Reverse Regressions and Long-Horizon Forecasting*

Min Wei[†] and Jonathan H. Wright[‡]

May 28, 2010

Abstract

Long-horizon predictive regressions in finance pose formidable econometric problems when estimated using available sample sizes. Hodrick (1992) proposed a remedy that is based on running a reverse regression of short-horizon returns on the long-run mean of the predictor. Unfortunately, this only allows the null of no predictability to be tested, and assumes stationary regressors. In this paper, we revisit long-horizon forecasting from reverse regressions, and argue that reverse regression methods avoid serious size distortions in long-horizon predictive regressions, even when there is some predictability and/or near unit roots. Meanwhile, the reverse regression methodology has the practical advantage of being easily applicable when there are many predictors. We apply these methods to forecasting excess bond returns using the term structure of forward rates, and find that there is indeed some return forecastability. But confidence intervals for the coefficients of the predictive regressions are about twice as wide as are obtained with the conventional approach to inference.

JEL Classification: C12, G12, G17
Keywords: Predictive Regressions; Long Horizons, Confidence Intervals; Small Sample Problems; Persistence

*We are grateful to Bob Hodrick for helpful comments on an earlier draft of this manuscript, entitled "Confidence Intervals for Long-Horizon Predictive Regressions via Reverse Regressions." The views expressed in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other employee of the Federal Reserve System.

[†]Division of Monetary Affairs, Federal Reserve Board, Washington DC 20551; min.wei@frb.gov.
[‡]Department of Economics, Johns Hopkins University, Baltimore MD 21218; wrightj@jhu.edu.

1 Introduction

Asset returns are widely thought to be somewhat forecastable, and perhaps more so at long than at short horizons. But inference in long-horizon predictive regressions is well known to be complicated by severe econometric problems in empirically relevant sample sizes. The problems arise because the predictors that are used are variables like the dividend yield or term spread that are highly persistent, while the regressor is an overlapping sum of short-term returns. This creates something akin to a spurious regression. This is compounded by the feedback effect, or absence of strict exogeneity—a shock to returns will in turn affect future values of the predictors. As a result, conventional t-statistics have rejection rates that are well above their nominal levels. The vast literature on the problems with long-horizon predictive regressions includes work such as Goetzmann and Jorion (1993), Elliott and Stock (1994), Stambaugh (1999), Valkanov (2003), Campbell and Yogo (2006) and Rossi (2007).

Hodrick (1992) proposed an approach to test the null hypothesis that a certain predictor does not help forecast long-horizon returns. His idea was instead of regressing the cumulative h-period returns onto the predictor at the start of the holding period, to regress the one-period return onto the sum of the predictors over the previous h periods. Under stationarity, for the coefficient in the first projection to be equal to zero is necessary and sufficient for the coefficient in the second projection to be equal to zero. However, the second regression has a persistent right-hand-side variable, but not a persistent left-hand side variable. Intuitively, this might mean that the size distortions of a test based on the second regression are much smaller.

Hodrick finds that this is indeed the case. Hodrick also considered a standard error for the original long-horizon predictive regression that uses the same reverse regression logic. The reverse regression approach to inference has become fairly widely used. For example, in their study of the long-horizon predictability of stock returns, Ang and Bekaert (2007) rely mainly on Hodrick standard errors 1B.

The methodology proposed by Hodrick (1992) however has two limitations. Firstly, its justification relies on stationarity. Secondly, it is only valid for testing the null of *no* predictability. Many researchers believe that there is some time series predictability in asset returns, even after controlling for econometric problems (see for example Campbell (2000)) and would like to test other hypotheses about the slope coefficient in a long-horizon predictive regression and, in particular, would like to form a confidence interval for this coefficient.

This paper revisits the use of reverse regressions in long-horizon asset return prediction, making a number of contributions. First, in the case with stationary regressors, we propose a methodology related to the reverse-regression, and show that it can be used more widely for inference on the parameter vector in a long-horizon regression, not just to test that the slope coefficient is equal to zero. Second, we derive the asymptotic distribution of the reverse regression Wald statistics if the predictors are highly persistent, modelled as having roots that are local to unity. In these results, we allow for some predictability in returns. Although the standard reverse regression does not give the correct size asymptotically in this case, we show that the size distortions are modest, provided that the degree of predictability is not too great. This contrasts with results for the usual forward regression Wald statistics, where asymptotic size distortions with near unit roots are enormous (Valkanov (2003)). Third, we assess the properties of the different reverse regression procedures in Monte-Carlo simulations with highly persistent predictors and some predictability. We find that the reverse regression procedures give confidence intervals with good coverage properties. In contrast, the standard forward regression with Newey-West standard errors has unreliable coverage. Finally, we apply the reverse regression to the prediction of excess bond returns, considering the regressions of Fama and Bliss (1987) and Cochrane and Piazzesi (2005). We find that the confidence intervals for the slope coefficients are roughly twice as wide as we would obtain from the usual long-horizon regression with Newey-West heteroskedasticity- and autocorrelation-robust standard errors. Using the reverse regression does *not* eliminate the clear empirical evidence for the predictability of bond returns using the term structure of forward rates. But, the p-value testing joint significance of all the slope coefficients goes from an eye-popping values around 10^{-6} in the ordinary regression to more reasonable numbers around or a bit below 0.01 in the reverse regression.

The reverse regression approach to inference applies regardless of whether there is a single predictor or multiple predictors. That is an advantage of this approach to inference relative to some others that have been proposed, such as the methods of Campbell and Yogo (2006), Torous, Valkanov and Yan (2004) and Rossi (2007) that are feasible only for a scalar predictor.

The plan for the remainder of the paper is as follows. Section 2 considers the case with stationary predictors and describes long-horizon regressions, reverse regressions, and the proposed extension to the reverse-regression methodology that allows us to test any hypothesis on the long-horizon slope coefficient, not just the null that it is equal to zero. Section 3 derives the limiting distribution of reverse-regression Wald tests if the predictors have roots that are local to unity. Section 4 contains Monte-Carlo simulations. Section 5 uses the reverse regression methodology to re-examine the forecasting of excess bond returns. Section 6 concludes.

2 Forward and Reverse Regressions

Let r_{t+1} denote the continuously compounded return on any asset from t to $t+1$ and let $r_{t+h}^{(h)} = r_{t+1} + r_{t+2} ... + r_{t+h}$ denote the h-period return. Let x_t be some px1 vector of predictors. Assume that $y_t = (r_t, x_t')'$ is covariance-stationary and that $A(L)y_t = \varepsilon_t$ where $A(L)$ is a lag polynomial with all roots outside the unit circle and ε_t is a martingale difference sequence with $2+\delta$ finite moments for some $\delta > 0$. Consider the standard long-horizon predictive regression:

$$r_{t+h}^{(h)} = \alpha^{(h)} + x_t'\beta^{(h)} + \varepsilon_{t+h}^{(h)} \tag{1}$$

Let $\hat{\beta}^{(h)}$ denote the OLS estimator of this regression. Researchers commonly estimate equation (1), using either Newey-West or Hansen-Hodrick standard errors (Newey and West (1987) and Hansen and Hodrick (1980)), to control for serial correlation in the errors.[1]

Alternative standard errors in equation (1) are given by Hodrick standard errors

[1] Throughout this paper, we will use the Newey-West standard errors as the "conventional" standard errors. Hansen-Hodrick standard errors are an alternative, but these can occasionally be imaginary numbers, which cannot happen with Newey-West. Otherwise, the two sets of standard errors have very similar properties in our Monte-Carlo analysis and empirical work. For this reason, we report Newey-West standard errors, and omit Hansen-Hodrick standard errors.

1B (Hodrick (1992)). This involves estimating the variance of $(\alpha^{(h)}, \beta^{(h)\prime})'$ in the forward regression (equation 1) as $W = (\Sigma\tilde{x}_t\tilde{x}_t')^{-1}\Sigma w_{t+1}w_{t+1}'(\Sigma\tilde{x}_t\tilde{x}_t')^{-1}$ where $w_{t+1} = (r_{t+1} - \bar{r})\Sigma_{i=0}^{h}\tilde{x}_{t-i}$, $\tilde{x}_t = (1, x_t')'$ and $\bar{r}$ is the sample mean of returns. Hodrick standard errors 1B are valid if and only if $\beta = 0$, because it is in this case alone that the sample variance of w_{t+1} is a consistent estimate of the zero-frequency spectral density of $x_t\varepsilon_{t+h}$. The Wald statistic testing the hypothesis that $\beta^{(h)} = 0$ using Hodrick standard errors 1B is $\hat{\beta}^{(h)\prime}W_{22}^{-1}\hat{\beta}^{(h)}$ where W is partitioned conformably with $(\alpha^{(h)}, \beta^{(h)\prime})'$ as $\begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix}$. This has a $\chi^2(p)$ asymptotic distribution under the null

Consider also the reverse regression of the one-period return on the h-period sum of the regressor:

$$r_{t+1} = \mu^{(h)} + x_t^{(h)\prime}\gamma^{(h)} + u_{t+1} \tag{2}$$

where $x_t^{(h)} = x_t + x_{t-1}... + x_{t-h+1}$. The coefficients in the forward and reverse regressions are related as

$$\beta^{(h)} = V_{xx}^{-1}Cov(r_{t+h}^{(h)}, x_t) = V_{xx}^{-1}\Sigma_{j=1}^{h}Cov(r_{t+j}, x_t) = V_{xx}^{-1}\Sigma_{j=1}^{h}Cov(r_{t+1}, x_{t+1-j})$$

$$= V_{xx}^{-1}Cov(r_{t+1}, x_t^{(h)}) = V_{xx}^{-1}V_{xx}^{(h)}\left(V_{xx}^{(h)}\right)^{-1}Cov(r_{t+1}, x_t^{(h)}) = V_{xx}^{-1}V_{xx}^{(h)}\gamma^{(h)} \tag{3}$$

where V_{xx} and $V_{xx}^{(h)}$ are the variance-covariance matrices of x_t and $x_t^{(h)}$, respectively, and the last equality on the first line uses the assumption of covariance-stationarity. A consequence of this is that $\beta^{(h)} = 0$ if and only if $\gamma^{(h)} = 0$. However, inference in the reverse regression is less prone to size distortions. Consequently, Hodrick (1992) also proposed testing the hypothesis that $\beta^{(h)} = 0$ by testing the implication that

$\gamma^{(h)} = 0$ in the reverse regression, equation (2). This can be implemented by the Wald statistic:

$$F_1 = (T-h)\hat{\gamma}^{(h)\prime}\hat{V}_{xx}^{(h)}Var(u_{t+1}x_t^{(h)})^{-1}\hat{V}_{xx}^{(h)}\hat{\gamma}^{(h)} \tag{4}$$

using a heteroskedasticity-robust estimate of the variance of $u_{t+1}x_t^{(h)}$, and where $\hat{V}_{xx}^{(h)} = \frac{1}{T-h}\sum_{t=h}^{T-1}\left(x_t^{(h)} - \overline{x}^{(h)}\right)\left(x_t^{(h)} - \overline{x}^{(h)}\right)'$. This statistic also has a $\chi^2(p)$ asymptotic distribution under the null. Note that Hodrick proposed the reverse regression in addition to his standard errors 1B, where the latter are alternative standard errors for the forward regression. Both apply only as tests of the hypothesis of no predictability, i.e. that $\beta^{(h)} = 0$.

2.1 Testing the hypothesis of some predictability via reverse regressions

However, the evidence for some predictability in asset returns at long horizons is quite strong, and we are instead perhaps more interested in testing other hypotheses about $\beta^{(h)}$, or forming a confidence set for it. The first contribution of this paper is to propose a method for inference on $\beta^{(h)}$ in a long-horizon regression that is based on the reverse regression logic, but that goes beyond just testing the null that $\beta^{(h)} = 0$. Like the work of Hodrick (1992), its formal justification relies on covariance-stationarity.

The idea is that from equation (3), under covariance-stationarity $\beta^{(h)} = V_{xx}^{-1}V_{xx}^{(h)}\gamma^{(h)}$, and so inference about from the reverse regression can be used for inference on $\beta^{(h)}$, taking account of the distribution of the x_ts. Since $\gamma^{(h)} = V_{xx}^{(h)-1}Cov(r_{t+1}, x_t^{(h)})$, we only need to adjust the numerator of the reverse regression, as

$$\beta^{(h)} = V_{xx}^{-1} Cov(r_{t+1}, x_t^{(h)}) \tag{5}$$

We now describe concretely how to use equation (5) for inference on $\beta^{(h)}$. Let $\theta_1 = Cov(r_{t+1}, x_t^{(h)})$ and $\theta_2 = V_{xx}$. Also let $\hat{\theta}_1 = (T-h)^{-1}\Sigma_{t=h}^{T-1} r_{t+1}(x_t^{(h)} - \bar{x}^{(h)}) = (T-h)^{-1}\Sigma_{t=1}^{T-h} r_{t+h}(x_{t+h-1}^{(h)} - \bar{x}^{(h)})$ and $\hat{\theta}_2 = (T-h)^{-1}\Sigma_{t=1}^{T-h}(x_t - \bar{x})(x_t - \bar{x})'$ be the sample counterparts where $\bar{r} = T^{-1}\Sigma_{t=1}^{T} r_t$, $\bar{x} = (T-h)^{-1}\Sigma_{t=1}^{T-h} x_t$ and $\bar{x}^{(h)} = (T-h)^{-1}\Sigma_{t=h}^{T-1} x_t^{(h)}$. We have $\beta^{(h)} = \theta_2^{-1}\theta_1$ and assume that

$$T^{1/2}(\hat{\theta} - \theta) \rightarrow_d N(0, V)$$

where $\theta = (\theta_1', vech(\theta_2)')'$, $\hat{\theta} = (\hat{\theta}_1', vech(\hat{\theta}_2)')'$ and V is 2π times the spectral density at frequency zero of $\begin{pmatrix} r_{t+h}(x_{t+h-1}^{(h)} - \bar{x}^{(h)}) \\ vech((x_t - \bar{x})(x_t - \bar{x})') \end{pmatrix}$, which can be partitioned conformably as $V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}$.

We can then use the delta method for inference on $\beta^{(h)}$, using the fact that $\beta^{(h)}$ is a nonlinear function of θ that is itself root-T consistently estimable and asymptotically normal. Concretely, consider the estimator

$$\tilde{\beta}^{(h)} = \hat{\theta}_2^{-1}\hat{\theta}_1 \tag{6}$$

Because the derivatives of $\beta^{(h)}$ with respect to θ_1 and $vech(\theta_2)$ are θ_2^{-1} and $-(\theta_1'\theta_2^{-1} \otimes \theta_2^{-1})D_p$, respectively, where D_p denotes the duplication matrix, it follows that:

$$T^{1/2}(\tilde{\beta}^{(h)} - \beta^{(h)}) \rightarrow_d N(0, \frac{\partial\beta^{(h)}}{\partial\theta} V \frac{\partial\beta^{(h)}}{\partial\theta}') \tag{7}$$

where $\frac{\partial\beta^{(h)}}{\partial\theta} = \begin{pmatrix} \theta_2^{-1} & -(\theta_1'\theta_2^{-1} \otimes \theta_2^{-1})D_p \end{pmatrix}$, allowing conventional Wald tests to be conducted.

Implementation of the proposed confidence intervals requires choosing a specific estimator of V, which is 2π times the spectral density at frequency zero of $\begin{pmatrix} r_{t+h}(x_{t+h-1}^{(h)} - \bar{x}^{(h)}) \\ vech((x_t - \bar{x})(x_t - \bar{x})') \end{pmatrix}$. We use a Newey-West estimator with lag length equal to h. We refer to this method a "reverse regression" estimate even though it does not require explicit estimation of equation (2), because it is based on assessing the covariance between one-period returns and the h-period sum of the predictor. We henceforth call this the *reverse-regression delta method* approach to inference for long-horizon forecasting.[2]

3 Reverse Regressions with Near Unit Roots and Some Predictability

The second contribution of the paper is to derive the asymptotic distributions of reverse regression tests when there is some predictability and the predictors are so persistent that it is suitable to model them as having roots local to unity. For this purpose, we consider the model

$$r_{t+1} = a_r + x_t'b + \varepsilon_{r,t+1}, \tag{8}$$

$$(I - \Phi L)x_t = a_x + B(L)\varepsilon_{x,t}, \tag{9}$$

and the following assumptions are made:

[2]The estimator in equation (6) is the multivariate analog of a ratio of two random variables, which could motivate forming a confidence set for it by the method of Fieller (1954). We have investigated this, and find that it does give modest improvements in the coverage of the confidence intervals, relative to the reverse-regression delta method. However, the method involves explicitly inverting the acceptance region of a confidence set, which makes it much harder to operationalize when the number of predictors is more than about 2 (as in the main application in this paper). For this reason, we do not consider this approach further.

A1. $(\varepsilon_{r,t}, \varepsilon'_{x,t})'$ is a martingale difference sequence with $2+\delta$ finite moments for some $\delta > 0$.

A2. $B(L)$ is a 1-summable matrix lag polynomial with all roots outside the unit circle.

A3. $\Phi = I + T^{-1}C$. The matrix $C = diag(c_1, c_2, ...c_p)$ is a fixed diagonal matrix where $c_i \leq 0 \ \forall i$. We write the matrix as c in the case of a scalar predictor ($p = 1$).

A4. $a_r = a_x = 0$ (an assumption that involves no loss of generality).

The near I(1) parameterization in A3 is not designed to be a literal description of the data generating process, as we do not believe that predictors become more persistent as the sample size increases. Rather it is a well-known device that is designed to provide a good approximation to the small sample behavior of estimators and test statistics when times series are highly persistent (Phillips (1987), Stock (1991, 1996)).

In this section, we derive the limiting distributions of reverse regression Wald tests under this near I(1) parameterization. It should be emphasized that our objective in doing this is *not* to provide alternative usable critical values. This is because the distributions will depend on the parameter C, which is in practice not known, and not consistently estimable.[3] Rather, what we are considering is how severe the size distortions will be when the researcher simply uses the incorrect conventional χ^2 critical values.

We consider the $h-$period predictive regression

$$r_{t+h}^{(h)} = \alpha^{(h)} + x_t'\beta^{(h)} + \varepsilon_{t+h}^{(h)} \tag{10}$$

[3]Some papers get around this problem by forming a confidence set for C and then appealing to the Bonferroni inequality (Campbell and Yogo (2006), Torous, Valkanov and Yan (2004) and Rossi (2007)). But these papers all consider just one predictor because if there are multiple predictors the number of nuisance parameters in C makes this strategy impractical.

but also let the horizon be an increasing function of the sample size to represent the idea that the forecast horizon is non-negligible relative to the sample size (as in Richardson and Stock (1989) and Stock (1996)). Thus, h is equal to $[\lambda T]$, where $[.]$ denotes the integer part. The setup of local to unit roots and a horizon that is an increasing function of the sample size was also considered by Valkanov (2003) and Rossi (2005, 2007).

Under these assumptions[4], it is well known from Phillips (1987) that

$$\begin{aligned} T^{-1/2}\Sigma_{t=1}^{[Tr]}\varepsilon_{r,t} &\rightarrow \sigma W(r), \\ T^{-1/2}x_{[Tr]} &\rightarrow \Omega^{1/2}J_c(r), \\ T^{-1/2}\left(x_{[Tr]}-\overline{x}\right) &\rightarrow \Omega^{1/2}J_c^{\mu}(r), \\ T^{-3/2}x_t^{(h)} &\rightarrow \Omega^{1/2}\bar{J}_C(r), \\ T^{-3/2}\left(x_t^{(h)}-\overline{x}^{(h)}\right) &\rightarrow \Omega^{1/2}\bar{J}_C^{\mu}(r), \end{aligned}$$

where $\sigma^2 = Var(\varepsilon_{r,t})$, $\Omega = B(1)Var(\varepsilon_{x,t})B(1)'$, $J_C(r)$ denotes an Ornstein-Uhlenbeck process defined by

$$dJ_C(r) = CJ_C(r) + dV(r),$$

$J_C^{\mu}(r) = J_C(r) - \frac{1}{1-\lambda}\int_0^{\lambda} J_C(s)ds$, $\bar{J}_C(r) = \int_{r-\lambda}^{r} J_C(s)ds$ (defined for $r \geq \lambda$), $\bar{J}_C^{\mu}(r) = \bar{J}_C(r) - \frac{1}{1-\lambda}\int_{\lambda}^{1} \bar{J}_C(s)ds$, and $V(r)$ and $W(r)$ are px1 and scalar standard Brownian motions with some correlation ρ (the long-run correlation between $\varepsilon_{x,t}$ and $\varepsilon_{r,t}$).

[4] We think of the case where $c_i < 0\ \forall i$ as the leading case, but can accommodate exact unit roots ($c_i = 0$) as well. The matrix $C^{-1}(\exp(C\lambda) - I)$ is diagonal, and if there are some exact unit roots, the relevant diagonal elements of this matrix are set equal to λ.

Following Rossi (2007), we have the following results:

$$
\begin{aligned}
\beta^{(h)} &= \Sigma_{i=0}^{h-1}\Phi^{i}b=(I-\Phi)^{-1}(I-\Phi^{h})b=[TC^{-1}(\exp(C\lambda)-I)+O(1)]b \\
\varepsilon_{t+h} &= \Sigma_{i=1}^{h}\varepsilon_{r,t+i}+\{\Sigma_{i=1}^{h-1}[\Sigma_{k=0}^{i-1}\Phi^{i}]B(1)\varepsilon_{x,t+h-i}\}'b+o_{p}(T^{-1/2})
\end{aligned}
$$

First consider the Wald statistic testing the hypothesis of no predictability ($\gamma^{(h)}=0$) in the reverse regression (equation 4). Theorem 1 provides the limiting distribution of this test statistic with near unit roots, under the null hypothesis of no predictability.

Theorem 1. Under the null $\beta^{(h)}=\gamma^{(h)}=0$, in the limit as T goes to infinity,

$$
F_{1}\rightarrow\left[\int_{\lambda}^{1}\bar{J}_{C}^{\mu}(r)dW(r)\right]'\left[\int_{\lambda}^{1}\bar{J}_{C}^{\mu}(r)\bar{J}_{C}^{\mu}(r)'dr\right]^{-1}\left[\int_{\lambda}^{1}\bar{J}_{C}^{\mu}(r)dW(r)\right]
$$

The proofs of the Theorems are collected in the appendix. The reverse-regression Wald statistic does not have a standard χ^2 null limiting distribution, and its distribution depends on the unknown local-to-unit root parameter that is not consistently estimated. So we cannot obtain usable critical values for the asymptotic distribution of F_1. Nonetheless for any particular parameter configurations, we can compute the size distortions when standard χ^2 critical values are used. These are reported in Table 1 in the case $p=1$ (scalar regressor) for different choices of c, ρ and λ, using the 5 percent nominal significance level. As can been seen from the table, the Hodrick test that uses conventional critical values is close to being asymptotically correctly sized, except when $c=0$. Even in the exact unit root case, the effective size is around 20 percent, at worst. This is in marked contrast to a t-test based on the forward regression which diverges at the rate $T^{1/2}$ for all values of c (Valkanov (2003)).

Next consider the Wald statistic testing the hypothesis that $\beta^{(h)} = \beta_0$ using Hodrick standard errors 1B:

$$F_2 = (\hat{\beta}^{(h)} - \beta_0)' W_{22}^{-1} (\hat{\beta}^{(h)} - \beta_0) \tag{11}$$

The test that compares F_2 with conventional χ^2 critical values is justified only as a test of the null hypothesis that $\beta^{(h)} = 0$, and only under covariance-stationarity. Here we are considering its properties with near unit roots and a more general null hypothesis. Theorem 2 provides the null limiting distribution of this test statistic.

Theorem 2. Under the null $\beta^{(h)} = \beta_0$, in the limit as T goes to infinity,

$$F_2 \to (\xi_1 + \xi_2 \beta_0)' \Xi^{-1} (\xi_1 + \xi_2 \beta_0)$$

where

$$\begin{aligned}
\xi_1 &= \left(\int_0^{1-\lambda} J_C^\mu(r) J_C^\mu(r)' dr\right)^{-1} \int_0^{1-\lambda} J_C^\mu(r)(W(r+\lambda) - W(r)) dr, \\
\xi_2 &= \sigma^{-1} \left(\int_0^{1-\lambda} J_c^\mu(r) J_c^\mu(r)' dr\right)^{-1} \left(\int_0^{1-\lambda} J_c^\mu(r) \overline{J}_c(r+\lambda)' dr\right) (\Omega^{1/2})' \tilde{C} - I, \\
\Xi &= \left[\int_0^{1-\lambda} J_c^\mu(r) J_c^\mu(r)' dr\right]^{-1} \left[\int_\lambda^1 \overline{J}_c^\mu(r) \overline{J}_c^\mu(r)' dr\right] \left[\int_0^{1-\lambda} J_c^\mu(r) J_c^\mu(r)' dr\right]^{-1}
\end{aligned}$$

and $\tilde{C} = [\exp(C\lambda) - I]^{-1} C$. Table 2 shows the simulated asymptotic rejection rates from comparing the test statistic F_2 to the 5 percent conventional χ^2 critical values in the case $p = 1$ (scalar regressor) for different choices of c, ρ, λ and β. The asymptotic size of this test is increasing in the degree of predictability (β_0), ρ, and c. For values of c at or below -10, the size distortions are mild: comparing F_2 to χ^2 critical values yields a test with an asymptotic size of around 15 percent or less. With roots closer still to the unit circle, the size distortions from treating F_2 as though it were χ^2

distributed get worse, and can be quite large with the combination of exact unit roots ($c = 0$), a high degree of predictability *and* a large correlation ρ.

Finally we conclude this section by considering the limiting distribution of the Wald statistic based on the reverse-regression delta method proposed in subsection 2.1 (equations 6 and 7). This Wald statistic testing the null $\beta^{(h)} = \beta_0$ is:

$$F_3 = (\hat{\theta}_2^{-1}\hat{\theta}_1 - \beta_0)'\{\begin{pmatrix} \hat{\theta}_2^{-1} & -(\hat{\theta}_1'\hat{\theta}_2^{-1} \otimes \hat{\theta}_2^{-1})D_p \end{pmatrix} \hat{V} \begin{pmatrix} \hat{\theta}_2^{-1} & -(\hat{\theta}_1'\hat{\theta}_2^{-1} \otimes \hat{\theta}_2^{-1})D_p \end{pmatrix}'\}^{-1}(\hat{\theta}_2^{-1}\hat{\theta}_1 - \beta_0) \tag{12}$$

It's null limiting distribution is provided in Theorem 3.

Theorem 3. Under the null $\beta^{(h)} = \beta_0$, in the limit as T goes to infinity,

$$F_3 \to (\beta^* - \beta_0)'\{\begin{pmatrix} \theta_2^{*-1} & -(\beta^{*\prime} \otimes \theta_2^{*-1})D_p \end{pmatrix} \begin{pmatrix} V_{11}^* & V_{12}^* \\ V_{21}^* & V_{22}^* \end{pmatrix} \begin{pmatrix} \theta_2^{*-1} & -(\beta^{*\prime} \otimes \theta_2^{*-1})D_p \end{pmatrix}'\}^{-1}(\beta^* - \beta_0)$$

where

$$\begin{aligned}
\beta^* &= \theta_2^{*-1}\theta_1^* \\
\theta_1^* &= \sigma\Omega^{1/2}\left[\left(\int_\lambda^1 \bar{J}_C^\mu(r)B_C(r)\,dr\right) + \int_\lambda^1 \bar{J}_C^\mu(r)dW(r)\right] \\
\theta_2^* &= \int_0^{1-\lambda} H_C(r)dr \\
V_{11}^* &= \sigma^2\{\int_{-\lambda}^{\lambda}\int_{\lambda-\min(0,\omega)}^{1-\max(0,\omega)} G_C(r,\omega)B_C(r)B_C(r+\omega)(1-\frac{|\omega|}{\lambda})drd\omega \\
&+ \int_\lambda^1 \int_{\max(\lambda-r,-\lambda)}^{\min(1-r,\lambda)} G_C(r,\omega)\left(1-\frac{|w|}{\lambda}\right)dW(r+\omega)dW(r) \\
&+ \int_\lambda^1 \int_{\max(\lambda-r,-\lambda)}^{\min(1-r,\lambda)} G_C(r,\omega)(1-\frac{|\omega|}{\lambda})dW(r+\omega)B_C(r)dr \\
&+ \int_{-\lambda}^{\lambda}\int_{\lambda-\min(0,\omega)}^{1-\max(0,\omega)} G_C(r,\omega)B_C(r+\omega)(1-\frac{|\omega|}{\lambda})dW(r)d\omega\} - \lambda^*\theta_1^*\theta_1^{*\prime}
\end{aligned}$$

$$
\begin{aligned}
V_{21}^{*} &= \sigma\{\int_{-\lambda}^{\lambda}\int_{-\min(0,\omega)}^{1-\lambda-\max(0,\omega)} vech(H_C(r))\bar{J}_C^{\mu}(r+\lambda+\omega)'\Omega^{1/2\prime}B_C(r+\lambda+\omega)(1-\frac{|\omega|}{\lambda})drd\omega \\
&\quad + \int_{0}^{1-\lambda}\int_{\max(-r,-\lambda)}^{\min(1-\lambda-r,\lambda)} vech(H_C(r))\bar{J}_C^{\mu}(r+\lambda+\omega)'\Omega^{1/2\prime}(1-\frac{|\omega|}{\lambda})dW(r+\lambda+\omega)dr\} \\
&\quad -\lambda^{*}vech\left(\theta_2^{*}\right)\theta_1^{*\prime} \\
V_{12}^{*} &= V_{21}^{*\prime} \\
V_{22}^{*} &= \int_{-\lambda}^{\lambda}\int_{\lambda-\min(0,\omega)}^{1-\max(0,\omega)} vech(H_C(r))vech(H_C(r+\omega))'dr\left(1-\frac{|w|}{\lambda}\right)dw-\lambda^{*}vech(\theta_2^{*})vech(\theta_2^{*})'
\end{aligned}
$$

$B_C(r) = \sigma^{-1}J_C(r)'(\Omega^{1/2})'\tilde{C}\beta_0$, $G_C(r,\omega) = \Omega^{1/2}\bar{J}_C^{\mu}(r)\bar{J}_C^{\mu}(r+\omega)'(\Omega^{1/2})'$, $H_C(r) = \Omega^{1/2}J_C^{\mu}(r)J_C^{\mu}(r)'(\Omega^{1/2})'$ and $\lambda^{*} = \lambda - \frac{4\lambda^2}{3}$. Table 3 shows the simulated asymptotic rejection rates from comparing the test statistic F_3 to the conventional χ^2 critical values in the case $p = 1$ (scalar regressor) for different choices of c, ρ, λ and β. The size of this test is again increasing in the degree of predictability ($\beta^{(h)}$), ρ, and c. But the size distortions are noticeably milder than for the test comparing F_2 to χ^2 critical values (shown earlier in Table 2). The asymptotic rejection rate the test comparing F_3 to conventional χ^2 critical values does not exceed 20 percent for any parameter configuration considered here. And it is below 16 percent for all simulations in which c is strictly negative.

4 Monte-Carlo Simulations

The motivation for considering the proposed approaches to inference via reverse regressions is that they may work better in small samples. Like the conventional forward regression methods, their justification is based on an assumption of stationarity, and methods that assume stationarity often fare poorly in the presence of a unit root, or a near unit root, at least in empirically relevant sample sizes. However, the results of the previous section suggest that they might in practice be quite robust to near non-stationarity. The intuition is that they back out the implied coefficient in the long-horizon regression from the correlation between one-period returns and a long-run sum of the predictor, which avoids a spurious regression. How well the reverse regression methods actually work in finite samples with nearly non-stationary predictors is the key practical question that we answer in a Monte-Carlo experiment.

In this experiment, returns and the predictor follow a VAR(1):

$$\begin{pmatrix} r_{t+1} \\ x_{t+1} \end{pmatrix} = \Phi \begin{pmatrix} r_t \\ x_t \end{pmatrix} + \begin{pmatrix} \varepsilon_{r,t+1} \\ \varepsilon_{x,t+1} \end{pmatrix}$$

where the errors are iid normal with mean zero and covariance matrix V_ε. Following Campbell (2001), set , $\Phi = \begin{pmatrix} 0 & \alpha \\ 0 & \phi \end{pmatrix}$ and $V_\varepsilon = \begin{pmatrix} \sigma_r^2 & \rho\sigma_r\sigma_x \\ \rho\sigma_r\sigma_x & \sigma_x^2 \end{pmatrix}$. As the units of measurement for returns and the predictors are arbitrary, we can normalize $\sigma_r = \sigma_x = 1$ without loss of generality, leaving three free parameters: α, ρ and ϕ.

The slope coefficient in the long-horizon regression is $\beta^{(h)} = \alpha\frac{1-\phi^h}{1-\phi}$. The population R-squared in this regression is

$$R^2 = \frac{\beta^{(h)2}}{\beta^{(h)2} + (1-\phi^2)\Sigma_{i=1}^h e_1'(\Sigma_{j=1}^i \Phi^{j-1})V_\varepsilon(\Sigma_{j=1}^i \Phi^{j-1})'e_1}$$

where $e_1 = (1,0)'$. So long as we fix the sign of α, R^2 will be a monotone increasing function of α (holding the other parameters fixed). Figure 1 plots the effective coverage of three different confidence intervals for the long-horizon slope coefficient ($\beta^{(h)}$) against the population R^2 for the case[5] where $\alpha \geq 0$ with different choices of

[5] The point of plotting coverage against population R^2 rather than α is just that this seems easier to interpret.

h and ρ. The coverage rates of the confidence sets are of course 1 minus the sizes of the test that $\beta^{(h)}$ is equal to its true value. The sample size is $T = 500$, which corresponds to about 40 years of monthly data, the nominal coverage is 95 percent, and the parameter ϕ is 0.98. The confidence intervals considered are: (i) the ordinary confidence intervals based on estimating equation (1), using Newey-West standard errors with a lag truncation parameter of h, (ii) the confidence interval based on estimating equation (1) using standard errors 1B of Hodrick (1992), and (iii) confidence intervals using the reverse-regression delta method (equations 6 and 7). The appendix of supplemental materials includes analogous figures for different choices of T, ϕ and where $\alpha \leq 0$. The results that we show in Figure 1 are representative of the results that apply in these other cases.

The confidence interval for $\beta^{(h)}$ formed using Newey-West standard errors has coverage that is considerably too low, regardless of whether there is no predictability or some predictability. In many cases, it has an effective coverage around 60 percent. The confidence interval formed using Hodrick standard errors 1B does much better. It gets the coverage about right in the case of no predictability ($\beta^{(h)} = 0$). Even with mild predictability, the coverage does not fall below about 80 percent. It is only when the predictability is considerable, that it can have coverage that is substantially too low. All this is consistent with the asymptotic results under near unit roots in the previous section. The confidence interval formed using the reverse-regression delta method approach proposed in this paper always has effective coverage of at least 80 percent, and usually a good bit more. The case where this fares better than Hodrick standard errors 1B is if the predictability is considerable (population R-squared of above 30 percent).

Although in this Monte-Carlo simulation, we know the true value of $\beta^{(h)}$, in practice, of course, the researcher does not know the data generating process and so it is important that the coverage of a confidence interval be as close as possible to the nominal level uniformly in reasonable values of $\beta^{(h)}$. In this regard, Figure 1 shows that the reverse-regression delta method does best.

Coverage is of course not the only criterion for a confidence interval; precision matters too. The median width of the alternative confidence intervals is shown in Figure 2 (as before, results for other parameter configurations are in the appendix of supplemental materials). Confidence intervals with higher coverage naturally tend to have higher width[6]. The two confidence intervals based on reverse regressions have comparable width, but both are wider than the Newey-West confidence intervals. That seems to be a price worth paying given that the conventional methods consistently fail to get an effective coverage rate that is even close to the nominal level.

5 Forecasting Excess Bond Returns

We now apply the reverse regression methodology to an important predictive regression in finance; the prediction of excess bond returns using the term structure of interest rates. Many authors have found predictability long-horizon excess bond returns. For example, Fama and Bliss (1987) found that the steeper is the yield curve, the higher are the subsequent excess returns on holding a long-maturity bond. In an

[6]This is what one would expect, given that these are symmetric one-dimensional confidence intervals constructed around the parameter estimates.

influential paper, Cochrane and Piazzesi (2005) argued that while the slope of the yield curve has some predictive power for bond returns, using a combination of forward rates gives better forecasting performance, and that a "tent-shaped" function of forward rates has remarkable predictive ability for excess bond returns with R-squared values up to 44 percent. These results represent strong evidence against the expectations hypothesis of the term structure. Yet one might wonder if they are—at least in part—an artefact of small-sample econometric problems.

Let $P_{n,t}$ be the price of an n-month zero-coupon bond in month t; the per annum continuously compounded yield on this bond is $z_{n,t} = -\frac{12}{n}\log(P_{n,t})$. The excess return (over the one-month riskfree rate) from buying this bond in month t and selling it in month $t+1$ is

$$r_{n,t+1} = \log(P_{n-1,t+1}) - \log(P_{n,t}) - z_{1,t}$$

where $z_{1,t}$ is the one-month yield. We can then construct the h-period excess return $r^{(h)}_{n,t+h} = \Sigma^{h}_{j=1} r_{n,t+j}$. This is very close to—though not exactly the same as—the excess return on holding an n-month zero-coupon bond for h months over the return on holding the h-month bond for that same holding period, considered by Cochrane and Piazzesi (2005) and others.

A basic premise of term structure analysis is that today's yield curve can be used to forecast future yield curves and, hence, the excess returns on long bonds. For example, Fama and Bliss (1987) argue that when the yield curve is steep, long-term bonds can be expected to subsequently have high excess returns. Accordingly, researchers project excess returns onto the term structure of interest rates at the start of the holding period, running regressions of the form

$$r^{(h)}_{n,t+h} = \alpha + x_t'\beta^{(h)} + \varepsilon_{t+h} \tag{13}$$

where x_t is some vector of yields or spreads at time t .

We considered estimates of $\beta^{(h)}$ formed from estimating equation (13) with the long-term bond maturity, n, ranging from 2 to 5 years and the holding period, h, of 12 months. End-of-month data on zero-coupon bond yields and riskfree rates from the Fama-Bliss dataset were used.[7] The sample period is 1964:01-2009:12.

We first used the spread between the five-year and one-month yield as the sole predictor, x_t. The top panel of Table 4 shows the forward regression estimates (equation 1) of $\beta^{(h)}$ along with Newey-West standard errors and Hodrick standard errors 1B. The Newey-West standard errors indicate statistical significance at the 5 percent level (except for $n = 2$). However, using Hodrick standard errors 1B, none of the slope coefficients is statistically significant at the 5 percent level.

The middle panel of Table 4 reports the reverse regression Wald test (equation 4). The hypothesis that $\beta^{(h)} = \gamma^{(h)} = 0$ is not rejected at the 5 percent level for any maturity n.

The bottom panel shows the reverse regression delta method estimates and standard errors of $\beta^{(h)}$ (from equations 6 and 7). These are significant at the 5 percent level only in the case $n = 5$. Overall, the use of the reverse regression methods shows only marginal evidence of a relationship between the slope of the yield curve and

[7] We have also used the yields from the dataset of Gürkaynak, Sack and Wright (2007), and obtained similar results. Note also that the Fama and Bliss dataset only gives yields at 1, 2, 3, 4 and 5 year maturities (in addition to short-term risk-free rates). Following Campbell and Shiller (1991) and others, we approximate the price of an $n - \frac{1}{12}$ year bond as $\exp(-(n - \frac{1}{12})z_{n,t})$.

subsequent excess bond returns.

5.1 Forecasting Excess Bond Returns with the Term Structure of Forward Rates

We next follow Cochrane and Piazzesi (2005) in estimating equation (13), using as the predictors the one-year yield, and the one-year forward rates ending in two, three, four, and five years. This regression has five predictors, which limits the number of approaches that are available to handle econometric inference in this context, but is not a problem for the reverse regression methodology. Table 5 shows p-values from the Wald test of the hypothesis that $\beta^{(h)} = 0$ using the Newey-West standard errors, Hodrick standard errors 1B, the reverse-regression Wald test and the reverse-regression delta method. In this case, all of the Wald tests are significant at conventional significance levels. But the Newey-West p-values are very extreme—we report them in scientific notation—and they around 10^{-9}! Meanwhile, the other Wald statistics that are all based on reverse regressions—give p-values between 0.1 and 1 percent.

Finally, in Table 6, we report the point estimates of the elements of $\beta^{(h)}$ and the associated standard errors for the regression of excess bond returns on the term structure of forward rates. The top panel shows the forward regression estimates and standard errors, along with Newey-West standard errors and Hodrick standard errors 1B. The bottom panel shows the reverse regression delta method estimates and standard errors. The two sets of point estimates are virtually identical, and show the "tent-shaped" pattern highlighted by Cochrane and Piazzesi (2005). However, the two sets of standard errors that are based on the reverse regression methodology are both considerably larger than the Newey-West standard errors. Typically, they are roughly twice as big.[8]

The regression of Cochrane and Piazzesi (2005) implies that the *ex-ante* risk premium on buying a five-year bond and going short a one-year bond are both large and volatile. Some find this surprising and implausible (Sack (2006)). In this regard, it seems relevant that the underlying parameters from the return prediction equation appear to be quite imprecisely estimated.

6 Conclusion

In this paper, we have revisited the use of reverse regressions for inference in long-horizon forecasting. The reverse regression methodology of Hodrick (1992) assumes stationary predictors and gives only as a test of the null hypothesis of no predictability. In this paper, we have evaluated the properties of reverse regression methodologies (including a new variant of the reverse regression) with some predictability and/or near unit roots. We find, both using local-to-unit root asymptotics and Monte-Carlo simulations, that the reverse regression with standard χ^2 critical values offers an approach to inference in long-horizon predictive regressions that avoids serious size distortions, while working easily with an arbitrary number of predictors. In marked contrast, conventional Wald tests using the standard errors of Newey and West (1987) or Hansen and Hodrick (1980) in a long-horizon forecasting regression reject the null far too often, and indeed diverge to infinity under the null in the presence of local-

[8]Bekaert and Hodrick (2001) and Bekaert, Hodrick and Marshall (2001) are other papers arguing that some–but not all–of the evidence against the Expectations Hypothesis of the term structure owes to small-sample problems. Those papers are however considering the tests of Campbell and Shiller (1991), not the forward rate regressions of Cochrane and Piazzesi (2005).

to-unit roots (Valkanov (2003)).

We have used reverse regressions to re-examine the predictability of excess bond returns using the term structure of interest rates, considered by Fama and Bliss (1987) and Cochrane and Piazzesi (2005). We continue to find some predictability of excess bond returns, contradicting the expectations hypothesis of the term structure, and indicating the existence of time-varying term premia. However, the standard errors on the equations for predicting excess bond returns are much larger than in the forward regression using conventional heteroskedasticity- and autocorrelation-robust standard errors.

Appendix: Proofs

Proof of Theorem 1. By definition

$$\begin{aligned}
\widehat{\gamma}^{(h)} &= \left(\frac{1}{T-h}\sum_{t=h}^{T-1}\left(x_t^{(h)}-\overline{x}^{(h)}\right)\left(x_t^{(h)}-\overline{x}^{(h)}\right)'\right)^{-1}\left(\frac{1}{T-h}\sum_{t=h}^{T-1}\left(x_t^{(h)}-\overline{x}^{(h)}\right)r_{t+1}\right) \\
&= \gamma^{(h)}+\left(\hat{V}_{xx}^{(h)}\right)^{-1}\hat{V}_{x\varepsilon}^{(h)},
\end{aligned}$$

where $\hat{V}_{xx}^{(h)}$ is as defined in the text and

$$\hat{V}_{x\varepsilon}^{(h)}=\frac{1}{T-h}\sum_{t=h}^{T-1}\left(x_t^{(h)}-\overline{x}^{(h)}\right)\varepsilon_{r,t+1}.$$

Letting $\hat{u}_{t+1}=r_{t+1}-\bar{r}$, under the null of $\gamma^{(h)}=0$, the Wald statistic is

$$\begin{aligned}
F_1 &= (T-h)\hat{V}_{x\varepsilon}^{(h)\prime}\hat{V}_{xx}^{(h)-1}\hat{V}_{xx}^{(h)}\left(\frac{1}{T-h}\sum_{t=h}^{T-1}\left(x_t^{(h)}-\overline{x}^{(h)}\right)\left(x_t^{(h)}-\overline{x}^{(h)}\right)'\hat{u}_{t+1}^2\right)^{-1}\hat{V}_{xx}^{(h)-1}\hat{V}_{xx}^{(h)}\hat{V}_{x\varepsilon}^{(h)} \\
&= (T-h)\hat{V}_{x\varepsilon}^{(h)\prime}\left(\frac{1}{T-h}\sum_{t=h}^{T-1}\left(x_t^{(h)}-\overline{x}^{(h)}\right)\left(x_t^{(h)}-\overline{x}^{(h)}\right)'\hat{u}_{t+1}^2\right)^{-1}\hat{V}_{x\varepsilon}^{(h)} \\
&= \left(\frac{T-h}{T}\right)^2\left(T^{-1}\hat{V}_{x\varepsilon}^{(h)}\right)'\left(T^{-4}\sum_{t=h}^{T-1}\left(x_t^{(h)}-\overline{x}^{(h)}\right)\left(x_t^{(h)}-\overline{x}^{(h)}\right)'\hat{u}_{t+1}^2\right)^{-1}T^{-1}\hat{V}_{x\varepsilon}^{(h)}
\end{aligned}$$

Under assumptions (A1) through (A4),

$$\begin{aligned}
T^{-4}\sum_{t=h}^{T-1}\left(x_t^{(h)}-\overline{x}^{(h)}\right)\left(x_t^{(h)}-\overline{x}^{(h)}\right)'\hat{u}_{t+1}^2 &= T^{-1}\sum_{t=h}^{T-1}T^{-3}\left(x_t^{(h)}-\overline{x}^{(h)}\right)\left(x_t^{(h)}-\overline{x}^{(h)}\right)'\hat{u}_{t+1}^2 \\
&\to \frac{\sigma^2}{1-\lambda}\Omega^{1/2}\left(\int_\lambda^1\overline{J}_c^\mu(r)\,\overline{J}_c^\mu(r)'\,ds\right)\left(\Omega^{1/2}\right)' \qquad (14) \\
T^{-1}\hat{V}_{x\varepsilon}^{(h)} &= \frac{T}{T-h}T^{-1/2}\sum_{t=h}^{T-1}T^{-3/2}\left(x_t^{(h)}-\overline{x}^{(h)}\right)\varepsilon_{r,t+1} \\
&\to \frac{1}{1-\lambda}\Omega^{1/2}\int_0^{1-\lambda}\overline{J}_c^\mu(r)\,dW(r) \qquad (15)
\end{aligned}$$

while $\frac{T-h}{T}\to 1-\lambda$. Hence, under the null of $\gamma^{(h)}=0$,

$$F_1\to\left[\int_\lambda^1\overline{J}_c^\mu(r)\,dW(r)\right]'\left[\int_\lambda^1\overline{J}_c^\mu(r)\,\overline{J}_c^\mu(r)'\,dr\right]^{-1}\left[\int_\lambda^1\overline{J}_c^\mu(r)\,dW(r)\right].$$

■

Proof of Theorem 2. By definition

$$\widehat{\beta}^{(h)}=\left(\hat{V}_{xx}\right)^{-1}\hat{V}_{xr}^{(h)} \qquad (16)$$

where

$$\hat{V}_{xx} = \frac{1}{T-h}\sum_{t=1}^{T-h}(x_t-\overline{x})(x_t-\overline{x})',$$
$$\hat{V}_{xr}^{(h)} = \frac{1}{T-h}\sum_{t=1}^{T-h}(x_t-\overline{x})\,r_{t+h}^{(h)}$$

Under Assumptions (A1) through (A4),

$$\begin{aligned}
T^{-1}\hat{V}_{xx} &= \frac{T}{T-h}\frac{1}{T}\sum_{t=1}^{T-h}T^{-1/2}(x_t-\overline{x})\,T^{-1/2}(x_t-\overline{x})' \\
&\to \frac{1}{1-\lambda}\Omega^{1/2}\left(\int_0^{1-\lambda}J_c^{\mu}(r)\,J_c^{\mu}(r)'\,dr\right)\left(\Omega^{1/2}\right)' \qquad (17)
\end{aligned}$$

Using Equation (8) we can write

$$\begin{aligned}
r_{t+h}^{(h)} &= x_{t+h-1}^{(h)\prime}b+\sum_{i=1}^{h}\varepsilon_{r,t+i} \\
&= \frac{1}{T}x_{t+h-1}^{(h)\prime}\tilde{C}\beta^{(h)}+O_p(T^{-1/2})+\sum_{i=1}^{h}\varepsilon_{r,t+i}
\end{aligned}$$

and hence

$$\begin{aligned}
T^{-1}\hat{V}_{xr}^{(h)} &= \frac{1}{T-h}\sum_{t=1}^{T-h}T^{-1/2}(x_t-\overline{x})\,T^{-3/2}x_{t+h-1}^{(h)\prime}\tilde{C}\beta^{(h)} \\
&\quad +\frac{1}{T-h}\sum_{t=1}^{T-h}T^{-1/2}(x_t-\overline{x})\,T^{-1/2}\sum_{i=1}^{h}\varepsilon_{r,t+i}+o_p(1) \\
&\to \frac{1}{1-\lambda}\Omega^{1/2}\left(\int_0^{1-\lambda}J_c^{\mu}(r)\,\bar{J}_c(r+\lambda)dr\right)\left(\Omega^{1/2}\right)'\tilde{C}\beta^{(h)} \\
&\quad +\frac{1}{1-\lambda}\Omega^{1/2}\sigma\left(\int_0^{1-\lambda}J_c^{\mu}(r)\left[W(r+\lambda)-W(r)\right]dr\right) \qquad (18)
\end{aligned}$$

Combining Equations (16), (17) and (18) we have

$$\begin{aligned}
\widehat{\beta}^{(h)} &\to \left(\Omega^{-1/2}\right)'\left(\int_0^{1-\lambda}J_c^{\mu}(r)\,J_c^{\mu}(r)'\,dr\right)^{-1}\left(\int_0^{1-\lambda}J_c^{\mu}(r)\,\overline{J}_c(r)'\,dr\right)\left(\Omega^{1/2}\right)'\tilde{C}\}\beta^{(h)}\Big\} \\
&\quad +\left(\Omega^{-1/2}\right)'\sigma\left(\int_0^{1-\lambda}J_c^{\mu}(r)\,J_c^{\mu}(r)'\,dr\right)^{-1}\left(\int_0^{1-\lambda}J_c^{\mu}(r)\,(W(r+\lambda)-W(r))\,dr\right) \\
&= \left(\Omega^{-1/2}\right)'\sigma\left[\xi_1+\xi_2\beta^{(h)}\right]+\beta^{(h)}
\end{aligned}$$

The Hodrick (1B) estimator of its variance-covariance matrix is given by

$$
\begin{aligned}
W_{22} &= \left((T-h)\hat{V}_{xx}\right)^{-1}\left(\sum_{t=h}^{T-1}\left(x_t^{(h)}-\overline{x}^{(h)}\right)\left(x_t^{(h)}-\overline{x}^{(h)}\right)'(r_{t+1}-\bar{r})^2\right)\left((T-h)\hat{V}_{xx}\right)^{-1} \\
&= \left(\frac{T-h}{T^2}\hat{V}_{xx}\right)^{-1}\left(T^{-4}\sum_{t=h}^{T-1}\left(x_t^{(h)}-\overline{x}^{(h)}\right)\left(x_t^{(h)}-\overline{x}^{(h)}\right)'(r_{t+1}-\bar{r})^2\right)\left(\frac{T-h}{T^2}\hat{V}_{xx}\right)^{-1} \\
&\to \sigma^2\left(\Omega^{-1/2}\right)'\left[\int_0^{1-\lambda}J_c^{\mu}(r)\,J_c^{\mu}(r)'\,dr\right]^{-1}\left[\int_\lambda^1\overline{J}_c^{\mu}(r)\,\overline{J}_c^{\mu}(r)'\,dr\right] \\
&\quad \left[\int_0^{1-\lambda}J_c^{\mu}(r)\,J_c^{\mu}(r)'\,dr\right]^{-1}\Omega^{-1/2} \\
&= \sigma^2\left(\Omega^{-1/2}\right)'\Xi\Omega^{-1/2}.
\end{aligned}
$$

Under the null $\beta^{(h)}=\beta_0$,

$$
\begin{aligned}
F_2 &= \left(\widehat{\beta}^{(h)}-\beta_0\right)'W_{22}^{-1}\left(\widehat{\beta}^{(h)}-\beta_0\right) \\
&\to \left(\xi_1+\xi_2\beta_0\right)'\Xi^{-1}\left(\xi_1+\xi_2\beta_0\right)
\end{aligned}
$$

■

Proof of Theorem 3. By definition

$$
\begin{aligned}
T^{-1}\widehat{\theta}_1 &= \frac{T^{-1}}{T-h}\sum_{t=h}^{T-1}(x_t^{(h)}-\bar{x}^{(h)})\left(x_t{}'b+\varepsilon_{r,t+1}\right) \\
&= \frac{T^{-1}}{T-h}\sum_{t=h}^{T-1}(x_t^{(h)}-\bar{x}^{(h)})\left(\frac{1}{T}x_t{}'\tilde{C}\beta^{(h)}+\varepsilon_{r,t+1}\right)+o_p(1) \\
&= \frac{T}{T-h}\frac{1}{T}\sum_{t=h}^{T-1}T^{-3/2}(x_t^{(h)}-\bar{x}^{(h)})T^{-1/2}x_t{}'\tilde{C}\beta^{(h)} \\
&\quad +\frac{T}{T-h}\sum_{t=h}^{T-1}T^{-3/2}(x_t^{(h)}-\bar{x}^{(h)})T^{-1/2}\varepsilon_{r,t+1}+o_p(1) \\
&\to \frac{\sigma}{1-\lambda}\Omega^{1/2}\left[\left(\int_\lambda^1\bar{J}_C^{\mu}(r)B_C(r)\,dr\right)+\int_\lambda^1\bar{J}_C^{\mu}(r)dW(r)\right]=\frac{1}{1-\lambda}\theta_1^* \qquad (19)
\end{aligned}
$$

and from Equation (17)

$$
\begin{aligned}
T^{-1}\widehat{\theta}_2 &= \frac{T}{T-h}T^{-2}\sum_{t=1}^{T-h}\left(x_t-\overline{x}\right)\left(x_t-\overline{x}\right)' \\
&\to \frac{1}{1-\lambda}\Omega^{1/2}\left(\int_0^{1-\lambda}J_c^{\mu}(r)\,J_c^{\mu}(r)'\,dr\right)\left(\Omega^{1/2}\right)'=\frac{1}{1-\lambda}\theta_2^* \qquad (20)
\end{aligned}
$$

The limiting distributions of the Newey-West estimators of the variance-covariance matrices are derived as follows:

$$
\begin{aligned}
var\left(\widehat{\theta}_1\right) &= \left(\frac{1}{T-h}\right)^2 \left[\sum_{s=-h}^{h} \sum_{t=h-\min(0,s)}^{T-1-\max(0,s)} (x_t^{(h)}-\bar{x}^{(h)})(x_{t+s}^{(h)}-\bar{x}^{(h)})' r_{t+1} r_{t+s+1} \left(1-\frac{|s|}{h+1}\right)\right. \\
&\quad \left. - \sum_{s=-h}^{h} \widehat{\theta}_1 \widehat{\theta}_1' (T-h-|s|) \left(1-\frac{|s|}{h+1}\right)\right] \\
&= \left(\frac{1}{T-h}\right)^2 \left[\sum_{s=-h}^{h} \sum_{t=h-\min(0,s)}^{T-1-\max(0,s)} (x_t^{(h)}-\bar{x}^{(h)})(x_{t+s}^{(h)}-\bar{x}^{(h)})' T^{-1} x_t{}' \tilde{C}\beta^{(h)} T^{-1} x_{t+s}{}' \tilde{C}\beta^{(h)} \left(1-\frac{|s|}{h+1}\right)\right. \\
&\quad + \sum_{s=-h}^{h} \sum_{t=h-\min(0,s)}^{T-1-\max(0,s)} (x_t^{(h)}-\bar{x}^{(h)})(x_{t+s}^{(h)}-\bar{x}^{(h)})' \varepsilon_{r,t+1} \varepsilon_{r,t+s+1} \left(1-\frac{|s|}{h+1}\right) \\
&\quad + \sum_{s=-h}^{h} \sum_{t=h-\min(0,s)}^{T-1-\max(0,s)} (x_t^{(h)}-\bar{x}^{(h)})(x_{t+s}^{(h)}-\bar{x}^{(h)})' T^{-1} x_t{}' \tilde{C}\beta^{(h)} \varepsilon_{r,t+s+1} \left(1-\frac{|s|}{h+1}\right) \\
&\quad \left. + \sum_{s=-h}^{h} \sum_{t=h-\min(0,s)}^{T-1-\max(0,s)} (x_t^{(h)}-\bar{x}^{(h)})(x_{t+s}^{(h)}-\bar{x}^{(h)})' T^{-1} x_{t+s}{}' \tilde{C}\beta^{(h)} \varepsilon_{r,t+1} \left(1-\frac{|s|}{h+1}\right)\right] \\
&\quad - \widehat{\theta}_1 \widehat{\theta}_1' \left(\frac{1}{T-h}\right)^2 \sum_{s=-h}^{h} (T-h-|s|) \left(1-\frac{|s|}{h+1}\right) \\
&= \left(\frac{1}{T-h}\right)^2 \left[\sum_{s=-h}^{h} \sum_{t=h-\min(0,s)}^{T-1-\max(0,s)} (x_t^{(h)}-\bar{x}^{(h)})(x_{t+s}^{(h)}-\bar{x}^{(h)})' T^{-1} x_t{}' \tilde{C}\beta^{(h)} T^{-1} x_{t+s}{}' \tilde{C}\beta^{(h)} \left(1-\frac{|s|}{h+1}\right)\right. \\
&\quad + \sum_{t=h}^{T-1} \sum_{s=\max(h-t,-h)}^{\min(T-1-t,h)} (x_t^{(h)}-\bar{x}^{(h)})(x_{t+s}^{(h)}-\bar{x}^{(h)})' \left(1-\frac{|s|}{h+1}\right) \varepsilon_{r,t+s+1} \varepsilon_{r,t+1} \\
&\quad + \sum_{t=h}^{T-1} \sum_{s=\max(h-t,-h)}^{\min(T-1-t,h)} (x_t^{(h)}-\bar{x}^{(h)})(x_{t+s}^{(h)}-\bar{x}^{(h)})' \left(1-\frac{|s|}{h+1}\right) \varepsilon_{r,t+s+1} T^{-1} x_t{}' \tilde{C}\beta^{(h)} \\
&\quad \left. + \sum_{s=-h}^{h} \sum_{t=h-\min(0,s)}^{T-1-\max(0,s)} (x_t^{(h)}-\bar{x}^{(h)})(x_{t+s}^{(h)}-\bar{x}^{(h)})' T^{-1} x_{t+s}{}' \tilde{C}\beta^{(h)} \varepsilon_{r,t+1} \left(1-\frac{|s|}{h+1}\right)\right] \\
&\quad - \widehat{\theta}_1 \widehat{\theta}_1' \left(\frac{1}{T-h}\right)^2 \sum_{s=-h}^{h} (T-h-|s|) \left(1-\frac{|s|}{h+1}\right)
\end{aligned}
$$

Noting that $T^{-3}(x_{[Tr]}^{(h)}-\bar{x}^{(h)})(x_{[T(r+\omega)]}^{(h)}-\bar{x}^{(h)})' \to G_C(r,\omega)$, $T^{-1} x_{[Tr]}{}' \tilde{C}\beta^{(h)} \to B_C(r)$,

$$
T^{-3.5} \sum_{s=\max(h-[Tr],-h)}^{\min(T-1-[Tr],h)} (x_{[Tr]}^{(h)}-\bar{x}^{(h)})(x_{[Tr]+s}^{(h)}-\bar{x}^{(h)})' \left(1-\tfrac{|s|}{h+1}\right) \varepsilon_{r,[Tr]+s+1} \to \int_{\max(\lambda-r,-\lambda)}^{\min(1-r,\lambda)} G_C(r,\omega) \left(1-\tfrac{|w|}{\lambda}\right) dW(r+\omega)
$$

and

$$\left(\tfrac{1}{T-h}\right)^2 \sum_{s=-h}^{h} (T-h-|s|)\left(1-\tfrac{|s|}{h+1}\right) \to (\tfrac{1}{1-\lambda})^2 2\int_0^\lambda (1-\lambda-s)(1-\tfrac{s}{\lambda}) = \tfrac{1}{(1-\lambda)^2}\lambda^*,$$

this means that

$$\begin{aligned} T^{-2}var\left(\widehat{\theta}_1\right) \quad &\to \quad \left(\frac{\sigma}{1-\lambda}\right)^2 \{\int_{-\lambda}^{\lambda}\int_{\lambda-\min(0,\omega)}^{1-\max(0,\omega)} G_C(r,\omega)B_C(r)B_C(r+\omega)(1-\frac{|\omega|}{\lambda})drd\omega \\ &+\int_{\lambda}^{1}\int_{\max(\lambda-r,-\lambda)}^{\min(1-r,\lambda)} G_C(r,\omega)\left(1-\frac{|w|}{\lambda}\right)dW(r+\omega)dW(r) \\ &+\int_{\lambda}^{1}\int_{\max(\lambda-r,-\lambda)}^{\min(1-r,\lambda)} G_C(r,\omega)(1-\frac{|\omega|}{\lambda})dW(r+\omega)B_C(r)dr \\ &+\int_{-\lambda}^{\lambda}\int_{\lambda-\min(0,\omega)}^{1-\max(0,\omega)} G_C(r,\omega)B_C(r+\omega)(1-\frac{|\omega|}{\lambda})dW(r)d\omega\} - \frac{1}{(1-\lambda)^2}\lambda^*\theta_1^*\theta_1^{*\prime} \\ &= \quad \frac{1}{(1-\lambda)^2}V_{11}^* \end{aligned} \tag{21}$$

$$\begin{aligned} var\left(vech\left(\widehat{\theta}_2\right)\right) \quad &= \quad \frac{1}{(T-h)^2}\sum_{s=-h}^{h}\sum_{t=1-\min(0,s)}^{T-h-\max(0,s)} vech\left((x_t-\overline{x})(x_t-\overline{x})'\right) vech\left((x_{t+s}-\overline{x})(x_{t+s}-\overline{x})\right)' \\ &\left(1-\frac{|s|}{h+1}\right) - \sum_{s=-h}^{h} vech\left(\widehat{\theta}_2\right) vech\left(\widehat{\theta}_2\right)'(T-h-|s|)\left(1-\frac{|s|}{h+1}\right)\Big] \\ T^{-2}var\left(vech\left(\widehat{\theta}_2\right)\right) \quad &\to \quad \left(\frac{1}{1-\lambda}\right)^2\int_{-\lambda}^{\lambda}\int_{-\min(0,\omega)}^{1-\lambda-\max(0,\omega)} vech(H_C(r))vech(H_C(r+\omega))'dr\left(1-\frac{|w|}{\lambda}\right)dw \\ &-\frac{1}{(1-\lambda)^2}\lambda^* vech\left(\theta_2^*\right) vech\left(\theta_2^*\right)' \\ &= \quad \frac{1}{(1-\lambda)^2}V_{22}^* \end{aligned} \tag{22}$$

$$
\begin{aligned}
cov\left(vech\left(\widehat{\theta}_2\right), \widehat{\theta}_1\right) &= \left(\frac{1}{T-h}\right)^2 \left(\sum_{s=-h}^{h} \sum_{t=1-\min(0,s)}^{T-h-\max(0,s)} vech\left(\left(x_t - \overline{x}\right)\left(x_t - \overline{x}\right)'\right)\right. \\
&\quad (x_{t+s+h-1}^{(h)} - \bar{x}^{(h)})'(T^{-1}x'_{t+s+h-1}\tilde{C}\beta^{(h)} + \varepsilon_{r,t+s+h})(1 - \frac{|s|}{h+1}) \\
&\quad \left. - \sum_{s=-h}^{h} vech\left(\widehat{\theta}_2\right)\widehat{\theta}_1'(T-h-|s|)\left(1 - \frac{|s|}{h+1}\right)\right) \\
&= \left(\frac{1}{T-h}\right)^2 \left(\sum_{s=-h}^{h} \sum_{t=1-\min(0,s)}^{T-h-\max(0,s)} vech\left(\left(x_t - \overline{x}\right)\left(x_t - \overline{x}\right)'\right)\right. \\
&\quad (x_{t+s+h-1}^{(h)} - \bar{x}^{(h)})'T^{-1}x_{t+s+h-1}{}'\tilde{C}\beta^{(h)}\left(1 - \frac{|s|}{h+1}\right) \\
&\quad + \sum_{t=1}^{T-h} \sum_{s=\max(1-t,-h)}^{\min(T-h-t,h)} vech\left(\left(x_t - \overline{x}\right)\left(x_t - \overline{x}\right)'\right)(x_{t+s+h-1}^{(h)} - \bar{x}^{(h)})'\varepsilon_{r,t+s+h}\left(1 - \frac{|s|}{h+1}\right) \\
&\quad \left. - \sum_{s=-h}^{h} vech\left(\widehat{\theta}_2\right)\widehat{\theta}_1'(T-h-|s|)\left(1 - \frac{|s|}{h+1}\right)\right) \\
T^{-2}cov\left(vech\left(\widehat{\theta}_2\right), \widehat{\theta}_1\right) &= \frac{\sigma}{(1-\lambda)^2}\left\{\int_{-\lambda}^{\lambda}\int_{-\min(0,\omega)}^{1-\lambda-\max(0,\omega)} vech(H_C(r))\bar{J}_C^{\mu}(r+\lambda+\omega)'\Omega^{1/2\prime}B_C(r+\lambda+\omega)\right. \\
&\quad (1 - \frac{|\omega|}{\lambda})drd\omega + \int_{0}^{1-\lambda}\int_{\max(-r,-\lambda)}^{\min(1-\lambda-r,\lambda)} vech(H_C(r))\bar{J}_C^{\mu}(r+\lambda+\omega)'\Omega^{1/2\prime}(1 - \frac{|\omega|}{\lambda}) \\
&\quad \left. dW(r+\lambda+\omega)dr\right\} - \frac{1}{(1-\lambda)^2}\lambda^* vech\left(\theta_2^*\right)\theta_1^{*\prime} \\
&= \frac{1}{(1-\lambda)^2}V_{21}^* \qquad (23) \\
T^{-2}cov\left(\widehat{\theta}_1, vech\left(\widehat{\theta}_2\right)\right) &= \frac{1}{(1-\lambda)^2}V_{21}^{*\prime} = \frac{1}{(1-\lambda)^2}V_{12}^* \qquad (24)
\end{aligned}
$$

Substituting Equations (19) through (24) into Equation (12) completes the proof. ■

References

Andrews, Donald W.K. (1991): Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation, *Econometrica*, 59, pp.817-858.

Ang, Andrew and Geert Bekaert (2007): Stock Return Predictability: Is it There? *Review of Financial Studies*, 20, pp.651-707.

Bekaert, Geert and Robert J. Hodrick and David Marshall (2001): Expectations Hypothesis Tests, *Journal of Finance*, 56, pp.1357-1394.

Bekaert, Geert, Robert J. Hodrick and David Marshall (2001): Peso Problem Explanations for Term Structure Anomalies, *Journal of Monetary Economics*, 48, pp.241-270.

Campbell, John Y. (2000): Asset Pricing at the Millennium, *Journal of Finance*, 55, pp.1515-1567.

Campbell, John Y. (2001): Why Long Horizons? A Study of Power Against Persistent Alternatives, *Journal of Empirical Finance*, 9, pp.459-491.

Campbell, John Y. and Robert J. Shiller (1991): Yield Spreads and Interest Rate Movements: A Bird's Eye View, *Review of Economic Studies*, 58, pp.495-514.

Campbell, John Y. and Motohiro Yogo (2006): Efficient Tests of Stock Return Predictability, *Journal of Financial Economics*, 81, pp.27-60.

Cochrane, John H. and Monika Piazzesi (2005): Bond Risk Premia, *American Economic Review*, 95, pp.138-160.

Elliott, Graham and James H. Stock (1994): Inference in Time Series Regression When the Order of Integration of a Regressor is Unknown, *Econometric Theory*, 10, pp.672-700.

Fama, Eugene F. and Robert R. Bliss (1987): The Information in Long-Maturity Forward Rates, *American Economic Review*, 77, pp.680-692.

Goetzmann, William and Philippe Jorion (1993): Testing the Predictive Power of Dividend Yields, *Journal of Finance*, 48, pp.663-679.

Gürkaynak, Refet S., Brian Sack and Jonathan H. Wright (2007): The U.S. Treasury Yield Curve: 1961 to the Present, *Journal of Monetary Economics*, 54, pp.2291-2304.

Hansen, Lars P. and Robert J. Hodrick (1980): Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis, *Journal of Political Economy*, 88, pp.829–853.

Hodrick, Robert J. (1992): Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement, *Review of Financial Studies*, 5, pp.357-386.

Newey, Whitney K. and Kenneth D. West (1987): A Simple, Positive Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica*, 55, pp.703-708.

Phillips, Peter C.B. (1987): Towards a Unified Asymptotic Theory for Autoregression, *Biometrika*, 74, pp.535-547.

Richardson, Matthew and James H. Stock (1989): Drawing Inferences from Statistics Based on Multiyear Returns, *Journal of Financial Economics*, 25, pp.323-348.

Rossi, Barbara (2005): Testing Long-Horizon Predictive Ability with High Persistence, and the Meese-Rogoff Puzzle, *International Economic Review,* 46, pp.61-92.

Rossi, Barbara (2007): Expectations Hypotheses Tests at Long Horizons, *Econometrics Journal*, 10, pp.1-26.

Sack, Brian (2006): Comment on "Can Central Banks Target Bond Prices?" in Monetary Policy in an Environment of Low Inflation: Proceedings of the Bank of Korea International Conference 2006, Bank of Korea, Seoul, Korea.

Stambaugh, Robert (1999): Predictive Regressions, *Journal of Financial Economics*, 54, pp.375-421.

Stock, James H. (1996): VAR, Error Correction and Pretest Forecasts at Long Horizons, *Oxford Bulletin of Economics and Statistics*, 58, pp.685-701.

Stock, James H. (1991): Confidence Intervals for the Largest Autoregressive Root in U.S. macroeconomic time series, *Journal of Monetary Economics*, 28, pp.435-459.

Torous, Walter, Rossen Valkanov and Shu Yan (2004): On Predicting Stock Returns with Nearly Integrated Explanatory Variables, *Journal of Business*, 77, pp.937-966.

Valkanov, Rossen (2003): Long-Horizon Regressions: Theoretical Results and Applications, *Journal of Financial Economics*, 68, pp.201-232.

Table 1: Asymptotic Probability that F_1 exceeds the $\chi^2(1)$ critical values

	c=-30	c=-25	c=-20	c=-15	c=-10	c=-5	c=0
				$\lambda = 0.02$			
$\rho = 0$	5.5	5.8	5.3	4.7	4.1	3.5	4.8
$\rho = 0.5$	4.7	5.1	4.7	4.7	4.7	5.5	10.5
$\rho = 0.9$	7.4	7.4	7.6	8.1	8.7	10.4	21.4
				$\lambda = 0.05$			
$\rho = 0$	4.1	3.9	3.6	3.5	3.3	3.8	4.2
$\rho = 0.5$	4.6	4.3	4.1	4.5	4.8	5.3	10.2
$\rho = 0.9$	6.5	7.0	6.8	7.0	8.0	9.5	21.0
				$\lambda = 0.1$			
$\rho = 0$	3.5	3.5	3.5	3.6	3.8	3.8	5.4
$\rho = 0.5$	3.9	4.2	4.3	4.1	4.4	5.5	9.1
$\rho = 0.9$	6.1	6.4	6.1	6.9	7.6	9.1	19.4

This table shows the probability that the limiting distribution of F_1 exceeds 3.84, the upper 5th percentile for the $\chi^2(1)$ distribution. Entries were obtained by simulating the limiting distribution derived in Theorem 1.

Table 2: Asymptotic Probability that F_2 exceeds the $\chi^2(1)$ critical values

	c=-30	c=-25	c=-20	c=-15	c=-10	c=-5	c=0
			$\beta = 0, \lambda = 0.02$				
$\rho = 0$	5.1	4.9	4.5	4.5	3.9	3.1	4.4
$\rho = 0.5$	4.8	4.8	4.8	4.7	5.0	5.6	11.0
$\rho = 0.9$	6.9	6.7	7.2	7.6	8.9	10.0	21.6
			$\beta = 0, \lambda = 0.05$				
$\rho = 0$	3.7	3.8	3.7	3	3.1	2.7	3.8
$\rho = 0.5$	4.3	4.0	4.0	3.9	4.6	4.6	10.6
$\rho = 0.9$	5.7	5.8	5.9	6.1	7.1	8.1	20.7
			$\beta = 0, \lambda = 0.1$				
$\rho = 0$	3.1	3.2	3.4	3.2	3.4	3.4	4.4
$\rho = 0.5$	3.8	3.6	3.8	4.0	4.3	4.8	10.4
$\rho = 0.9$	5.8	5.9	5.7	6.5	6.6	8.4	21.1
			$\beta = 0.2, \lambda = 0.02$				
$\rho = 0$	4.8	4.8	4.6	3.8	4.1	3.2	4.9
$\rho = 0.5$	6.5	6.4	6.1	6.4	6.8	7.5	15.3
$\rho = 0.9$	10.2	10.3	10.8	11.3	11.9	15.2	29.9
			$\beta = 0.2, \lambda = 0.05$				
$\rho = 0$	4.0	4.2	3.4	3.6	3.1	3.1	4.9
$\rho = 0.5$	6.2	5.8	5.8	6.1	6.7	7.4	14.1
$\rho = 0.9$	9.6	9.5	9.8	10.2	10.6	13.6	28.0
			$\beta = 0.2, \lambda = 0.1$				
$\rho = 0$	3.4	3.1	3.4	3.2	3.1	3.4	5.1
$\rho = 0.5$	5.9	5.5	5.7	5.7	5.6	7.4	14.2
$\rho = 0.9$	11.1	10.3	10.2	9.8	10.7	13.5	27.1
			$\beta = 0.5, \lambda = 0.02$				
$\rho = 0$	5.4	5.6	4.7	4.6	4.8	4.5	7.1
$\rho = 0.5$	11.0	10.8	9.8	10.0	9.9	10.7	21.3
$\rho = 0.9$	16.2	16.2	16.5	16.8	17.8	21.1	42.3
			$\beta = 0.5, \lambda = 0.05$				
$\rho = 0$	4.4	4.7	4.4	4.2	4.7	3.9	6.3
$\rho = 0.5$	10.1	10.2	9.9	9.7	9.6	11.4	21.6
$\rho = 0.9$	16.9	16.3	16.1	16.7	17.2	20.6	38.8
			$\beta = 0.5, \lambda = 0.1$				
$\rho = 0$	5.0	4.4	4.3	4.3	4.1	4.6	6.9
$\rho = 0.5$	10.8	10.0	9.9	10.2	10.4	11.9	20.9
$\rho = 0.9$	18.0	17.7	16.4	17.4	16.9	20.0	37.5

This table shows the probability that the limiting distribution of F_2 exceeds 3.84, the upper 5th percentile for the $\chi^2(1)$ distribution. Entries were obtained by simulating the limiting distribution derived in Theorem 2.

Table 3: Asymptotic Probability that F_3 exceeds the $\chi^2(1)$ critical values

	c=-30	c=-25	c=-20	c=-15	c=-10	c=-5	c=0
			$\beta = 0, \lambda = 0.02$				
$\rho = 0$	4.4	4.3	3.7	3.7	3.9	3.7	4.0
$\rho = 0.5$	4.7	4.9	4.8	4.9	4.4	4.7	8.7
$\rho = 0.9$	7.9	8.2	7.8	8.3	8.9	10.0	15.7
			$\beta = 0, \lambda = 0.05$				
$\rho = 0$	3.5	3.5	3.0	3.2	3.0	3.0	3.8
$\rho = 0.5$	4.1	3.6	3.5	3.4	2.9	3.7	6.4
$\rho = 0.9$	6.3	6.0	5.9	6.3	6.3	7.6	13.3
			$\beta = 0, \lambda = 0.1$				
$\rho = 0$	2.4	2.7	2.7	2.6	3.1	4.4	5.5
$\rho = 0.5$	2.5	2.5	2.6	2.8	3.4	4.1	7.3
$\rho = 0.9$	2.9	3.3	3.3	3.2	4.0	5.4	10.4
			$\beta = 0.2, \lambda = 0.02$				
$\rho = 0$	3.8	4.1	3.9	4.0	4.0	4.6	5.9
$\rho = 0.5$	5.5	5.7	5.7	5.7	5.6	6.8	10.8
$\rho = 0.9$	8.3	8.8	8.6	9.4	10.9	11.8	17.0
			$\beta = 0.2, \lambda = 0.05$				
$\rho = 0$	3.3	3.6	3.8	4.0	4.2	4.5	5.3
$\rho = 0.5$	6.1	6.2	5.9	5.4	5.8	6.9	9.4
$\rho = 0.9$	9.6	9.4	9.8	10.5	11.5	12.2	15.8
			$\beta = 0.2, \lambda = 0.1$				
$\rho = 0$	3.8	3.7	3.5	3.6	4.0	4.6	6.1
$\rho = 0.5$	5.3	5.2	5.1	5.6	6.1	7.3	9.0
$\rho = 0.9$	9.0	8.4	8.0	9.0	9.6	10.7	13.1
			$\beta = 0.5, \lambda = 0.02$				
$\rho = 0$	2.7	2.7	2.6	2.7	2.5	2.3	4.3
$\rho = 0.5$	5.6	5.1	5.0	5.1	5.2	6.3	10.7
$\rho = 0.9$	7.7	8.3	8.4	8.6	9.5	10.5	17.9
			$\beta = 0.5, \lambda = 0.05$				
$\rho = 0$	4.9	5.0	5.4	5.6	5.5	5.9	7.9
$\rho = 0.5$	8.8	8.6	8.3	8.1	8.8	10.1	13.3
$\rho = 0.9$	11.9	12.3	12.6	12.9	14.2	13.9	18.9
			$\beta = 0.5, \lambda = 0.1$				
$\rho = 0$	6.5	6.8	6.8	6.9	6.5	7.8	10.4
$\rho = 0.5$	10.7	9.8	9.9	9.7	10.2	12.0	14.4
$\rho = 0.9$	15.1	14.8	14.6	15.3	15.3	15.5	19.4

This table shows the probability that the limiting distribution of F_3 exceeds 3.84, the upper 5th percentile for the $\chi^2(1)$ distribution. Entries were obtained by simulating the limiting distribution derived in Theorem 3.

Table 4: Regression of n-year excess bond returns on the yield curve slope

	n=2	n=3	n=4	n=5
Panel A: Forward Regression Estimate and SEs				
Estimator	0.41	0.85	1.26	1.74
	(0.28)	(0.39)**	(0.51)**	(0.61)***
	[0.46]	[0.62]	[0.80]*	[0.91]*
Panel B: Reverse Regression Wald Statistics				
Test statistic	0.61	1.53	2.15	3.24*
Panel C: Reverse Regression Delta Method Estimate and SEs				
Estimator	0.40	0.84	1.24	1.72
	(0.40)	(0.54)	(0.69)*	(0.81)**

Notes: This table reports the results of a regression of n-year excess bond returns on the slope of the yield curve with a holding period of $h = 12$ months. The top panel shows the forward regression estimate (equation 1) along with the Newey-West standard errors in round brackets and Hodrick standard errors 1B in square brackets. The middle panel shows the reverse regression Wald statistics testing the hypothesis that $\gamma^{(h)} = 0$ in equation (2). The bottom panel shows the estimates and standard errors (in round brackets) using the reverse regression delta method (equations 6 and 7). The data are Fama-Bliss yields spanning 1964:01 to 2009:12. One, two, and three asterisks denote significance at the 10, 5, and 1 percent levels, respectively.

Table 5: Cochrane-Piazzesi regression of n-year excess bond returns on forward rates: p-values from various Wald tests for the joint significance of the forward rates

	n=2	n=3	n=4	n=5
Newey-West	$3*10^{-6}$	$4*10^{-6}$	$5*10^{-7}$	$1*10^{-6}$
Hodrick Standard Errors 1B	0.019	0.014	0.008	0.015
Reverse Regression Wald	0.008	0.005	0.001	0.004
Reverse Regression Delta Method	0.008	0.004	0.002	0.002

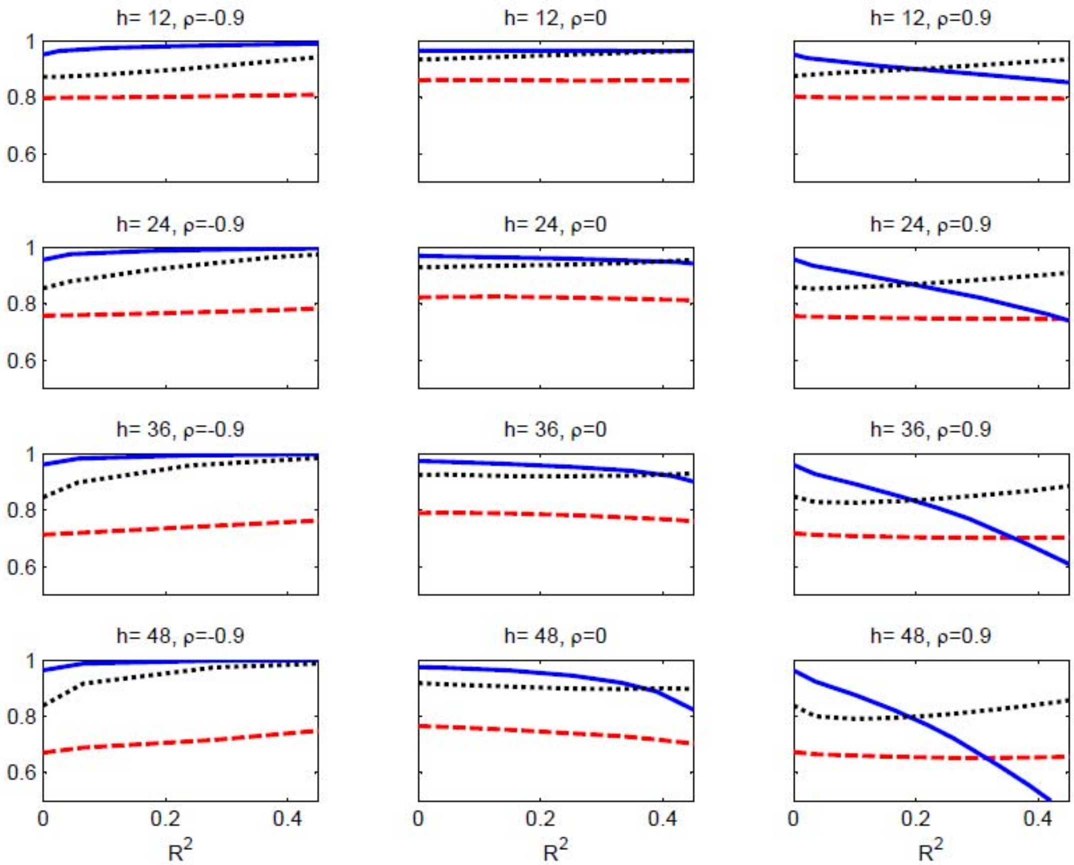
Notes: This table reports the p-values obtained from Wald tests of the hypothesis that the slope coefficients are jointly equal to zero in a regression of excess bond returns on one-year forward rates ending 1, 2, 3, 4 and 5 years hence with a holding period of $h = 12$ months. The Wald statistics are compared with $\chi^2(5)$ critical values. The Wald statistics are based on (i) the forward regression (equation 1) with Newey-West standard errors, (ii) the forward regression with Hodrick standard errors 1B, (iii) the reverse regression Wald statistic, and (iv) the reverse regression delta method (equations 6 and 7). The data are Fama-Bliss yields spanning 1964:01 to 2009:12. One, two, and three asterisks denote significance at the 10, 5, and 1 percent levels, respectively.

Table 6: Cochrane-Piazzesi regression of n-year excess bond returns on forward rates: Alternative Estimators and Standard Errors

	n=2	n=3	n=4	n=5
	Panel A: Forward Regression and Standard Errors			
1 Year Yield	-0.65	-1.36	-1.98	-2.64
	(0.39)*	(0.54)**	(0.67)***	(0.79)***
	[0.80]	[1.04]	[1.41]*	[1.58]*
1-2 Year Forward	0.00	0.23	0.33	0.70
	(0.67)	(0.89)	(1.07)	(1.26)
	[1.06]	[1.40]	[1.94]	[2.25]
2-3 Year Forward	1.66	2.62	3.30	3.62
	(0.66)**	(0.88)***	(1.10)***	(1.33)***
	[1.16]	[1.64]*	[2.03]*	[2.31]
3-4 Year Forward	0.79	1.03	1.71	1.88
	(0.44)*	(0.62)*	(0.82)**	(0.91)**
	[0.62]	[0.85]	[1.23]	[1.44]
4-5 Year Forward	-1.43	-2.07	-2.86	-2.97
	(0.46)***	(0.62)***	(0.74)***	(0.85)***
	[0.58]**	[0.83]**	[1.07]***	[1.26]**
	Panel B: Reverse Regression Delta Method			
1 Year Yield	-0.66	-1.37	-1.99	-2.67
	(0.65)	(0.89)*	(1.14)*	(1.31)**
1-2 Year Forward	0.03	0.27	0.41	0.82
	(1.04)	(1.45)	(1.85)	(2.17)
2-3 Year Forward	1.68	2.64	3.32	3.66
	(1.05)*	(1.41)*	(1.75)*	(2.09)*
3-4 Year Forward	0.79	1.03	1.73	1.90
	(0.68)	(0.92)	(1.21)	(1.40)
4-5 Year Forward	-1.46	-2.12	-2.95	-3.12
	(0.53)***	(0.72)***	(0.91)***	(1.03)***

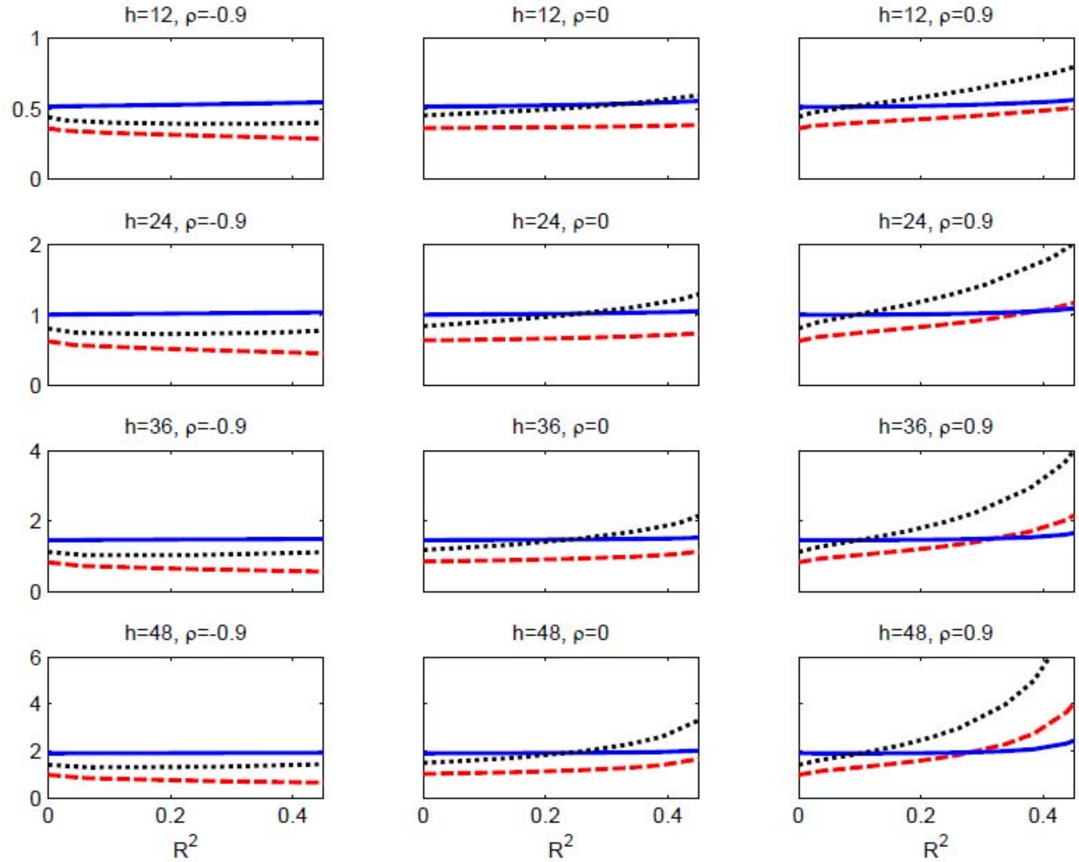
Notes: This table reports the results of a regression of n-year excess bond returns on the one-year forward rates ending 1, 2, 3, 4 and 5 years hence. The holding period is $h = 12$ months. The top panel shows the forward regression estimate (equation 1) along with the Newey-West standard errors in round brackets and Hodrick standard errors 1B in square brackets. The bottom panel shows the estimates and standard errors (in round brackets) using the reverse regression delta method (equations 6 and 7). The data are Fama-Bliss yields spanning 1964:01 to 2009:12. One, two, and three asterisks denote significance at the 10, 5, and 1 percent levels, respectively.

Figure 1: Effective Coverage of Alternative Confidence Intervals (Nominal Level: 95 Percent)



Notes: The figure shows the effective coverage of alternative confidence intervals for the slope coefficient in an *h*-step ahead predictive regression in the Monte-Carlo simulation described in the text. In this figure, the sample size is T=500 and the autoregressive parameter is ϕ=0.98. The confidence intervals are as follows: (i) Blue Solid Line—the ordinary confidence intervals based on estimating equation (1), using Newey-West standard errors with a lag truncation parameter of h, (ii) Red Dashed Line—the confidence interval based on estimating equation (1) using standard errors 1B of Hodrick (1992), and (iii) Black Dotted Line—the confidence intervals using the reverse-regression delta method proposed in this paper (equations 6 and 7). In all cases, the coverage is plotted against R^2 which is a monotone function of α, given the normalization $\alpha \geq 0$. Other simulations are included in the appendix of supplemental materials.

Figure 2: Median Width of Alternative Confidence Intervals (Nominal Level: 95 Percent)



Notes: The figure shows the median width of alternative confidence intervals for the slope coefficient in an h-step ahead predictive regression in the Monte-Carlo simulation described in the text. In this figure, the sample size is T=500 and the autoregressive parameter is ϕ=0.98. The confidence intervals are as follows: (i) Blue Solid Line—the ordinary confidence intervals based on estimating equation (1), using Newey-West standard errors with a lag truncation parameter of h, (ii) Red Dashed Line—the confidence interval based on estimating equation (1) using standard errors 1B of Hodrick (1992), and (iii) Black Dotted Line—the confidence intervals using the reverse-regression delta method proposed in this paper (equations 6 and 7). In all cases, the median width is plotted against R^2 which is a monotone function of α, given the normalization $\alpha \geq 0$. Other simulations are included in the appendix of supplemental materials.