{u}_{2,t} &= \ln(1-L)u_{2,t}1_{\{t>0\}} = -\sum_{j=1}^{t-1}\frac{1}{j}L^j u_{2,t} = -\sum_{k=0}^{\infty}C_{2k}\sum_{r=k+1}^{k+t-1}\left(\frac{1}{r-k}\right)\varepsilon_{t-r},\\
\max_{t\le T}E\tilde{u}^2_{2,t} &= \max_{t\le T}\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\sum_{r=k+1}^{k+t-1}\sum_{s=j+1}^{j+t-1}\left(\frac{1}{r-k}\right)\left(\frac{1}{s-j}\right)C_{2j}E[\varepsilon_{t-r}\varepsilon'_{t-s}]C'_{2k}\\
&\le \|\Sigma\|\left(\sum_{k=0}^{\infty}\|C_{2k}\|\right)^2\left(\sum_{v=1}^{\infty}\left(\frac{1}{v}\right)^2\right) < \infty.
\end{aligned}$$

Next, since $\tilde{u}_{2,t} = 0$, $t \le 0$ and $\sum_{j=1}^{T-1}\frac{1}{j} = O\left(\ln(T)\right)$ (Gradstein and Ryzhik 1994, eqn. 0.131),

$$\max_{t\le T}\tilde{\tilde{u}}^2_{2,t} \quad = \quad \max_{t\le T}\sum_{j=1}^{t-1}\sum_{k=1}^{t-1}\frac{1}{k}\frac{1}{j}E\left|\tilde{u}_{2,t-j}\tilde{u}_{2,t-k}\right| \le \max_{t\le T}E|\tilde{u}_{2,t}|^2\left(\sum_{j=1}^{T-1}\frac{1}{j}\right)^2 = O(\ln(T)^2),$$

showing (A.2). Let $\psi_{\delta^*_{T,j}}$ and $\psi_{\bar{\delta}_j}$, j =0,1,2,... denote the Maclaurin coefficients in the expansion of $(1-L)^{\delta^*_T}$ and $(1-L)^{\bar{\delta}_T}$, respectively. Noting that $\tilde{u}_{2,t}=0$, $t\leq 0$, $|\psi_{\delta^*_{T,j}}|\leq|\psi_{\bar{\delta}_j}|$, where $\psi_{\bar{\delta}_j}$ is non-random, and $u^*_{2,t,T}=\sum_{j=0}^{t-1}\psi_{\delta^*_{T,j}}\tilde{\tilde{u}}_{2t-j}$, (A.3) then follows since

$$\begin{aligned}\max_{t\leq T}E(u^*_{2,t,T})^2 &= \max_{t\leq T}E\left|\sum_{j=0}^{t-1}\sum_{k=0}^{t-1}\psi_{\delta^*_{T,j}}\tilde{\tilde{u}}_{2,t-j}\psi_{\delta^*_{T,k}}\tilde{\tilde{u}}_{2,t-k}\right|\leq\max_{t\leq T}\sum_{j=0}^{t-1}\left|\psi_{\delta^*_{T,j}}\right|\sum_{k=0}^{t-1}\left|\psi_{\delta^*_{T,k}}\right|E\left|\tilde{\tilde{u}}_{2,t-j}\tilde{\tilde{u}}_{2,t-k}\right| \\ &\leq E|\tilde{\tilde{u}}_{2,t}|^2\left(\sum_{j=0}^{T-1}|\psi_{\bar{\delta},j}|\right)^2=O\left(\left(\ln(T)T^{\bar{\delta}}\right)^2\right),\end{aligned}$$

since $\sum_{j=0}^{T-1}\psi_{\bar{\delta},j}\approx\sum_{j=0}^{T-1}j^{\bar{\delta}-1}=O(T^{\bar{\delta}})$ (Gradstein and Ryzhik 1994, eqn. 0.121).

Proof of Lemma 6

First, (A.6) follows by Markov's inequality and $E\left|T^{-1}\sum_{t=1}^{T-1}\underline{\tilde{u}}^2_{2,t}\right|<\infty$, which holds since

$$\begin{aligned}T^{-1}\sum_{t=1}^{T-1}\underline{\tilde{u}}^2_{2,t} &= T^{-1}\sum_{t=1}^{T-1}\tilde{u}^2_{2,t}-\left(T^{-1}\sum_{t=1}^{T-1}\tilde{u}_{2,t}\right)^2>0\text{ and therefore} \\ E\left[\left(T^{-1}\sum_{t=1}^{T-1}\tilde{u}_{2,t}\right)^2\right] &\leq T^{-1}\sum_{t=1}^{T-1}E\tilde{u}^2_{2,t}\leq\max_{t\leq T}E\tilde{u}^2_{2,t}<\infty.\end{aligned}\tag{A.20}$$

Employing the Cauchy-Schwartz inequality, similar argument shows (A.5). For (A.4) write

$$T^{-1/2}\sum_{t=1}^{T-1}\underline{\tilde{u}}_{2,t}\varepsilon_{1,t+1}=T^{-1/2}\sum_{t=1}^{T-1}\tilde{u}_{2,t}\varepsilon_{1,t+1}-\left(T^{-1}\sum_{t=1}^{T-1}\tilde{u}_{2,t}\right)T^{-1/2}\sum_{t=1}^{T-1}\varepsilon_{1,t+1}.\tag{A.21}$$

The second term on the RHS is O_p (1) by (A.20) and application of the standard central limit theorem. For the first term, since $u_{2,t}$ is predetermined, by the Law of Iterative Expectations,

$$\begin{aligned}E\left[T^{-1/2}\sum_{t=1}^{T-1}\tilde{u}_{2,t}\varepsilon_{1,t+1}\right]^2 &= T^{-1}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}E\left[\tilde{u}_{2,t}\tilde{u}_{2,s}\varepsilon_{1,t+1}\varepsilon_{1,s+1}\right]=T^{-1}\sum_{t=1}^{T-1}E\left[\tilde{u}^2_{2,t}\varepsilon^2_{1,t+1}\right] \\ &\leq \max_{t\leq T}E\left[\tilde{u}^2_{2,t}\right]\Sigma_{11}<\infty.\end{aligned}$$

Next, (A.10) follows by similar argument as (A.6) since

$$E\left[\left(T^{-1}\sum_{t=1}^{T-1}u^*_{2,t,T}\right)^2\right]\leq T^{-1}\sum_{t=1}^{T-1}E\left[\left(u^*_{2,t,T}\right)^2\right]\leq\max_{t\leq T}E\left[\left(u^*_{2,t,T}\right)^2\right]=O\left(\ln(T)^2T^{2\bar{\delta}}\right).\tag{A.22}$$

(A.7) follows from the application of the Cauchy-Schwarz inequality since

$$\begin{aligned}E\left|T^{-1}\sum_{t=1}^{T-1}\underline{u}^*_{2,t,T}\varepsilon_{1,t+1}\right| &\leq \left[T^{-1}\sum_{t=1}^{T-1}E\left(\underline{u}^*_{2,t,T}\right)^2\right]^{1/2}\left[E\left(T^{-1}\sum_{t=1}^{T-1}\varepsilon^2_{1,t+1}\right)\right]^{1/2} \\ &\leq \left[\max_{t\leq T}\left(\underline{u}^*_{2,t,T}\right)^2\right]^{1/2}\Sigma^{1/2}_{1,1}=O(\ln(T)T^{\bar{\delta}}).\end{aligned}$$

Parts (A.8) and (A.9) follow by similar argument.

Proof of Lemma 8

For **(A.33)**: Since convergence in probability is implied by MSE convergence, we need only show that

$$lim_{T\to\infty} E\left[T^{-1}\sum_{t=1}^{T-1}\underline{u}_{2,t}\tilde{\varepsilon}_{1,t+1}\right] = A \qquad \text{and that} \qquad \lim_{T\to\infty}\text{var}\left(T^{-1}\sum_{t=1}^{T-1}\underline{u}_{2,t}\tilde{\varepsilon}_{1,t+1}\right) = 0.$$

Substituting $u_{2,t} = \sum_{j=0}^{\infty} C_{2j}\varepsilon_{t-j}$ from (6) and $\tilde{\varepsilon}_{1,t+1} = ln(1-L)\left(\varepsilon_{1,t+1}1_{\{t+1>0\}}\right) = -\sum_{j=1}^{t}\frac{1}{j}\varepsilon_{1,t+1-j}$ and noting that $E\left[\varepsilon_{t-j}\varepsilon_{1,t-k}\right] = 0$ for $j \neq k$,

$$E\left[\varepsilon_{t-j}\varepsilon_{2,t-j}\right] = E\left\{\begin{bmatrix}\varepsilon_{1,t-j}\\ \varepsilon_{2,t-j}\end{bmatrix}\varepsilon_{2,t-j}\right\} = \begin{bmatrix}\Sigma_{12}\\ \Sigma_{22}\end{bmatrix} \equiv \Sigma_{.,1},$$

and $lim_{T\to\infty}E\left[T^{-1}\sum_{t=1}^{T-1}\bar{u}_2\tilde{\varepsilon}_{1,t+1}\right] = 0$ by (A.11) and because $\bar{u}_2 \to_p E\left[u_{2,t}\right] = 0$, we obtain

$$\begin{aligned}
lim_{T\to\infty}E\left[T^{-1}\sum_{t=1}^{T-1}\underline{u}_{2,t}\tilde{\varepsilon}_{1,t+1}\right] &= lim_{T\to\infty}E\left[T^{-1}\sum_{t=1}^{T-1}u_{2,t}\tilde{\varepsilon}_{1,t+1}\right]\\
&= lim_{T\to\infty}T^{-1}\sum_{t=1}^{T-1}E\left[u_{2,t}\tilde{\varepsilon}_{1,t+1}\right] =\\
&= lim_{T\to\infty}T^{-1}\sum_{t=1}^{T-1}E\left[\left(\sum_{j=0}^{\infty}C_{2,j}\varepsilon_{t-j}\right)\left(-\sum_{k=0}^{t-1}\frac{1}{k+1}\varepsilon_{1,t-k}\right)\right] =\\
&= -lim_{T\to\infty}T^{-1}\sum_{t=1}^{T-1}\sum_{j=0}^{\infty}C_{2,j}\sum_{k=0}^{t-1}\frac{1}{k+1}E\left[\varepsilon_{t-j}\varepsilon_{1,t-k}\right]\\
&= -lim_{T\to\infty}T^{-1}\sum_{t=1}^{T-1}\sum_{j=0}^{t-1}\frac{1}{j+1}C_{2,j}\Sigma_{.,1} \equiv A
\end{aligned}$$

and this limit is finite since its argument is bounded:

$$\begin{aligned}
\left|T^{-1}\sum_{t=1}^{T-1}\sum_{j=0}^{t-1}\frac{1}{j+1}C_{2,j}\Sigma_{.,1}\right| &\le T^{-1}\sum_{t=1}^{T-1}\sum_{j=0}^{t-1}\left|\frac{1}{j+1}C_{2,j}\Sigma_{.,1}\right| \le T^{-1}\sum_{t=1}^{T}\sum_{j=0}^{T}\left|\frac{1}{j+1}C_{2,j}\Sigma_{.,1}\right|\\
&= \sum_{j=0}^{T}\left|\frac{1}{j+1}C_{2,j}\Sigma_{.,1}\right| = \sum_{j=0}^{T}\left\|\frac{1}{j+1}C_{2,j}\Sigma_{.,1}\right\| \le \sum_{j=0}^{T}\frac{1}{j+1}\left\|C_{2,j}\right\|\left\|\Sigma_{.,1}\right\| \quad \text{(A.23)}
\end{aligned}$$

which is finite by (7), where $\|\cdot\|$ denotes a matrix norm.

Next we turn to the variance. Using Cauchy-Schwarz and Holder's Inequality:

$$\begin{aligned}
E\left[\left|T^{-1}\sum_{t=1}^{T-1}\underline{u}_{2,t}\tilde{\varepsilon}_{1,t+1}\right|^2\right] &= E\left[\left|T^{-1}\sum_{t=1}^{T-1}\underline{u}_{2,t}\underline{\tilde{\varepsilon}}_{1,t+1}\right|^2\right]\\
&\le E\left[\left(T^{-1}\sum_{t=1}^{T-1}\underline{u}_{2,t}^2\right)\left(T^{-1}\sum_{t=1}^{T-1}\underline{\tilde{\varepsilon}}_{1,t+1}^2\right)\right]\\
&\le \left\{E\left|T^{-1}\sum_{t=1}^{T-1}\underline{u}_{2,t}^2\right|^2\right\}^{1/2}\left\{E\left|T^{-1}\sum_{t=1}^{T-1}\underline{\tilde{\varepsilon}}_{1,t+1}^2\right|^2\right\}^{1/2},
\end{aligned}$$

where by the argument of (A.20) $E\left|T^{-1}\sum_{t=1}^{T-1}\underline{u}_{2,t}^2\right|^2$, $E\left|T^{-1}\sum_{t=1}^{T-1}\underline{\tilde{\varepsilon}}_{1,t+1}^2\right|^2 < \infty$

The proof of (A.15) is omitted since it follows by the same arguments as the proofs of (A.5) and (A.6). Likewise, the proofs of (A.16)-(A.19) follow closely those of (A.7)-(A.10).

Proof of Theorem 1

Define $\hat{\delta}_T = (\hat{d} - d)$, where $-\hat{\delta}_T$ is the integration order of the second-stage regressor. By assumption

$$T^{\alpha}\hat{\delta}_T = T^{\alpha}(\hat{d} - d) = O_p(1). \tag{A.24}$$

Using demeaned fitted and true models $\underline{y}_{t+1} = \hat{\beta}_1\underline{\hat{u}}_{2,t} + \underline{\hat{\varepsilon}}_{1t+1}$ and $\underline{y}_{t+1} = \beta_1\underline{u}_{2,t} + \underline{\varepsilon}_{1t+1}$,

$$\sqrt{T}(\hat{\beta}_1 - \beta_1) = \left(T^{-1}\sum_{t=1}^{T-1}\underline{\hat{u}}_{2,t}^2\right)^{-1}\left(T^{-1/2}\sum_{t=1}^{T-1}\underline{\hat{u}}_{2,t}\varepsilon_{1,t+1} + \beta_1 T^{-1/2}\sum_{t=1}^{T-1}(\underline{u}_{2,t} - \underline{\hat{u}}_{2,t})\underline{\hat{u}}_{2,t}\right). \tag{A.25}$$

Let $\bar{\delta} > 0$ and let the indicator $I_{\bar{\delta}}$ take the value 1 if $|\hat{\delta}_T| < \bar{\delta}$ and zero otherwise. Let $\eta > 0$. Since $\hat{d} \to_p d$, for large T, $P(I_{\bar{\delta}}=0)=P(|\hat{\delta}_T| > \bar{\delta}) < \eta$. Thus, $I_{\bar{\delta}} \to_p 1$ and

$$\sqrt{T}(\hat{\beta}_1 - \beta_1) = I_{\bar{\delta}}\sqrt{T}(\hat{\beta}_1 - \beta_1) + (1 - I_{\bar{\delta}})\sqrt{T}(\hat{\beta}_1 - \beta_1) = I_{\bar{\delta}}\sqrt{T}(\hat{\beta}_1 - \beta_1) + o_p(1),$$

where the last term is o_p (1) since $(1 - I_{\bar{\delta}})\sqrt{T}(\hat{\beta}_1 - \beta_1)$ =0 when $I_{\bar{\delta}} = 1$, and $P(I_{\bar{\delta}} = 1) \to 1$. Therefore in what follows below we will assume $|\hat{\delta}_T| < \bar{\delta}$ without loss of generality.

Next, applying an exact second order Taylor series expansion to the function $(1 - L)^{\hat{\delta}_T}$ with argument $\hat{\delta}_T$ about zero and where δ_T^* lies between 0 and $\hat{\delta}_T$ gives

$$\begin{aligned}(1 - L)^{\hat{\delta}_T} &= 1 + \hat{\delta}_T\ln(1 - L) + \frac{1}{2}\hat{\delta}_T^2\ln(1 - L)^2(1 - L)^{\delta_T^*} \text{ and}\\ \hat{u}_{2,t} &= (1 - L)^{\hat{\delta}_T}u_{2,t}1_{\{t>0\}} = u_{2,t}1_{\{t>0\}} + \hat{\delta}_T\tilde{u}_{2,t} + \frac{1}{2}\hat{\delta}_T^2 u_{2,t,T}^*\end{aligned} \tag{A.26}$$

where $u_{2,t,T}^*$, $\tilde{\tilde{u}}_{2,t}$, and $\tilde{u}_{2,t}$ are defined in Lemma 5.

Next, we turn to the first term in the numerator of $\sqrt{T}(\hat{\beta}_1 - \beta_1)$ in (A.25). Using (A.26) to substitute for $\hat{u}_{2,t}$, we have

$$T^{-1/2}\sum_{t=1}^{T-1}\underline{\hat{u}}_{2,t}\varepsilon_{1,t+1} = T^{-1/2}\sum_{t=1}^{T-1}\underline{u}_{2,t}\varepsilon_{1,t+1} + R_{1,T} \to_d N(0, \mathrm{var}[u_{2,t}]\Sigma_{11}) \tag{A.27}$$

by Davidson (2000, Theorem 6.2.3, p. 124), since $u_t\varepsilon_{1t+1}$ is a strictly stationary Martingale difference Sequence,[15] and since, by Lemma 6 (A.4) and (A.7),

$$R_{1,T} = \hat{\delta}_T T^{-1/2}\sum_{t=1}^{T-1}\underline{\tilde{u}}_{2,t}\varepsilon_{1,t+1} + \frac{1}{2}\hat{\delta}_T^2 T^{-1/2}\sum_{t=1}^{T-1}\underline{u}_{2,t,T}^*\varepsilon_{1,t+1} = O(T^{-\alpha}) + O_p(\ln(T)T^{1/2-2\alpha+\bar{\delta}}) = o_p(1)$$

for $\alpha > \frac{1}{4}(1 + 2\bar{\delta})$, again with $\bar{\delta}$ arbitrarily small.

[15] Note that $u_{2,t}$ is a pre-determined short-memory linear process and ε_{1t+1} is an i.i.d. series so that the asymptotic normality result employed here is quite standard.

The behavior of the second term in the numerator of $\sqrt{T}(\hat{\beta}_1-\beta_1)$ in (A.25) is given by

$$\begin{aligned}\beta_1 T^{-1/2}\sum_{t=1}^{T-1}(\underline{u}_{2,t}-\hat{\underline{u}}_{2,t})\hat{\underline{u}}_{2,t} &= -\beta_1 T^{-1/2}\sum_{t=1}^{T-1}\left(\hat{\delta}_T\tilde{\underline{u}}_{2,t}+\frac{1}{2}\hat{\delta}_T^2\underline{u}_{2,t,T}^*\right)\left(\underline{u}_{2,t}+\hat{\delta}_T\tilde{\underline{u}}_{2,t}+\frac{1}{2}\hat{\delta}_T^2\underline{u}_{2,t,T}^*\right)\\ &= \beta_1 T^{-1/2}\hat{\delta}_T\sum_{t=1}^{T-1}\tilde{\underline{u}}_{2,t}\underline{u}_{2,t}+\beta_1 R_{2,T}=\beta_1 O_p(T^{1/2-\alpha}),\end{aligned}$$

giving the order of magnitude of the contamination term B_T in (9), where $R_{2,T}$ is defined as,

$$R_{2,T}=T^{-1/2}\left[\hat{\delta}_T^2\sum_{t=1}^{T-1}\tilde{u}_{2,t}^2+\frac{1}{2}\hat{\delta}_T^2\sum_{t=1}^{T-1}u_{2,t,T}^*u_{2,t}+\hat{\delta}_T^3\sum_{t=1}^{T-1}u_{2,t,T}^*\tilde{u}_{2,t}+\frac{1}{4}\hat{\delta}_T^4\sum_{t=1}^{T-1}(u_{2,t,T}^*)^2\right].$$

For $\alpha>\frac{1}{4}(1+2\bar{\delta})$, and by Lemma 6, we have $R_{2,T}=o_p\,(1)$.

For the denominator of $\sqrt{T}(\hat{\beta}_1-\beta_1)$ in equation (A.25) we have

$$T^{-1}\sum_{t=1}^{T-1}\hat{\underline{u}}_{2,t}^2=T^{-1}\sum_{t=1}^{T-1}\underline{u}_{2,t}^2+R_{3,T}\to_p \mathrm{var}[u_{2,t}] \tag{A.28}$$

by standard argument, since by Lemma 6,

$$\begin{aligned}R_{3,T} &= \hat{\delta}_T^2 T^{-1}\sum_{t=1}^{T-1}\tilde{\underline{u}}_{2,t}^2+\frac{1}{4}\hat{\delta}_T^4 T^{-1}\sum_{t=1}^{T-1}\left(\underline{u}_{2,t,T}^*\right)^2+2\hat{\delta}_T T^{-1}\sum_{t=1}^{T-1}\tilde{\underline{u}}_{2,t}\underline{u}_{2,t}+\hat{\delta}_T^2 T^{-1}\sum_{t=1}^{T-1}\underline{u}_{2,t,T}^*\underline{u}_{2,t}\\ &+ \hat{\delta}_T^3 T^{-1}\sum_{t=1}^{T-1}\underline{u}_{2,t,T}^*\tilde{\underline{u}}_{2,t}=O_p(T^{-2\alpha})+O_p\left(T^{2\bar{\delta}-4\alpha}\ln(T)^2\right)+O_p(T^{-\alpha})\\ &+ O_p\left(T^{\bar{\delta}-2\alpha}\ln(T)\right)+O_p\left(T^{\bar{\delta}-3\alpha}\ln(T)\right)=o_p(1),\end{aligned}$$

for $\alpha>\bar{\delta}$. Combining the above results shows Theorem 1.

Proof of Corollary 2

$\underline{y}_{t+1}=\hat{\beta}_1\hat{\underline{u}}_{2,t}+\hat{\varepsilon}_{1t+1}$ and $\hat{\underline{u}}_{2,t}=\underline{u}_{2,t}+\hat{\delta}_T\tilde{\underline{u}}_{2,t}+\frac{1}{2}\hat{\delta}_T^2\underline{u}_{2,t,T}^*$. Therefore,

$$\begin{aligned}\hat{\varepsilon}_{1t+1} &= \underline{y}_{t+1}-\hat{\beta}_1\hat{\underline{u}}_{2,t}=\varepsilon_{1t+1}-(\hat{\beta}_1-\beta_1)\underline{u}_{2,t}-\hat{\delta}_T\hat{\beta}_1\tilde{\underline{u}}_{2,t}-\frac{1}{2}\hat{\delta}_T^2\hat{\beta}_1\underline{u}_{2,t,T}^*\\ \hat{\sigma}^2 &= T^{-1}\sum_{t=1}^{T-1}\hat{\varepsilon}_{1t+1}^2=T^{-1}\sum_{t=1}^{T-1}\varepsilon_{1t+1}^2+R_{4,T}\end{aligned}$$

by standard argument since, for $\alpha>\frac{\bar{\delta}}{2}$, by (A.24), Lemma 6 (A.4) - (A.10), and (10)

$$\begin{aligned}R_{4,T} &= (\hat{\beta}_1-\beta_1)^2T^{-1}\sum_{t=1}^{T-1}\underline{u}_{2,t}^2+\hat{\beta}_1^2\hat{\delta}_T^2T^{-1}\sum_{t=1}^{T-1}\tilde{\underline{u}}_{2,t}^2+\frac{1}{4}\hat{\beta}_1^2\hat{\delta}_T^4T^{-1}\sum_{t=1}^{T-1}(u_{2,t,T}^*)^2\\ &- 2(\hat{\beta}_1-\beta_1)T^{-1}\sum_{t=1}^{T-1}\underline{u}_{2,t}\varepsilon_{1t+1}-2\hat{\delta}_T\hat{\beta}_1T^{-1}\sum_{t=1}^{T-1}\tilde{\underline{u}}_{2,t}\varepsilon_{1t+1}-\hat{\delta}_T^2\hat{\beta}_1T^{-1}\sum_{t=1}^{T-1}u_{2,t,T}^*\varepsilon_{1t+1}\\ &+ 2\hat{\delta}_T(\hat{\beta}_1-\beta_1)\hat{\beta}_1T^{-1}\sum_{t=1}^{T-1}\underline{u}_{2,t}\tilde{\underline{u}}_{2,t}+\hat{\delta}_T^2\left(\hat{\beta}_1-\beta_1\right)\hat{\beta}_1T^{-1}\sum_{t=1}^{T-1}\underline{u}_{2,t}u_{2,t,T}^*\\ &+ \hat{\delta}_T^3\hat{\beta}_1^2T^{-1}\sum_{t=1}^{T-1}\tilde{\underline{u}}_{2,t}u_{2,t,T}^*=o_p(1).\end{aligned}$$

Proof of Corollary 3

Result (a) follows from Theorem 1 and Corollary 2 by standard arguments (note that the contamination term B_T is not present under the null $H_o : \beta_1 = 0$). For (**b**) note that under $H_A : \beta_1 \neq 0$, we have $\hat{\beta}_1 - \beta_1 = O_p(T^{-\alpha})$ by (10). Therefore

$$\begin{aligned} T^{-1/2}t &= \hat{\sigma}^{-1}\left(T^{-1}\sum_{t=1}^{T-1}\underline{\hat{u}}_{2,t}^2\right)^{1/2}\hat{\beta}_1 = \hat{\sigma}^{-1}\left(T^{-1}\sum_{t=1}^{T-1}\underline{\hat{u}}_{2,t}^2\right)^{1/2}T^{1/2}\beta_1 + \hat{\sigma}^{-1}\left(T^{-1}\sum_{t=1}^{T-1}\underline{\hat{u}}_{2,t}^2\right)^{1/2}(\hat{\beta}_1 - \beta_1) \\ &\to {}_p\Sigma_{11}^{-1/2}\mathrm{var}[u_{2,t}]^{1/2}\beta_1 \end{aligned}$$

since the second term is $o_p\ (1)$ on account of the consistency of $\hat{\beta}_1$ for β_1.

Proof of Theorem 4

Define $\hat{\delta}_T^y = \hat{d}_y - d_y$ and let $\bar{\delta} > 0$. By the arguments of Theorem 1 assume $\hat{d}_x, \hat{d}_y < \bar{\delta}$ without loss of generality. The denominator of $\hat{\beta}_1$ in (14) is unchanged relative to Theorem 1. Using (12) to substitute for y_{t+1}, that $T^{(\alpha_y-1)}$ times the numerator is given by

$$T^{(\alpha_y-1)}\sum_{t=1}^{T-1}(1-L)^{\hat{d}_y}y_{t+1}\underline{\hat{u}}_{2,t} = \sum_{t=1}^{T-1}(1-L)^{\hat{d}_y-d_y}\left(\varepsilon_{1t+1}1_{\{t>0\}}\right)\underline{\hat{u}}_{2,t}.$$

Following Theorem 1, take an exact second order Taylor series expansion of $\hat{\delta}_T^y$, with $0 \leq \delta_T^{*y} \leq \hat{\delta}_T^y$, to obtain

$$(1-L)^{\hat{\delta}_T^y} = 1 + \hat{\delta}_T^y \ln(1-L) + \frac{1}{2}(\hat{\delta}_T^y)^2\left[\ln(1-L)\right]^2(1-L)^{\delta_T^{*y}}. \tag{A.29}$$

Then, employing the definitions of Lemma 7,

$$(1-L)^{\hat{d}_y-d_y}\left(\varepsilon_{1t+1}1_{\{t>0\}}\right)\underline{\hat{u}}_{2,t} = \left(\varepsilon_{1t+1} + \hat{\delta}_T^y\tilde{\varepsilon}_{1,t+1} + (\hat{\delta}_T^y)^2\varepsilon_{1,t+1,T}^*\right)\underline{\hat{u}}_{2,t}.$$

Therefore, $T^{(\alpha_y-1)}$ times the numerator is given by the three terms:

$$\begin{aligned} T^{(\alpha_y-1)}\sum_{t=1}^{T-1}[(1-L)^{\hat{d}_y-d_y}\varepsilon_{1t+1}\underline{\hat{u}}_{2,t}] &= T^{(\alpha_y-1)}\sum_{t=1}^{T-1}\varepsilon_{1t+1}\underline{\hat{u}}_{2,t} + \hat{\delta}_T^y T^{(\alpha_y-1)}\sum_{t=1}^{T-1}\tilde{\varepsilon}_{1,t+1}\underline{\hat{u}}_{2,t} \\ &+ (\hat{\delta}_T^y)^2 T^{(\alpha_y-1)}\sum_{t=1}^{T-1}\varepsilon_{1,t+1,T}^*\underline{\hat{u}}_{2,t}. \end{aligned} \tag{A.30}$$

The behavior of the first term in (A.30) is derived in (A.27), from which we can see that

$$T^{(\alpha_y-1)}\sum_{t=1}^{T-1}\varepsilon_{1t+1}\underline{\hat{u}}_{2,t} \to_d 1_{\{\alpha_y=1/2\}}\mathrm{var}\left[u_{2,t}\right]\xi_\beta \tag{A.31}$$

where ξ_β is specified in (18).

Substituting (A.26) for $\underline{u}_{2,t}$, the second term in (A.30) is given by

$$\hat{\delta}_T^y T^{(\alpha_y-1)}\sum_{t=1}^{T-1}\tilde{\varepsilon}_{1,t+1}\underline{\hat{u}}_{2,t} = \hat{\delta}_T^y T^{(\alpha_y-1)}\sum_{t=1}^{T-1}\tilde{\varepsilon}_{1,t+1}\underline{u}_{2,t} + \hat{\delta}_T^y\hat{\delta}_T T^{(\alpha_y-1)}\sum_{t=1}^{T-1}\tilde{\varepsilon}_{1,t+1}\underline{\tilde{u}}_{2,t} + \hat{\delta}_T^y\hat{\delta}_T^2 T^{(\alpha_y-1)}\sum_{t=1}^{T-1}\tilde{\varepsilon}_{1,t+1}\underline{u}_{2,t,T}^* \tag{A.32}$$

For the first term, using Lemma 8 (A.33), and noting that $|A| < \infty$ by (A.23),

$$T^{-(1-\alpha)}\hat{\delta}_T^y \sum_{t=1}^{T-1} \tilde{\varepsilon}_{1,t+1}\underline{u}_{2,t} = T^{\alpha}\hat{\delta}_T^y T^{-1} \sum_{t=1}^{T-1} \tilde{\varepsilon}_{1,t+1}\underline{u}_{2,t} \rightarrow_d A\xi_{\delta^y},$$

where the distribution of ξ_{δ^y} is specified in (15) and (17) respectively. By (A.15) and (A.16), the remaining two terms in (A.32) are $O_p\left(T^{-\alpha_x}\right)$ and $O_p\left(T^{\left(\bar{\delta}_x-2\alpha_x\right)}\ln\left(T\right)\right) = o_p\left(1\right)$ respectively.

Then the third main term in (A.30) is given by:

$$\begin{aligned}\left(\hat{\delta}_T^y\right)^2 T^{(\alpha_y-1)} \sum_{t=1}^{T-1} \varepsilon_{1,t+1,T}^*\underline{\hat{u}}_{2,t} &= \left(\hat{\delta}_T^y\right)^2 T^{(\alpha_y-1)} \sum_{t=1}^{T-1} \varepsilon_{1,t+1,T}^*\underline{u}_{2,t} + \left(\hat{\delta}_T^y\right)^2 \hat{\delta}_T T^{(\alpha_y-1)} \sum_{t=1}^{T-1} \varepsilon_{1,t+1,T}^*\underline{\tilde{u}}_{2,t} \\ &+ \frac{1}{2}\left(\hat{\delta}_T^y\right)^2 \hat{\delta}_T^2 T^{(\alpha_y-1)} \sum_{t=1}^{T-1} \varepsilon_{1,t+1,T}^*\underline{u}_{2,t,T}^* \qquad \text{(A.33)}\end{aligned}$$

By (A.17), (A.18), and (A.19) the three terms in (A.33) are $O_p\left(T^{\left(\bar{\delta}^y-\alpha_y\right)}ln\left(T\right)\right) = o_p(1)$, $O_p\left(T^{\left(\bar{\delta}^y-\alpha_y-\alpha_x\right)}ln\left(T\right)\right) = o_p(1)$, and $O_p\left(T^{\left(\bar{\delta}^y+\bar{\delta}^x-\alpha_y-2\alpha_x\right)}ln\left(T\right)^2\right) = o_p\left(1\right)$, respectively.

References

Agiakloglou, Christos, Paul Newbold and Mark Wohar (1993). Bias in an estimator of the fractional difference parameter. *Journal of Time Series Analysis* **14**(3), 235–246.

Andrews, D W K and P Guggenberger (2003). A biased reduced log-periodogram regression estimator for the long memory parameter. *Econometrica* **71**, 675–712.

Baillie, R T, C-F Chung and M A Tieslau (1996). Analyzing inflation by the fractionally integrated ARFIMA-GARCH model. *Journal of Applied Econometrics* **11**, 23–40.

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

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

Beran, J (1995). Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. *Journal of the Royal Statistical Society B* **57**, 659–672.

Campbell, Bryan and Jean-Marie Dufour (1997). Exact nonparametric tests of orthogonality and random walk in the presence of a drift parameter. *International Economic Review* **38**(1), 151–173.

Cavanagh, Christopher L., Graham Elliott and James H. Stock (1995). Inference in models with nearly integrated regressors. *Econometric Theory* **11**, 1131–1147.

Chung, C-F and R T Baillie (1993). Small sample bias in conditional sum-of-squares estimators of fractionally integrated ARMA models. *Empirical Economics* **18**, 791–806.

Davidson, J (2000). *Econometric Theory*. Blackwell Publishers Malden and Mass.

Dolado, J, J Gonzalo and L Mayoral (2002). A fractional dickey-fuller test for unit roots. *Econometrica* **70**, 1963–2006.

Fox, R. and M. S. Taqqu (1986). Large sample properties of parameter estimates for strongly dependent stationary gaussian time series. *Annals of Statistics* **14**, 517–532.

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

Gradstein, I. S. and I. M. Ryzhik (1994). *Table of Integrals, Series and Products*. Academic Press. Boston.

Jansson, Michael and Marcelo J. Moreira (2006). Optimal inference in regression models with nearly integrated regressors. *Econometrica* **74**, 681–714.

Koustas, Z and A Serletis (2005). Rational bubbles or persistent deviations from market fundamentals?. *Journal of Banking and Finance* **29**, 2523–2539.

Liu, Wei and Alex Maynard (2005). Testing forward rate unbiasedness allowing for persistent regressors. *Journal of Empirical Finance* **12**, 613–628.

Mankiw, N.G. and M. Shapiro (1986). Do we reject too often? Small sample properties of tests of rational expectations models. *Economics Letters* **20**, 139–145.

Marinucci, D and P M Robinson (1999). Alternative forms of fractional Brownian motion. *Journal of Statistical Planning and Inference* **80**, 111–122.

Maynard, Alex (2006). The forward premium anomaly: statistical artifact or economic puzzle? new evidence from robust tests. *Canadian Journal of Economics* **39**(4), 1244–1281.

Maynard, Alex and Katsumi Shimotsu (forthcoming). Covariance-based orthogonality tests for regressors with unknown persistence. *Econometric Theory*.

Maynard, Alex and Peter C. B. Phillips (2001). Rethinking an old empirical puzzle: Econometric evidence on the forward discount anomaly. *Journal of Applied Econometrics* **16**(6), 671–708.

Nielsen, M and P H Frederiksen (2005). Finite sample comparison of parametric, semiparametric, and wavelet estimators of fractional integration. *Econometric Reviews* **24**, 405–443.

Phillips, P. C. B. (1999). Discrete fourier transforms of fractional processes. Cowles Foundation Discussion Paper #1243, Yale University.

Phillips, Peter C. B. and Victor Solo (1992). Asymptotics for linear processes. *Annals of Statistics* **20**, 971–1001.

Rapach, D E and M E Wohar (2006). In-sample vs. out-of-sample tests of stock return predictability in the context of data mining. *Journal of Empirical Finance* **13**, 231–247.

Richardson, M. and J. Stock (1989). Drawing inferences from statistics based on multi-year asset returns. *Journal of Financial Economics* **25**, 323–348.

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

Rossi, Barbara (2005). Expectations hypotheses tests at long horizons. Mimeo, Duke University.

Shimotsu, K and P C B Phillips (2005). Exact local whittle estimation of fractional integration. *Annals of Statistics* **33**, 1890–1933.

Sowell, F B (1992). Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics* **53**, 165–188.

Stambaugh, Robert F. (1999). Predictive regressions. *Journal of Financial Economics* **54**, 375–421.

Sun, Y and P C B Phillips (2003). Nonlinear log-periodogram regression estimator for perturbed fractional processes. *Journal of Econometrics* **115**, 355–389.

Tanaka, Katsuto (1999). The nonstationary fractional unit root. *Econometric Theory* **15**(4), 549–582.

Torous, Walter, Rossen Valkanov and Shu Yan (2005). On predicting stock returns with nearly integrated explanatory variables. *Journal of Business* **78**(1), 937–966.

Valkanov, R. (2003). Long-horizon regressions: theoretical results and applications. *Journal of Financial Economics* **68**, 201–232.

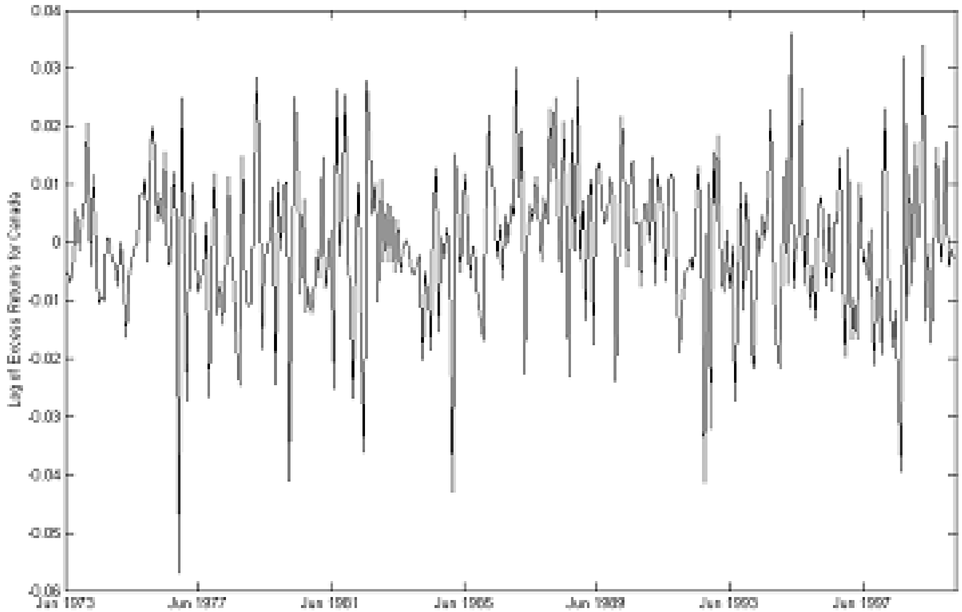

Figure 1: Log of Excess Returns for the Canadian Dollar vis-à-vis the US Dollar (1973-2000)

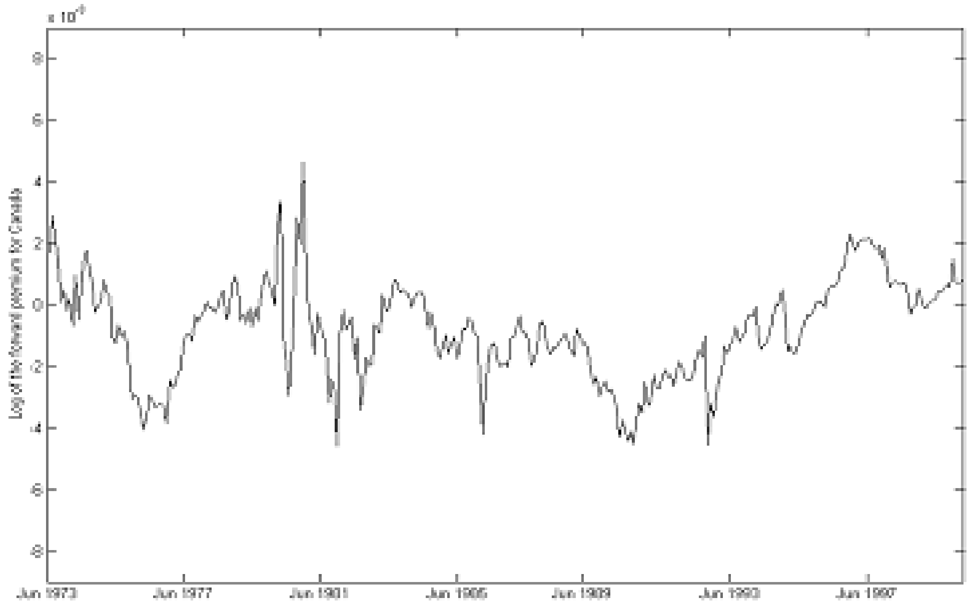

Figure 2: Log of the Forward Premium for the Canadian Dollar vis-à-vis the US Dollar (1973-2000)

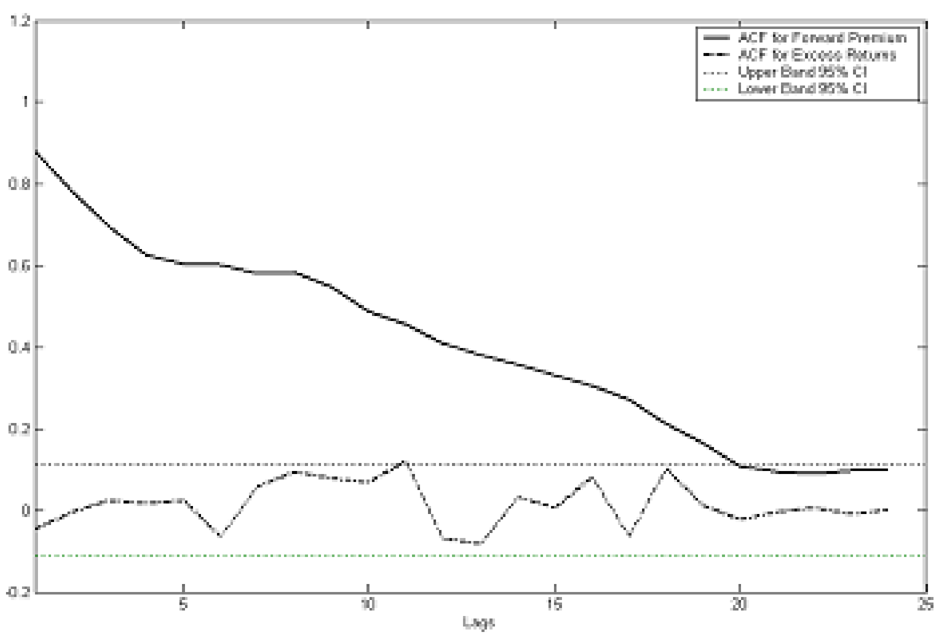

Figure 3: Sample Autocorrelations for Canadian Excess Returns and Forward Premium With 95% Confidence Intervals about Zero

Table 1
Unbalanced Regression without Differencing
Regressor is Fractional Noise; Dependent Variable is Short Memory

Table 1a
Proportion of Rejections in a 5% Nominal Test of the Null Hypothesis that $b_1 = 0$
True Value of $b_1 = 0$, Sample Size=350

ρ /d	-0.95	-0.90	-0.80	-0.40	0.00	0.40	0.80	0.90	0.95
0.40	0.0600	0.0577	0.0593	0.0520	0.0450	0.0420	0.0470	0.0507	0.0507
0.50	0.0703	0.0700	0.0647	0.0533	0.0443	0.0473	0.0550	0.0627	0.0690
0.60	0.0983	0.0943	0.0883	0.0550	0.0427	0.0507	0.0757	0.0890	0.0960
0.70	0.1357	0.1240	0.1080	0.0587	0.0440	0.0557	0.1117	0.1247	0.1343
0.80	0.1883	0.1647	0.1400	0.0677	0.0457	0.0667	0.1450	0.1707	0.1867
0.90	0.2273	0.2033	0.1727	0.0760	0.0517	0.0777	0.1807	0.2197	0.2423
0.95	0.2553	0.2317	0.1860	0.0780	0.0483	0.0847	0.2000	0.2457	0.2710
1.000	0.2770	0.2533	0.2017	0.0800	0.0473	0.0857	0.2167	0.2633	0.2913

Table 1b
Bias of the Estimate of b_1
True Value of $b_1 = 0$, Sample Size=350

ρ /d	-0.95	-0.90	-0.80	-0.40	0.00	0.40	0.80	0.90	0.95
0.40	0.7001	0.6662	0.5960	0.3080	0.0203	-0.2677	-0.5630	-0.6405	-0.6809
0.50	0.8845	0.8390	0.7466	0.3757	0.0091	-0.3551	-0.7255	-0.8212	-0.8706
0.60	0.9882	0.9351	0.8289	0.4096	-0.0020	-0.4077	-0.8204	-0.9263	-0.9806
0.70	0.9766	0.9223	0.8156	0.3982	-0.0106	-0.4097	-0.8176	-0.9230	-0.9763
0.80	0.8645	0.8145	0.7194	0.3492	-0.0142	-0.3665	-0.7282	-0.8237	-0.8709
0.90	0.6975	0.6565	0.5797	0.2807	-0.0130	-0.2961	-0.5898	-0.6692	-0.7079
0.95	0.6081	0.5724	0.5056	0.2448	-0.0115	-0.2567	-0.5146	-0.5850	-0.6189
1.000	0.5208	0.4902	0.4332	0.2098	-0.0100	-0.2177	-0.4408	-0.5020	-0.5310

Table 1c
Variance of the Estimate of b_1
True Value of $b_1 = 0$, Sample Size=350

ρ /d	-0.95	-0.90	-0.80	-0.40	0.00	0.40	0.80	0.90	0.95
0.40	2.29422	2.31768	2.35168	2.38560	2.33915	2.27288	2.22209	2.21883	2.22216
0.50	1.71949	1.73744	1.76483	1.79524	1.76253	1.71194	1.67128	1.66344	1.66440
0.60	1.21465	1.22060	1.22929	1.22687	1.20446	1.17843	1.18086	1.18033	1.18525
0.70	0.81844	0.81071	0.80037	0.76638	0.75296	0.74634	0.79357	0.81039	0.82101
0.80	0.52854	0.51057	0.48965	0.44447	0.43467	0.43862	0.50608	0.53609	0.55052
0.90	0.32904	0.31239	0.29019	0.24245	0.23006	0.24059	0.30314	0.33540	0.35056
0.95	0.25544	0.24076	0.22198	0.17576	0.16194	0.17378	0.22907	0.25962	0.27290
1.000	0.19685	0.18333	0.16811	0.12547	0.11203	0.12333	0.17105	0.19841	0.20938

Notes: The table shows simulation results from the standard predictability regression without rebalancing under the null hypothesis ($b_1 = 0$) based on the following regression:

$$y_{t+1} = c_1 + b_1 x_t + e_{1t+1} \tag{34}$$

The regressor x_t is integrated of order d and given by

$$(1-L)^d x_t = c_2 + e_{yt}, \tag{35}$$

where $e_t = \begin{pmatrix} e_{1t} & e_{2t} \end{pmatrix}' \sim$ i.i.d. $N(0,\Sigma)$, and $\rho = \Sigma_{12} \big/ \sqrt{\Sigma_{11}\Sigma_{22}}$ denotes the residual correlation.

Notes for tables 1-2:
Throughout, the true value of b_1 is equal to 0. Values for c_1 and c_2 are set equal to 0, while the standard deviations of the innovations in equations (34) and (35) above, have been estimated from the exchange rate data for Germany where the forward premium has been fractionally differenced with $d = 0.80$. The resulting values for the standard deviations of ε_{1t} and ε_{2t} are 0.03294 and 0.000942, respectively. To calculate correlated residuals we use the Cholesky factorization of the desired correlation matrix.

Table 2
Unbalanced Regression without Differencing
Regressor is an ARFIMA(1,0.80,0) Process

Table 2a

Proportion of Rejections in a 5% Nominal Test of the Null Hypothesis that $b_1 = 0$

True Value of $b_1 = 0$, Sample Size=350

ρ / ϕ	-0.95	-0.90	-0.80	-0.40	0.00	0.40	0.80	0.90	0.95
-0.99	0.0197	0.0213	0.0287	0.0397	0.0497	0.0463	0.0267	0.0230	0.0220
-0.95	0.0913	0.0850	0.0800	0.0537	0.0517	0.0553	0.0770	0.0823	0.0907
-0.80	0.1213	0.1127	0.1037	0.0643	0.0467	0.0557	0.1010	0.1147	0.1247
-0.40	0.1703	0.1533	0.1277	0.0657	0.0453	0.0647	0.1317	0.1553	0.1710
0.00	0.1883	0.1647	0.1400	0.0677	0.0457	0.0667	0.1450	0.1707	0.1867
0.40	0.1943	0.1770	0.1470	0.0700	0.0440	0.0707	0.1460	0.1827	0.2013
0.80	0.2110	0.1903	0.1603	0.0717	0.0520	0.0810	0.1633	0.1993	0.2160
0.95	0.2203	0.2120	0.1637	0.0727	0.0510	0.0817	0.1733	0.2163	0.2270
0.99	0.2560	0.2323	0.1903	0.0833	0.0470	0.0810	0.1857	0.2273	0.2553

Table 2b

Bias of the Estimate of b_1

True Value of $b_1 = 0$, Sample Size=350

ρ / ϕ	-0.95	-0.90	-0.80	-0.40	0.00	0.40	0.80	0.90	0.95
-0.99	0.0325	0.0305	0.0262	0.0052	-0.0068	-0.0127	-0.0283	-0.0338	-0.0358
-0.95	0.7950	0.7501	0.6617	0.3159	-0.0153	-0.3365	-0.6684	-0.7546	-0.7979
-0.80	1.0554	0.9959	0.8797	0.4244	-0.0165	-0.4449	-0.8868	-1.0016	-1.0592
-0.40	1.1106	1.0472	0.9253	0.4494	-0.0167	-0.4691	-0.9341	-1.0563	-1.1169
0.00	0.8645	0.8145	0.7194	0.3492	-0.0142	-0.3665	-0.7282	-0.8237	-0.8709
0.40	0.5526	0.5202	0.4589	0.2215	-0.0114	-0.2363	-0.4670	-0.5283	-0.5584
0.80	0.2083	0.1958	0.1724	0.0826	-0.0057	-0.0899	-0.1765	-0.1998	-0.2112
0.95	0.1158	0.1088	0.0958	0.0460	-0.0032	-0.0498	-0.0978	-0.1111	-0.1175
0.99	0.0239	0.0225	0.0200	0.0096	-0.0006	-0.0099	-0.0199	-0.0229	-0.0241

Table 2c

Variance of the Estimate of b_1

True Value of $b_1 = 0$, Sample Size=350

ρ / ϕ	-0.95	-0.90	-0.80	-0.40	0.00	0.40	0.80	0.90	0.95
-0.99	0.13639	0.14753	0.16445	0.22524	0.25382	0.22961	0.16235	0.13853	0.12816
-0.95	0.58076	0.60825	0.65784	0.80035	0.84368	0.78585	0.64009	0.59124	0.56670
-0.80	0.84047	0.85316	0.88078	0.96557	0.97829	0.92816	0.86841	0.84784	0.83704
-0.40	0.87308	0.85362	0.83286	0.79103	0.77707	0.76959	0.84981	0.88344	0.89903
0.00	0.52854	0.51057	0.48965	0.44447	0.43467	0.43862	0.50608	0.53609	0.55052
0.40	0.21792	0.20832	0.19693	0.17212	0.16816	0.17251	0.20512	0.22076	0.22919
0.80	0.03266	0.03059	0.02807	0.02321	0.02233	0.02335	0.02938	0.03234	0.03417
0.95	0.01070	0.00990	0.00893	0.00717	0.00668	0.00710	0.00916	0.01032	0.01094
0.99	0.00073	0.00066	0.00065	0.00040	0.00035	0.00040	0.00055	0.00070	0.00074

Notes: The table shows simulation results from the standard predictability regression without rebalancing under the null hypothesis ($b_1 = 0$) based on the following regression:

$$y_{t+1} = c_1 + b_1 x_t + e_{1t+1}.$$

The regressor x_t is integrated of order $d = 0.80$ throughout and given by

$$(1 - \phi L)(1 - L)^d x_t = c_2 + e_{yt},$$

where $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})' \sim$ i.i.d.$N(0, \Sigma)$. The values under the heading ρ/ϕ are the corresponding autoregressive coefficients (ϕ), while the values to the right of this heading yield the residual correlation coefficients (ρ).

Table 3
Original Process is Fractional Noise

Table 3a
Proportion of Rejections in a 5% Nominal Test of the Null Hypothesis that $\beta_1 = 0$
Sample Size=350, True Value of $\beta_1 = 0$

ρ /d	-0.95	-0.90	-0.80	-0.40	0.00	0.40	0.80	0.90	0.95
0.40	0.05933	0.05400	0.05367	0.04967	0.04867	0.04867	0.04667	0.04967	0.05033
0.50	0.05867	0.05667	0.05500	0.05133	0.04900	0.04867	0.04767	0.04933	0.05000
0.60	0.05833	0.05400	0.05533	0.05167	0.05000	0.05067	0.04800	0.04900	0.05167
0.70	0.06033	0.05467	0.05633	0.05067	0.05133	0.05000	0.04833	0.04967	0.05200
0.80	0.05633	0.05400	0.05367	0.05100	0.05400	0.04933	0.04700	0.05033	0.04967
0.90	0.05633	0.05367	0.05533	0.05133	0.05367	0.05033	0.04700	0.05067	0.05200
0.95	0.05500	0.05600	0.05467	0.05200	0.05467	0.05000	0.04733	0.04867	0.05100
1.000	0.05367	0.05267	0.05300	0.05367	0.05367	0.05133	0.05300	0.04733	0.04967

Table 3b
Bias of the Estimate of β_1 using the Two-Step Procedure
Sample Size=350, True Value of $\beta_1 = 0$

ρ /d	-0.95	-0.90	-0.80	-0.40	0.00	0.40	0.80	0.90	0.95
0.40	0.1531	0.1507	0.1422	0.0932	0.0370	-0.0245	-0.0946	-0.1156	-0.1277
0.50	0.1503	0.1482	0.1401	0.0936	0.0391	-0.0212	-0.0910	-0.1119	-0.1241
0.60	0.1430	0.1416	0.1348	0.0928	0.0424	-0.0150	-0.0829	-0.1035	-0.1158
0.70	0.1375	0.1369	0.1316	0.0946	0.0474	-0.0079	-0.0751	-0.0958	-0.1081
0.80	0.1302	0.1309	0.1278	0.0967	0.0540	0.0009	-0.0645	-0.0846	-0.0968
0.90	0.1169	0.1196	0.1198	0.0951	0.0586	0.0112	-0.0484	-0.0662	-0.0774
0.95	0.1084	0.1120	0.1136	0.0908	0.0581	0.0155	-0.0384	-0.0543	-0.0645
1.000	0.0993	0.1037	0.1060	0.0826	0.0544	0.0177	-0.0286	-0.0417	-0.0503

Table 3c
Variance of the Estimate of β_1 using the Two-Step Procedure
Sample Size=350, True Value of $\beta_1 = 0$

ρ /d	-0.95	-0.90	-0.80	-0.40	0.00	0.40	0.80	0.90	0.95
0.40	3.59836	3.60396	3.59999	3.51075	3.42144	3.39318	3.44707	3.48076	3.50570
0.50	3.57555	3.58211	3.57901	3.49147	3.40264	3.37472	3.42513	3.45733	3.48142
0.60	3.53193	3.53852	3.53572	3.45165	3.36702	3.34177	3.38954	3.41916	3.44155
0.70	3.43130	3.43802	3.43818	3.36923	3.29336	3.27361	3.31432	3.33653	3.35345
0.80	3.23513	3.23913	3.23852	3.19041	3.13405	3.12308	3.15497	3.16652	3.17656
0.90	2.88472	2.87871	2.86471	2.83797	2.80650	2.80929	2.83142	2.82905	2.83516
0.95	2.64168	2.62779	2.60618	2.58968	2.56308	2.57730	2.58937	2.57956	2.58617
1.000	2.35700	2.33481	2.30983	2.30710	2.27721	2.30580	2.30102	2.28715	2.29478

Notes: The table shows results based on a 2-step estimation procedure with the true model given as:

$$y_{t+1} = \beta_0 + \beta_1(1-L)^d x_t + \varepsilon_{1t+1} \quad (36)$$
$$(1-L)^d x_t = c_2 + \varepsilon_{2t} \quad (37)$$

Here, the CSS estimator is used in the first step to estimate the parameter d. In the second step, y_{t+1} is regressed on the fractional difference of x_t using the estimated value of d obtained in step 1.

Notes for tables 3-7 :

Throughout, the true value of β_1 is equal to 0. Values for β_0 and c_2 are set equal to 0, while the standard deviations of the innovations in equations (36) and (37) above, have been estimated from the exchange rate data for Germany where the forward premium has been fractionally differenced with $d = 0.80$. The resulting values for the standard deviations of ε_{1t} and ε_{2t} are 0.03294 and 0.000942, respectively. To calculate correlated residuals we use the Cholesky factorization of the desired correlation matrix.

Table 4

2-step Procedure Using CSS Estimator where the Original Process is an ARFIMA(1,*d*,0) process

Table 4a

Proportion of Rejections of the True Null Hypothesis β=0

ρ/ϕ	-0.95	-0.90	-0.80	-0.40	0.00	0.40	0.80	0.90	0.95
-0.99	0.0470	0.0460	0.0490	0.0450	0.0457	0.0513	0.0477	0.0503	0.0487
-0.95	0.0447	0.0477	0.0483	0.0493	0.0503	0.0440	0.0443	0.0443	0.0423
-0.80	0.0457	0.0480	0.0493	0.0490	0.0477	0.0450	0.0423	0.0423	0.0440
-0.40	0.0470	0.0493	0.0467	0.0510	0.0500	0.0510	0.0470	0.0453	0.0410
0.00	0.0540	0.0510	0.0497	0.0540	0.0547	0.0523	0.0507	0.0487	0.0497
0.40	0.0550	0.0567	0.0547	0.0523	0.0547	0.0530	0.0560	0.0553	0.0520
0.80	0.0473	0.0490	0.0507	0.0483	0.0503	0.0513	0.0597	0.0563	0.0573
0.95	0.0537	0.0567	0.0570	0.0493	0.0493	0.0517	0.0547	0.0577	0.0590
0.99	0.1387	0.1260	0.1093	0.0763	0.0593	0.0647	0.1153	0.1383	0.1503

Table 4b

Mean Estimated Value of β using two step-procedure

ρ/ϕ	-0.95	-0.90	-0.80	-0.40	0.00	0.40	0.80	0.90	0.95
-0.99	-0.1888	-0.1796	-0.1623	-0.0838	-0.0020	0.0795	0.1611	0.1805	0.1897
-0.95	-0.1712	-0.1646	-0.1503	-0.0850	-0.0131	0.0626	0.1365	0.1546	0.1640
-0.80	-0.1295	-0.1265	-0.1185	-0.0756	-0.0231	0.0344	0.0919	0.1070	0.1153
-0.40	0.0263	0.0195	0.0103	-0.0158	-0.0329	-0.0456	-0.0501	-0.0490	-0.0470
0.00	0.1700	0.1568	0.1336	0.0521	-0.0260	-0.0976	-0.1577	-0.1709	-0.1773
0.40	0.2416	0.2293	0.1993	0.0920	-0.0169	-0.1182	-0.2118	-0.2353	-0.2480
0.80	0.2546	0.2385	0.2113	0.1005	-0.0075	-0.1150	-0.2240	-0.2515	-0.2612
0.95	0.3033	0.2889	0.2588	0.1310	0.0025	-0.1286	-0.2596	-0.2904	-0.3068
0.99	0.4550	0.4307	0.3849	0.2012	0.0100	-0.1855	-0.3739	-0.4236	-0.4497

Table 4c

Variance Estimated Value of β using two step-procedure

ρ/ϕ	-0.95	-0.90	-0.80	-0.40	0.00	0.40	0.80	0.90	0.95
-0.99	0.1549	0.1520	0.1478	0.1319	0.1245	0.1313	0.1504	0.1535	0.1541
-0.95	0.6954	0.6965	0.7001	0.7066	0.7150	0.7087	0.6910	0.6864	0.6862
-0.80	1.2362	1.2419	1.2515	1.2799	1.2975	1.2865	1.2358	1.2221	1.2180
-0.40	2.8195	2.8390	2.8705	2.9546	2.9881	2.9549	2.8472	2.8135	2.7966
0.00	3.4081	3.4209	3.4463	3.5145	3.5635	3.5640	3.4820	3.4536	3.4350
0.40	3.0321	3.0172	3.0064	3.0215	3.0633	3.0794	3.0823	3.0578	3.0590
0.80	1.3795	1.3701	1.3633	1.3592	1.3754	1.4260	1.4281	1.4261	1.4225
0.95	0.8057	0.8017	0.7874	0.7684	0.7694	0.7846	0.8180	0.8258	0.8294
0.99	0.2248	0.2162	0.2089	0.1920	0.1867	0.1933	0.2141	0.2240	0.2304

Table 4d

Mean Estimated Value of d using two step-procedure

ρ/ϕ	-0.95	-0.90	-0.80	-0.40	0.00	0.40	0.80	0.90	0.95
-0.99	0.7926	0.7926	0.7925	0.7920	0.7917	0.7919	0.7925	0.7925	0.7925
-0.95	0.7953	0.7953	0.7953	0.7950	0.7946	0.7947	0.7953	0.7953	0.7953
-0.80	0.7949	0.7950	0.7950	0.7946	0.7941	0.7942	0.7948	0.7950	0.7951
-0.40	0.7967	0.7969	0.7967	0.7965	0.7958	0.7957	0.7971	0.7973	0.7974
0.00	0.7999	0.7998	0.7997	0.7994	0.7986	0.7988	0.8005	0.8008	0.8010
0.40	0.8059	0.8057	0.8068	0.8060	0.8045	0.8055	0.8071	0.8072	0.8075
0.80	0.8122	0.8120	0.8117	0.8127	0.8125	0.8128	0.8133	0.8128	0.8131
0.95	0.8061	0.8060	0.8056	0.8067	0.8069	0.8062	0.8071	0.8070	0.8065
0.99	0.7951	0.7950	0.7949	0.7947	0.7948	0.7943	0.7940	0.7941	0.7944

Notes: The results here are based on a 2-step procedure with the true model given as:

$$y_{t+1} = c_1 + \beta(1\text{-}L)^{0.80}x_t + \varepsilon_{1t+1}, \qquad (1\text{-}\phi L)(1\text{-}L)^{0.80}x_t = c_2 + \varepsilon_{2t}$$

Here, d has been obtained from estimation of an ARFIMA(1,d,0) model using the CSS estimator.

Table 5
2-step Procedure Using Semi-Parametric Estimator
where the Original Process is an ARFIMA(1,0.80,0) process.

Table 5a: Proportion of Rejections in a 5% Nominal Test of the Null Hypothesis that $\beta_1 = 0$

ρ / ϕ	-0.95	-0.90	-0.80	-0.40	0.00	0.40	0.80	0.90	0.95
-0.99	0.0527	0.0497	0.0503	0.0460	0.0513	0.0540	0.0493	0.0453	0.0470
-0.95	0.0493	0.0487	0.0487	0.0557	0.0613	0.0560	0.0480	0.0457	0.0440
-0.80	0.0490	0.0497	0.0493	0.0570	0.0543	0.0560	0.0510	0.0493	0.0487
-0.40	0.0517	0.0560	0.0563	0.0517	0.0537	0.0520	0.0483	0.0487	0.0500
0.00	0.0530	0.0540	0.0587	0.0530	0.0523	0.0497	0.0497	0.0490	0.0470
0.40	0.0513	0.0513	0.0483	0.0560	0.0507	0.0487	0.0463	0.0500	0.0513
0.80	0.0553	0.0523	0.0503	0.0553	0.0557	0.0553	0.0460	0.0473	0.0483
0.95	0.0523	0.0560	0.0567	0.0613	0.0557	0.0557	0.0490	0.0497	0.0513
0.99	0.0587	0.0593	0.0583	0.0573	0.0560	0.0540	0.0590	0.0547	0.0540

Table 5b: Bias of the Estimate of β_1 using the Two-Step Procedure

ρ / ϕ	-0.95	-0.90	-0.80	-0.40	0.00	0.40	0.80	0.90	0.95
-0.99	-0.1597	-0.1510	-0.1338	-0.0684	0.0010	0.0697	0.1332	0.1470	0.1547
-0.95	-0.1407	-0.1343	-0.1201	-0.0615	0.0017	0.0664	0.1224	0.1362	0.1427
-0.80	-0.0999	-0.0957	-0.0847	-0.0407	0.0049	0.0512	0.0889	0.0985	0.1033
-0.40	0.0485	0.0499	0.0533	0.0410	0.0240	0.0077	-0.0200	-0.0286	-0.0323
0.00	0.1875	0.1852	0.1789	0.1107	0.0397	-0.0257	-0.1111	-0.1384	-0.1552
0.40	0.2708	0.2649	0.2492	0.1500	0.0489	-0.0460	-0.1664	-0.2049	-0.2305
0.80	0.1567	0.1505	0.1406	0.0956	0.0432	-0.0045	-0.0728	-0.1004	-0.1177
0.95	0.0863	0.0771	0.0774	0.0576	0.0314	0.0078	-0.0134	-0.0312	-0.0422
0.99	0.0253	0.0261	0.0071	0.0003	0.0045	-0.0120	-0.0072	-0.0120	-0.0035

Table 5c: Variance of the Estimate of β_1 using the Two-Step Procedure

ρ / ϕ	-0.95	-0.90	-0.80	-0.40	0.00	0.40	0.80	0.90	0.95
-0.99	0.1320	0.1272	0.1203	0.1105	0.1077	0.1067	0.1181	0.1175	0.1211
-0.95	0.6792	0.6824	0.6871	0.7076	0.7172	0.7125	0.6824	0.6731	0.6720
-0.80	1.2122	1.2253	1.2433	1.2744	1.2748	1.2630	1.2158	1.1964	1.1896
-0.40	2.7153	2.7253	2.7464	2.7775	2.7069	2.6937	2.7069	2.6913	2.6805
0.00	3.1164	3.1267	3.1536	3.1518	3.0743	3.0546	3.0963	3.0913	3.0767
0.40	2.5361	2.5530	2.5592	2.5335	2.5040	2.4821	2.4817	2.4769	2.4769
0.80	1.4127	1.4001	1.3764	1.4066	1.4280	1.3959	1.3551	1.3227	1.3360
0.95	0.9257	0.9104	0.9057	0.9200	0.9398	0.8992	0.8575	0.8321	0.8376
0.99	0.1224	0.1232	0.1227	0.1320	0.1555	0.1330	0.1225	0.1239	0.1424

Table 5d: Bias of the Estimate of d

ρ / ϕ	-0.95	-0.90	-0.80	-0.40	0.00	0.40	0.80	0.90	0.95
-0.99	0.0228	0.0227	0.0225	0.0239	0.0239	0.0238	0.0234	0.0238	0.0243
-0.95	0.0324	0.0331	0.0326	0.0326	0.0328	0.0333	0.0330	0.0342	0.0340
-0.80	0.0326	0.0332	0.0332	0.0329	0.0334	0.0340	0.0336	0.0346	0.0345
-0.40	0.0309	0.0314	0.0316	0.0319	0.0320	0.0325	0.0322	0.0333	0.0332
0.00	0.0291	0.0297	0.0302	0.0307	0.0301	0.0309	0.0305	0.0321	0.0321
0.40	0.0638	0.0644	0.0649	0.0652	0.0643	0.0656	0.0660	0.0670	0.0673
0.80	0.3109	0.3110	0.3121	0.3129	0.3141	0.3172	0.3161	0.3152	0.3145
0.95	0.4106	0.4116	0.4129	0.4169	0.4192	0.4211	0.4158	0.4141	0.4130
0.99	0.3209	0.3197	0.3187	0.3203	0.3245	0.3251	0.3191	0.3210	0.3225

Notes: The results reported above are based on a 2-step estimation procedure with the true model given as:

$$y_{t+1} = \beta_0 + \beta_1(1-L)^{0.80}x_t + \varepsilon_{1t+1}, \qquad (1-\phi L)(1-L)^{0.80}x_t = c_2 + \varepsilon_{2t}$$

Here we use the log periodogram regression based estimator of Andrews and Guggenberger (2003) to obtain d. We apply a taper equal to $(1-L)^{0.50}x_t$, and set $m = T^{0.75}$.

Table 6
2-step Procedure Using CSS Estimator
where the Original Process is an ARFIMA(1,d,0) process.
A Misspecified ARFIMA(0,d,0) Model is Fit Instead

Table 6a: Proportion of Rejections in a 5% Nominal Test of the Null Hypothesis that $\beta_1 = 0$

ρ / ϕ	-0.95	-0.90	-0.80	-0.40	0.00	0.40	0.80	0.90	0.95
-0.99	0.0580	0.0557	0.0560	0.0497	0.0520	0.0533	0.0543	0.0570	0.0583
-0.95	0.0647	0.0627	0.0583	0.0543	0.0497	0.0570	0.0560	0.0560	0.0573
-0.80	0.0607	0.0603	0.0580	0.0553	0.0520	0.0513	0.0570	0.0547	0.0567
-0.40	0.0617	0.0613	0.0583	0.0523	0.0507	0.0520	0.0463	0.0447	0.0510
0.00	0.0563	0.0540	0.0537	0.0510	0.0540	0.0493	0.0470	0.0503	0.0497
0.40	0.0567	0.0560	0.0500	0.0523	0.0510	0.0523	0.0460	0.0473	0.0470
0.80	0.0583	0.0580	0.0543	0.0583	0.0550	0.0570	0.0500	0.0490	0.0513
0.95	0.0547	0.0577	0.0583	0.0627	0.0610	0.0570	0.0510	0.0480	0.0517
0.99	0.0603	0.0583	0.0587	0.0617	0.0587	0.0550	0.0570	0.0547	0.0557

Table 6b: Bias of the Estimate of β_1 using the Two-Step Procedure

ρ / ϕ	-0.95	-0.90	-0.80	-0.40	0.00	0.40	0.80	0.90	0.95
-0.99	0.04373	0.04141	0.0367	0.01202	-0.0056	-0.01894	-0.04031	-0.04953	-0.05184
-0.95	0.24854	0.23349	0.20518	0.09622	-0.00514	-0.09974	-0.20299	-0.23235	-0.24714
-0.80	0.28688	0.27085	0.24026	0.11969	0.00249	-0.11163	-0.23311	-0.26615	-0.28338
-0.40	0.22077	0.21329	0.19634	0.11879	0.03239	-0.06098	-0.16072	-0.18787	-0.20266
0.00	0.13018	0.13089	0.12778	0.09667	0.05402	0.00088	-0.06445	-0.08462	-0.09681
0.40	0.07516	0.08023	0.08545	0.0865	0.07487	0.04487	-0.00466	-0.0216	-0.03279
0.80	0.03648	0.0377	0.04501	0.05669	0.06702	0.05627	0.02617	0.01226	0.00351
0.95	0.02759	0.02369	0.03183	0.03497	0.04544	0.03257	0.03208	0.02132	0.01359
0.99	0.02199	0.02255	0.0036	-0.00131	0.00409	-0.00956	-0.00429	-0.0052	-0.00011

Table 6c: Variance of the Estimate of β_1 using the Two-Step Procedure

ρ / ϕ	-0.95	-0.90	-0.80	-0.40	0.00	0.40	0.80	0.90	0.95
-0.99	0.24955	0.24892	0.24678	0.25670	0.26589	0.25530	0.24477	0.23383	0.23192
-0.95	1.18432	1.18581	1.19177	1.21323	1.21843	1.19456	1.16523	1.16062	1.16447
-0.80	1.84372	1.85378	1.86892	1.86181	1.83358	1.79712	1.78885	1.79446	1.80431
-0.40	3.10619	3.11243	3.11170	3.04014	2.96201	2.94127	2.99533	3.02120	3.03980
0.00	3.23513	3.23913	3.23852	3.19041	3.13405	3.12308	3.15497	3.16652	3.17656
0.40	2.78884	2.78663	2.77597	2.75323	2.73689	2.73924	2.75371	2.74969	2.75169
0.80	1.59337	1.58225	1.55409	1.57588	1.58700	1.57021	1.56168	1.54124	1.54587
0.95	0.99162	0.97326	0.96443	0.97360	0.96264	0.94104	0.92604	0.90271	0.90023
0.99	0.12944	0.12912	0.12368	0.14531	0.16119	0.13928	0.12705	0.12963	0.15215

Table 6d: Bias of the Estimate of d

ρ / ϕ	-0.95	-0.90	-0.80	-0.40	0.00	0.40	0.80	0.90	0.95
-0.99	-0.83737	-0.83715	-0.83710	-0.83714	-0.83738	-0.83692	-0.83749	-0.83827	-0.83829
-0.95	-0.56793	-0.56763	-0.56764	-0.56722	-0.56697	-0.56616	-0.56645	-0.56698	-0.56739
-0.80	-0.46677	-0.46665	-0.46672	-0.46640	-0.46591	-0.46497	-0.46543	-0.46599	-0.46644
-0.40	-0.22631	-0.22624	-0.22610	-0.22568	-0.22518	-0.22440	-0.22469	-0.22520	-0.22556
0.00	0.00028	0.00022	0.00012	0.00026	0.00072	0.00139	0.00159	0.00146	0.00129
0.40	0.28915	0.28882	0.28846	0.28848	0.28941	0.29045	0.29100	0.29089	0.29062
0.80	0.55457	0.55493	0.55548	0.55815	0.55995	0.56047	0.55742	0.55582	0.55511
0.95	0.53212	0.53245	0.53380	0.53870	0.54085	0.53861	0.53420	0.53161	0.53068
0.99	0.34042	0.34045	0.34213	0.34307	0.34551	0.34709	0.34220	0.34279	0.34307

Notes: The results reported above are based on a 2-step estimation procedure with the true model given as:

$$y_{t+1} = \beta_0 + \beta_1 (1-L)^{0.80} x_t + \varepsilon_{1t+1}, \qquad (1-\phi L)(1-L)^{0.80} x_t = c_2 + \varepsilon_{2t}$$

Note that the true model is an ARFIMA(1,0.80,0), where the values of ϕ and ρ appear under the heading ρ / ϕ. The CSS estimator is used to incorrectly estimate a misspecified ARFIMA(0,d,0) model.

Table 7
2-Step Procedure Using CSS Estimator
where the Original Process is Fractional Noise
An Over-specified ARFIMA(1,*d*,0) Model has been Fit to the Original Process

Table 7a: Proportion of Rejections in a 5% Nominal Test of the Null Hypothesis that $\beta_1 = 0$

ρ /d	-0.95	-0.90	-0.80	-0.40	0.00	0.40	0.80	0.90	0.95
0.40	0.0570	0.0550	0.0547	0.0493	0.0480	0.0473	0.0453	0.0470	0.0490
0.50	0.0543	0.0547	0.0557	0.0510	0.0490	0.0470	0.0480	0.0460	0.0483
0.60	0.0560	0.0563	0.0560	0.0520	0.0490	0.0457	0.0480	0.0480	0.0490
0.70	0.0550	0.0557	0.0573	0.0517	0.0490	0.0477	0.0493	0.0473	0.0500
0.80	0.0540	0.0527	0.0570	0.0503	0.0517	0.0497	0.0460	0.0480	0.0500
0.90	0.0527	0.0537	0.0550	0.0507	0.0537	0.0507	0.0450	0.0463	0.0487
0.95	0.0520	0.0543	0.0547	0.0513	0.0543	0.0497	0.0453	0.0483	0.0490
1.00	0.0520	0.0527	0.0510	0.0533	0.0547	0.0500	0.0510	0.0503	0.0503

Table 7b: Bias of the Estimate of β_1 using the Two-Step Procedure

ρ /d	-0.95	-0.90	-0.80	-0.40	0.00	0.40	0.80	0.90	0.95
0.40	0.1794	0.1776	0.1660	0.1061	0.0395	-0.0357	-0.1149	-0.1375	-0.1522
0.50	0.2094	0.2048	0.1909	0.1203	0.0405	-0.0446	-0.1398	-0.1666	-0.1822
0.60	0.2509	0.2439	0.2245	0.1371	0.0421	-0.0578	-0.1682	-0.2023	-0.2200
0.70	0.2793	0.2721	0.2509	0.1543	0.0442	-0.0670	-0.1919	-0.2265	-0.2471
0.80	0.3080	0.2998	0.2748	0.1668	0.0465	-0.0747	-0.2083	-0.2503	-0.2714
0.90	0.2846	0.2798	0.2613	0.1663	0.0496	-0.0621	-0.1908	-0.2255	-0.2453
0.95	0.2612	0.2579	0.2416	0.1571	0.0505	-0.0511	-0.1676	-0.1990	-0.2166
1.00	0.2348	0.2321	0.2184	0.1387	0.0471	-0.0393	-0.1429	-0.1726	-0.1882

Table 7c: Variance of the Estimate of β_1 using the Two-Step Procedure

ρ /d	-0.95	-0.90	-0.80	-0.40	0.00	0.40	0.80	0.90	0.95
0.40	3.5254	3.5332	3.5409	3.4855	3.4065	3.3631	3.3851	3.4066	3.4310
0.50	3.4957	3.5096	3.5228	3.4570	3.3710	3.3268	3.3600	3.3932	3.4090
0.60	3.4419	3.4584	3.4661	3.4056	3.3153	3.2681	3.2983	3.3268	3.3463
0.70	3.3231	3.3359	3.3467	3.2908	3.2127	3.1555	3.1954	3.2126	3.2273
0.80	3.0637	3.0764	3.0996	3.0939	3.0467	2.9985	2.9837	2.9896	2.9943
0.90	2.7097	2.7204	2.7207	2.7524	2.7305	2.6889	2.6661	2.6550	2.6580
0.95	2.4844	2.4852	2.4759	2.5089	2.4998	2.4694	2.4586	2.4360	2.4377
1.00	2.2266	2.2214	2.2062	2.2458	2.2234	2.2160	2.2029	2.1685	2.1630

Table 7d: Bias of the Estimate of d

ρ /d	-0.95	-0.90	-0.80	-0.40	0.00	0.40	0.80	0.90	0.95
0.40	-0.0068	-0.0072	-0.0076	-0.0078	-0.0077	-0.0079	-0.0067	-0.0065	-0.0068
0.50	-0.0396	-0.0402	-0.0401	-0.0404	-0.0405	-0.0404	-0.0402	-0.0398	-0.0396
0.60	-0.0657	-0.0664	-0.0662	-0.0667	-0.0674	-0.0668	-0.0658	-0.0653	-0.0647
0.70	-0.0869	-0.0871	-0.0868	-0.0869	-0.0864	-0.0858	-0.0854	-0.0858	-0.0859
0.80	-0.0926	-0.0923	-0.0914	-0.0919	-0.0921	-0.0918	-0.0909	-0.0916	-0.0914
0.90	-0.0790	-0.0790	-0.0793	-0.0802	-0.0788	-0.0792	-0.0785	-0.0777	-0.0771
0.95	-0.0762	-0.0768	-0.0770	-0.0777	-0.0763	-0.0771	-0.0753	-0.0753	-0.0742
1.00	-0.0804	-0.0814	-0.0814	-0.0816	-0.0807	-0.0814	-0.0798	-0.0808	-0.0803

Notes: The results reported above are based on a 2-step estimation procedure with the true model given as:

$$y_{t+1} = \beta_0 + \beta_1(1-L)^{0.80}x_t + \varepsilon_{1t+1}, \quad (1-L)^{0.80}x_t = c_2 + \varepsilon_{2t}$$

The true model is an ARFIMA(0,d,0), where the true value of d appears under the heading ρ /d. Here an over-parametrized ARFIMA(1,d,0) model is estimated instead of an ARFIMA (0,d,0) model using the CSS estimator.

Table 8
OLS Estimates from the FRUH Regressions
No Differencing Applied

Sample (1973-2000):		Canada	France	Germ.	Japan	UK
Dependent Variable						
$(s_{t+1} - f_t)$	$\hat{c}1$	-0.0021	-0.0028	0.0019	0.0030	-0.0046
		[0.0087]	[0.1615]	[0.3819]	[0.1921]	[0.0365]
	$\hat{b}1$	-2.1356	-1.8457	-1.7150	-1.0215	-2.4554
		[0.00000]	[0.0007]	[0.0139]	[0.0245]	[0.0001]
P-Value Q (5)		0.6536	0.3263	0.4990	0.6949	0.7448
P-Value Q (10)		0.4505	0.4247	0.3699	0.3927	0.8352
P-Value Q (20)		0.1819	0.7184	0.3379	0.3884	0.1694
P-Value Q^2 (1)		0.4604	0.1538	0.0679	0.2713	0.0002
P-Value Q^2 (5)		0.8043	0.6157	0.4921	0.0323	0.0051
P-Value Q^2 (10)		0.9754	0.8866	0.5621	0.1454	0.0598
Kurtosis		5.0991	3.8087	3.5706	4.5902	4.8989
Skewness		-0.4996	-0.2928	-0.0753	0.4510	-0.1917
Jarque-Bera		72.2844	13.3337	4.6578	44.7048	50.1924
	$\hat{c}2$	-0.0021	-0.0028	0.0019	0.0030	-0.0046
Dependent Variable		[0.0087]	[0.0020]	[0.3819]	[0.1921]	[0.0365]
$\Delta\ s_{t+1}$	$\hat{b}2$	-1.1356	-0.8457	-0.7150	-0.0215	-1.4554
		[0.00000]	[0.0007]	[0.0139]	[0.0245]	[0.0001]
P-Value Q (5)		0.6536	0.3263	0.4990	0.6949	0.7448
P-Value Q (10)		0.4505	0.4247	0.3699	0.3927	0.8352
P-Value Q (20)		0.1819	0.7184	0.3379	0.3884	0.1694
P-Value Q^2 (1)		0.4604	0.1538	0.0679	0.2713	0.0002
P-Value Q^2 (5)		0.8043	0.6157	0.4921	0.0323	0.0051
P-Value Q^2 (10)		0.9754	0.8866	0.5621	0.1454	0.0598
Kurtosis		5.0991	3.8087	3.5706	4.5902	4.8989
Skewness		-0.4996	-0.2928	-0.0753	0.4510	-0.1917
Jarque-Bera		72.2844	13.3337	4.6578	44.7048	50.1924

Notes: The independent variables throughout are a constant and the forward premium. The OLS estimates of the constant and the slope parameter are given by $\hat{c}_i$ and $\hat{b}_i$, respectively where $i = 1, 2$. The quantities appearing in brackets are p-values. When the dependent variable is the change in the spot rate, we use a two-sided test of the null hypothesis that $b_1 = 1$. The remaining p-values are associated with the null hypothesis that the given coefficient is equal to zero.
P-Value Q (j) refers to the p-value associated with the Ljung-Box Q statistic based on the hypothesis that the first j autocorrelations of the estimated residuals are zero. P-Value $Q^2(j)$ refers to the same thing for the squared residuals.

Table 9
OLS Estimates of the FRUH-type Regressions with Fractional Differencing: CSS Estimator used to Estimate *d*

Table 9

Sample (1973-2000):		Canada	France	Germ.	Japan	UK
	$\hat{\beta}_0$	-0.0004	0.0000	-0.0010	0.0004	0.0001
Dependent Variable		[0.5993]	[0.9596]	[0.9803]	[0.5987]	[0.8415]
$(s_{t+1} - f_t)$	$\hat{\beta}_1$	-4.2539	-2.1508	-2.4152	0.0142	-3.3031
		[0.00001]	[0.0104]	[0.2190]	[0.9784]	[0.0038]
	$\hat{d}$	**0.7152**	**0.5837**	**0.9734**	**0.5865**	**0.4362**
		{0.1284}	{0.0637}	{0.1126}	{0.0918}	{0.2070}
P-Value Q (5)		0.7828	0.1229	0.3049	0.2551	0.6338
P-Value Q (10)		0.5104	0.2642	0.2454	0.1628	0.8063
P-Value Q (20)		0.2074	0.5626	0.2086	0.1720	0.1629
P-Value Q^2 (1)		0.1687	0.2085	0.1267	0.2654	0.0016
P-Value Q^2 (5)		0.6178	0.7034	0.5365	0.0157	0.0127
P-Value Q^2 (10)		0.9214	0.9464	0.5943	0.0814	0.0961
Kurtosis		4.4661	3.6490	3.4670	4.4756	4.3616
Skewness		-0.4197	-0.3048	-0.0837	0.3462	-0.1271
Jarque-Bera		38.1733	10.6039	3.2921	35.5359	25.6618

Notes: The independent variables throughout are a constant and the fractional difference of the forward premium. The OLS estimates of the constant and the slope parameter are given by $\hat{\beta}_0$ and $\hat{\beta}_1$, respectively. The quantities appearing in brackets are p-values associated with the hypothesis that the given coefficient is zero. The first-stage estimate of d is obtained via the CSS estimator for an ARFIMA(p,d,q) model. The quantities appearing in braces under the estimates of d are numerical standard errors calculated from the outer product of the numerical gradient vector.

P-Value $Q(j)$ refers to the p-value associated with the Ljung-Box Q statistic based on the hypothesis that the first j autocorrelations of the estimated residuals are zero. P-Value $Q^2(j)$ refers to the same thing for the squared residuals.

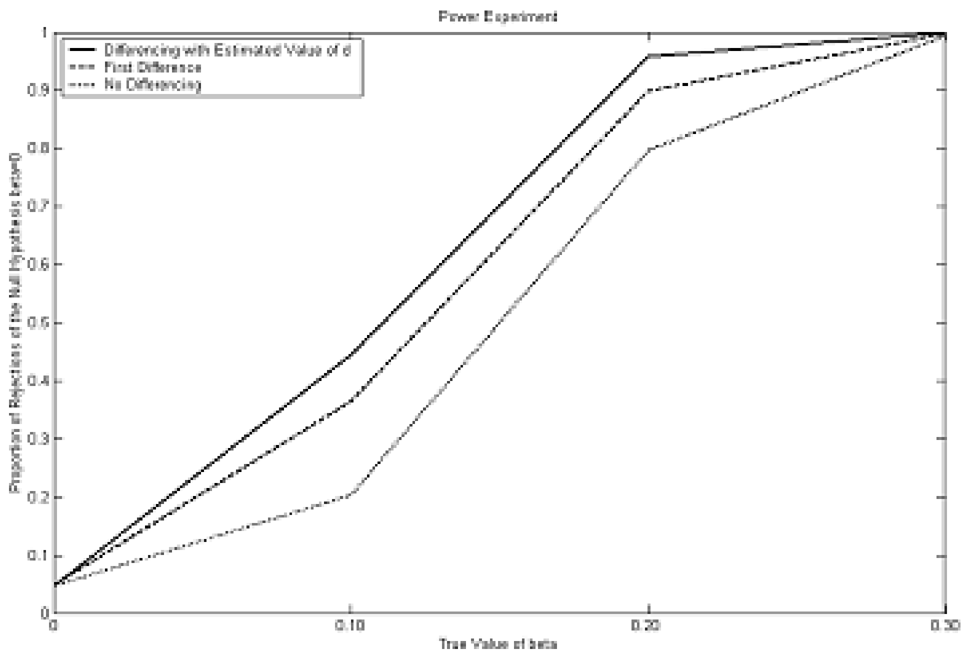

Figure 4: Power to Reject the Null Hypothesis that $\beta_1 = 0$. Dependent Variable is Short Memory with a Long Memory Regressor ($d = 0.40$)