Expectations Hypotheses Tests and Predictive Regressions at Long Horizons

Barbara Rossi†

October 2004

Work in progress   Comments welcome

Abstract

Many rational expectations models state that an economic variable is determined as the present value of future variables. The restrictions imposed by these models have traditionally been tested on Vector Autoregressions in which variables appear either in levels (or cointegrating relationships) or in first differences, based on unit root pre-tests. Commonly applied test statistics may lead to over-rejections in small samples in the presence of highly persistent variables. Similarly, long horizon predictive regressions may suffer from this problem too. We propose to do inference by using an alternative method based on local-to-unity asymptotic approximations, where the horizon of interest is a fixed fraction of the sample size. In this framework, these tests become tests on the roots of the process. We show that the proposed method has the correct size, and it is easy to use. We apply this method to test long horizon Predictive Regressions, Uncovered Interest Rate Parity, the Expectation Hypothesis of the Term Structure, the Permanent Income Hypothesis and the net present value theory of the Current Account in a unifying framework.

Keywords: expectation hypotheses; present value models; long-horizon; local to unity roots.

JEL classification: F300; F400.

†Department of Economics, Duke University, Durham NC27708 USA. E-mail: brossi@econ.duke.edu. Tel.: (919) 660 1801. Fax: (919) 684 8974. First draft: April 2003.

I am grateful to T. Bollerslev and G. Tauchen for constructive suggestions, and to P. Dudenhefer, A. Inoue, A. Tarozzi and participants at the Duke Financial Econometrics Lunch for useful comments. All mistakes are mine.

1. Introduction

Many rational expectation models state that an economic variable is determined as the present value of future variables. Examples of such models include: the Permanent Income Hypothesis (PIH), where consumption is the discounted sum of future changes in income; the Uncovered Interest Rate Parity condition (UIRP), where the interest rate differential between two currencies is the expected sum of future changes in exchange rates; the Expectation Hypothesis of the Term Structure (EH-TS), where the long-term interest rate equals the average expected short-term interest rate; and models of Asset Pricing determination, where asset prices are determined as the present value of the stream of future dividends. We will refer to these models collectively as "Expectations Hypothesis" (EH) models. While these models constitute the foundation of modern theoretical economics (Obstfeld and Rogoff (1996)), the empirical evidence is mixed.[1] Similar considerations apply to long horizon predictive regressions.

This literature faces the common problem of testing an economic restriction at long horizons in the presence of possibly non-stationary time series data. The econometric methodology used to test these models shares two common features. The first is that, since valid statistical tests must be carried out on stationary time series, researchers assume *exact* unit roots in some variables and then use first differences or cointegrating relationships as stationary variables. While this practice was justified by results of unit root pre-tests and has become standard in applied work, it is practically difficult to distinguish between a unit root process and a mean-reverting but highly persistent one in small samples (see Cavanagh, Elliott and Stock (1995)). Similarly, tests for cointegration suffer from size problems (see Elliott (1998)). The second feature is that they (directly or indirectly) perform inference at horizons that are large relative to the sample size, and possibly infinity.

The question that naturally arises is whether these econometric procedures are robust to inference at long horizons in the presence of roots close to (but not necessarily equal to) unity. Since even very small deviations from unit roots can potentially have big effects at long horizons, we investigate EH tests at long horizons that do *not* rely on *exact* unit roots *nor* exact cointegration.[2] This paper proposes a method to test economic restrictions at long horizons that is robust to uncertainty on

[1] Tests on the PIH find some empirical support (Campbell (1987)), but similar tests applied to the current account (CA) are less satisfactory (Sheffrin and Woo (1990)); results for EH models are mixed as well (Bekaert and Hodrick (2001), Campbell and Shiller (1987)). An additional example, not considered in the present paper, is the monetary model of exchange rate determination; see Kilian (1999), Mark (1995), Mark and Sul (2002) and Engel and West (2002).

[2] Even in those cases in which there is an economic reason to expect some economic variables to have unit roots (for example, in the RBC literature, a random walk technology implies that output and consumption are difference-stationary), it is still practically difficult in small samples to distinguish between shocks that have permanent effects and shocks that have transitory but long-lasting effects. It is desirable that the empirical conclusions be robust to these potential problems.

the roots. Furthermore, researchers routinely use delta methods (Sheffrin and Woo (1990)) or the bootstrap (Bekaert and Hodrick (2001) and Cochrane and Piazzesi (2002)) to deal with the non-linear constraints that originate at long horizons, and to evaluate R^2 measures of predictive ability.[3] The method proposed in this paper does not require delta method approximations and easier to use than the bootstrap.

Although this paper focuses on expectations hypotheses, its methods can be applied to predictive regressions as well. The analysis of long horizon predictability in small samples has been the focus of recent research, both in-sample (see Valkanov (2003)) and out of sample (see Clark and McCracken (2002) and Rossi (2002)). We will provide empirical applications and discuss theoretical results for in-sample long horizon predictive ability too. Our method is computationally easier to use than Valkanov (2003), as it does not require Monte Carlo simulations but only the use of critical values that are already available.

2. An example

To clarify the problems of conventional inference, let's consider the Expectation Hypothesis in the Foreign Exchange market (EH-FX). It states that the long-term interest rate differential between the two countries, $i_{t,h}$, is an unbiased predictor of the expected change in the short-term bilateral spot exchange rate at horizon h, $s_{t+h} - s_t$, plus a constant risk premium, c_h.[4] Thus, letting η_t denote an unforecastable error given information at time t:

$$E_t \sum_{j=1}^{h} \Delta s_{t+j} = c_h + i_{t,h} + \eta_t \tag{1}$$

To test the theory, we estimate a VAR for the vector $w_t' = (\Delta s_t, i_{t,h})'$: $w_t = A w_{t-1} + \epsilon_t$. The restriction imposed by (1) is:

$$i_1' \frac{1}{h} \sum_{j=1}^{h} A^j - i_2' = i_1' \frac{1}{h} A \left(I - A^h\right) (I - A)^{-1} - i_2' = 0 \tag{2}$$

Some features of this approach (or maintained hypotheses) are that: *(i)* the exchange rate s_t has an exact unit root; *(ii)* $I - A$ is invertible; *(iii)* the constraint is non-linear; *(iv)* the Wald test has a chi-square asymptotic distribution provided the sample is large. However, Bekaert and Hodrick

[3] Non-linear constraints originate at finite horizons; at infinite horizons, it is possible to rewrite the constraint so that the restriction becomes linear. See Campbell (1987).

[4] Note that the objective of the paper is not to study time variation in risk premia, although time varying risk premia could be generated by misspecifying the order of integration of the variables. Incorrectly selecting the order of integration leads to small persistent components that are time-varying and, thus, leads to size distortions in hypotheses tests. Thus, our method is robust to time-varying risk premia that originate from misspecifying the order of integration of the variables. See Bekaert and Hodrick (2001) and Bekaert, Wei and Xing (2002).

(2001) and Bekaert, Wei and Xing (2002) show that the Wald test has considerable size distortions. This might be due to the fact that (some of) the maintained hypotheses may not hold.

In fact, suppose that the DGP for $z_t' = (s_t, i_{t,h})'$ is such that:

$$z_t = \Phi z_{t-1} + \epsilon_t, \quad \Phi = \begin{pmatrix} \rho_1 & \beta \\ 0 & \rho_2 \end{pmatrix} \tag{3}$$

where ρ_1 and ρ_2 are both close to unity. Note that, in this case, $A = \begin{pmatrix} \rho_1 - 1 & \beta \\ 0 & \rho_2 \end{pmatrix}$, so that testing the hypothesis $i_1'A = i_2'$ is a test of the joint hypothesis that $\beta = 1$ and that there is a unit root in the data; rejections of the null hypothesis might then be due to a highly persistent root which is not equal to unity rather than to $\beta \neq 1$. Interestingly, in this case, even if ρ_1 were equal to one (so that first differencing would be appropriate), the persistence in the interest rate differential would complicate estimation and inference. In fact, in that case:

$$A^h = \begin{pmatrix} 0 & \beta\rho_2^{(h-1)} \\ 0 & \rho_2^h \end{pmatrix} \tag{4}$$

so that the estimate of the coefficient in (1) would be downward biased (see Cavanagh, Elliott and Stock (1995) and Stambaugh (1999)). In this context, persistence has a first order effect on the estimate of the parameter of interest. Note that a 90% confidence interval for the largest root in short-term interest rates is $(0.958, 1.006)$, so the problem considered in this paper is empirically relevant.

3. The econometric model

Let the data generating process (hereafter DGP) follow a multivariate local-to-unity process (cfr. Stock and Watson (1996), Stock (1991, 1996) and Phillips (1987)):

$$(I - \Phi L)\, w_t = \Theta(L)\epsilon_t \tag{5}$$

where w_t is a $(n+1) \times 1$ vector of variables partitioned as $w_t = (w_{1t}, w_{2t}')'$, where w_{1t} is a scalar and w_{2t} is $(n \times 1)$ vector (and all matrices and vectors are partitioned accordingly). Following the literature, the DGP is demeaned and without time trend. I is a $n \times n$ identity matrix, ϵ_t is a martingale difference sequence with finite fourth moments, such that: $E(\epsilon_t\epsilon_t'|\epsilon_{t-1}, \epsilon_{t-2}, ..., \epsilon_1) = I$ and $\Theta(L) = \sum_{i=0}^{\infty} \Theta_i L^i$ is a matrix polynomial in the lag operator with $\sum_{i=0}^{\infty} i\,|\Theta_i| < \infty$; Θ_0 and $\Lambda \equiv \Theta(1)$ are invertible.

To obtain better approximations for the statistics of interest in small samples, we decompose Φ into its eigenvalues (Λ) and eigenvectors (V) such that $\Phi = V\Lambda V^{-1}$, and use a local-to-unity assumption:

$$\Lambda = I + \frac{1}{T}\mathbf{C} \tag{6}$$

where T is the sample size, $\mathbf{C}$ is a $(n+1)\times(n+1)$ diagonal matrix (diagonality rules out processes for $V^{-1}w_t$ that behave like I(2) in small samples, see Elliott (1998)). It is extremely convenient to decompose the DGP into its eigenvalues and eigenvectors in order to perform inference at long horizons. In fact, if $\Phi = V\Lambda V^{-1}$, then $\Phi^h = V\Lambda^h V^{-1}$.

This paper also assumes that, as $T\to\infty$, the horizon of prediction is a fraction of the sample size:

$$\frac{h}{T} \underset{T\to\infty}{\to} \delta \tag{7}$$

This condition is introduced here in order to obtain better approximations to the asymptotic distribution of the test statistics when the horizon of prediction is big relative to the sample size (see Richardson and Stock (1989), Stock (1996) and Phillips (1998)).

Under these assumptions, the behavior at long horizons is described by the sum of forecastable ($\widehat{w}_{t+h}$) and unforecastable ($e_{t,t+h}$) components:[5]

$$w_{t+h} = \underbrace{\Phi^h w_t}_{\widehat{w}_{t+h}} + \underbrace{\sum_{j=0}^{h-1}\Phi^j\Theta(I)\,\epsilon_{t+h-j}}_{e_{t,t+h}} + O_p(1) \tag{8}$$

A special case: Long horizon predictive regressions

An important special case of the above DGP that has been widely used in the long horizon predictive ability literature (see, among others, Campbell (2001), Campbell and Yogo (2002), Cavanagh, Elliott and Stock (1995) and Valkanov (2002)) is the following:

$$\begin{aligned} y_{1,t+1} &= \beta w_{2t} + \epsilon_{1,t+1} \\ b(L)^{-1}(1-\rho L)\,w_{2,t+1} &= \epsilon_{2,t+1} \end{aligned} \tag{9}$$

where $\epsilon_{1,t+1}$ and $\epsilon_{2,t+1}$ can be contemporaneously (but not serially) correlated. If we let $y_{1,t}$ be the first-difference of $w_{1,t}$, then (9) corresponds to (5) with $\Phi = [1,\beta;0,\rho]$ and $\Theta(L)$ block diagonal.[6]

In empirical applications, researchers are interested in estimating and finding confidence intervals for β_h in:

$$\Sigma_{j=1}^{h} y_{1,t+j} = \beta_h w_{2,t} + \xi_{t,h} \tag{10}$$

where $\beta_h \equiv \beta\Sigma_{j=0}^{h-1}\rho^j$, $\xi_{t,h} \equiv \beta\Sigma_{j=0}^{h-1}\left(\Sigma_{k=0}^{j-1}\rho^k\right) b(1)\,\epsilon_{2,t+h-j} + \Sigma_{j=0}^{h-1}\epsilon_{1t+h-j}$ and $\Sigma_{j=1}^{h} y_{1,t+h} = w_{1,t+h} - w_{1,t}$. When ρ is close to unity, it is difficult to find good approximations to the asymptotic distribution in small samples. The recent literature found interesting and useful results. Valkanov (2002) derived asymptotic distributions for test statistics for long horizon regressions and discussed

[5] See Rossi (2001) for a proof of this result.

[6] Note that $\Theta(L)$ is a diagonal matrix with 1 and $b(L)$ on the main diagonal. See the Appendix for a proof of (10).

them in a unifying framework. However, these asymptotic distributions need to be simulated conditional upon the observed DGP, and are computationally intensive. As a consequence, researchers routinely rely on the bootstrap (e.g. Cochrane and Piazzesi (2002)).

We construct confidence intervals with the correct coverage by using the fact that $\widetilde{\beta}_h \equiv \beta_h/T$ can be approximated by:

$$\widetilde{\beta}_\delta (c) \equiv \beta \int_0^\delta e^{cs} ds = \beta \frac{e^{c\delta} - 1}{c\delta} \tag{11}$$

For small $c\delta$, this is a monotone increasing function of c as $\beta \frac{e^{c\delta}-1}{\delta c} \simeq \beta (1 + c\delta/2)$ (by applying a Taylor expansion around $c = 0$). Given that the estimate of β from the first equation in (9), $\widehat{\beta}$, is consistent, the confidence interval for $\widetilde{\beta}_h$ is simply obtained by transforming a confidence interval for c as follows. Let the confidence interval for c be (c_0, c_1). This can be easily obtained by an ADF test on w_{2t} and using Stock (1991) tables. Then, the confidence interval for $\widetilde{\beta}_h$ is: $\left(\widehat{\beta}_\delta (c_0), \widehat{\beta}_\delta (c_1)\right) = \left(\widehat{\beta} \frac{e^{c_0\delta}-1}{c_0\delta}, \widehat{\beta} \frac{e^{c_1\delta}-1}{c_1\delta}\right)$. The advantage of this method is that it very easy to implement, as it does not require simulations, but only the use of critical values that are already available.[7] In this framework, our method has the correct size asymptotically and it is fast to use. Similarly, one can obtain a median unbiased estimate of $\widetilde{\beta}_\delta (c)$ by using a median unbiased estimate of c.

A statistic that plays an important role in predictive regressions is the coefficient of determination, or R^2. Our method can be used to construct confidence intervals for R^2 as well. Write (9) in its VAR companion form, and let A denote the VAR companion matrix (see Appendix 1). Then the population R^2, which depends upon c, is:

$$R_h^2 (c) = \frac{i_1' \overline{A}_{1,h} \left[\Sigma_{j=1}^{t-1} \overline{A}_{0,j} \Omega \overline{A}_{0,j}'\right] \overline{A}_{1,h}' i_1}{i_1' \left[\Sigma_{j=1}^{h-1} \overline{A}_{0,j} \Omega \overline{A}_{0,j}'\right] i_1 + i_1' \overline{A}_{1,h} \left[\Sigma_{j=1}^{t-1} \overline{A}_{0,j} \Omega \overline{A}_{0,j}'\right] \overline{A}_{1,h}' i_1} \underset{T\to\infty}{\to} R_\delta^2 (c) \tag{12}$$

where $\Omega = E(\epsilon\epsilon')$ and $\overline{A}_{p,q} \equiv \Sigma_{k=p}^q A^k$.[8] Thus, a confidence interval for the coefficient of determination can simply be obtained as $\left(R_h^2 (c_0) ; R_h^2 (c_1)\right)$ for any given estimate of all parameters other than ρ.

[7] One could also use methods other than Stock (1991) to obtain a confidence interval for c, for example Elliott and Jansson (2003), although the critical values of their test depend on a nuisance parameter, the coefficient of determination.

[8] In our simple example:
$R_\delta^2 (c) \equiv \frac{i_1' A_\delta \left(\int_{s=0}^\tau A_s \Omega A_s'\right) A_\delta' i_1}{i_1' \left(\int_{s=0}^\delta A_s \Omega A_s'\right) i_1 + i_1' A_\delta \left(\int_{s=0}^\tau A_s \Omega A_s'\right) A_\delta' i_1}$, where $A_\delta \equiv \begin{pmatrix} 0 & \beta \int_0^\delta e^{cs} ds \\ 0 & \beta \int_0^\delta e^{cs} ds \end{pmatrix}$, $A_s \equiv \begin{pmatrix} 1 & \beta \int_0^s e^{cj} dj \\ 0 & \beta \int_0^s e^{cj} dj \end{pmatrix}$.
We report the distribution in the appendix rather than in the main text because the researcher does not need it. It is important for our purposes that $R_h^2 (c)$ be a monotone function of c. This is shown in the Appendix.

4. The economic hypotheses

This section describes the economic hypotheses of interest. It shows that all of them impose constraints on Φ^h. We will denote such constraints by $q\left(\Phi^h\right)$, with a superscript that refers to the model that we are considering. Let $\mathbf{0}$ denote a (1×2) vector of zeros, $\mathbf{I}_1$ and $\mathbf{I}_2$ denote, respectively, the first and second columns of a (2×2) identity matrix. Finally, let E_t denote the expectation conditional on information at time t.

The *Uncovered Interest Rate Parity* (UIRP) states that the interest rate differential between two countries is the conditional expected value of the rate of depreciation of the high interest rate currency relative to the low interest rate currency. Let $w_t = [e_t, i_{t,h}]'$, where e_t is the logarithm of the bilateral nominal exchange rate at time t and $i_{t,h}$ is the continuously compounded zero-coupon h-period interest rate differential at time t. Then:[9]

$$UIRP: E_t\left(e_{t+h} - e_t\right) = \alpha + i_{t,h} \tag{13}$$

$$q^{UIRP}\left(\Phi^h\right) \equiv \mathbf{I}_1'\left(\Phi^h - I\right) - \mathbf{I}_2' = \mathbf{0} \tag{14}$$

The *Permanent Income Hypothesis* (PIH). Let $w_t = [y_t, s_t]'$, where is y_t income and s_t is savings, and let φ denote the discount factor (the inverse of one plus the interest rate). Then:

$$PIH: s_t = -\left(1-\varphi\right)\lim_{h\to\infty}\sum_{i=0}^{h}\varphi^i E_t\left(y_{t+i} - y_t\right) \tag{15}$$

$$q^{PIH}\left(\Phi^h\right) \equiv \mathbf{I}_1'\left(1 - \left(1-\varphi\right)\lim_{h\to\infty}\sum_{i=0}^{h}\left(\varphi\Phi\right)^i\right) - \mathbf{I}_2' = \mathbf{0} \tag{16}$$

The *Expectation Hypothesis of the Term Structure* (EH-TS) states that the long-term, n-period interest rate, $i_t^{(n)}$, equals the expected average of the short-term interest rate, $i_{t+jm}^{(m)}$, plus a constant term premium. Let $w_t = [i_t^{(m)}, i_t^{(n)}]'$; then:

$$EH-TS: i_t^{(n)} - i_t^{(m)} = a_n + \frac{1}{n/m}E_t\sum_{j=0}^{n/m-1}\left(i_{t+jm}^{(m)} - i_t^{(m)}\right) \tag{17}$$

$$q^{EH-TS}\left(\Phi^h\right) \equiv \mathbf{I}_1'\left(\lim_{n/m\to\eta}\sum_{i=0}^{n/m-1}\Phi^i\right) - \mathbf{I}_2' = \mathbf{0} \tag{18}$$

The *Granger Causality* (GC) implication. As discussed in Campbell (1987), a weak implication of the PIH for VARs is that savings must Granger cause future changes in income. If the two are

[9] Sometimes the constraint is written as:

$E_t\left(e_{t+h} - e_t\right) = \alpha + h\ i_{t,h}$, and justified by saying the the average change in the exchange rate is equal to the interest rate differential. However, this is not what the IRPC predicts (cfr. Obstfeld and Rogoff (1995)). Also, if covered interest parity holds, the UIRP is equivalent to the proposition that the logarithm of the forward exchange rate is an unbiased predictor of the logarithm of the future spot rate. Our method could be applied to this case as well.

independent ("$\perp$" denotes independence), then there is no evidence of Granger Causality:[10]

$$No\ GC : E_t\left(y_{t+h} - y_{t+1}\right) \perp s_t \tag{19}$$

$$q^{GC}\left(\Phi^h\right) \equiv \mathbf{I}_1'\left(\sum_{i=0}^{h}\Phi^i\right)\mathbf{I}_2 = 0 \tag{20}$$

The *Random Walk Hypothesis* (RWH) is a joint test of a unit root ($\Phi_{11} = 1$) and no GC ($\Phi_{12} = 0$). Thus, the null hypothesis is:

$$RW : E_t\left(y_{t+h} - y_{t+1}\right) \perp \{s_\tau, y_\tau\}_{\tau=1}^{t} \tag{21}$$

$$q^{RW}\left(\Phi^h\right) \equiv \mathbf{I}_1'\left(\Phi^h - I\right) = \mathbf{0} \tag{22}$$

All expectation hypotheses involve restrictions on Φ^h or some functions of it. That is, the restrictions depend on the largest root matrix of the process, which governs the behavior of the series at long horizons. Note that the DGP has two roots close to unity but in general the system is not I(1) (because the first difference of one variable depends on the level of the second variable, which is I(1), and, according to theory, in general this parameter may not be zero). Thus, Φ is not diagonal. Let V and Λ be the its eigenvector and eigenvalue matrices, respectively: $\Phi = V\Lambda V^{-1}$. As in Stock and Watson (1988), we assume that the roots are distinct, we let the real part of largest roots drive the long run behavior of the process,[11] and we model them as local to unity:

$$\Lambda = \mathrm{Re}\left(\Lambda\right) + i\,\mathrm{Im}\left(\Lambda\right);\ \ \mathrm{Re}\left(\Lambda\right) = \begin{pmatrix} e^{c_1/T} & 0 \\ 0 & e^{c_2/T} \end{pmatrix} \tag{23}$$

Thus, at long horizons:

$$\mathrm{Re}\left(\Lambda^h\right) \underset{h/T\to\delta}{\to} \begin{pmatrix} e^{c_1\delta} & 0 \\ 0 & e^{c_2\delta} \end{pmatrix} \equiv \Lambda_{C,\delta} \tag{24}$$

As long as we allow the same deviation from a unit root for both real parts associated with the same complex conjugates pair, then our approximation will be real as well:[12]

$$\Phi^h = V\Lambda^h V^{-1} = V\left(\mathrm{Re}\left(\Lambda\right) + i\,\mathrm{Im}\left(\Lambda\right)\right)^h V^{-1} \tag{25}$$

We will approximate the situation in which the researcher is interested in an infinite horizon as one in which $\delta = 1$.

Note that both the eigenvectors and the eigenvalues can be consistently estimated. In fact, say that there are two roots close to unity. We know that $T\left(\widehat{\Phi} - \Phi\right) = O_p\left(1\right)$ and $T\left(\widehat{\Lambda} - \Lambda\right) = O_p\left(1\right)$

[10] Unless savings is itself an exact linear function of current and lagged income. Note that, if (20) holds, the system is I(1).

[11] For example, this rules out nonstationary seasonality.

[12] Note that we construct a confidence region for $\mathrm{Re}(\Lambda)$, we add the imaginary part, and only after that we take the exponential and multiply it by the eigenvectors. This ensures that the result is real.

Thus, $T\left(\widehat{\Phi}-\Phi\right)=T\left(\widehat{V}\widehat{\Lambda}\widehat{V}^{-1}-V\Lambda V^{-1}\right)=\widehat{V}\widehat{\Lambda}\left(\widehat{V}^{-1}-V^{-1}\right)+\widehat{V}\left(\widehat{\Lambda}-\Lambda\right)V^{-1}+\left(\widehat{V}-V\right)\Lambda V^{-1}$ so $T\left(\widehat{V}-V\right)=O_p(1)$. However, Λ^h is not, and it drives the asymptotic distribution.

5. Some small Monte Carlo illustrative examples

We compare the performance of the method proposed in this paper with the methods usually used in the literature. A strand of the literature uses Wald tests, assuming that the eigenvalues in the process $(\Delta y', x')'$ are less than one in absolute value and $h \to \infty$ (as in Campbell (1987) and Sheffrin and Woo (1990)). Under these assumptions, the limiting distribution of the Wald test statistic under the null hypothesis is a chi-square. We expect this asymptotic distribution to provide good approximations when x is stationary and y has an exact unit root. More recently, people have been interested in inference for h large but finite (e.g. Bekaert and Hodrick (2001)). In this case, the constraint involves non-linear restrictions.

In a first Monte Carlo experiment, we consider the case in which the researcher knows that there is one root equal to one (as in Cavanagh, Elliott and Stock (1985) and Valkanov (2002)):[13]

$$y_t = Ay_{t-1} + \epsilon_t \tag{26}$$

where $A \equiv \begin{pmatrix} 0 & \beta \\ 0 & \rho \end{pmatrix}$, $\beta = 1$, $y_t = (\Delta y_{1t}, y_{2t})'$, $\rho = \left(1+\frac{c}{T}\right)$ and $\epsilon_t \sim N(0, I)$. Thus, $y_{t+h} = A^h y_t + \Sigma_{j=0}^{h-1} A^j \epsilon_{t+h-j}$. Note that $A^h = \begin{pmatrix} 0 & \beta\rho^{h-1} \\ 0 & \rho^h \end{pmatrix}$ and that $E_t(y_{1t+h} - y_{1t+1}) = \Sigma_{j=1}^{h} E_t \Delta y_{1t+j} = \gamma_h y_{2t}$ for $\gamma_h = \beta\left(\Sigma_{j=1}^{h}\rho^{j-1}\right)$.

We consider the following cases:

(1) *Predictive regressions and UIRP*. Let $y_t = (\Delta e_t, i_t)'$. The restrictions imposed by the EH-UIRP, which we write as in Bekaert and Hodrick (2001) as $\frac{1}{h}\Sigma_{j=1}^{h} E_t \Delta y_{1t+j} = y_{2t}$, become:

$$\gamma_{h,0}^{UIRP} \equiv \beta\left(\frac{1}{h}\Sigma_{j=1}^{h}\rho^{j-1}\right) = 1 \tag{27}$$

where $\gamma_{h,0}^{UIRP}$ is exactly $\widetilde{\beta}_h/\delta$ in (10). Thus, we can use the approximation in (11).

We compare the performance of our method (referred to as "*big h*") with the Wald test used in the literature. Results are reported in Table 1. The table reports, for various values of ρ and δ, the value of $\widetilde{\beta}_h/\delta$ along with the percentage of times the true value of $\widetilde{\beta}_h$ lies outside its confidence interval, either below (labeled "*big* h_L") or above (labeled "*big* h_U"). Ideally, these percentages should be 0.05. It also reports their sum (labeled "*big h*") and the percentage of times the Wald test

[13] These authors allow for serial correlation in ϵ_{2t} (but not in ϵ_{1t}); for simplicity we don't, but the results would be the same (provided one takes into account the effects of the serial correlation).

with or without a Newey and West (1987) correction for serial correlation rejects the null hypothesis (labeled "*Wald*" and "$Wald_{NW}$" respectively). Ideally, these percentages should be 0.10. The Wald test involves non-linear constraints, and we used numerical derivatives to evaluate the gradient of the constraint. We used critical values corresponding to a χ^2 distribution with the number of degrees of freedom equal to the number of restrictions. The sample size is 200, the number of Monte Carlo replications is 1,000 and $\beta = 1$.

The table shows that at small horizons the Wald test rejection rate is close to nominal, but, as the horizon increases, rejection rates increase. For example, when $\rho = 0.99$, the Wald test rejects 22% of the times at a horizon of 10 quarters (i.e. $\delta = 0.05$), 30% at horizons of 20 quarters (i.e. $\delta = 0.10$) and even more at longer horizons. On the other hand, while the method proposed in this paper has the correct coverage at longer horizons, it has bad coverage at short horizons, due to the nature of the approximation. Thus, as the horizon increases, our approximation works better than the usual chi-square distribution of large sample Wald tests (which has the correct size only for very small horizons). Comparing columns "*big h*" and "*Wald*", the asymptotic distribution proposed in this paper is better in terms of coverage/size for horizons bigger than 10 ($\delta = 0.15$).[14]

Note that, as the process becomes more stationary (i.e. as ρ becomes smaller), the performance of the Wald test worsens. This is to be expected, since the Wald test relies on an exact unit root (and, as a consequence, it uses variables in first differences), so, the more the value of the root is different from one, the bigger the size distortions of the Wald test will be. As a consequence, our method is better than Wald tests even at smaller horizons (for example, when $\rho = 0.95$ it is better for horizons equal or bigger than 6 ($\delta = 0.03$). The table also shows the coverage of confidence intervals for R^2. The pattern is similar to the one described above.

(2) *Granger causality.* Let $y_t = (\Delta Y_t, s_t)'$. The restriction imposed by PIH models for the question: "does savings GC output?" is: $E_t\left(Y_{t+h} - Y_{t+1}\right) = \beta\left(\Sigma_{j=1}^{h}\rho^{j-1}\right)s_t$, which becomes:

$$\gamma_{h,0}^{GC} \equiv \frac{1}{\rho}\beta\left(\Sigma_{j=1}^{h}\rho^{j-1}\right) = 0 \tag{28}$$

thus $\frac{1}{h}\gamma_{h,0}^{GC} \to \gamma_{\delta,0}^{GC} = \beta\frac{1}{\delta}\int_0^{\delta} e^{cs}ds$. We assume that $\beta = 0$ (so that there is no detectable one-period ahead Granger causality) and explore what happens as h increases.

(3) *PIH.* Let again $y_t = (\Delta Y_t, s_t)'$. The restriction imposed by PIH models is (see Campbell, 1987): $s_t = -\Sigma_{j=1}^{\infty}\varphi^i E_t \Delta Y_{t+j} = -\Sigma_{j=1}^{\infty}\varphi^i\beta\rho^{j-1}s_t$ so that:

$$\gamma_{\infty,0}^{PIH} \equiv -\beta\frac{1}{\rho}\left(\Sigma_{j=1}^{\infty}\varphi^j\rho^j\right) = 1 \tag{29}$$

[14] Results obtained by constructing confidence intervals based on (11) and those obtained by simulating (11) over a confidence set for c and then taking its maximum and minimum values (which would be more appropriate if we were not sure about monotonicity) gave very similar results in terms of coverage for $\delta = 0.10$ and higher. Also, the coverage of confidence intervals based on $\beta\frac{e^{c\delta}-1}{\delta c}$ and on approximating the integral $\frac{1}{\delta}\int_0^{\delta} e^{cs}ds$ were almost identical.

Let φ be approximated by e^{-r}.[15] Thus, $\gamma_{\infty,0}^{PIH} \rightarrow -\beta \int_0^{\delta} e^{(c-r)s} ds$. We generate the DGP under the null hypothesis and then compare the coverage of normal sampling Wald tests with that of our method.

INSERT TABLE 1

In a second Monte Carlo experiment, we use the eigenvalue/eigenvector decomposition (25) with one root close to unity and the other much smaller than unity (this is a system where one variable is in first differences and the other is in levels):

$$y_t = A y_{t-1} + V \epsilon_t \tag{30}$$

where the DGP is calibrated on a bivariate VAR for Δs_{t+1} and i_{t+1}^{diff} : $V = \begin{pmatrix} 1 & 0.95 \\ 0.03 & 1 \end{pmatrix}$, $\Lambda_{(\rho)} \equiv \begin{pmatrix} 0 & 0 \\ 0 & \rho \end{pmatrix}$, $A = V \Lambda_{(\rho)} V^{-1}$, $y_t = (y_{1t}, y_{2t})'$, $\rho = \left(1 + \frac{c}{T}\right)$ and $\epsilon_t \sim N(0, I)$. Noting that

$$V^{-1} y_t = \Lambda_{(\rho)} V^{-1} y_{t-1} + \epsilon_t \tag{31}$$

and that the root close to unity enters the second rows only, we construct a confidence interval for ρ by inverting ADF for $i_2' V^{-1} y_t$. By construction, we impose that the residuals of (31) are uncorrelated; if this were not the case, one should simply correct for that, as in Stock and Watson (1988). A^h is approximated by $V \Lambda_{(1)} V^{-1} \rho^h$ and we report the confidence interval for $A_{12} = \rho^h i_1' V \Lambda_{(1)} V^{-1} i_2'$.[16]

INSERT TABLE 2

6. Empirical evidence

The previous section showed that the chi-squared asymptotic distribution rejects the null hypothesis too often relative to our method. Thus, in this section, we apply our method to a variety of empirical applications to analyze whether we find more favorable evidence for Expectations Hypotheses models and for predictive ability at long horizons. Appendix 2 provides a detailed description of the data.

We first provide some summary statistics in Table 3. Notice the high persistence of the time series variables.

[15] By using standard Taylor approximations, $ln(1+r) \approx r$ for small r. Hence $-ln(1+r) = ln(\varphi)$ and, by taking exponentials, $e^{-r} \approx \varphi$. For simplicity, as in Campbell and Shiller (1987), we assume that r is the average value of the interest rate and it is known. However, parameter estimation error in its estimation will be asymptotically relevant.

[16] Pesavento and Rossi (2003) explore how the method works if there is more than one root close to unity.

INSERT TABLE 3

First, we present some evidence on predictive regressions, with the empirical application used in Fama and Bliss (1987) and Cochrane and Piazzesi (2003). Note the downward bias in R^2 that, under the assumptions of this paper, is due to the serial correlation in the forward-spot differential, whose estimate is 0.90. As a consequence, the confidence interval for R^2 is wide and its upper bound can be as high as 50%.

INSERT TABLE 4

Second, we consider the empirical evidence on the EH-TS from the McCulloch and Kwon (1993) database.[17] We consider a bivariate VAR with the short-term interest rate in levels and the spread between the short rate and a long-term interest rate at different maturities. Figure 1 shows the values of the smallest and largest eigenvalues in this bivariate VAR as a function of the maturity of the long-term interest rate. The high persistence at long horizons (say, bigger than 50 months) is clearly visible.

INSERT FIGURE 1 AND TABLE 5

We apply the method proposed in this paper, as well as the standard Wald test used in the literature. Table 5 reports the results. Note that our method rejects the null hypothesis suggested by economic theory, as the confidence interval for the parameter does not contain 1 at any horizon. The Wald test also rejects the null hypothesis (the value of the statistic is always bigger than 4.61, its critical value). Similar conclusions hold for the CA and the PIH models, as tables 6 and 7 show.

INSERT TABLE 6 AND 7

Finally, let us consider the UIRP. Table 8 reports the confidence intervals based on our method (labeled "*big h*") and the results of a Wald test. Both methods strongly reject the restrictions imposed by the economic model. Thus, while we found in the Monte Carlo simulations that Wald tests may over-reject the null hypothesis, we are unable not to reject it by using our method. The problem is that there is too little information content in the VAR regression (26), and the estimate of β is too small to provide any significant predictive content.

[17] We show results based on a sample from 1946:12 to 1985:8. Results based on the whole McCullom and Kwon database (until 1991:2) are very similar and are not reported.

INSERT TABLE 8

6. Conclusions

The methods discussed in this paper apply to situations in which the researcher is interested in evaluating predictive regressions, expectations hypotheses and present value models at long horizons in the presence of small sample sizes. We show that inference at long-horizons depends crucially on the highest roots of the process. Most economic time-series variables are highly persistent, but it is difficult to distinguish between an exact unit root and a mean-reverting, highly persistent process in small samples. This leads to problematic size distortions in Wald tests. This paper provides a method for dealing with this problem, and applies it to test the empirical validity of well-known core economic theories.

High persistence is a feature of most time series variables used in all areas of macroeconomics, finance and international macroeconomics. Thus, the methods proposed in this paper can more generally be applied to many other economically interesting applications, among which impulse-response functions (as in Pesavento and Rossi (2003)), testing forecastable co-movements in Real Business Cycle models, and so forth.

References

Bekaert, G. and R. J. Hodrick (2001): "Expectation Hypotheses Tests", *Journal of Finance* 56(4), pp. 115-38.

Bekaert, G., M. Wei and Y. Xing (2002): "Uncovered Interest Rate Parity and the Term Structure", *NBER Working Paper* No. 8795.

Campbell, J. Y. (1987): "Does Saving Anticipate Declining Labor Income? An Alternative Test of the Permanent Income Hypothesis", *Econometrica* 55(6), pp. 1249-73.

Campbell, J. Y. (2001): "Why long horizons? A study of power against persistent alternatives", *Journal of Empirical Finance* 8, pp. 459-491.

Campbell, J. Y. and R. Shiller (1987): "Cointegration and tests of present value models", *Journal of Political Economy* 87, 95(5), pp. 1062-1088.

Campbell, J. Y., and R. Shiller (1991): "Yield spreads and interest rate movements: a bird's eye view", *Review of Economic Studies* 58(3), pp. 495-514.

Campbell, J.Y., and M. Yogo (2002): "Efficient Tests of Stock Return Predictability", *mimeo,* Harvard University.

Cavanagh, C. L., G. Elliott and Stock, J. H. (1995): "Inference in Models with nearly Integrated Regressors", *Econometric Theory* 11, pp. 1131-47.

Clark, T. and M. McCracken (2002): " Evaluating Long Horizon Forecasts", *mimeo*, Federal Reserve Bank of Kansas City.

Cochrane, J. and M. Piazzesi (2002): "Bond Risk Premia", *NBER W.P. 9178.*

Diebold, F. X. and C. Li (2002): "Forecasting the Term Structure of Government Bond Yields", *mimeo*, University of Pennsylvania.

Elliott, G. (1998): "On the Robustness of Cointegration Methods when Regressors almost have Unit Roots", *Econometrica.*

Elliott, G. and M. Jansson (2003): "Testing for Unit Roots with Stationary Covariates", *Journal of Econometrics* 115, pp. 75-89.

Engel, C. and K. West (2002): "Exchange Rates and Fundamentals", *mimeo*, University of Wisconsin-Madison.

Fama, E. and R. Bliss (1987): "The Information in Long-Maturity Forward Rates", *American Economic Review* 77, pp. 680-92.

Kilian, L. (1999): "Exchange Rates and Monetary Fundamentals: What Do We Learn From Long-Horizon Regressions?", *Journal of Applied Econometrics 14(5)*, pp. 491-510.

Mark, N. C. (1995): "Exchange Rates and Fundamentals: Evidence on Long-Horizon Predictability", *American Economic Review*, 85(1), pp. 201-218.

Mark, N. C. and D. Sul (2002): "Asymptotic power advantages of long-horizon regression tests", *mimeo.*

McCulloch, J. and H. Kwon (1993): "U.S. Term Structure Data, 1947-1991", Working Paper 93-6, Ohio State University.

Newey, W. and K. West (1987): "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix", *Econometrica* 55, 1987, pp. 703-708.

Obstfeld, M. and K. Rogoff (1996): *Foundations of International Macroeconomics*, MIT Press.

Pesavento, E. and B. Rossi (2003): "Small sample confidence intervals for multivariate impulse response functions at long horizons", *mimeo.*

Phillips, P. C. B. (1987): "Towards a Unified Asymptotic Theory for Autoregression", *Biometrika*, 74(3), pp.535-47.

Phillips, P. C. B. (1998): "Impulse Response and Forecast Error Variance Asymptotics in Nonstationary VARs", *Journal of Econometrics.*

Richardson, M. and J. H. Stock (1989): "Drawing Inferences from Statistics Based on Multiyear Returns", *Journal of Financial Economics.*

Rossi, B. (2002): "Testing Long-Horizon Predictive Ability, and the Meese-Rogoff Puzzle", *Duke University Working Paper 02-10.*

Sheffrin, S. M., and W. T. Woo (1990): "Present value tests of an intertemporal model of the current account", *Journal of International Economics* 29, pp. 237-53.

Stambaugh, R. (1999): "Predictive regressions", *Journal of Financial Economics* 54, pp. 375-421.

Stock, J. H. (1996): "VAR, Error Correction and Pretest Forecasts at Long Horizons", *Oxford Bullettin of Economics and Statistics*, 58(4).

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

Stock, J. H. and M. W. Watson (1988): "Testing for Common Trends", *Journal of the American Statistical Association*, 83(404), pp. 1097-1107.

Stock, J. H., and M. W. Watson (1996): "Confidence sets in regressions with highly serially correlated regressors", *mimeo.*

Valkanov, R. (2002): "Long-horizon regressions: theoretical results and applications", *mimeo*, UCLA.

Appendix 1. Proofs.

Proof of (10).

Rewrite (9) as:

$$\begin{aligned}\Delta w_{1,t+1}-\beta w_{2t} &= \epsilon_{1,t+1}\\ (1-\rho L)\,w_{2,t+1} &= b(L)\,\epsilon_{2,t+1}\end{aligned}$$

that is:

$$\underbrace{\begin{pmatrix}1-L & -\beta L\\ 0 & 1-\rho L\end{pmatrix}}_{\equiv(I-\Phi L)}\begin{pmatrix}w_{1,t+1}\\ w_{2,t+1}\end{pmatrix}=\underbrace{\begin{pmatrix}1 & 0\\ 0 & b(L)\end{pmatrix}}_{\equiv\Theta(L)}\underbrace{\begin{pmatrix}\epsilon_{1,t+1}\\ \epsilon_{2,t+1}\end{pmatrix}}_{\equiv\epsilon_{t+1}}$$

where $\Phi=\begin{pmatrix}1 & \beta\\ 0 & \rho\end{pmatrix}$. Note that $y_{1,t+1}=\Delta w_{1,t+1}$ so that $\Sigma_{j=1}^{h}y_{1,t+j}=\Sigma_{j=1}^{h}\Delta w_{1,t+j}=w_{1,t+h}-w_{1,t}$. By applying (8):

$$\begin{pmatrix}w_{1,t+h}\\ w_{2,t+h}\end{pmatrix}=\Phi^{h}\begin{pmatrix}w_{1,t}\\ w_{2,t}\end{pmatrix}+\Sigma_{j=0}^{h-1}\Phi^{j}\Theta(I)\,\epsilon_{t+h-j}+O_p(1)\,.$$

Note also that $\Phi^{j}=\begin{pmatrix}1 & \beta\Sigma_{k=0}^{j-1}\rho^{k}\\ 0 & \rho^{j}\end{pmatrix}$; thus:

$$\begin{pmatrix}w_{1,t+h}\\ w_{2,t+h}\end{pmatrix}=\begin{pmatrix}1 & \beta\Sigma_{k=0}^{h-1}\rho^{k}\\ 0 & \rho^{h}\end{pmatrix}\begin{pmatrix}w_{1,t}\\ w_{2,t}\end{pmatrix}+\Sigma_{j=0}^{h-1}\begin{pmatrix}1 & \beta\Sigma_{k=0}^{j-1}\rho^{k}\\ 0 & \rho^{j}\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & b(1)\end{pmatrix}\epsilon_{t+h-j}+O_p(1)\,.$$

so that:

$$\begin{aligned}w_{1,t+h}-w_{1,t} &= \underbrace{\beta\Sigma_{k=0}^{h-1}\rho^{k}}_{\equiv\beta_h}w_{2,t}+\underbrace{\Sigma_{j=0}^{h-1}\left(\epsilon_{1,t+h-j}+\beta b(1)\,\Sigma_{k=0}^{j-1}\rho^{k}\epsilon_{2,t+h-j}\right)}_{\equiv\xi_{t,h}}\\ w_{2,t+h} &= \rho^{h}w_{2,t}+\Sigma_{j=0}^{h-1}\rho^{j}b(1)\,\epsilon_{2,t+h-j}\end{aligned}$$

The asymptotic distribution of R^2 in the predictive regression, eq. (12).

The DGP: $(I-\Phi L)\,w_t=\Theta(L)\,\epsilon_t$, $\Theta(L)\equiv(I+\Theta_1 L+...+\Theta_p L^p)$, $\Phi=\begin{pmatrix}0 & \beta\\ 0 & \rho\end{pmatrix}=B\rho$, $B\equiv\begin{pmatrix}0 & \beta/\rho\\ 0 & 1\end{pmatrix}$, $\Theta_i=\begin{pmatrix}1 & 0\\ 0 & \Theta_{22,i}(L)\end{pmatrix}$. As $\sum_{k=1}^{h}w_{t+k}=\sum_{k=1}^{h}\Phi^{k}w_t+\sum_{k=1}^{h}\sum_{j=0}^{k-1}\Phi^{j}\Theta(I)\,\epsilon_{t+k-j}+$

$O_p(h)$, we consider only the first two components after the equality, which are the asymptotically relevant ones. Let $\widetilde{\epsilon}_t \equiv \Theta(I)\epsilon_t$ and $Var(\widetilde{\epsilon}_t) = \Theta(I)\Sigma_\epsilon \Theta(I)' \equiv \Omega$.

$$\sum_{k=1}^{h}\sum_{j=0}^{k-1}\Phi^j\Theta(I)\epsilon_{t+k-j} = \widetilde{\epsilon}_{t+1}+$$
$$\widetilde{\epsilon}_{t+2}+\Phi\widetilde{\epsilon}_{t+1}+$$
$$+\widetilde{\epsilon}_{t+k}+\Phi\widetilde{\epsilon}_{t+k-1}+\Phi^2\widetilde{\epsilon}_{t+k-1}+...+\Phi^{h-1}\widetilde{\epsilon}_{t+1}=$$
$$=\widetilde{\epsilon}_{t+k}+(I+\Phi)\widetilde{\epsilon}_{t+k-1}+\left(I+\Phi+\Phi^2\right)\widetilde{\epsilon}_{t+k-2}+...+\left(I+\Phi+..+\Phi^{h-1}\right)\widetilde{\epsilon}_{t+1}$$
$$Var\left(\sum_{k=1}^{h}\sum_{j=0}^{k-1}\Phi^j\Theta(I)\epsilon_{t+k-j}\right) = \Omega+(I+\Phi)\Omega(I+\Phi)'+...+\left(I+\Phi+..+\Phi^{h-1}\right)\Omega\left(I+\Phi+..+\Phi^{h-1}\right)' =$$
$$=\sum_{k=1}^{h-1}\left(\sum_{l=0}^{k}\Phi^l\right)\Omega\left(\sum_{l=0}^{k}\Phi^l\right)' \equiv \overline{\Omega}_{h-1}$$

$$Var\left(\sum_{k=1}^{h}\Phi^k w_t\right) = \left(\sum_{k=1}^{h}\Phi^k\right)Var(w_t)\left(\sum_{k=1}^{h}\Phi^k\right), \text{ where:}$$
(a) $Var(w_t) = Var\left(\sum_{k=1}^{t}\sum_{j=0}^{k-1}\Phi^j\Theta(I)\epsilon_{t-j}\right) = \Omega+(I+\Phi)\Omega(I+\Phi)'+...$
$+\left(I+\Phi+..+\Phi^{t-1}\right)\Omega\left(I+\Phi+..+\Phi^{t-1}\right)' \equiv \overline{\Omega}_{t-1}$
(b) $\Phi^k = \rho^k B$ implies that $\sum_{k=1}^{h}\Phi^k = \left(\sum_{k=1}^{h}\rho^k\right)B$, so that:
$$Var\left(\sum_{k=1}^{h}\Phi^k w_t\right) = \left(\sum_{k=1}^{h}\rho^k\right)^2 B\overline{\Omega}_{t-1}B'$$

Thus the coefficient of determination for the first regression is:
$$R_h^2 = \frac{I_1'\left(\sum_{k=1}^{h}\rho^k\right)^2 B\overline{\Omega}_{t-1}B'I_1}{I_1'\left(\sum_{k=1}^{h}\rho^k\right)^2 B\overline{\Omega}_{t-1}B'I_1+I_1'\overline{\Omega}_{h-1}I_1} = \left(1+\frac{I_1'\overline{\Omega}_{h-1}I_1}{I_1'\left(\sum_{k=1}^{h}\rho^k\right)^2 B\overline{\Omega}_{t-1}B'I_1}\right)^{-1}$$
As $T\to\infty$, note that: $\sum_{l=0}^{m-1}\Phi^l = \left(I+\Phi+..+\Phi^{m-1}\right) = \begin{pmatrix} 1 & \beta\sum_{k=0}^{m-1}\rho^k \\ 0 & \sum_{k=0}^{m}\rho^k \end{pmatrix}$
$$\underset{T\to\infty}{\to}\begin{pmatrix} 1 & \beta T\int_0^{[(m-1)/T]}e^{cs}ds \\ 0 & T\int_0^{[m/T]}e^{cs}ds \end{pmatrix}$$
$$\overline{\Omega}_{n-1} = \sum_{k=1}^{n-1}\left(\sum_{l=0}^{k}\Phi^l\right)\Omega\left(\sum_{l=0}^{k}\Phi^l\right)'$$
$$= T\int_0^{[(n-1)/T]}\begin{pmatrix} 1 & \beta T\int_0^{[(m-1)/T]}e^{cs}ds \\ 0 & T\int_0^{[m/T]}e^{cs}ds \end{pmatrix}\Omega\begin{pmatrix} 1 & \beta T\int_0^{[(m-1)/T]}e^{cs}ds \\ 0 & T\int_0^{[m/T]}e^{cs}ds \end{pmatrix}' dm =$$
$$= T\int_0^{[(n-1)/T]}\begin{pmatrix} 1 & \beta T\int_0^{[(m-1)/T]}e^{cs}ds \\ 0 & T\int_0^{[m/T]}e^{cs}ds \end{pmatrix}\Omega\begin{pmatrix} 1 & \beta T\int_0^{[(m-1)/T]}e^{cs}ds \\ 0 & T\int_0^{[m/T]}e^{cs}ds \end{pmatrix}' dm$$

Let $\psi_{(m)} = \int_0^{[(m-1)/T]}e^{cs}ds$
$$I_1'\left(\sum_{k=1}^{h}\rho^k\right)^2 B\overline{\Omega}_{t-1}B'I_1 = \left(\sum_{k=1}^{h}\rho^k\right)^2 I_1'B\overline{\Omega}_{t-1}B'I_1 =$$
$$= \left(\sum_{k=1}^{h}\rho^k\right)^2 T\int_0^{[(n-1)/T]}I_1'\begin{pmatrix} 0 & \beta/\rho \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & \beta T\psi_{(m)} \\ 0 & T\psi_{(m)} \end{pmatrix}\times$$
$$\times\begin{pmatrix} \omega_{11} & \Omega_{12} \\ \Omega_{12}' & \Omega_{22} \end{pmatrix}\begin{pmatrix} 1 & 0 \\ \beta'\psi_{(m)}T & T\psi_{(m)} \end{pmatrix}\begin{pmatrix} 0 & 0 \\ \beta'/\rho & 1 \end{pmatrix}I_1 =$$

$$= \left(\sum_{k=1}^{h} \rho^k\right)^2 T \int_0^{[(n-1)/T]} \left(T^2 \frac{\psi^2_{(m)}}{\rho^2} \beta\Omega_{22}\beta'\right) dm$$

$$I_1' \overline{\Omega}_{h-1} I_1 = I_1' T \int_0^{[(n-1)/T]} \begin{pmatrix} 1 & \beta T\psi_{(h)} \\ 0 & T\psi_{(h)} \end{pmatrix} \times$$
$$\times \begin{pmatrix} \omega_{11} & \Omega_{12} \\ \Omega_{12}' & \Omega_{22} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ T\psi_{(h)}\beta' & T\psi_{(h)} \end{pmatrix} I_1 dh =$$
$$= T \int_0^{[(n-1)/T]} \left(\omega_{11} + \beta T\psi_{(h)}\Omega_{12}' + \left(\Omega_{12} + \beta T\psi_{(h)}\Omega_{22}\right) T\psi_{(h)}\beta'\right)$$
$$= T \int_0^{[(n-1)/T]} T^2 \psi^2_{(h)} \left(\beta\Omega_{22}\beta'\right) dh$$

$$R_h^2 = \left(1 + \frac{\int_0^{[(n-1)/T]} \psi^2_{(h)} \left(\beta\Omega_{22}\beta'\right) dh}{\left(\sum_{k=1}^{h} \rho^k\right)^2 \int_0^{[(n-1)/T]} \left(\frac{\psi^2_{(m)}}{\rho^2} \beta\Omega_{22}\beta'\right) dm}\right)^{-1} \simeq \left(1 + \frac{1}{\frac{1}{[(n-1)/T]} \int_0^{[(n-1)/T]} \psi^2_{(m)} dm}\right)^{-1}$$

The above formula clearly shows that R^2 is a monotone function of c.

To implement the test, note that the DGP can be rewritten as $\Theta(L)^{-1}(I - \Phi L) w_t = \epsilon_t$ where $\Theta(L)^{-1}$ can be approximated by $\left(I - \Theta_1 L - \Theta_2 L^2 - ... - \Theta_p L^p\right)$. Thus:

$$\begin{pmatrix} w_t \\ w_{t-1} \\ ... \\ w_{t-p+1} \end{pmatrix} = \underbrace{\begin{pmatrix} A_1 & A_2 & ... & A_p \\ I & \mathbf{0} & ... & \mathbf{0} \\ ... & ... & ... & ... \\ \mathbf{0} & ... & I & \mathbf{0} \end{pmatrix}}_{\equiv A} \begin{pmatrix} w_{t-1} \\ w_{t-2} \\ ... \\ w_{t-p} \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ \mathbf{0} \\ ... \\ \mathbf{0} \end{pmatrix}$$

where $A_1 = -(\Phi + \Theta_1)$, $A_j = (\Theta_{j-1}\Phi + \Theta_j)$ for $j = 2, ..p-1$, $A_p = \Theta_{p-1}\Phi$ and where $\Theta(L)$ can be consistently estimated by a VAR regression of Δw_t onto its lagged values.

Appendix 2

The data used in this paper are the following. For Predictive Regressions, we use the data in Cochrane and Piazzesi (2002), available on John Cochrane website. The original monthly data are from CRSP, the sample is from 1964:1 to 2002:12.

For the EHTS, we use the McCulloch and Kwon (1993) quarterly database, 1946:12-1985:8. Results using the full McCulloch and Kwon database (1946:12-1991:2) were similar and are not reported.

For the CA, we use the same data as in Sheffrin and Woo (1990), obtained from Datastream. In order to explain the mnemonics, let us focus on the U.S. data. The net output series is GDP minus investment and government purchases, and subsequently transformed in real per-capita terms by dividing by the GDP deflator and population. The CA is GNP minus consumption, investment and government spending, and subsequently transformed in real per-capita terms by dividing by the GNP deflator and population. We follow Sheffrin and Woo (1990) in the decision not to use CA data, as they require an arbitrary allocation of some "errors and omissions". Data are quarterly, from 1978Q4 to 2001Q4, limited only by data availability. Mnemonics are as follows: USGDP...B (GDP), USGNP...B (GNP), USGFCF..B (investment), USCNGOV.B (government consumption), USGDP..CE (GDP deflator), USGNP..CE (GNP deflator), USCNPER.B (consumption), USPOPTOT (population, annual data, linearly interpolated, mid-year estimate attributed to second quarter). For Canada and UK we use the same mnemonics (starting with "CN" and "UK") except that the deflator is - - GDPIPDE (we could not find GNP deflator). For Canada, data are from 1981Q1 to 2002Q4 and for UK 1968Q1 to 2001Q2. We used a discount factor $R = e^{-0.04}$.

For the PIH, we use quarterly data from the Bureau of Economic Analysis, NIPA tables, from 1946Q1 to 2002Q1. We follow Campbell (1987) and construct labor income by adding wage and salary disbursement (line 2 in Table 2.1, Personal Income and Its Disposition), other labor income (line 9) and adding a fraction (equal to the ratio of labor income as above relative to total disposable income, line 1) of Proprietors' income (line 10). We subtracted Personal tax and nontax payments (line 24) and Personal contributions for social insurance (line 23). We divided by the population (line 34) and by the Personal consumption expenditure chain-type price index (line 24 in Table 7.4. Chain-Type Quantity and Price Indexes for Personal Consumption Expenditures by Major Type of Product). Nominal data are in thousands of U.S. Dollars. We obtained a measure of nondurable consumption by subtracting expenditure on clothing and shoes (line 8 in Table 2.2. Personal Consumption Expenditures by Major Type of Product) from nondurable goods consumption (line 6 in the same table). We obtained per capita real variables as above, but using the Nondurable goods deflator (line 29 in Table 7.4. Chain-Type Quantity and Price Indexes for Personal Consumption Expenditures by Major Type of Product).

For the UIRP, we use monthly (logarithm) exchange rates and Eurocurrency interest rates from Datastream. The dollar-based exchange rates are calculated from the quoted sterling exchange rates with are closing middle rates provided by Reuters (all data before 31.12.93 are from the Financial Times). The mnemonics are as follows. The 3-month German, U.K. and U.S. interest rates are, respectively, ECWGM3M, ECUKP3M and ECUKP3M; the German and U.S. exchange rates in terms of the U.K. Pound (which has a nominal value of 1) are DMARKER and USDOLLR. The sample period is February 1975 to April 2002. We also use the same data as in Bekaert, Wei and Xing (2002); the data were kindly provided by the authors.

Figure 1

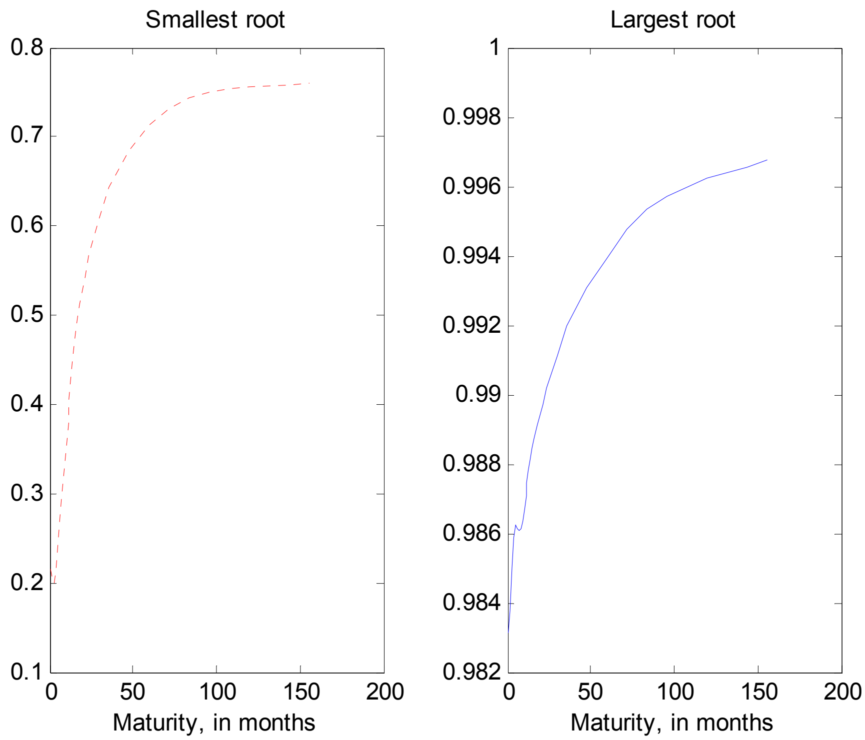
Smallest root
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
50
100
150
200
Maturity, in months
Largest root
1
0.998
0.996
0.994
0.992
0.99
0.988
0.986
0.984
0.982
0
50
100
150
200
Maturity, in months

Table 1: Monte Carlo results

ρ	h	$\widetilde{\beta}_h/\delta$	$Big\ h_L$	$Big\ h_U$	$Big\ h$	$Wald_{NW}$	$Wald$	R^2_L	R^2_U
0.99	1	1.00	0.45	0.30	0.75	0.10	0.10	0.06	0.06
	2	1.00	0.36	0.20	0.56	0.14	0.13	0.06	0.06
	4	0.99	0.23	0.11	0.33	0.21	0.19	0.06	0.06
	6	0.98	0.15	0.08	0.23	0.25	0.22	0.06	0.06
	8	0.97	0.11	0.07	0.18	0.27	0.23	0.06	0.06
	10	0.96	0.08	0.06	0.14	0.30	0.25	0.06	0.06
	12	0.95	0.07	0.07	0.14	0.33	0.25	0.06	0.06
	14	0.94	0.07	0.06	0.13	0.34	0.26	0.06	0.06
	16	0.93	0.07	0.06	0.13	0.35	0.26	0.06	0.06
	18	0.92	0.07	0.06	0.13	0.38	0.26	0.06	0.06
	20	0.91	0.07	0.06	0.13	0.38	0.26	0.06	0.06
	40	0.83	0.07	0.06	0.13	0.46	0.23	0.06	0.06
	80	0.69	0.06	0.06	0.12	0.48	0.16	0.06	0.06
	120	0.58	0.06	0.06	0.12	0.50	0.12	0.06	0.06
0.98	1	1.00	0.45	0.25	0.70	0.09	0.09	0.05	0.05
	2	0.99	0.33	0.16	0.49	0.18	0.16	0.05	0.05
	4	0.96	0.16	0.09	0.25	0.36	0.33	0.05	0.05
	6	0.94	0.09	0.08	0.17	0.46	0.41	0.05	0.06
	8	0.92	0.08	0.07	0.15	0.52	0.45	0.05	0.06
	10	0.89	0.07	0.06	0.13	0.57	0.48	0.05	0.06
	12	0.87	0.06	0.06	0.12	0.62	0.51	0.05	0.06
	14	0.85	0.06	0.05	0.11	0.65	0.52	0.05	0.06
	16	0.83	0.06	0.05	0.11	0.65	0.53	0.05	0.06
	18	0.81	0.06	0.05	0.11	0.68	0.54	0.05	0.06
	20	0.79	0.06	0.05	0.11	0.69	0.54	0.05	0.06
	40	0.64	0.05	0.05	0.10	0.76	0.53	0.05	0.06
	80	0.43	0.05	0.05	0.10	0.77	0.41	0.05	0.06
	120	0.32	0.05	0.05	0.10	0.79	0.31	0.05	0.06

Table 1 (continued)

ρ	h	$\widetilde{\beta}_h/\delta$	$Big\ h_L$	$Big\ h_U$	$Big\ h$	$Wald_{NW}$	$Wald$	R_L^2	R_U^2
0.95	1	1.00	0.50	0.19	0.69	0.11	0.11	0.06	0.07
0.95	2	0.98	0.34	0.10	0.44	0.26	0.24	0.05	0.07
0.95	4	0.93	0.16	0.05	0.21	0.62	0.58	0.05	0.06
0.95	6	0.88	0.11	0.04	0.15	0.76	0.73	0.05	0.06
0.95	8	0.84	0.10	0.04	0.14	0.84	0.80	0.05	0.06
0.95	10	0.80	0.09	0.05	0.14	0.88	0.84	0.05	0.06
0.95	12	0.77	0.08	0.05	0.13	0.91	0.87	0.05	0.06
0.95	14	0.73	0.08	0.05	0.13	0.92	0.88	0.05	0.06
0.95	16	0.70	0.08	0.05	0.13	0.92	0.89	0.05	0.06
0.95	18	0.67	0.08	0.05	0.13	0.94	0.90	0.05	0.06
0.95	20	0.64	0.07	0.05	0.12	0.94	0.90	0.05	0.06
0.95	40	0.44	0.07	0.05	0.12	0.97	0.90	0.05	0.06
0.95	80	0.25	0.07	0.05	0.12	0.96	0.83	0.05	0.06
0.95	120	0.17	0.07	0.06	0.11	0.96	0.76	0.05	0.06
0.90	1	1.00	0.61	0.08	0.69	0.09	0.09	0.04	0.08
0.90	2	0.95	0.41	0.05	0.46	0.38	0.36	0.04	0.07
0.90	4	0.86	0.20	0.03	0.23	0.90	0.89	0.04	0.07
0.90	6	0.78	0.13	0.03	0.16	0.99	0.99	0.04	0.07
0.90	8	0.71	0.11	0.03	0.14	1.00	1.00	0.04	0.07
0.90	10	0.65	0.10	0.04	0.14	1.00	1.00	0.04	0.07
0.90	12	0.60	0.09	0.04	0.13	1.00	1.00	0.04	0.07
0.90	14	0.55	0.09	0.04	0.13	1.00	1.00	0.04	0.07
0.90	16	0.51	0.09	0.04	0.13	1.00	1.00	0.04	0.07
0.90	18	0.47	0.08	0.04	0.12	1.00	1.00	0.04	0.07
0.90	20	0.44	0.08	0.04	0.12	1.00	1.00	0.04	0.07
0.90	40	0.25	0.08	0.04	0.12	1.00	1.00	0.04	0.07
0.90	80	0.12	0.08	0.04	0.12	1.00	1.00	0.04	0.07
0.90	120	0.08	0.08	0.04	0.12	1.00	1.00	0.04	0.07

Table 2: Monte Carlo results

ρ	δ	A_{12}	$Big\ h_L$	$Big\ h_R$	$Big\ h$	$Wald$
0.99	0.005	0.97	0.26	0.36	0.62	0.11
	0.01	0.96	0.17	0.27	0.44	0.17
	0.02	0.94	0.11	0.15	0.26	0.15
	0.03	0.92	0.09	0.10	0.20	0.17
	0.04	0.90	0.08	0.08	0.17	0.17
	0.05	0.88	0.08	0.08	0.16	0.18
	0.06	0.87	0.08	0.07	0.15	0.19
	0.07	0.85	0.08	0.07	0.15	0.18
	0.08	0.83	0.08	0.07	0.15	0.18
	0.09	0.82	0.07	0.06	0.14	0.18
	0.1	0.80	0.07	0.06	0.14	0.18
	0.2	0.65	0.07	0.07	0.14	0.13
	0.4	0.44	0.07	0.06	0.13	0.13
	0.6	0.29	0.07	0.06	0.13	0.16
0.975	0.005	0.95	0.25	0.33	0.57	0.09
	0.01	0.93	0.14	0.19	0.33	0.14
	0.02	0.88	0.08	0.09	0.18	0.15
	0.03	0.84	0.05	0.07	0.13	0.16
	0.04	0.80	0.05	0.07	0.12	0.16
	0.05	0.76	0.05	0.06	0.12	0.17
	0.06	0.72	0.06	0.07	0.12	0.17
	0.07	0.69	0.05	0.07	0.12	0.18
	0.08	0.65	0.05	0.07	0.12	0.18
	0.09	0.62	0.05	0.07	0.12	0.18
	0.1	0.59	0.05	0.07	0.12	0.18
	0.2	0.36	0.05	0.07	0.11	0.13
	0.4	0.13	0.05	0.06	0.11	0.12
	0.6	0.05	0.05	0.06	0.12	0.15

Table 2 (continued)

ρ	δ	A_{12}	$Big\ h_L$	$Big\ h_R$	$Big\ h$	$Wald$
0.96	0.005	0.94	0.27	0.28	0.54	0.11
	0.01	0.90	0.15	0.16	0.31	0.14
	0.02	0.83	0.06	0.08	0.15	0.14
	0.03	0.77	0.05	0.07	0.13	0.15
	0.04	0.71	0.05	0.06	0.12	0.16
	0.05	0.65	0.05	0.06	0.11	0.16
	0.06	0.60	0.05	0.06	0.11	0.17
	0.07	0.55	0.05	0.06	0.11	0.16
	0.08	0.51	0.05	0.06	0.11	0.16
	0.09	0.47	0.05	0.06	0.11	0.15
	0.1	0.43	0.05	0.06	0.11	0.14
	0.2	0.19	0.05	0.05	0.10	0.13
	0.4	0.04	0.05	0.05	0.10	0.15
	0.6	0.01	0.05	0.05	0.10	0.16
0.9425	0.005	0.92	0.26	0.27	0.53	0.11
	0.01	0.87	0.15	0.13	0.28	0.15
	0.02	0.77	0.07	0.07	0.14	0.14
	0.03	0.69	0.06	0.06	0.12	0.16
	0.04	0.61	0.05	0.07	0.12	0.17
	0.05	0.54	0.05	0.06	0.11	0.18
	0.06	0.48	0.05	0.06	0.11	0.18
	0.07	0.43	0.05	0.06	0.11	0.19
	0.08	0.38	0.05	0.06	0.11	0.18
	0.09	0.34	0.05	0.06	0.11	0.18
	0.1	0.30	0.05	0.06	0.11	0.17
	0.2	0.09	0.05	0.06	0.11	0.16
	0.4	0.01	0.05	0.06	0.11	0.16
	0.6	0.00	0.06	0.06	0.11	0.18

Table 3: Descriptive statistics

Regression	Φ	Λ	V^{-1}
Fama – Bliss			
$VAR(rx_{t+1}^{(12)}, f_t - s_t)$	$\begin{pmatrix} 0.93 & 0.03 \\ 0.03 & 0.85 \end{pmatrix}$	$\begin{pmatrix} 0.93 & 0 \\ 0 & 0.84 \end{pmatrix}$	$\begin{pmatrix} 0.95 & -0.33 \\ 0.29 & 0.94 \end{pmatrix}$
$VAR(rx_{t+1}^{(24)}, f_t - s_t)$	$\begin{pmatrix} 0.96 & -0.01 \\ 0.02 & 0.85 \end{pmatrix}$	$\begin{pmatrix} 0.96 & 0 \\ 0 & 0.86 \end{pmatrix}$	$\begin{pmatrix} 0.98 & 0.13 \\ 0.19 & 0.99 \end{pmatrix}$
UIRP			
$VAR(\Delta s_{t+1}, i_{t+1}^{diff})$	$\begin{pmatrix} 0.003 & -0.00 \\ -4.24 & 0.96 \end{pmatrix}$	$\begin{pmatrix} 0.022 & 0 \\ 0 & 0.99 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.00 \\ 4.41 & 1 \end{pmatrix}$
$VAR(s_{t+1}, i_{t+1}^{diff})$	$\begin{pmatrix} 0.98 & -0.00 \\ 0.00 & 0.97 \end{pmatrix}$	$\begin{pmatrix} 0.98 & 0 \\ 0 & 0.97 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.07 \\ 0.30 & 1 \end{pmatrix}$
Term Structure			
Spread $3-12$	$\begin{pmatrix} 0.96 & -0.13 \\ 0.02 & 0.59 \end{pmatrix}$	$\begin{pmatrix} 0.954 & 0 \\ 0 & 0.603 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.39 \\ -0.05 & 1 \end{pmatrix}$
Spread $3-36$	$\begin{pmatrix} 0.95 & -0.07 \\ 0.00 & 0.80 \end{pmatrix}$	$\begin{pmatrix} 0.951 & 0 \\ 0 & 0.811 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.47 \\ -0.04 & 1 \end{pmatrix}$
Spread $3-60$	$\begin{pmatrix} 0.95 & -0.06 \\ 0.01 & 0.88 \end{pmatrix}$	$\begin{pmatrix} 0.937 & 0 \\ 0 & 0.899 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.78 \\ -0.37 & 1 \end{pmatrix}$
CA			
Canada, $VAR(\Delta y, CA)$	$\begin{pmatrix} 0.25 & 0 \\ 0.90 & 0.94 \end{pmatrix}$	$\begin{pmatrix} 0.25 & 0 \\ 0 & 0.94 \end{pmatrix}$	$\begin{pmatrix} -0.61 & 0 \\ 0.79 & -1 \end{pmatrix}$
Canada, $VAR(y, CA)$	$\begin{pmatrix} 1.01 & -0.02 \\ 0.02 & 0.91 \end{pmatrix}$	$\begin{pmatrix} 1.01 & 0 \\ 0 & 0.91 \end{pmatrix}$	$\begin{pmatrix} 0.98 & 0.17 \\ 0.21 & 0.98 \end{pmatrix}$
U.S., $VAR(\Delta y, CA)$	$\begin{pmatrix} -0.24 & -0.05 \\ 1.003 & 1.013 \end{pmatrix}$	$\begin{pmatrix} -0.19 & 0 \\ 0 & 0.96 \end{pmatrix}$	$\begin{pmatrix} -0.76 & 0.05 \\ 0.64 & -0.999 \end{pmatrix}$
U.S., $VAR(y, CA)$	$\begin{pmatrix} 0.96 & -0.07 \\ -0.01 & 0.914 \end{pmatrix}$	$\begin{pmatrix} 0.98 & 0 \\ 0 & 0.898 \end{pmatrix}$	$\begin{pmatrix} 0.98 & 0.76 \\ -0.21 & 0.65 \end{pmatrix}$
U.K., $VAR(\Delta y, CA)$	$\begin{pmatrix} -0.11 & -0.01 \\ 0.717 & 0.883 \end{pmatrix}$	$\begin{pmatrix} -0.1 & 0 \\ 0 & 0.87 \end{pmatrix}$	$\begin{pmatrix} -0.81 & 0.01 \\ 0.59 & -1 \end{pmatrix}$
U.K., $VAR(y, CA)$	$\begin{pmatrix} 1 & -0.06 \\ 0 & 0.82 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 0.82 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.29 \\ -0.04 & 0.96 \end{pmatrix}$
PIH			
$VAR(\Delta y, s)$	$\begin{pmatrix} -0.21 & -0.04 \\ 0.70 & 0.98 \end{pmatrix}$	$\begin{pmatrix} -0.19 & 0 \\ 0 & 0.96 \end{pmatrix}$	$\begin{pmatrix} -0.86 & 0.03 \\ 0.51 & -1 \end{pmatrix}$
$VAR(y, s)$	$\begin{pmatrix} 1 & -0.05 \\ 0 & 0.95 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 0.95 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.69 \\ -0.03 & 0.73 \end{pmatrix}$

TABLE 4: Fama Bliss Regressions, 1964-2001

12 months

$$rx_t^n = \alpha + \beta \left(f_t^{n-1 \to n} - y_t^1 \right)$$

maturity	α	β	R^2	*Big h* R_L^2	*Big h* R_U^2
2	0.0387	0.9394	0.1384	0.0085	0.4830
3	−0.1440	1.2411	0.1409	0.0085	0.3756
4	−0.4116	1.4949	0.1529	0.1474	0.6128
5	−0.1125	1.0947	0.0601	0.0855	0.5453

24 months

$$rx_t^n = \alpha + \beta \left(f_t^{n-1 \to n} - y_t^1 \right)$$

maturity	α	β	R^2	*Big h* R_L^2	*Big h* R_U^2
3	0.2199	0.9713	0.0846	0.0001	0.2332
4	0.0712	1.3788	0.0990	0.0001	0.1411
5	−0.5391	1.8888	0.1407	0.0217	0.3755

Table 5: The EHTS for the U.S.

h	$Big\ h_L$	$Big\ h_U$	$Wald$ p-value	h	$Big\ h_L$	$Big\ h_U$	$Wald$ p-value
1	-1.16	-1.15	0	17	-2.48	-2.22	0
2	-1.47	-1.44	0	18	-2.43	-2.18	0
3	-1.88	-1.82	0	21	-2.32	-2.06	0
4	-2.21	-2.13	0	24	-2.22	-1.97	0
5	-2.40	-2.29	0	30	-2.02	-1.78	0
6	-2.50	-2.36	0	36	-1.85	-1.62	0
7	-2.57	-2.41	0	48	-1.63	-1.42	0
8	-2.64	-2.46	0	60	-1.48	-1.29	0
9	-2.69	-2.50	0	72	-1.38	-1.20	0
10	-2.73	-2.52	0	84	-1.31	-1.15	0
11	-2.74	-2.52	0	96	-1.29	-1.13	0.01
12	-2.72	-2.49	0	108	-1.28	-1.12	0.01
13	-2.68	-2.44	0	120	-1.28	-1.12	0.01
14	-2.63	-2.39	0	144	-1.28	-1.12	0.01
15	-2.58	-2.33	0	156	-1.27	-1.11	0.01
16	-2.53	-2.27	0				

Table 6: The Current Account[18]

	Canada	$T = 87$	*U.S.*	$T = 90$	*U.K.*	$T = 132$
h	*Big* h_L	*Big* h_U	*Big* h_L	*Big* h_U	*Big* h_L	*Big* h_U
1	0.05	0.06	-0.04	-0.03	-0.0654	-0.0549
3	0.12	0.17	-0.11	-0.09	-0.1745	-0.1034
5	0.16	0.29	-0.19	-0.13	-0.2589	-0.1082
7	0.17	0.40	-0.26	-0.15	-0.3225	-0.0951
9	0.17	0.52	-0.34	-0.17	-0.3690	-0.0767
15	0.13	0.85	-0.57	-0.17	-0.4334	-0.0316
20	0.09	1.12	-0.77	-0.16	-0.4317	-0.0132
25	0.06	1.39	-0.97	-0.13	-0.4032	-0.0051
30	0.04	1.65	-1.18	-0.11	-0.3615	-0.0019
40	0.01	2.16	-1.59	-0.07	-0.2690	-0.0003
50	0.00	2.65	-2.03	-0.04	-0.1877	0
60	0.00	3.12	-2.47	-0.02	-0.1257	0

[18] P-values of Wald test statistics are equal to zero for every currency and every horizon.

Table 7: The PIH model for the U.S.

h	$Big\ h_L$	$Big\ h_U$	$Wald$ p-value
1	-0.03	-0.03	0
2	-0.06	-0.05	0
4	-0.10	-0.08	0
6	-0.15	-0.10	0
8	-0.19	-0.11	0
10	-0.22	-0.11	0
20	-0.33	-0.09	0
30	-0.38	-0.05	0
40	-0.38	-0.03	0
50	-0.36	-0.01	0
60	-0.32	-0.01	0
70	-0.28	0.00	0
100	-0.17	0.00	0

Table 8: the UIRP for D-Mark/$

h	$Big\ h_L$	$Big\ h_U$	$Wald$ p-value
1	0.01	0.01	0
3	0.02	0.03	0
6	0.05	0.05	0
10	0.07	0.09	0
13	0.09	0.12	0
16	0.10	0.15	0
20	0.11	0.19	0
23	0.12	0.22	0
26	0.13	0.25	0
30	0.14	0.30	0
60	0.18	0.65	0
100	0.18	1.25	0
200	0.19	3.62	0

Notes to tables.

Note to Table 1. This table contains a Monte Carlo comparison of our method and the Wald test. We compare one minus the coverage of our method relative to the size of the Wald test, both of which should ideally be 0.10. The sample size is T=200 and the number of Monte Carlo replications is 1000. The table reports: the root close to unity ("ρ"), the horizon ("h"), the value of the parameter of interest ("$\widetilde{\beta}_h/\delta$"), the percentage of samples in which the true value of the parameter lays below the lower bound and above the upper bound of the confidence interval proposed in this paper, based on Stock (1991) method (labeled "*Big* h_L" and "*Big* h_U" respectively); one minus the Monte Carlo coverage of our method (labeled "*Big* h"); the size of the Wald test for testing $[1,0]$ $A(I-A^k)$ $((I-A)^{-1})$ $-[0,1]=[0,0]$ with and without a Newey and West (1987) correction for serial correlation (labeled "$Wald_{NW}$" and "$Wald$" respectively). The confidence intervals proposed in this paper is calculated as follows: the parameter of interest is $\beta\frac{e^{c\delta}-1}{\delta c}$, where β is consistently estimated by a the corresponding coefficient in a VAR of order one, $(I-AL)\,y_t$, $A=\begin{pmatrix} 0 & \beta \\ 0 & \rho \end{pmatrix}$. The confidence interval relies on a median unbiased confidence interval of c obtained by inverting an ADF test statistic. The Wald test employed a Newey-West HAC covariance matrix estimator robust to $(h-1)$ serial correlation. The gradient of the constraint is calculated by using numerical derivatives. Finally, R_L^2 and R_U^2 respectively report the percentage of samples in which the true value of R^2 lays below the lower bound and above the upper bound of the confidence interval proposed in this paper.

Note to Table 2. As per Table 1, with the exception that the value of the parameter of interest is labeled "A_{12}" and the number of Monte Carlo replications is only 100. The Wald test is not corrected for serial correlation.

Note to Table 3. The table reports, for each regression described in the first column, the estimated values of Φ (corrected for serial correlation, as in Stock and Watson (1988)), its eigenvalues Λ and eigenvectors V^{-1} such that $\Phi=V\Lambda V^{-1}$. In the UIRP, for the regression: $\Delta s_{t+1}=\alpha+\beta i_{t+1}^{diff}$, then the serial correlation in the residuals is 0.977.

Note to Table 4. The table reports Fama and Bliss regressions based on Cochrane and Piazzesi (2002) data. "rx_{t+1}^n" is the excess log return, "y_t^1" is the spot log yield and "$f_t^{n-1\to n}$" is the log forward rate at time t for loans between time t+n-1 and t+n. See Cochrane and Piazzesi (2002) for a detailed description of the variables and of the regressions. R^2 is the (unadjusted) R^2 of the regression and "*Big* h R_L^2" and "*Big* h R_U^2" are the lower and upper bounds of the confidence interval for the (unadjusted) R^2 by using the method described in this paper (12).

Note to Tables 5 to 8. In the tables, we report the two-sided confidence interval for β_h proposed in this paper, based on Stock (1991) method (labeled "Big h_L" and "Big h_R" respectively) and p-values of Wald tests of the expectations hypotheses (labeled "Wald p-value"). In Tables 6 and 7, the confidence interval ("Big h_L", "Big h_R") should contain "-1" if the economic theory is correct; in Tables 5 and 8, it should contain "1".

Appendix 1. Proofs

Derivation of eq. (10).

Rewrite (9) as:

$$\begin{aligned}\Delta w_{1,t+1} - \beta w_{2t} &= \epsilon_{1,t+1} \\ (1-\rho L)\, w_{2,t+1} &= b(L)\,\epsilon_{2,t+1}\end{aligned}$$

that is:

$$\underbrace{\begin{pmatrix} 1-L & -\beta L \\ 0 & 1-\rho L \end{pmatrix}}_{\equiv (I-\Phi L)} \begin{pmatrix} w_{1,t+1} \\ w_{2,t+1} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & b(L) \end{pmatrix}}_{\equiv \Theta(L)} \underbrace{\begin{pmatrix} \epsilon_{1,t+1} \\ \epsilon_{2,t+1} \end{pmatrix}}_{\equiv \epsilon_{t+1}}$$

where $\Phi = \begin{pmatrix} 1 & \beta \\ 0 & \rho \end{pmatrix}$. Note that $y_{1,t+1} = \Delta w_{1,t+1}$ so that $\Sigma_{j=1}^{h} y_{1,t+j} = \Sigma_{j=1}^{h} \Delta w_{1,t+j} = w_{1,t+h} - w_{1,t}$. By applying (8):

$$\begin{pmatrix} w_{1,t+h} \\ w_{2,t+h} \end{pmatrix} = \Phi^h \begin{pmatrix} w_{1,t} \\ w_{2,t} \end{pmatrix} + \Sigma_{j=0}^{h-1} \Phi^j \Theta(I)\, \epsilon_{t+h-j} + O_p(1)\,.$$

Note also that $\Phi^j = \begin{pmatrix} 1 & \beta \Sigma_{k=0}^{j-1} \rho^k \\ 0 & \rho^j \end{pmatrix}$; thus:

$$\begin{pmatrix} w_{1,t+h} \\ w_{2,t+h} \end{pmatrix} = \begin{pmatrix} 1 & \beta \Sigma_{k=0}^{h-1} \rho^k \\ 0 & \rho^h \end{pmatrix} \begin{pmatrix} w_{1,t} \\ w_{2,t} \end{pmatrix} + \Sigma_{j=0}^{h-1} \begin{pmatrix} 1 & \beta \Sigma_{k=0}^{j-1} \rho^k \\ 0 & \rho^j \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & b(1) \end{pmatrix} \epsilon_{t+h-j} + O_p(1)\,.$$

so that:

$$\begin{aligned} w_{1,t+h} - w_{1,t} &= \underbrace{\beta \Sigma_{k=0}^{h-1} \rho^k}_{\equiv \beta_h} w_{2,t} + \underbrace{\Sigma_{j=0}^{h-1} \left(\epsilon_{1,t+h-j} + \beta b(1)\, \Sigma_{k=0}^{j-1} \rho^k \epsilon_{2,t+h-j} \right)}_{\equiv \xi_{t,h}} \\ w_{2,t+h} &= \rho^h w_{2,t} + \Sigma_{j=0}^{h-1} \rho^j b(1)\, \epsilon_{2,t+h-j} \end{aligned}$$

The asymptotic distribution of R^2 in the predictive regression ()

The DGP: $(I-\Phi L)\, w_t = \Theta(L)\, \epsilon_t$, $\Theta(L) \equiv (I + \Theta_1 L + ... + \Theta_p L^p)$, $\Phi = \begin{pmatrix} 0 & \beta \\ 0 & \rho \end{pmatrix} = B\rho$, $B \equiv \begin{pmatrix} 0 & \beta/\rho \\ 0 & 1 \end{pmatrix}$, $\Theta_i = \begin{pmatrix} 1 & 0 \\ 0 & \Theta_{22,i}(L) \end{pmatrix}$. As $\sum_{k=1}^{h} w_{t+k} = \sum_{k=1}^{h} \Phi^k w_t + \sum_{k=1}^{h} \sum_{j=0}^{k-1} \Phi^j \Theta(I)\, \epsilon_{t+k-j} +$

$O_p(h)$, we consider only the first two components after the equality, which are the asymptotically relevant ones. Let $\widetilde{\epsilon}_t \equiv \Theta(I)\epsilon_t$ and $Var(\widetilde{\epsilon}_t) = \Theta(I)\Sigma_\epsilon\Theta(I)' \equiv \Omega$.

$$\sum_{k=1}^{h}\sum_{j=0}^{k-1}\Phi^j\Theta(I)\epsilon_{t+k-j} = \widetilde{\epsilon}_{t+1}+$$
$$\widetilde{\epsilon}_{t+2}+\Phi\widetilde{\epsilon}_{t+1}+$$
$$+\widetilde{\epsilon}_{t+k}+\Phi\widetilde{\epsilon}_{t+k-1}+\Phi^2\widetilde{\epsilon}_{t+k-1}+...+\Phi^{h-1}\widetilde{\epsilon}_{t+1} =$$
$$= \widetilde{\epsilon}_{t+k}+(I+\Phi)\widetilde{\epsilon}_{t+k-1}+\left(I+\Phi+\Phi^2\right)\widetilde{\epsilon}_{t+k-2}+...+\left(I+\Phi+..+\Phi^{h-1}\right)\widetilde{\epsilon}_{t+1}$$
$$Var\left(\sum_{k=1}^{h}\sum_{j=0}^{k-1}\Phi^j\Theta(I)\epsilon_{t+k-j}\right) = \Omega+(I+\Phi)\Omega(I+\Phi)'+...+\left(I+\Phi+..+\Phi^{h-1}\right)\Omega\left(I+\Phi+..+\Phi^{h-1}\right)' =$$
$$= \sum_{k=1}^{h-1}\left(\sum_{l=0}^{k}\Phi^l\right)\Omega\left(\sum_{l=0}^{k}\Phi^l\right)' \equiv \overline{\Omega}_{h-1}$$

$Var\left(\sum_{k=1}^{h}\Phi^k w_t\right) = \left(\sum_{k=1}^{h}\Phi^k\right)Var(w_t)\left(\sum_{k=1}^{h}\Phi^k\right)$, where:

(a) $Var(w_t) = Var\left(\sum_{k=1}^{t}\sum_{j=0}^{k-1}\Phi^j\Theta(I)\epsilon_{t-j}\right) = \Omega+(I+\Phi)\Omega(I+\Phi)'+...+\left(I+\Phi+..+\Phi^{t-1}\right)\Omega\left(I+\Phi+..+\Phi^{t-}\right.$
$\overline{\Omega}_{t-1}$

(b) $\Phi^k = \rho^k B$ implies that $\sum_{k=1}^{h}\Phi^k = \left(\sum_{k=1}^{h}\rho^k\right)B$, so that:
$Var\left(\sum_{k=1}^{h}\Phi^k w_t\right) = \left(\sum_{k=1}^{h}\rho^k\right)^2 B\overline{\Omega}_{t-1}B'$

Thus the coefficient of determination for the first regression is:
$$R_h^2 = \frac{I_1'\left(\sum_{k=1}^{h}\rho^k\right)^2 B\overline{\Omega}_{t-1}B'I_1}{I_1'\left(\sum_{k=1}^{h}\rho^k\right)^2 B\overline{\Omega}_{t-1}B'I_1+I_1'\overline{\Omega}_{h-1}I_1} = \left(1+\frac{I_1'\overline{\Omega}_{h-1}I_1}{I_1'\left(\sum_{k=1}^{h}\rho^k\right)^2 B\overline{\Omega}_{t-1}B'I_1}\right)^{-1}$$
As $T\to\infty$, note that: $\sum_{l=0}^{m-1}\Phi^l = \left(I+\Phi+..+\Phi^{m-1}\right) = \begin{pmatrix} 1 & \beta\sum_{k=0}^{m-1}\rho^k \\ 0 & \sum_{k=0}^{m}\rho^k \end{pmatrix} \underset{T\to\infty}{\to} \begin{pmatrix} 1 & \beta T\int_0^{[(m-1)/T]}e^{cs}ds \\ 0 & T\int_0^{[m/T]}e^{cs}ds \end{pmatrix}$
$$\overline{\Omega}_{n-1} = \sum_{k=1}^{n-1}\left(\sum_{l=0}^{k}\Phi^l\right)\Omega\left(\sum_{l=0}^{k}\Phi^l\right)' = T\int_0^{[(n-1)/T]}\begin{pmatrix} 1 & \beta T\int_0^{[(m-1)/T]}e^{cs}ds \\ 0 & T\int_0^{[m/T]}e^{cs}ds \end{pmatrix}\Omega\begin{pmatrix} 1 & \beta T\int_0^{[(m-1)/T]}e^{cs}ds \\ 0 & T\int_0^{[m/T]}e^{cs}ds \end{pmatrix}'$$
$$= T\int_0^{[(n-1)/T]}\begin{pmatrix} 1 & \beta T\int_0^{[(m-1)/T]}e^{cs}ds \\ 0 & T\int_0^{[m/T]}e^{cs}ds \end{pmatrix}\Omega\begin{pmatrix} 1 & \beta T\int_0^{[(m-1)/T]}e^{cs}ds \\ 0 & T\int_0^{[m/T]}e^{cs}ds \end{pmatrix}' dm$$

Let $\psi_{(m)} = \int_0^{[(m-1)/T]}e^{cs}ds$
$$I_1'\left(\sum_{k=1}^{h}\rho^k\right)^2 B\overline{\Omega}_{t-1}B'I_1 = \left(\sum_{k=1}^{h}\rho^k\right)^2 I_1'B\overline{\Omega}_{t-1}B'I_1 =$$
$$= \left(\sum_{k=1}^{h}\rho^k\right)^2 T\int_0^{[(n-1)/T]}I_1'\begin{pmatrix} 0 & \beta/\rho \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & \beta T\psi_{(m)} \\ 0 & T\psi_{(m)} \end{pmatrix}\begin{pmatrix} \omega_{11} & \Omega_{12} \\ \Omega_{12}' & \Omega_{22} \end{pmatrix}\begin{pmatrix} 1 & 0 \\ \beta'\psi_{(m)}T & T\psi_{(m)} \end{pmatrix}\begin{pmatrix} 0 & 0 \\ \beta'/\rho & 1 \end{pmatrix}$$
$$= \left(\sum_{k=1}^{h}\rho^k\right)^2 T\int_0^{[(n-1)/T]}\left(T^2\frac{\psi_{(m)}^2}{\rho^2}\beta\Omega_{22}\beta'\right)dm$$

$$I_1'\overline{\Omega}_{h-1}I_1 = I_1'T\int_0^{[(n-1)/T]}\begin{pmatrix} 1 & \beta T\psi_{(h)} \\ 0 & T\psi_{(h)} \end{pmatrix}\begin{pmatrix} \omega_{11} & \Omega_{12} \\ \Omega_{12}' & \Omega_{22} \end{pmatrix}\begin{pmatrix} 1 & 0 \\ T\psi_{(h)}\beta' & T\psi_{(h)} \end{pmatrix}I_1 dh =$$
$$= T\int_0^{[(n-1)/T]}\left(\omega_{11}+\beta T\psi_{(h)}\Omega_{12}'+\left(\Omega_{12}+\beta T\psi_{(h)}\Omega_{22}\right)T\psi_{(h)}\beta'\right) = T\int_0^{[(n-1)/T]}T^2\psi_{(h)}^2\left(\beta\Omega_{22}\beta'\right)dh$$

$$R_h^2 = \left(1 + \frac{\int_0^{[(n-1)/T]} \psi_{(h)}^2 \left(\beta\Omega_{22}\beta'\right) dh}{\left(\sum_{k=1}^{h} \rho^k\right)^2 \int_0^{[(n-1)/T]} \left(\frac{\psi_{(m)}^2}{\rho^2} \beta\Omega_{22}\beta'\right) dm}\right)^{-1} \simeq \left(1 + \frac{1}{\frac{1}{[(n-1)/T]} \int_0^{[(n-1)/T]} \psi_{(m)}^2 dm}\right)^{-1}$$

The above formula clearly shows that R^2 is a monotone function of c.

Expectation hypothesis from economic theory usually imply that:

$$x_t = [0, 1] \begin{pmatrix} \Delta y_t \\ x_t \end{pmatrix} = \left(\frac{1}{b}\right) \sum_{j=1}^{h} a^j E_t \Delta y_{t+j} \tag{32}$$

where a and b are scalars. For example, in the CA literature[19], x is the CA, y is the net output and $a = -\left(\frac{1}{1+r}\right), b = 1$; in the PIH literature[20], x is saving, y is income and $a = -\left(\frac{1}{1+r}\right), b = 1$; in the EHTS literature[21], x is the interest rate differential, y is the nominal exchange rate and $a = 1, b = h$; in the UIRP literature[22], x is the long-term interest rate differential, y is short term interest rate and $a = 1, b = h$.

By combining (**??**) and (32), we obtain:

[19] See Sheffrin and Woo (1990) and Otto (1992).

[20] See Campbell (1987).

[21] See Bekaert and Hodrick (2001).

[22] See Campbell and Shiller (1987), Bekaert and Hodrick (2001).