The Long and the Short of it:
Long-Memory Regressors and Predictive Regressions

Alex Maynard
Department of Economics
University of Toronto
150 St. George St. rm N304
Toronto, Ontario
CANADA M5S 3G7
Phone: 416-978-4358
Fax: 416-978-6713
E-mail:
amaynard@chass.utoronto.ca

Aaron Smallwood
Department of Economics
University of Oklahoma
Norman, OK 73072
Phone: 405-325-2643
Fax: 405-325-5842
Email:
Aaron.Smallwood-1@ou.edu

Mark E. Wohar*
Department of Economics
University of Nebraska at Omaha
RH 512K
Omaha, NE 68182-0286
Phone: 402-554-3712
Fax: 402-554-2853
Email:
mwohar@mail.unomaha.edu

April 11, 2005

Abstract

Persistent regressors pose a common problem in predictive regressions. Tests of the forward rate unbiasedness hypothesis (FRUH) constitute a prime example. Standard regression tests that strongly reject the FRUH have been questioned on the grounds of potential long-memory in the forward premium. Researchers have argued that this could create a regression imbalance thereby invalidating standard statistical inference. To address this concern we employ a two-step procedure that rebalances the predictive equation, while still permitting us to impose the FRUH. We derive large sample results and conduct a comprehensive simulation study to validate our procedure. The simulations demonstrate the good small sample performance of our two-stage procedure, and its robustness to possible errors in the first stage estimation of the memory parameter. By contrast, the simulations for standard regression tests show the potential for significant size distortion, validating the concerns of previous researchers. Our empirical application to excess returns suggests less evidence against the FRUH than found using the standard, but possibly questionable, t-tests.

JEL classifications: C22, C12, F31

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

*Corresponding author. We thank Arthur Lewbel and Katsumi Shimotsu for useful discussion. Maynard thanks the SSHRC for research funding. We take full responsibility for any remaining errors.

I. 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 in excess returns regressions, where returns typically exhibit little or no persistence.

The problem of size distortion has led to a large empirical literature as well as the development of econometric techniques designed to address these issues, with much of the literature concentrated in the context of local-to-unity models (Cavanagh et al. 1995). However, persistence can also manifest itself in the form of long-memory. Evidence of long-memory has been documented in several predictive regressors including the forward premium (Baillie and Bollerslev, 1994, Maynard and Phillips, 2001, Maynard, 2003), volatilities (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 considerations suggest 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. In other words, both the left and right hand side variables are stationary, despite the fact that the right hand side variable has a root near one. Thus, while size distortion is of central importance in predictive regressions with near unit roots, problems of regression imbalance may arguably be avoided. This is no longer the case when predictive regressors have long-memory, since imbalance may exist when a short-memory dependent 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, while still allowing for rich short-memory dynamics. Then, in the second stage, the predictive regression is rebalanced by fractionally differencing the regressor. Although this alters the definition of the regression coefficient under the alternative hypothesis, it maintains the same interpretation under the null, 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 unusually robust to estimation error or even modest mis-specification in the first stage. This is due to the fact that a fairly high degree of residual long-memory must be present in order to produce significant size distortion. We see this as an important practical benefit, since the memory parameter can be difficult to estimate in small samples (see Nielsen and Frederiksen, 2004, for a survey).

As an empirical application, 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. Results from this regression, which indicate a strong and rather counterintuitive rejection of the FRUH, have been called into question on account of this long-memory behavior (Baillie and Bollerslev, 2000, and Maynard and Phillips, 2001).

Employing our two stage rebalancing procedure, we regress the excess currency return on the fractionally differenced forward premium. We test the FRUH based on a zero restriction in the rebalanced regression, thus avoiding the potential for size distortion. Our tests yield less evidence against the FRUH than do standard tests. In particular, we find that the standard t-statistic in an OLS regression of excess returns on the forward premium is large in absolute value and significantly negative for every country in our sample. In contrast, when our two step procedure is employed, we fail to reject the FRUH for two of the five cases in our sample, and in every case, the conventional p-values associated with the hypothesis of unbiasedness increases relative to the case where no fractional differencing is applied.

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 with ideas for future research. An appendix contains the proofs affiliated with the asymptotic properties of our two-step procedure.

2. Background

Perhaps the most puzzling set of predictive regression results come from tests of the forward rate unbiasedness hypothesis (FRUH). These empirical results 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 forward rate 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 $b_2= b_1+1=1$, and satisfies $b_1=0$ 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 specifications 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, mis-predicting not only the magnitude of the exchange rate movement, but even the direction of change.

Recent literature has questioned the validity of the inference on the basis of persistence in the forward premium, suggesting that this may induce bias and size distortion. Similar issues arise in other predictability regressions, such as those involving the regression of stock returns on interest rates or dividend and/or earnings price ratios. Proposed corrections have generally been undertaken employing an autoregressive or near unit root model for the regressor.[1] However, evidence exists

[1] For corrections to tests of the FRUH, see Newbold et. al (1998), Bekeart and Hodrick (2001), and Liu and Maynard (2005). For corrections in the stock return predictability literature, see Stambaugh (1999), Rapach and Wohar (2004), Torous, Volkanov, and Yan, (2005), and the references within.

that suggests many predictive regressors, such as the forward premium (Baillie and Bollerslev, 1994), may be subject to long memory rather than near unit root behavior. The statistical properties of long-memory and near-unit root processes are different, and it is unclear to what extent corrective procedures based on autoregressive or local-to-unity assumptions carry over to a long-memory context. It has also been argued that fractionally integrated regressors may cause inference problems in predictive regressions qualitatively similar to those found in regressions with near unit roots. For example, the long memory behavior of the forward premium has been suggested as a possible resolution of the forward discount anomaly (Baillie and Bollerslev, 2000, Maynard and Phillips 2001). To date, few studies have specifically attempted to design tests of predictability that account for the type of regression imbalance considered here. In this paper, we provide a valid test for predictability in the context of long-memory.

We also address a second important issue that arises with long-memory regressors. When regressors display long-memory, predictive regressions, such as (2), may suffer from a statistical imbalance since the return variables on the LHS are generally short-memory. For example, under the FRUH, the forward rate 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 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 conceivably be derived. However, the apparent imbalance in (2) can cause fundamental problems under the alternative hypothesis as characterized by this regression specification. In fact, if the order of integration of the RHS variable exceeds 0.5, then the regression attempts to explain a stationary dependent variable with a nonstationary regressor. Since the RHS has a tendency to wander off, whereas the LHS variable does not, Maynard and Phillips (2001) argue that b_2=0 is the only possible parameter value consistent with this statistical unbalance and show that the OLS estimate of b_2 converges to zero, but has a nonstandard distribution.[2] In other words, there can be no linear relationship between short and long-memory (with d>0.5) variables.

On the basis of the above discussion, one might be tempted to declare victory without further tests on the grounds that the imbalance in (2) implies b_1=0. However, it is crucial at this point to recall that our ultimate interest is 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. This imbalance thus calls for a test that not only maintains correct size, but also allows for realistic and reasonable alternative specifications.

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

[2] Maynard and Phillips (2001) focus on the imbalance in (3), whereas as we focus on (2). Only an imbalance in the latter equation is compatible with the FRUH. Thus (2) seems a more natural choice.

been proposed.[3] We contribute to this literature by providing a simple intuitive two-step predictability test in the presence of long-memory regressors, which remains valid under the null hypothesis and sensible under the alternative hypothesis.

3. Econometric Methodology

Our two step procedure is intended to rebalance predictive regressions that 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 of d. 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 the t-statistic from the rebalanced regression achieves a standard normal asymptotic distribution. In section 3.2, we discuss the estimation alternatives for d that are available in the first stage.

3.1 Two-Stage Test procedure

We model x_t as a Type I fractionally integrated process

$$x_t = (1-L)^{-d}\,(u_{2,t} 1_{\{t>0\}}) \tag{4}$$

where $1_{\{t>0\}}$ is an indicator function and in which $u_{2,t}$ is a general linear process

$$u_{2,t} = C_2(L)\varepsilon_t = \sum_{j=0}^{\infty} C_{2j}\varepsilon_{t-j} \text{ where} \tag{5}$$

$$\varepsilon_t = \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \sim \text{i.i.d}(0,\Sigma) \text{ and } \sum_{\text{j}=0}^{\infty} j^2 \left\| C_{2j} \right\|^2 < \infty. \tag{6}$$

[3] Maynard (2004) employs nonparametric sign tests that remain valid under long-memory assumptions, but apart from this paper, we know of no other work that attempts to deal with this issue.

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 results when $C_2(L)=[0\ \theta(L)/\phi(L)]$, where $\theta(L)$ and $\phi(L)$ 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.

Next, we model y_t as a linear function of the fractionally differenced predictor

$$y_{t+1} = \beta_0 + \beta_1 (1-L)^d x_t + \varepsilon_{1,t+1} \tag{7}$$

and consider tests of the hypothesis $H_0 : \beta_1 = 0$. Note, for example, that the FRUH holds (i.e. $E_t s_{t+1} - f_t = 0$) if $\beta_1 = 0$ but is violated for $\beta_1 \neq 0$, which implies a predictable excess return.

We propose a two-stage estimation procedure. The first stage consists of obtaining a consistent estimate ($\hat{d}$) for d, with convergence rate T^{α}, where $\alpha > 1/4$. Using the estimate $\hat{d}$, 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 (7) 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.

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

$$B_T = \left(T^{-1} \sum_{t=1}^{T-1} \underline{\hat{u}}_{2,t}^2 \right)^{-1} T^{-1/2} \sum_{t=1}^{T-1} \left(\underline{u}_{2t} - \underline{\hat{u}}_{2,t} \right) \underline{\hat{u}}_{2,t} = O_p\left(T^{1/2-\alpha}\right),$$

$$\underline{u}_{2,s} = u_{2,s} - T^{-1} \sum_{t=1}^{T-1} u_{2,t} \text{ and } \hat{u}_{2,t} = (1-L)^{\hat{d}} x_t.$$

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 addition 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. Similar results hold for two-stage estimators in micro-econometrics (Newey and McFadden 1994, Section 6).

The asymptotic properties of the second stage t-statistic for predictability test (β_1=0) are established in corollary 1.2, which shows that the test statistic is standard normal under the null (β_1=0) and diverges at rate $T^{1/2}$ under the alternative ($\beta_1 \neq 0$). It thus provides a solid basis for testing this hypothesis. First, corollary 1.1 shows that the residual variance is estimated consistently.

Corollary 1.1 $\hat{\sigma}^2 = T^{-1}\sum_{t=1}^{T-1}\hat{\varepsilon}_{1t+1}^2 \rightarrow_p \Sigma_{11}$, where $\hat{\varepsilon}_{1t+1} = y_{t+1} - \hat{\beta}_0 - \hat{\beta}_1\hat{u}_{2,t}$.

Corollary 1.2 (a) If the null hypothesis $\beta_1 = 0$ holds, then $t \rightarrow_p N(0,1)$. (b) If the alternative $\beta_1 \neq 0$ holds then $T^{-1/2}t \rightarrow_p \Sigma_{11}{}^{-1/2}\,\mathrm{var}[u_{2,t}]^{1/2}\beta_1$, with $t = \hat{\sigma}^{-1}\left(T^{-1}\sum_{t=1}^{T-1}\underline{\hat{u}}^2{}_{2,t}\right)^{1/2} T^{1/2}\hat{\beta}_1$.

The results of the theorem demonstrate that the two-stage procedure results in a consistent estimate of β_1 for all possible values of this parameter. Further, if the predictability test can be formulated in terms of a zero restriction on β_1, then the asymptotic distribution of the standard t-test is standard normal, and thus achieves the same asymptotic distribution had d been known. In section 4, we demonstrate the robust small sample properties of the two-stage estimator advocated here. Before doing so, we briefly discuss the choice of the estimator for d.

3.2 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 estimation techniques exist for the first stage calculation of d, which range from parametric, to semi-parametric, and wavelet based estimators. Parametric estimators include exact maximum likelihood estimators (see Sowell, 1992) and approximations of the exact MLE in both the time domain (constrained sum of squares, CSS, also known as approximate MLE or AMLE) and frequency domain (Fox and Taqqu, 1986). The properties of the exact MLE estimator and Fox and Taqqu's estimator are identical. For $-½<d<½$, these estimators converge at rate $T^{1/2}$ and are asymptotically normal. For our purpose, the frequency based approximation to MLE and MLE have an established probability law only for stationary and invertible processes. The typical response is to difference the data when it is believed $d>½$ and to add 1 to the resulting estimate.

In the time domain, the CSS estimator has become popular because of its relative simplicity and robustness to non-stationary processes. The CSS estimates are the set of parameters that maximize the approximate maximum likelihood function, ψ, which is given as follows,

$$\begin{aligned} &\psi(\mu,\phi',\theta',d,\sigma^2) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=2}^{T} a_t^{\,2}, \\ &a_t = \frac{\phi(L)}{\theta(L)}(1-L)^d(x_t-\mu), \end{aligned} \tag{11}$$

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. 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 $d> -½$. 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, 2004), 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 mis-specification have been developed. These estimators include log periodogram regression (LPR) based estimators (Geweke and Porter-Hudak, 1983, Robinson, 1995, and Andrews and Guggenburger, 2003), and Whittle type estimators including the local Whittle estimator (Robinson, 1995), and the exact Whittle estimator (Phillips, 1999, and Shimotsu and Phillips, 2002). The exact Whittle estimator is an attractive alternative, as Shimotsu and Phillips show that it is asymptotically normal for any value of d.

The attractiveness of the LPR based estimators lie in their incredible simplicity. These estimators are based on the properties of the spectral density function of a long memory fractional process. In particular, the log of the spectral density function for a long memory fractional process, including an ARFIMA process, satisfies

$$\log f(\omega) \sim \log[g(\omega)] - 2d\log(\omega) \tag{12}$$

where "~" denotes asymptotic equivalence as $\omega \rightarrow 0$, and $g(\omega)$ is an even function that is continuous at zero and finite. An estimator of d can easily be obtained using the sample analogues of the quantities in (12), where the periodogram at the first m Fourier frequencies replaces the spectral density function. A value for the bandwidth parameter, m, must be chosen, and in what follows below, we use $m=T^{0.50}$ and $m=T^{0.65}$, which are typical choices in the literature (see Nielsen and Frederiksen, 2004). The remaining issue concerns the first term, $log[g(\omega)]$, in (12). The original LPR based estimator of Geweke and Porter-Hudak (GPH, 1983) replaces this quantity with a constant. Then, the log periodogram at the first m frequencies is regressed on a constant and $-2log(\omega_j)$, $j=1,\ldots,m$. However, the approximation of $log[g(\omega)]$ may not be innocuous and may in fact lead to a sizeable small sample bias as shown by Agiakloglou, Newbold, and Wohar (1992). Andrews and Guggenberger (2003) have suggested that a decrease in the small sample bias can be accomplished by approximating the term first term in (12) with a constant and the polynomial $\sum_{r=0}^{R} \omega_j^{2r}$, $j = 1,\ldots,m$. Recently, Nielsen and Frederiksen (2004) have shown the use of the biased reduced LPR (BRLPR) estimator of Andrews and Guggenberger (2003) does indeed result in substantial mitigation of 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, (2004) we also set R=1 throughout.[4]

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. 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 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, even when the model is mis-specified. 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

$$\begin{aligned} y_{t+1} &= \beta_0 + \beta_1(1-L)^d x_t + \varepsilon_{1t+1} \\ x_t &= (1-L)^{-d} u_{2t},\ u_{2t} = (1-\phi L)^{-1} \varepsilon_{2t}\ , \\ \varepsilon_t &= \left(\varepsilon_{1t}, \quad \varepsilon_{2t}\right)' \sim i.i.d.N(0,\Sigma) \end{aligned} \qquad (13)$$

[4] Andrews and Guggenberger (2003) have shown that their estimator is consistent and asymptotically normal for -1/2<*d*<1/2. Although it appears likely that it remains consistent for *d<1*, given the potential for non-stationarity, we follow Sun and Phillips (2003), and apply the linear filter $(1\text{-}L)^{0.50}$ prior to using the BRLPR estimation technique. The final estimate of *d* is then given by the BRLPR estimate plus 0.50.

where $|\phi| < 1$, and Σ is a positive definite matrix with potentially non-zero diagonal elements. To generate data, we first draw the residual vector ε_t, whose elements are correlated with correlation coefficient equal to $\rho = \Sigma_{12} / \sqrt{\Sigma_{11}\Sigma_{22}}$. The sequence u_{2t} is then generated using the AR structure given above. Finally, the variable x_t is created by multiplying the disturbance sequence by the Cholesky factorization of the Toeplitz matrix of exact autocovariances for the desired value of d.[5] For the majority of the simulation experiments, 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. A stylized fact regarding exchange rate dynamics and the FRUH regressions is the excessive volatility of the dependent variables relative to the forward premium (Baillie and Bollerslev, 2000). Therefore, we set the standard deviations of ε_{1t} and ε_{2t} equal to 0.0329 and 0.0009 respectively.[6] In this paper, one of our major concerns is with accessing the size of the empirical tests, and in this regard, the tests are independent of the selection of the relative magnitudes of the standard deviations of the two innovation series. Finally, we allow the correlation coefficient across the residuals to vary from -0.95 to 0.95.

Tables 1-2 motivate the problem by demonstrating the size distortion that results when long-memory regressors are included in predictive regressions. They show simulation results under the null hypothesis (b_1=0) for a standard predictability regression, using the fitted model $\hat{y}_{t+1} = \hat{b}_0 + \hat{b}_1 x_t$, in which our two-step procedure is not applied. In our empirical application this would correspond to the traditional tests of FRUH when the forward premium displays long-memory. Our objective here is to observe the consequences of not adequately accounting for long

[5] See Sowell (1992) for details about the autocovariances of an ARFIMA process. To create a non-stationary series, we first create a series with a differencing parameter equal to d-1, and then integrate this series.
[6] These values result from the German data for the FRUH regressions discussed below. The values used for standard deviations of ε_{1t} and ε_{2t} are the standard deviations of the log of excess returns and the fractionally differenced log of the 1 month forward premium with a differencing parameter equal to 0.80 respectively.

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 yield the integration order (d) of the regressor. Across the top of each Table, we report the correlation coefficients (ρ) between the simulated residuals.

Table 1a shows rejection rates under the null hypothesis for a standard predictability test when the regressor is I(d). The test retains the correct size when the regressor is stationary (regardless of correlation), but 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 30% in a nominal 5% test. These results are similar to those of Mankiw and Shapiro (1986), who analyze size distortion in predictive regression with near unit root regressors. Tables 1b and 1c contain the simulated biases and 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 infinite variances, for most of the processes considered in these tables.

We next consider the effects of adding short run dynamics to the system in Table 2. In this case, we allow x_t to follow various ARFIMA(1,0.40,0) specifications. The results are similar to those reported in Table 1, although it is interesting to note that the inclusion of short term dynamics can influence predictive tests. In particular, every process considered in Table 2 is stationary. Nonetheless, rejection rates of the true null of no predictability can be as high as 26% in a nominal 5% test. Again, we reach the same conclusion as above 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).[7] 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 ϕ=0.99, we see that the true hypothesis that β=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 set of tables is dramatically reduced by rebalancing, and we no longer observe a declining variance as x_t becomes more persistent. 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. 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, but fix d to be equal to 0.80. The last panel of the table documents the exceptional performance of the BRLPR estimator. 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 (see equation 13). 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

[7] The CSS estimator of d is remarkably accurate when there are no ARMA components, and thus we omit the mean estimated values of d from Tables 3. These results are available upon request.

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 long-memory model in the first stage is misspecified. 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. Table 7 considers the opposite scenario, in which the true process is fractional noise, but an over parameterized ARFIMA(1,*d*,0) model is estimated. In Table 6, we fix *d* equal to 0.80 and consider the complications that result as ϕ increases 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, the differencing parameter 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. In Table 7, we we allow the value of *d* to vary from 0.40 to 1.00, and analyze the effects from fitting an ARFIMA(1,*d*,0) model when the regressor is actually fractional noise. It is clear, from the last panel of the table, that *d* is underestimated, as the algorithm will routinely select large autoregressive parameters rather than the correct value of *d*. However, the bias is reasonable, resulting in quite an accurate second stage test, with approximately an empirical size of 5%.

We close this section by commenting on the implications of misspecification. Under the null that β_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 and 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 (13), we allowed β_1 to vary from 0 to 1 with a step size of 0.10, and tested the hypothesis that β_1=0. For brevity, we chose a value of ρ=0.80 and set the standard deviations of both disturbance sequences equal to unity. We allowed the regressor, x_t, to follow 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 a 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.[8] 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 β_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 beat by the application of our two-step procedure for every value of β_1.[9]

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 mis-specification.

[8] 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.
[9] 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.

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 procedure to the FRUH.

5. Empirical Application Results

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:[10]

$$s_{t+1} - f_t = \beta_0 + \beta_1 (1-L)^d (f_t - s_t) + \varepsilon_{1t+1}. \tag{14}$$

If excess returns are I(0), as both intuition and empirical evidence suggest, then the regression in (14) 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, and are the same data as employed in Liu and Maynard (2005). We use the one month forward and spot US dollar price of the foreign currency, where the data are collected on the last day of each month. See Liu and Maynard (2005) for precise details.

As a benchmark, Table 8 yields regression results for the standard FRUH equations shown in equations (2) and (3). Under unbiasedness, we expect b_1=1 and b_2=0. Using the probability values for these hypotheses, we encounter a strong rejection of the unbiasedness hypothesis. In

[10] Baillie and Bollerslev (2000) also analyze the effects of long memory in the conditional variance. GARCH effects were less pronounced in our monthly data than they appear to be in daily data, and our results were not affected by the inclusion of a GARCH in mean term in (14). We thus concentrate on the simple form of equation (14), but see the inclusion of GARCH effects as an interesting extension to our analysis.

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.

We next consider the possibility that the rejection of the FRUH results from long memory in the forward premium by rebalancing the FRUH regression with excess returns as the dependent variable, using two different estimation techniques. Table 9 presents our results using the CSS estimator. It is interesting to note that our findings are very much in line with previous authors (Baillie and Bollerslev, 1994, and Maynard and Phillips, 2001) in that we find significant evidence of long memory dynamics in the forward premium. Using the numerical standard errors as our guide, we see that 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 (14). First we note 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 exceed the same values in Table 8. We continue to reject the hypothesis that β_2=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 at least two countries in our sample, and we conclude that when our two-step procedure is implemented, substantially less evidence against unbiasedness is uncovered.

Table 10 contains results for our two-step procedure using the BRLPR estimator of Andrews and Guggenberger (2003). The first panel contains results using m=$T^{0.50}$, with the latter panel presenting our results with m=$T^{0.65}$. While the estimated value of d can be erratic, which is not surprising given the large variance of the BRLPR estimator, we reach the exact same conclusions as we did in Table 9 regarding the unbiasedness hypothesis. Again, in three cases we reject unbiasedness at the 1% level with somewhat larger p-values than are recorded in Table 8, while we fail to reject unbiasedness for Germany and Japan for any conventionally sized test.

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 may be of concern in the near unit root case (Maynard and Shimotsu, 2004), 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 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 a consistent estimator of the long-memory parameter. This is then used to rebalance the second stage predictive regression by fractionally differencing the regressor. A full set of asymptotic results are provided. In particular, the t-statistic in the second stage regression 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 FRUH. We find that the forward premium is typically subject to long memory, while the standard regressands in the FRUH regressions appear to be I(0). We demonstrate that the use of our technique is able to reverse a strong rejection of unbiasedness for two of the five countries in our sample.

Appendix

Lemma 1. 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

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

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

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

Lemma 2 Using the same definitions in the statement of Theorem 1and Lemma 1, the following convergence rates apply

a) $T^{1/2}\sum_{t=1}^{T-1}\underline{\tilde{u}}_{2,t}\varepsilon_{1,t+1} = O_p(1)$, **b)** $T^1\sum_{t=1}^{T-1}\underline{\tilde{u}}_{2,t}\underline{u}_{2,t} = O_p(1)$, **c)** $T^1\sum_{t=1}^{T-1}\underline{\tilde{u}}_{2,t}^{\,2} = O_p(1)$,

d) $T^1\sum_{t=1}^{T-1}\underline{u}^*_{2,t,T}\varepsilon_{1,t+1} = O_p(ln(T)T^{\bar{\delta}})$ **e)** $T^1\sum_{t=1}^{T-1}\underline{u}^*_{2,t,T}\underline{u}_{2,t} = O_p(ln(T)T^{\bar{\delta}})$

f) $T^1\sum_{t=1}^{T-1}\underline{u}^*_{2,t,T}\underline{\tilde{u}}_{2,t} = O_p(ln(T)T^{\bar{\delta}})$ **g)** $T^1\sum_{t=1}^{T-1}\underline{u}^{*\,2}_{2,t,T} = O_p(ln(T)^2T^{2\bar{\delta}})$.

Proofs

Proof of Lemma 1 (A.1) follows by (6) and the series expansion $ln(x) = \sum_{j=1}^{\infty}(-1)^{j-1}\frac{(x-1)^j}{j}$:

$$\tilde{u}_{2,t} = ln(1\text{-L})u_{2,t}1_{\{t>0\}} = -\sum_{j=1}^{t-1}\frac{1}{j}L^j u_{2,t} = -\sum_{k=0}^{\infty}C_{2k}\sum_{r=k+1}^{k+t-1}\left(\frac{1}{r-k}\right)\varepsilon_{t-r},$$

$$max_{t\leq T}E\,\tilde{u}_{2,t}^2 = max_{t\leq T}\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}C_{2k}C_{2j}\sum_{r=k+1}^{k+t-1}\sum_{s=j+1}^{j+t-1}\left(\frac{1}{r-k}\right)\left(\frac{1}{s-k}\right)E[\varepsilon_{t-r}\varepsilon'_{t-s}].$$

$$\leq \Sigma\left(\sum_{k=0}^{\infty}C_{2k}\right)^2\left(\sum_{r=1}^{\infty}\left(\frac{1}{r-k}\right)^2\right) < \infty.$$

Since $\tilde{u}_{2,t}$=0, $t\leq 0$ and $\sum_{j=1}^{T-1}\frac{1}{j} = O(ln(T))$ (Gradshtein and Ryzhik, 1994, eqn. 0.131),

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

showing (A.2). Let $\psi_{\delta^*_{T,j}}$ and $\psi_{\bar{\delta}_j}$, j=0,1,2,… denote the MA(∞) coefficients in the expansion of

$(1-L)^{-\delta^*_T}$ and $(1-L)^{-\bar{\delta}_T}$, respectively. Noting that $\tilde{u}_{2,t} = 0$, $t \le 0$, $|\psi_{\delta^*_{T,j}}| \le |\psi_{\bar{\delta}_j}|$, 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

$$max_{t\le T}\, E\,(u^*_{2,t,\mathrm{T}})^2 = max_{t\le T}\; E\left|\sum_{j=0}^{t-1}\sum_{k=0}^{t-1} \psi_{\delta^*_{T,j}} \tilde{\tilde{u}}_{2,t-j} \psi_{\delta^*_{T,k}} \tilde{\tilde{u}}_{2,t-k}\right| \le$$

$$max_{t\le T} \sum_{j=0}^{t-1}\sum_{k=0}^{t-1} \left[E\left(|\psi_{\delta^*_{T,j}}|^2 |\tilde{\tilde{u}}_{2,t-j}|^2\right)\right]^{1/2} \left[E\left(|\psi_{\delta^*_{T,k}}|^2 |\tilde{\tilde{u}}_{2,t-k}|^2\right)\right]^{1/2}$$

$$\le max_{t\le T} \sum_{j=0}^{t-1}\sum_{k=0}^{t-1} \left[E\left(|\psi_{\bar{\delta}_{T,j}}|^2 |\tilde{\tilde{u}}_{2,t-j}|^2\right)\right]^{1/2} \left[E\left(|\psi_{\bar{\delta}_{T,k}}|^2 |\tilde{\tilde{u}}_{2,t-k}|^2\right)\right]^{1/2}$$

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

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 2 First, (c) follows since

$$T^1 \sum_{t=1}^{T-1} \underline{\tilde{u}}^2_{2,t} = T^{-1}\sum_{t=1}^{T-1} \tilde{u}^2_{2,t} - \left(T^{-1}\sum_{t=1}^{T-1} \tilde{u}_{2,t}\right)^2 \text{ and}$$

$$E\left(T^{-1}\sum_{t=1}^{T-1} \tilde{u}_{2,t}\right)^2 \le T^{-1}\sum_{t=1}^{T-1} E\tilde{u}^2_{2,t} \le max_{t\le T}\; E\tilde{u}^2_{2,t} < \infty. \tag{A.4}$$

Employing the Cauchy Schwartz inequality, similar argument shows (b). For (a) write

$$T^{-1/2}\sum_{t=1}^{T-1} \underline{\tilde{u}}_{2,t}\varepsilon_{1,t+1} = T^{-1/2}\sum_{t=1}^{T-1} \tilde{u}_{2,t}\varepsilon_{1,t+1} - \left(T^{-1}\sum_{t=1}^{T-1}\tilde{u}_{2,t}\right) T^{-1/2}\sum_{t=1}^{T}\varepsilon_{1,t+1}. \tag{A.5}$$

The second term on the RHS is $O_p(1)$ by (A.4) and application of the standard CLT. For the first term, since $\tilde{u}_{2,t}$ is predetermined, by the Law of Iterative Expectations,

$$E\left[T^{-1/2}\sum_{t=1}^{T-1}\tilde{u}_{2,t}\varepsilon_{1,t+1}\right]^2 = T^{-1}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1} E[\tilde{u}_{2,t}\tilde{u}_{2,s}\varepsilon_{1,t+1}\varepsilon_{1,s+1}] = T^{-1}\sum_{t=1}^{T-1} E[\tilde{u}^2_{2,t}\varepsilon^2_{1,t+1}]$$

$$\le max_{t\le T}\; E[\tilde{u}^2_{2,t}]\Sigma_{11} < \infty.$$

(g) follows by similar argument as (c) since

$$E\left[\left(T^{-1}\sum_{t=1}^{T-1}u^{*}{}_{2,t,T}\right)^{2}\right]\le E\left[T^{-1}\sum_{t=1}^{T-1}(u^{*}{}_{2,t,T})^{2}\right]\le\ max_{t\le T}\ E[(u^{*}{}_{2,t,T})^{2}]=\mathrm{O}(ln(T)^{2}T^{2\bar{\delta}}).\quad\text{(A.6)}$$

Then (d) follows since

$$E\left|T^{-1}\sum_{t=1}^{T-1}\underline{u}^{*}{}_{2,t,T}\varepsilon_{1,t+1}\right|\le max_{t\le T}\ \left(E\,|\,\underline{u}^{*}{}_{2,t,T}\,|^{2}\right)^{1/2}\left(E\,|\,\varepsilon_{1,t+1}\,|^{2}\right)^{1/2}=O_{p}(\ln(T)\,T^{\bar{\delta}}).$$

Parts (e) and (f) follow by the same argument.

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

$$T^{\alpha}\hat{\delta}_{T}=-T^{\alpha}(\hat{d}-d)=O_{p}(1).\qquad\text{(A.7)}$$

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

$$\sqrt{T}(\hat{\beta}_1-\beta_1)=\left(T^{-1}\sum_{t=1}^{T-1}\underline{\hat{u}}^{2}{}_{2,t}\right)^{-1}\left(T^{-1/2}\sum_{t=1}^{T-1}\underline{\hat{u}}^{2}{}_{2,t}\varepsilon_{1,t+1}+\beta_1T^{-1/2}\sum_{t=1}^{T-1}(\underline{u}_{2,t}-\underline{\hat{u}}_{2,t})\underline{\hat{u}}_{2,t}\right).\qquad\text{(A.8)}$$

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 $\hat{d}\rightarrow_p\ d$, for large T, $P(I_{\bar{\delta}}=0)=P(|\,\hat{\delta}_T\,|>\bar{\delta})<\eta$. Thus, $I_{\bar{\delta}}\rightarrow_p 1$ 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 $P(I_{\bar{\delta}}=1)\rightarrow 1$. 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 (Rudin, 1976, p.110, Thm 5.15) to the function $(1-L)^{\hat{\delta}_T}$ with argument $\hat{\delta}_T$ about zero and where $\delta^{*}{}_T$ lies between 0 and $\hat{\delta}_T$ gives

$$(1-L)^{\hat{\delta}_T}=1+\hat{\delta}_T\ln(1-L)+\frac{1}{2}\hat{\delta}^{2}{}_T\ln(1-L)^{2}(1-L)^{\delta^{*}{}_T}\ \text{and}$$

$$\hat{u}_{2,t}=(1-L)^{\hat{\delta}_T}u_{2,t}=u_{2,t}+\hat{\delta}_T\tilde{u}_{2,t}+\frac{1}{2}\hat{\delta}^{2}{}_Tu^{*}{}_{2,t,T}$$

where $u^{*}{}_{2,t,T}$, $\tilde{\tilde{u}}_{2,t}$, and $\tilde{u}_{2,t}$ are defined in Lemma 1.

Next, we turn to the first term in the numerator of $\sqrt{T}(\hat{\beta}_1 - \beta_1)$ in (A.8). We have

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

by standard arguments since by Lemma 2 parts a and d

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

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

The behavior of the second term in the numerator of $\sqrt{T}(\hat{\beta}_1 - \beta_1)$ in (A.8) is given by

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

$$= \beta_1 T^{-1/2}\hat{\delta}_T\sum_{t=1}^{T-1}\underline{\tilde{u}}_{2,t}\underline{u}_{2,t} + \beta_1 R_{2,T} = \beta_1 O_p(T^{1/2-\alpha}),$$

where $R_{2,T}$ is defined as,

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

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

For the denominator of $\sqrt{T}(\hat{\beta}_1 - \beta_1)$ in equation (A.8) we have

$$T^{-1}\sum_{t=1}^{T-1}\underline{\hat{u}}^2{}_{2,t} = T^{-1}\sum_{t=1}^{T-1}\underline{u}^2{}_{2,t} + R_{3,T} \rightarrow_p \mathrm{var}[u_{2,t}] \qquad \text{(A.9)}$$

by standard argument since by Lemma 2,

$$R_{3,T} = \hat{\delta}^2{}_T T^{-1}\sum_{t=1}^{T-1}\underline{\tilde{u}}^2{}_{2,t} + \frac{1}{4}\hat{\delta}^4{}_T T^{-1}\sum_{t=1}^{T-1}\underline{u}^*{}_{2,t,T} + 2\hat{\delta}_T T^{-1}\sum_{t=1}^{T-1}\underline{\tilde{u}}_{2,t}\underline{u}_{2,t} + \hat{\delta}^2{}_T T^{-1}\sum_{t=1}^{T-1}\underline{u}^*{}_{2,t,T}\underline{u}_{2,t}$$

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

$$+ O_p\left(T^{\bar{\delta}-2\alpha}\ln(T)\right) + O_p\left(T^{\bar{\delta}-3\alpha}\ln(T)\right) = o_p(1),$$

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

Proof of Corollary 1.1 $\underline{y}_{t+1} = \hat{\beta}_1 \hat{\underline{u}}_{2,t} + \hat{\varepsilon}_{1t+1}$ and $\hat{\underline{u}}_{2,t} = \underline{u}_{2,t} + \hat{\delta}_T \tilde{\underline{u}}_{2,t} + \frac{1}{2}\hat{\delta}^2{}_T \underline{u}^*{}_{2,t,T}$.

Therefore,

$$\hat{\varepsilon}_{1t+1} = y_{t+1} - \hat{\beta}_1 \hat{\underline{u}}_{2,t} = \varepsilon_{1t+1} - (\hat{\beta}_1 - \beta_1)\underline{u}_{2,t} - \hat{\delta}_T \hat{\beta}_1 \tilde{\underline{u}}_{2,t} - \frac{1}{2}\hat{\delta}^2{}_T \hat{\beta}_1 \underline{u}^*{}_{2,t,T}$$

$$\hat{\sigma}^2 = T^{-1}\sum_{t=1}^{T-1}\hat{\varepsilon}^2{}_{1t+1} = T^{-1}\sum_{t=1}^{T-1}\varepsilon^2{}_{1t+1} + R_{4,T}$$

by standard argument since for $\alpha > \frac{\bar{\delta}}{2}$ by (A.7), Lemma 2 (a) - (g), and (10)

$$R_{4,T} = (\hat{\beta}_1 - \beta_1)^2 T^{-1}\sum_{t=1}^{T-1}\underline{u}^2{}_{2,t} + \hat{\beta}_1{}^2\hat{\delta}^2{}_T T^{-1}\sum_{t=1}^{T-1}\tilde{\underline{u}}^2{}_{2,t} + \frac{1}{4}\hat{\beta}_1{}^2\hat{\delta}^4{}_T T^{-1}\sum_{t=1}^{T-1}(u^*{}_{2,t,T})^2$$

$$-2(\hat{\beta}_1 - \beta_1)T^{-1}\sum_{t=1}^{T-1}\underline{u}_{2,t}\varepsilon_{1t+1} - 2\hat{\delta}_T\hat{\beta}_1 T^{-1}\sum_{t=1}^{T-1}\tilde{\underline{u}}_{2,t}\varepsilon_{1t+1} - \hat{\delta}^2{}_T\hat{\beta}_1 T^{-1}\sum_{t=1}^{T-1}u^*{}_{2,t,T}\varepsilon_{1t+1}$$

$$+2\hat{\delta}_T(\hat{\beta}_1 - \beta_1)\hat{\beta}_1 T^{-1}\sum_{t=1}^{T-1}\underline{u}_{2,t}\tilde{\underline{u}}_{2,t} + \hat{\delta}^2{}_T\left(\hat{\beta}_1 - \beta_1\right)\hat{\beta}_1 T^{-1}\sum_{t=1}^{T-1}\underline{u}_{2,t}u^*{}_{2,t,T} + \hat{\delta}^3{}_T\hat{\beta}_1{}^2 T^{-1}\sum_{t=1}^{T-1}\tilde{\underline{u}}_{2,t}u^*{}_{2,t,T} = o_p(1).$$

Proof of Corollary 1.2 Result (a) follows from Theorem 1 and corollary 1.1 by standard arguments (note that the bias term is not present under the null H_o: β_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

$$T^{-1/2}t = \hat{\sigma}^{-1}\left(T^{-1}\sum_{t=1}^{T-1}\hat{\underline{u}}^2{}_{2,t}\right)^{1/2}\hat{\beta}_1 = \hat{\sigma}^{-1}\left(T^{-1}\sum_{t=1}^{T-1}\underline{u}^2{}_{2,t}\right)^{1/2} T^{1/2}\beta_1$$

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

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

FIGURES

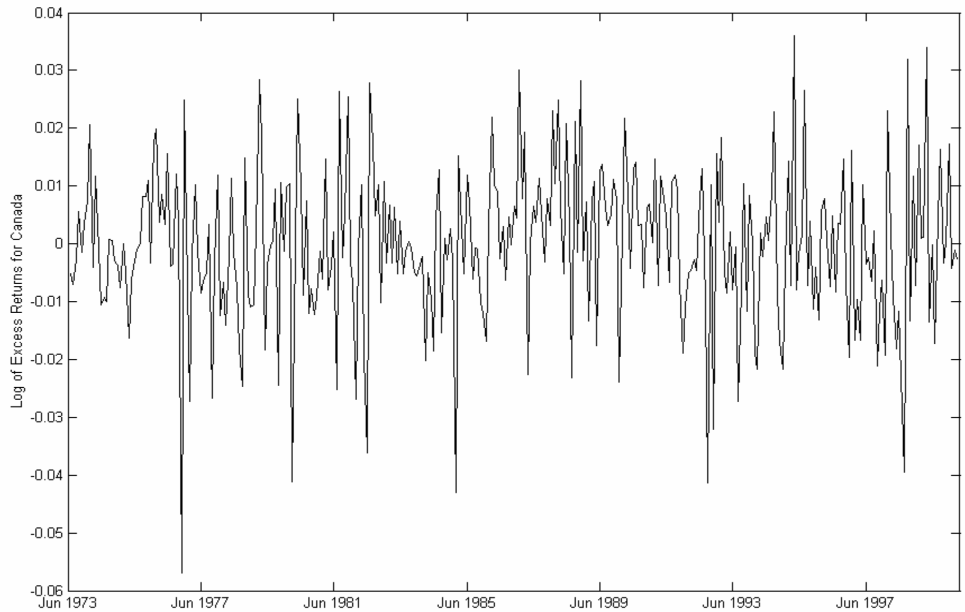

Figure 1
Log of Excess Returns for the Canadian Dollar vis-à-vis the US Dollar (1973-2000)

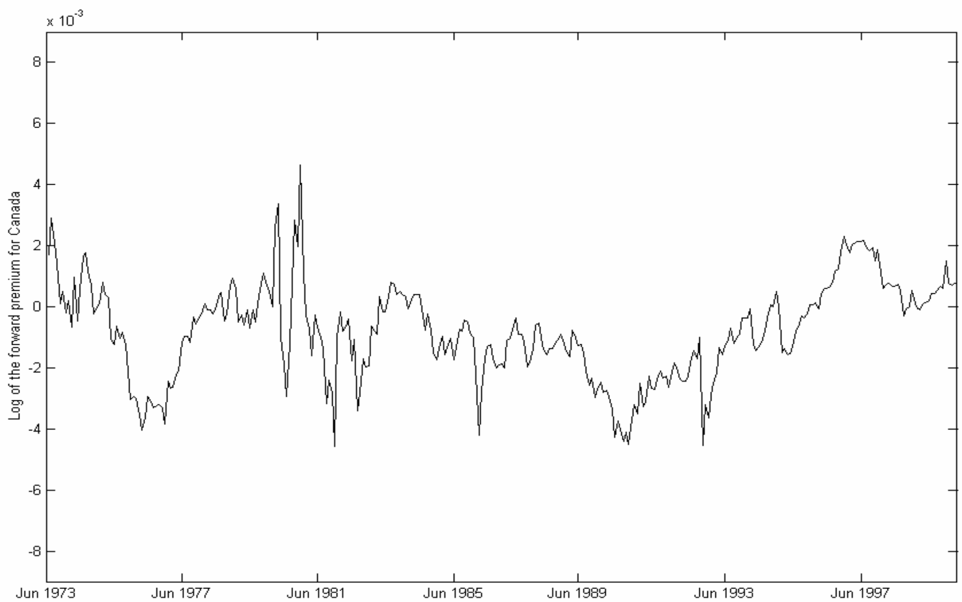

Figure 2
Log of the Forward Premium for the Canadian Dollar vis-à-vis the US Dollar (1973-2000)

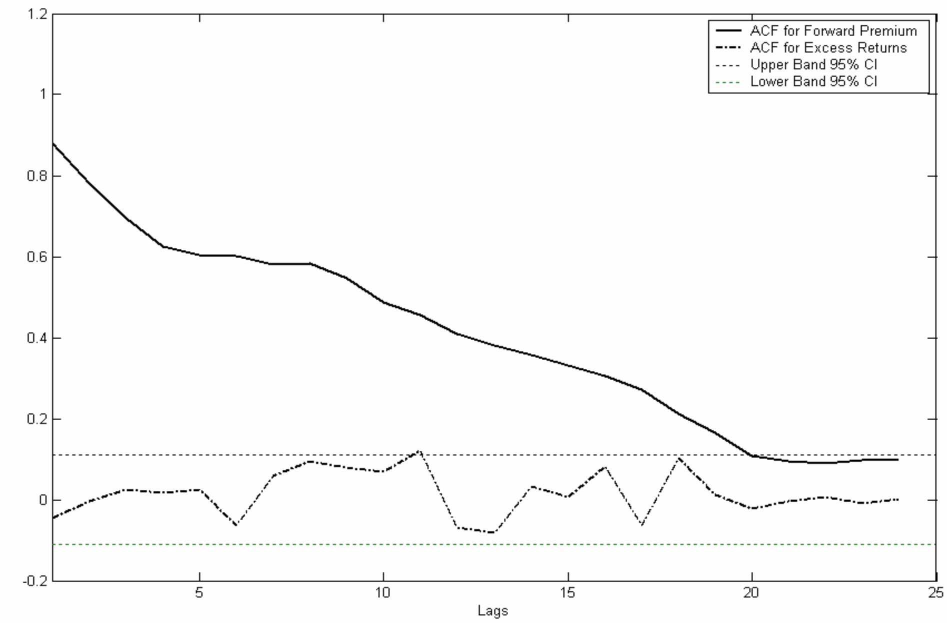

Figure 3
Sample Autocorrelations for Canadian Excess Returns and Forward Premium
With 95% Confidence Intervals about Zero

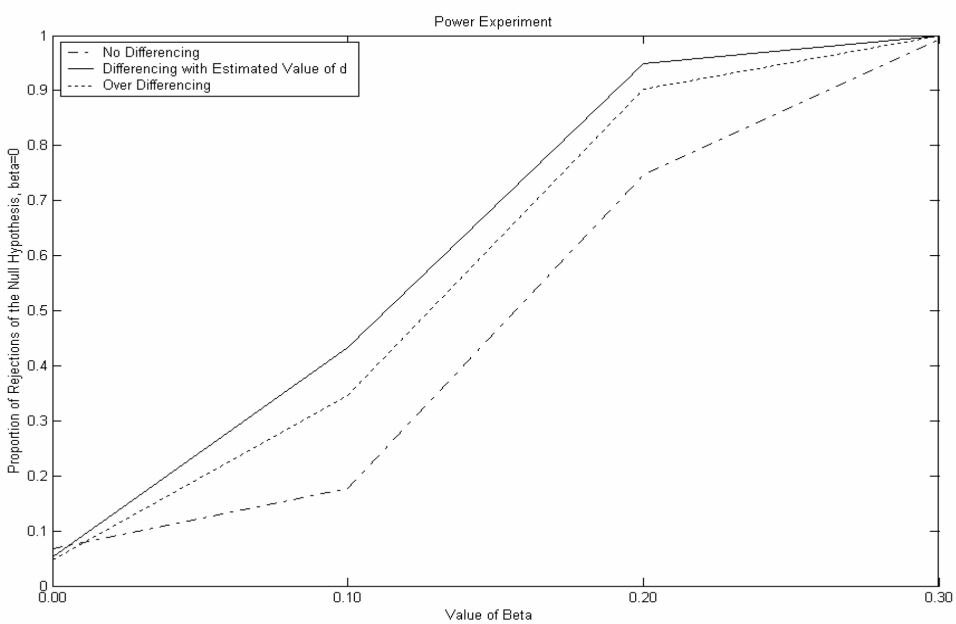

Figure 4
Power to Reject the Null Hypothesis that β_1=0
Dependent Variable is Short Memory with a Long Memory Regressor (d=0.40)

References

Agiakloglou, C., Newbold, P., and M. Wohar, 1992. Bias in an estimator of the fractional difference parameter. *Journal of Time Series Analysis*, **14**, 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., and T. Bollerslev, T, 1994. The long memory of the forward premium. *Journal of International Money and Finance*, **13 (5)**, 565--571.

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

Bekaert. G. and R. Hodrick, 2001. Expectations hypothesis tests. *Journal of Finance*, **56**, 1357-1394.

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, B. and J-M Dufour, 1997, Exact nonparametric tests of orthogonality and random walk in the presence of a drift parameter. *International Economic Review,* **38**, 151-173.

Cavanagh, C.L., Elliott, G., and J.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.

Fox, R., and M. Taqqu, 1986. Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *The Annals of Statistics*, **14**, 517-532.

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

Gradshtein, I. S. and I. M. Ryzhik, 1994. *Tables of Integrals, Series and Products*. Academic Press. Boston.

Koustas, Z., and A. Serletis, 2005. Rational bubbles or persistent deviations from market fundamentals? *Journal of Banking and Finance, forthcoming*.

Liu, W. and A. Maynard, 2005. Testing forward rate unbiasedness allowing for persistent regressors. *Journal of Empirical Finance, forthcoming*.

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

Maynard, A., 2003. Testing the forward rate unbiasedness: On regression in levels and in returns, *Review of Economics and Statistics*, **85 (2)**, 313-327.

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

Maynard, A., and K. Shimotsu, 2004. Covariance-based orthogonality tests for regressors with unknown persistence. *Mimeo, University of Toronto*.

Newbold, P., Wohar, M., Rayner, T., Kellard, N., and C. Ennew, 1998. Two puzzles in the analysis of foreign exchange market efficiency. *International Review of Financial Analysis*, **7**, 95-111.

Newey, W.K. and D. McFadden, 1994. Large sample estimation and hypothesis testing. In: *Handbook of Econometrics* (F. Engle and D.L. McFadden, Eds.). Vol. IV. Pp. 2111-2245. Elsevier.

Nielsen, M., and P.H. Frederiksen, 2004. Finite sample comparison of parametric, semiparametric, and wavelet estimators of fractional integration, *Cornell University Working Paper, 1-54.*

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

Rapach, D.E., and M.E. Wohar, 2004. In-sample vs. out-of-sample tests of stock return predictability in the context of data mining, Manuscript

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

Rudin, W., 1976. *Principles of Mathematical Analysis*. 3 ed., McGraw-Hill, Inc.

Shimotsu, K. and P.C.B. Phillips, 2002. Exact local Whittle estimation of fractional integration. *Cowles Foundation Discussion Paper #1367, Yale University*.

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

Stambaugh, R.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.

Torous, W., Valkanov, R., and S. Yan, 2005. On predicting stock returns with nearly integrated explanatory variables. *The Journal of Business*, forthcoming, **78 (1).**

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
0.20	0.0460	0.0483	0.0487	0.0513	0.0490	0.0503	0.0537	0.0530	0.0513
0.40	0.0497	0.0523	0.0503	0.0507	0.0517	0.0540	0.0600	0.0583	0.0600
0.60	0.0973	0.0910	0.0840	0.0587	0.0550	0.0620	0.0900	0.0987	0.1057
0.80	0.2023	0.1873	0.1623	0.0820	0.0570	0.0790	0.1590	0.1933	0.2033
0.90	0.2603	0.2350	0.2080	0.0920	0.0557	0.0913	0.1997	0.2337	0.2540
0.95	0.2823	0.2550	0.2150	0.0900	0.0550	0.0957	0.2143	0.2547	0.2777
1.000	0.3003	0.2743	0.2307	0.0917	0.0547	0.0957	0.2247	0.2703	0.2953

Table 1b
Bias of $\hat{b}_1$
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
0.20	0.3013	0.2830	0.2475	0.1115	-0.0189	-0.1452	-0.2677	-0.2971	-0.3112
0.40	0.7092	0.6714	0.5959	0.2955	-0.0041	-0.3041	-0.6007	-0.6741	-0.7107
0.60	1.0326	0.9787	0.8716	0.4429	0.0091	-0.4290	-0.8628	-0.9721	-1.0274
0.80	0.8930	0.8454	0.7524	0.3821	0.0063	-0.3732	-0.7423	-0.836	-0.8840
0.90	0.7063	0.6682	0.5939	0.3004	0.0027	-0.2990	-0.5847	-0.6568	-0.6940
0.95	0.6089	0.5761	0.5117	0.2585	0.0015	-0.2599	-0.5031	-0.5641	-0.5955
1.000	0.5159	0.4882	0.4336	0.2190	0.0007	-0.2221	-0.4255	-0.4762	-0.5021

Table 1c
Variance of $\hat{b}_1$
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
0.20	3.1439	3.1520	3.1735	3.2777	3.3613	3.3682	3.2637	3.2188	3.1918
0.40	2.2559	2.2671	2.2942	2.4148	2.5057	2.5158	2.3773	2.3244	2.2958
0.60	1.2069	1.2018	1.2077	1.2713	1.2983	1.2992	1.2387	1.2274	1.2241
0.80	0.5211	0.5031	0.4846	0.4634	0.4480	0.4583	0.4968	0.5197	0.5330
0.90	0.3205	0.3035	0.2827	0.2462	0.2305	0.2459	0.2915	0.3106	0.3229
0.95	0.2461	0.2318	0.2128	0.1755	0.1614	0.1771	0.2190	0.2352	0.2444
1.000	0.1864	0.1744	0.1588	0.1237	0.1115	0.1262	0.1615	0.1755	0.1820

Notes: The table shows simulation results from the standard predictability regression without rebalancing

$$y_{t+1} = c_1 + b_1 x_t + \varepsilon_{1t+1} \qquad (1)$$

under the null hypothesis (b_1=0) *when the regressor x_t is integrated of order d and given by*

$$(1-L)^d x_t = c_2 + \varepsilon_{2\tau}, \quad (2)$$

where $\varepsilon_t = (\varepsilon_{1t}, \ \varepsilon_{2t})' \sim i.i.d.N(0, \Sigma)$, *and* $\rho = \Sigma_{12} / \sqrt{\Sigma_{11} \Sigma_{22}}$ *denotes the residual correlation.*

***Notes for tables 1-2*:**

Throughout, the true value of β 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 (1) and (2) 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 Based On Partial Differencing
Regressor is an ARFIMA(1,0.40,0) 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	0.0433	0.0403	0.0447	0.0447	0.0470	0.0470	0.0410	0.0423	0.0400
-0.95	0.0410	0.0407	0.0437	0.0500	0.0490	0.0440	0.0367	0.0353	0.0360
-0.80	0.0403	0.0417	0.0463	0.0520	0.0497	0.0430	0.0400	0.0380	0.0400
-0.40	0.0490	0.0470	0.0493	0.0547	0.0553	0.0493	0.0480	0.0497	0.0507
0.00	0.0497	0.0523	0.0503	0.0507	0.0517	0.0540	0.0600	0.0583	0.0600
0.40	0.0623	0.0590	0.0580	0.0527	0.0517	0.0500	0.0660	0.0657	0.0663
0.80	0.0937	0.0877	0.0853	0.0610	0.0543	0.0593	0.0750	0.0907	0.0963
0.95	0.1170	0.1097	0.0947	0.0677	0.0543	0.0567	0.1003	0.1123	0.1243
0.99	0.2640	0.2403	0.2060	0.0913	0.0570	0.0917	0.1973	0.2407	0.2627

Table 2b

Bias of $\hat{b}_1$

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	-0.2197	-0.2089	-0.1888	-0.0973	-0.0009	0.0938	0.1880	0.2105	0.2211
-0.95	-0.0711	-0.0699	-0.0662	-0.0430	-0.0125	0.0219	0.0521	0.0591	0.0630
-0.80	0.0954	0.0871	0.0722	0.0216	-0.0203	-0.0571	-0.0946	-0.1038	-0.1077
-0.40	0.5366	0.5062	0.4464	0.2127	-0.0166	-0.2427	-0.4646	-0.5187	-0.5453
0.00	0.7092	0.6714	0.5959	0.2955	-0.0041	-0.3041	-0.6007	-0.6741	-0.7107
0.40	0.6704	0.6356	0.5659	0.2861	0.0041	-0.2807	-0.5627	-0.6329	-0.6682
0.80	0.4280	0.4062	0.3626	0.1860	0.0050	-0.1773	-0.3567	-0.4021	-0.4252
0.95	0.3165	0.3003	0.2683	0.1390	0.0048	-0.1309	-0.2633	-0.2975	-0.3149
0.99	0.1467	0.1381	0.1227	0.0635	0.0005	-0.0633	-0.1218	-0.1368	-0.1445

Table 2c

Variance of $\hat{b}_1$

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	0.2398	0.2369	0.2343	0.2193	0.2105	0.2171	0.2382	0.2404	0.2394
-0.95	1.0079	1.0191	1.0401	1.0962	1.1233	1.0944	1.0168	0.9954	0.9876
-0.80	1.6492	1.6705	1.7071	1.8174	1.8716	1.8308	1.6860	1.6436	1.6251
-0.40	2.7047	2.7301	2.7798	2.9535	3.0571	3.0313	2.8354	2.7649	2.7272
0.00	2.2559	2.2671	2.2942	2.4148	2.5057	2.5158	2.3774	2.3245	2.2958
0.40	1.2430	1.2421	1.2460	1.2840	1.3169	1.3260	1.2775	1.2645	1.2592
0.80	0.2954	0.2897	0.2833	0.2723	0.2646	0.2679	0.2812	0.2907	0.2969
0.95	0.1247	0.1207	0.1160	0.1058	0.1009	0.1046	0.1160	0.1235	0.1278
0.99	0.0183	0.0165	0.0148	0.0117	0.0101	0.0123	0.0162	0.0178	0.0187

Notes: The table shows simulation results from the standard predictability regression without rebalancing

$$y_{t+1} = c_1 + b_1 x_t + \varepsilon_{1t+1} \qquad (1)$$

under the null hypothesis (b_1=0) when the regressor x_t is integrated of order d=0.4 *throughout and given by*

$$(1-\phi L)(1-L)^{0.40} x_t = c_2 + \varepsilon_{2t}, \qquad (2)$$

where $\varepsilon_t = (\varepsilon_{1t}, \quad \varepsilon_{2t})' \sim i.i.d.N(0, \Sigma)$. *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.*

Table 3
Application of the 2-step Procedure using the Time Domain Estimator for *d*: Original Process is Fractional Noise

Table 3a
Proportion of Rejections in a 5% Nominal Test of the Null Hypothesis that β_1=0
Sample Size=350, True Value of β_1=0

ρ/d	-0.95	-0.90	-0.80	-0.40	0.00	0.40	0.80	0.90	0.95
0.40	0.0557	0.0550	0.0533	0.0537	0.0523	0.0523	0.0533	0.0520	0.0517
0.50	0.0567	0.0540	0.0540	0.0550	0.0540	0.0523	0.0543	0.0523	0.0553
0.60	0.0553	0.0557	0.0530	0.0543	0.0547	0.0523	0.0523	0.0520	0.0537
0.70	0.0557	0.0563	0.0527	0.0547	0.0560	0.0530	0.0523	0.0513	0.0537
0.80	0.0550	0.0563	0.0527	0.0547	0.0550	0.0527	0.0520	0.0517	0.0530
0.90	0.0550	0.0550	0.0530	0.0547	0.0550	0.0533	0.0520	0.0517	0.0527
0.95	0.0553	0.0550	0.0530	0.0550	0.0550	0.0537	0.0520	0.0513	0.0527
1.000	0.0553	0.0550	0.0533	0.0553	0.0550	0.0530	0.0517	0.0513	0.0523

Table 3b
Bias of $\hat{\beta}_1$ using two step-procedure
Sample Size=350, True Value of β_1=0

ρ/d	-0.95	-0.90	-0.80	-0.40	0.00	0.40	0.80	0.90	0.95
0.40	0.1605	0.1483	0.1260	0.0460	-0.0272	-0.0960	-0.1585	-0.1718	-0.1772
0.50	0.1736	0.1605	0.1368	0.0503	-0.0296	-0.1047	-0.1741	-0.1890	-0.1953
0.60	0.1510	0.1392	0.1178	0.0409	-0.0293	-0.0950	-0.1545	-0.1670	-0.1720
0.70	0.1347	0.1240	0.1041	0.0344	-0.0291	-0.0878	-0.1408	-0.1517	-0.1558
0.80	0.1274	0.1169	0.0980	0.0315	-0.0287	-0.0843	-0.1341	-0.1441	-0.1478
0.90	0.1246	0.1143	0.0958	0.0307	-0.0283	-0.0829	-0.1310	-0.1407	-0.1443
0.95	0.1240	0.1138	0.0954	0.0305	-0.0281	-0.0823	-0.1303	-0.1399	-0.1434
1.000	0.1242	0.114	0.0955	0.0306	-0.0280	-0.0822	-0.1300	-0.1396	-0.1432

Table 3c
Variance of $\hat{\beta}_1$ using two step-procedure
Sample Size=350, True Value of β_1=0

ρ/d	-0.95	-0.90	-0.80	-0.40	0.00	0.40	0.80	0.90	0.95
0.40	3.5096	3.5108	3.5126	3.5328	3.5617	3.5700	3.5325	3.5191	3.5131
0.50	3.6275	3.6282	3.6289	3.6464	3.6766	3.6927	3.6699	3.6594	3.6540
0.60	3.6090	3.6125	3.6198	3.6540	3.6860	3.6958	3.6594	3.6442	3.6351
0.70	3.6090	3.6125	3.6198	3.6540	3.6860	3.6958	3.6594	3.6442	3.6351
0.80	3.6020	3.6070	3.6162	3.6557	3.6898	3.6972	3.6554	3.6384	3.6287
0.90	3.5999	3.6053	3.6152	3.6566	3.6918	3.6983	3.6545	3.6365	3.6267
0.95	3.5991	3.6042	3.6143	3.6568	3.6926	3.6984	3.6538	3.6356	3.6253
1.000	3.5996	3.6048	3.6152	3.657	3.6925	3.6987	3.6552	3.6362	3.6264

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

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

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 (1) and (2) 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 in a 5% Nominal Test of the Null Hypothesis that β_1=0

ρ/ϕ	-0.95	-0.90	-0.80	-0.40	0.00	0.40	0.80	0.90	0.95
-0.99	0.0470	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
Bias of $\hat{\beta}_1$ 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 of $\hat{\beta}_1$ 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*

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:

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

Here, 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,*d*,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	0.0453	0.0463	0.0500	0.0447	0.0463	0.0513	0.0483	0.0503	0.0477
-0.95	0.0437	0.0463	0.0497	0.0497	0.0513	0.0460	0.0460	0.0433	0.0423
-0.80	0.0483	0.0470	0.0487	0.0513	0.0507	0.0477	0.0437	0.0433	0.0437
-0.40	0.0497	0.0517	0.0493	0.0527	0.0487	0.0477	0.0483	0.0483	0.0457
0.00	0.0600	0.0580	0.0497	0.0523	0.0543	0.0497	0.0560	0.0517	0.0527
0.40	0.0610	0.0557	0.0520	0.0523	0.0520	0.0547	0.0600	0.0590	0.0547
0.80	0.0567	0.0567	0.0567	0.0467	0.0510	0.0543	0.0643	0.0633	0.0617
0.95	0.0570	0.0547	0.0517	0.0473	0.0520	0.0500	0.0623	0.0567	0.0570
0.99	0.0537	0.0520	0.0513	0.0490	0.0510	0.0533	0.0570	0.0540	0.0533

Table 5b

Bias of $\hat{\beta}_1$ 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	-0.1854	-0.1767	-0.1601	-0.0832	-0.0021	0.0781	0.1563	0.1758	0.1853
-0.95	-0.1514	-0.1457	-0.1351	-0.0753	-0.0117	0.0548	0.1165	0.1344	0.1420
-0.80	-0.0963	-0.0945	-0.0936	-0.0600	-0.0207	0.0226	0.0617	0.0747	0.0806
-0.40	0.0806	0.0718	0.0511	0.0057	-0.0329	-0.0614	-0.0913	-0.1005	-0.1017
0.00	0.2323	0.2165	0.1834	0.0722	-0.0283	-0.1141	-0.2076	-0.2386	-0.2487
0.40	0.3615	0.3416	0.3020	0.1366	-0.0197	-0.1592	-0.3067	-0.3492	-0.3680
0.80	0.3632	0.3465	0.3083	0.1429	-0.0210	-0.1638	-0.3119	-0.3471	-0.3669
0.95	0.3067	0.2900	0.2529	0.1081	-0.0256	-0.1413	-0.2667	-0.2993	-0.3154
0.99	0.2050	0.1925	0.1654	0.0676	-0.0203	-0.1049	-0.1913	-0.2154	-0.2246

Table 5c

Variance of $\hat{\beta}_1$ 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	0.1580	0.1543	0.1521	0.1352	0.1277	0.1354	0.1506	0.1536	0.1537
-0.95	0.7007	0.7023	0.7043	0.7149	0.7278	0.7267	0.6949	0.6877	0.6931
-0.80	1.2397	1.2441	1.2479	1.2910	1.3097	1.3106	1.2433	1.2291	1.2349
-0.40	2.8059	2.8156	2.8252	2.9155	2.9375	2.9120	2.8209	2.8050	2.8053
0.00	3.3419	3.3452	3.3279	3.3752	3.4331	3.4130	3.3740	3.3505	3.3347
0.40	2.8822	2.8787	2.8753	2.8979	2.9441	2.9599	2.9505	2.9163	2.8955
0.80	2.1923	2.1905	2.1835	2.1557	2.1548	2.1977	2.2515	2.2382	2.2203
0.95	2.5194	2.5231	2.5130	2.5133	2.5342	2.5661	2.5859	2.5513	2.5322
0.99	3.1674	3.1733	3.1886	3.1926	3.3368	3.2996	3.1995	3.1801	3.1852

Table 5d

Mean Estimated Value of *d*

ρ/ϕ	-0.95	-0.90	-0.80	-0.40	0.00	0.40	0.80	0.90	0.95
-0.99	0.8003	0.8013	0.8001	0.7975	0.7954	0.7963	0.7990	0.7991	0.7997
-0.95	0.8069	0.8073	0.8077	0.8045	0.8016	0.8011	0.8052	0.8053	0.8049
-0.80	0.8072	0.8077	0.8081	0.8045	0.8019	0.8018	0.8054	0.8055	0.8053
-0.40	0.8074	0.8079	0.8079	0.8041	0.8017	0.8019	0.8056	0.8055	0.8055
0.00	0.8073	0.8079	0.8078	0.8038	0.8015	0.8019	0.8058	0.8055	0.8054
0.40	0.8161	0.8165	0.8166	0.8122	0.8100	0.8105	0.8144	0.8143	0.8139
0.80	1.0164	1.0170	1.0179	1.0141	1.0115	1.0115	1.0160	1.0159	1.0155
0.95	1.2615	1.2632	1.2637	1.2611	1.2591	1.2579	1.2634	1.2625	1.2616
0.99	1.6629	1.6630	1.6619	1.6609	1.6611	1.6604	1.6611	1.6607	1.6602

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 $(1-L)^{0.50}$ to x_t, and set $m=T^{0.65}$.

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

Table 6a

Proportion of Rejections in a 5% Nominal Test of the Null Hypothesis that β_1=0

ρ/ϕ	-0.95	-0.90	-0.80	-0.40	0.00	0.40	0.80	0.90	0.95
-0.99	0.0617	0.0563	0.0513	0.0487	0.0470	0.0483	0.0540	0.0597	0.0607
-0.95	0.0550	0.0517	0.0553	0.0520	0.0540	0.0500	0.0540	0.0573	0.0567
-0.80	0.0560	0.0553	0.0560	0.0543	0.0497	0.0470	0.0527	0.0513	0.0510
-0.40	0.0530	0.0510	0.0513	0.0523	0.0523	0.0507	0.0497	0.0520	0.0507
0.00	0.0550	0.0543	0.0520	0.0537	0.0543	0.0530	0.0517	0.0513	0.0520
0.40	0.0530	0.0537	0.0523	0.0553	0.0560	0.0583	0.0520	0.0543	0.0497
0.80	0.0527	0.0540	0.0533	0.0540	0.0567	0.0553	0.0510	0.0523	0.0507
0.95	0.0520	0.0537	0.0513	0.0540	0.0570	0.0567	0.0513	0.0537	0.0513
0.99	0.0527	0.0530	0.0513	0.0537	0.0550	0.0550	0.0543	0.0510	0.0510

Table 6b

Bias of $\hat{\beta}_1$ 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	0.0486	0.0455	0.0378	0.0185	0.0021	-0.0188	-0.0367	-0.0418	-0.0451
-0.95	0.2324	0.2167	0.1892	0.0884	-0.0098	-0.1062	-0.2000	-0.2249	-0.2374
-0.80	0.2745	0.2558	0.2219	0.0987	-0.0198	-0.1347	-0.2437	-0.2716	-0.2852
-0.40	0.2118	0.1962	0.1676	0.0654	-0.0295	-0.1195	-0.2019	-0.2206	-0.2289
0.00	0.1280	0.1176	0.0990	0.0331	-0.0265	-0.0816	-0.1306	-0.1404	-0.1443
0.40	0.0810	0.0740	0.0612	0.0167	-0.0232	-0.0589	-0.0889	-0.0939	-0.0953
0.80	0.0605	0.0543	0.0437	0.0072	-0.0246	-0.0521	-0.0726	-0.0752	-0.0754
0.95	0.0514	0.0451	0.0344	0.0000	-0.0275	-0.0511	-0.0683	-0.0700	-0.0693
0.99	0.0622	0.0549	0.0427	0.0038	-0.0286	-0.0570	-0.0790	-0.0816	-0.0813

Table 6c

Variance of $\hat{\beta}_1$ 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	0.2859	0.2848	0.2908	0.3041	0.2961	0.2918	0.2936	0.2903	0.2865
-0.95	1.1828	1.1874	1.1997	1.2299	1.2459	1.2227	1.1775	1.1693	1.1686
-0.80	1.8519	1.8570	1.8642	1.9003	1.9269	1.9134	1.8608	1.8456	1.8412
-0.40	3.2177	3.2215	3.2289	3.2618	3.2865	3.2885	3.2485	3.2328	3.2240
0.00	3.4919	3.4959	3.5040	3.5409	3.5720	3.5743	3.5293	3.5108	3.5008
0.40	3.3402	3.3541	3.3766	3.4388	3.4663	3.4417	3.3675	3.3430	3.3303
0.80	3.2723	3.2882	3.3145	3.3891	3.4127	3.3697	3.2839	3.2610	3.2510
0.95	3.3217	3.3346	3.3569	3.4288	3.4625	3.4310	3.3468	3.3228	3.3113
0.99	3.4413	3.4484	3.4624	3.5126	3.5511	3.5483	3.4856	3.4614	3.4491

Table 6d

Mean Estimated Value of *d*

ρ/ϕ	-0.95	-0.90	-0.80	-0.40	0.00	0.40	0.80	0.90	0.95
-0.99	-0.0167	-0.0165	-0.0166	-0.0181	-0.0194	-0.0186	-0.0160	-0.0157	-0.0158
-0.95	0.2369	0.2372	0.2368	0.2360	0.2354	0.2358	0.2376	0.2379	0.2378
-0.80	0.3376	0.3377	0.3375	0.3367	0.3360	0.3365	0.3380	0.3380	0.3381
-0.40	0.5761	0.5760	0.5759	0.5754	0.5746	0.5750	0.5762	0.5765	0.5766
0.00	0.7964	0.7964	0.7962	0.7957	0.7953	0.7955	0.7962	0.7964	0.7965
0.40	1.1045	1.1045	1.1043	1.1038	1.1037	1.1038	1.1042	1.1044	1.1045
0.80	1.5457	1.5456	1.5454	1.5450	1.5447	1.5450	1.5453	1.5456	1.5457
0.95	1.6672	1.6671	1.6670	1.6664	1.6661	1.6662	1.6667	1.6668	1.6669
0.99	1.7805	1.7804	1.7802	1.7795	1.7791	1.7791	1.7797	1.7800	1.7802

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 appear under the heading ρ/ϕ. The CSS estimator is used to incorrectly estimate a mis-specified 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 β_1=0

ρ/d	-0.95	-0.90	-0.80	-0.40	0.00	0.40	0.80	0.90	0.95
0.40	0.0537	0.0527	0.0520	0.0530	0.0530	0.0500	0.0533	0.0530	0.0487
0.50	0.0533	0.0527	0.0533	0.0530	0.0550	0.0523	0.0567	0.0530	0.0523
0.60	0.0530	0.0510	0.0520	0.0533	0.0523	0.0527	0.0550	0.0550	0.0517
0.70	0.0513	0.0510	0.0520	0.0533	0.0543	0.0540	0.0550	0.0560	0.0527
0.80	0.0517	0.0523	0.0507	0.0543	0.0527	0.0517	0.0523	0.0567	0.0550
0.90	0.0530	0.0510	0.0533	0.0550	0.0540	0.0520	0.0570	0.0580	0.0570
0.95	0.0543	0.0510	0.0533	0.0537	0.0543	0.0530	0.0567	0.0597	0.0580
1.00	0.0547	0.0523	0.0517	0.0533	0.0527	0.0530	0.0553	0.0617	0.0597

Table 7b

Bias of $\hat{\beta}_1$ using 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	0.2166	0.2030	0.1757	0.0716	-0.0244	-0.1174	-0.2030	-0.2219	-0.2332
0.50	0.2963	0.2760	0.2425	0.1066	-0.0244	-0.1480	-0.2681	-0.2966	-0.3107
0.60	0.3287	0.3082	0.2692	0.1228	-0.0232	-0.1597	-0.2922	-0.3247	-0.3415
0.70	0.3697	0.3471	0.3055	0.1421	-0.0205	-0.1697	-0.3279	-0.3643	-0.3823
0.80	0.4390	0.4157	0.3632	0.1726	-0.0162	-0.1995	-0.3762	-0.4225	-0.4461
0.90	0.4835	0.4559	0.4018	0.1953	-0.0124	-0.2170	-0.4076	-0.4540	-0.4774
0.95	0.5016	0.4735	0.4135	0.1928	-0.0134	-0.2223	-0.4185	-0.4644	-0.4903
1.00	0.5110	0.4851	0.4251	0.2019	-0.0210	-0.2298	-0.4314	-0.4832	-0.5042

Table 7c

Variance of $\hat{\beta}_1$ using 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	3.4317	3.4353	3.4480	3.5092	3.5559	3.5584	3.4759	3.4533	3.4294
0.50	3.4009	3.4056	3.4186	3.4933	3.5495	3.5499	3.4685	3.4393	3.4177
0.60	3.3582	3.3658	3.3757	3.4584	3.5206	3.5162	3.4557	3.4193	3.3873
0.70	3.2756	3.2789	3.2911	3.3971	3.4815	3.4914	3.3735	3.3373	3.3140
0.80	3.1449	3.1557	3.1860	3.3309	3.3866	3.3787	3.2529	3.2375	3.2121
0.90	3.0855	3.0918	3.0932	3.2404	3.3006	3.2944	3.1764	3.1727	3.1523
0.95	3.0409	3.0292	3.0403	3.1491	3.2635	3.2238	3.1245	3.1171	3.1091
1.00	2.9601	2.9722	2.9767	3.1077	3.1889	3.1565	3.0847	3.0518	3.0333

Table 7d

Mean Estimated Value of *d*

ρ/d	-0.95	-0.90	-0.80	-0.40	0.00	0.40	0.80	0.90	0.95
0.40	0.3697	0.3686	0.3677	0.3687	0.3675	0.3680	0.3678	0.3687	0.3678
0.50	0.4305	0.4313	0.4317	0.4295	0.4307	0.4307	0.4304	0.4295	0.4295
0.60	0.5030	0.5027	0.5038	0.5033	0.5028	0.5031	0.5033	0.5021	0.5025
0.70	0.5786	0.5786	0.5777	0.5765	0.5761	0.5764	0.5755	0.5768	0.5768
0.80	0.6486	0.6488	0.6488	0.6477	0.6474	0.6483	0.6473	0.6482	0.6479
0.90	0.7238	0.7226	0.7224	0.7226	0.7212	0.7230	0.7248	0.7241	0.7248
0.95	0.7588	0.7596	0.7574	0.7584	0.7576	0.7597	0.7586	0.7606	0.7601
1.00	0.7971	0.7967	0.7967	0.7961	0.7940	0.7959	0.7988	0.7970	0.7967

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

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

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):	Country:	Canada	France	Germ.	Japan	UK
Dependent Variable						
Δs_{t+1}	$\hat{c}_1$	-0.0021	-0.0028	0.0019	0.0030	-0.0046
		[0.0087]	[0.1615]	[0.3819]	[0.1921]	[0.0365]
	$\hat{b}_1$	-1.1356	-0.8457	-0.7150	-0.0215	-1.4554
		[0.00000]	[0.0007]	[0.0139]	[0.0245]	[0.0001]
	$\hat{c}_2$	-0.0021	-0.0028	0.0019	0.0030	-0.0046
Dependent Variable		[0.0087]	[0.0020]	[0.3819]	[0.1921]	[0.0365]
$(s_{t+1}-f_t)$	$\hat{b}_2$	-2.1356	-1.8457	-1.7150	-1.0215	-2.4554
		[0.00000]	[0.0007]	[0.0139]	[0.0245]	[0.0001]

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.

Table 9
OLS Estimates from the FRUH-type Regressions
Differencing Applied from ARFIMA(*p*,*d*,*q*) Model
Differencing Parameter Estimated from CSS Estimator

Sample (1973-2000):	Country:	Canada	France	Germ.	Japan	UK
	$\hat{\beta}_0$	-0.0004	0.0000	-0.0010	0.0004	0.0001
Dependent Variable		[0.9596]	[0.9803]	[0.5987]	[0.8415]	[0.9557]
$(s_{t+1}-f_t)$	$\hat{\beta}_1$	-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}

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.

Table 10
OLS Estimates of the FRUH-type Regressions with Fractional Differencing: First Stage Semi-Parametric used to Estimate *d*

Table 10a
Differencing Parameter Estimated from BRLP Regression Estimator with $m=T^{0.50}$

Sample (1973-2000):	Country:	Canada	France	Germ.	Japan	UK
	$\hat{\beta}_0$	-0.0004	-0.0007	-0.0010	0.0004	-0.0006
Dependent Variable		[0.5677]	[0.6974]	[0.5987]	[0.8415]	[0.7389]
$(s_{t+1}-f_t)$	$\hat{\beta}_1$	-4.6593	-2.3097	-2.0775	0.1532	-3.3995
		[0.0000]	[0.0019]	[0.2631]	[0.7456]	[0.0008]
	$\hat{d}$	**0.5702**	**0.2472**	**1.3045**	**0.8684**	**0.3026**

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 LP Regression based estimator of Andrews and Guggenberger (2003). A taper of $(1-L)^{0.50}$ has been applied to the forward premium and the number of periodogram ordinates used has been set equal to $T^{0.50}$.

Table 10b
Differencing Parameter Estimated from BRLP Regression Estimator with $m=T^{0.65}$

Sample (1973-2000):	Country:	Canada	France	Germ.	Japan	UK
	$\hat{\beta}_0$	-0.0004	-0.0002	-0.0010	0.0006	-0.0009
Dependent Variable		[0.5677]	[0.9115]	[0.5987]	[0.7642]	[0.6171]
$(s_{t+1}-f_t)$	$\hat{\beta}_1$	-3.0268	-2.3039	-2.1961	-0.3500	-3.3516
		[0.0018]	[0.0051]	[0.2511]	[0.5184]	[0.0005]
	$\hat{d}$	**1.0171**	**0.4421**	**1.1679**	**0.3436**	**0.2594**

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 LP Regression based estimator of Andrews and Guggenberger (2003). A taper of $(1-L)^{0.50}$ has been applied to the forward premium and the number of periodogram ordinates used has been set equal to $T^{0.65}$.

***Notes for Table 10**: Numerical standard errors do not exist for the BRLPR given their simple calculation. Further, the asymptotic standard error applies to the stationary processes, and do not consider the application of our taper. Thus, we do not include an estimate of the standard error of d.*