Efficient Prediction of Excess Returns*

Jon Faust[†] **Jonathan H. Wright**[‡]

*We thank Frank Diebold, Graham Elliott, Bruce Hansen, Art Goldberger, Atsushi Inoue, Ulrich Müller, Jack Porter, Shinichi Sakata, Jim Stock, Mark Watson and two anonymous referees for helpful comments. All remaining errors are our own. 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.

[†]Department of Economics, Johns Hopkins University, Baltimore MD 21218, phone: (410) 516-7614, e-mail: faustj@jhu.edu.

[‡]Department of Economics, Johns Hopkins University, Baltimore MD 21218, phone: (410) 516-5728, e-mail: wrightj@jhu.edu.

Efficient Prediction of Excess Returns

Abstract: It is well known that augmenting a standard linear regression model with variables that are correlated with the error term but uncorrelated with the original regressors will increase asymptotic efficiency of the original coefficients. We argue that in the context of predicting excess returns, valid augmenting variables exist and are likely to yield substantial gains in estimation efficiency and, hence, predictive accuracy. The proposed augmenting variables are *ex post* measures of an unforecastable component of excess returns: *ex post* errors from macroeconomic survey forecasts, the surprise components of asset price movements around macroeconomic news announcements, or even the weather. These "surprises" cannot be used directly in forecasting—they are not observed at the time that the forecast is made—but can nonetheless improve forecasting accuracy by reducing parameter estimation uncertainty. We derive formal results about the benefits and limits of this approach and apply it to standard examples of forecasting excess bond and equity returns. We find substantial improvements in out-of-sample forecast accuracy for standard excess bond return regressions; gains for forecasting excess stock returns are much smaller.

KEYWORDS: Excess returns, efficiency, predictive regression, term premiums, seemingly unrelated regression.

JEL Classifications: C22, C53, E17, E43.

1. Introduction

Many empirical papers in finance explore the predictability of excess returns using a simple regression-based approach: estimate a regression for future excess returns based on current predictors and then measure the degree of predictive power of the estimated model. We derive a method to increase efficiency in the estimation step and show that this method can lead to substantial gains in measured forecastability.

The key idea is familiar from first-year econometrics. If we take any regression and augment it with regressors that are correlated with the error term, but are known to be uncorrelated with the original regressors in population, we increase asymptotic efficiency of the estimates of the original coefficients without compromising consistency. The augmenting variables are not of direct interest, but *soak up* some residual variance, increasing precision of the estimates of the coefficients that are of interest. This idea is an example of the familiar principle that system estimation imposing correct cross-equation restrictions is more efficient than single equation estimation,[1] but we are not aware systematic treatment of the sort we are proposing in forecasting context.

We argue that forecasting excess returns provides an excellent opportunity for gains from this approach. The standard predictive regression is of the form,

$$y_t = \beta' x_{t-h} + \varepsilon_t$$

where y_t is the excess return from $t-h$ to t and the x_{t-h} is observed at $t-h$. In classic

[1]As other examples, Hansen (1995) shows how to construct more powerful unit root tests by augmenting the autoregression with additional covariates that are known to be stationary, and Eichenbaum, Hansen and Singleton (1988) show how the power of tests of moment restrictions in a GMM context can be enhanced by imposing other moment restrictions.

applications, the excess return is for government bonds or equity portfolios, and the predictors are term spreads or dividend yields at $t-h$ (e.g., Fama and Bliss (1987), Fama and French (1988), Campbell and Shiller (1988, 1991), Cochrane and Piazzesi (2005)). The variables in such regressions in practice have several distinct properties: excess returns are highly variable and have a large unpredictable component. The predictors are smooth relative to the excess returns and much less variable. In short, we expect large residual variance relative to the variance of the regressors. In finite samples, this is equivalent to saying that the coefficient standard errors will be large.

Our goal is to soak up some of the residual variance to improve the estimates of β. Formally, a valid augmenting variable, w_t must satisfy what we call an *identifying assumption* of being uncorrelated with the regressors, x_{t-h}. If w_t is also correlated with ε_t, then augmenting the estimated regression with w_t will increase the asymptotic efficiency of the estimate of β.

We consider various sets of augmenting variables: In particular, the *ex-post* errors in published macroeconomic forecasts, and the surprise components of macroeconomic news announcements that are released between $t-h$ and t. Whether or not these measures satisfy our identifying assumptions is an important question. We derive results about the implications of small violations of the assumption of unpredictability and present tests of this assumption. In the end, however, we argue that out of sample testing presents a stringent test of the empirical benefits of the approach.

Applying our approach to the prediction of excess returns, we show that the efficiency gain in pseudo out-of-sample prediction exercises can be quite substantial for excess bond returns—a 5 to 30 percent reduction in root mean square prediction error. Gains for forecasting excess stock returns are only found in some cases, and even then are very modest. We also consider forecasting returns on orange juice futures using the weather as an augmenting variable, finding that the technique increases the precision of the slope coefficient in this context too. In no case do we find that the method substantially degrades predictive performance.

In section 2, we describe the econometric methodology in a stylized model and section 3 lays out the specific inference procedures that we apply. Sections 4, 5 and 6 then report the applications to excess bond returns, excess equity returns and orange juice futures returns, respectively. Section 7 concludes.

2. Methodology

2.1 The baseline case

This section sketches the formal logic of our idea in a simplified framework; the next section formally explores a more general case. The baseline data generating process (DGP) is a linear model for the scalar variable to be forecasted, y, a $(p \times 1)$ vector of predictors, x, and a $(k \times 1)$ vector of augmenting variables, w. We suppress any deterministic elements for simplicity. The DGP is then,

$$y_t = \beta' x_{t-1} + \varepsilon_t, \tag{1}$$

$$x_t = A x_{t-1} + v_t \tag{2}$$

$$w_t = \Gamma x_{t-1} + u_t \tag{3}$$

$t = 1, \ldots, T$. The variables y_t, x_t and w_t enter the information set at $t; \varepsilon$, v , and u are never directly observed. Assume that the process for x is stationary and that x_0

is known and fixed. The shock vector, $(\varepsilon_t, v_t', u_t')'$, is iid with $2+\delta$ finite moments for some $\delta > 0$, and its expectation conditional on x_{t-1} is 0.

We assume that $E(w_t|x_{t-1}) = 0$ which implies that $\Gamma = 0$. Note that ε_t and v_t can be correlated, but ε_t is independent of x_s, $s < t$. Thus, x is *predetermined* but not *strictly exogenous* in (1). We can also write

$$\varepsilon_t = \phi' w_t + \xi_t \tag{4}$$

where $E(\xi_t|w_t, x_{t-1}) = 0$. For the augmenting variables to be correlated with the forecast error, ϕ must be nonzero.

Without the restriction that $\Gamma = 0$, equations (1)-(3) form a classical case of seemingly unrelated regressions (SUR). Because the regressors in each equation of this system are identical, OLS and SUR estimates of β would coincide. But imposing the restriction that $\Gamma = 0$, it is straightforward to see that the SUR estimator of β is a Gaussian pseudo-maximum-likelihood estimator and is a different OLS estimator: the estimator of β in the augmented regression:

$$y_t = \beta' x_{t-1} + \phi' w_t + \xi_t \tag{5}$$

obtained by substituting equation (4) into equation (1) (Goldberger (1970)). Let $\hat{\beta}$ and $\tilde{\beta}$ be the OLS estimators of β in the unaugmented regression (equation (1)) and the augmented regression (equation (5)), respectively.

Define Ω to be the variance covariance matrix of $(\varepsilon_t, v_t', u_t')'$ and partition this matrix conformably as $\begin{pmatrix} \Omega_{\varepsilon\varepsilon} & \Omega_{v\varepsilon}' & \Omega_{w\varepsilon}' \\ \Omega_{v\varepsilon} & \Omega_{vv} & \Omega_{uv}' \\ \Omega_{u\varepsilon} & \Omega_{uv} & \Omega_{uu} \end{pmatrix}$. Note that in (4), $\phi = \Omega_{uu}^{-1}\Omega_{u\varepsilon}$. Theorem 1 compares the asymptotic distributions of $\hat{\beta}$ and $\tilde{\beta}$.

Theorem 1. $T^{1/2}(\hat{\beta}-\beta) \rightarrow_d N(0, \sigma^2\Sigma_{xx}^{-1})$ and $T^{1/2}(\tilde{\beta}-\beta) \rightarrow_d N(0, \sigma^2(1-\lambda)\Sigma_{xx}^{-1})$ as $T \rightarrow \infty$, where $\Sigma_{xx} = E(x_t x_t')$, $\sigma^2 = \Omega_{\varepsilon\varepsilon}$ and $\lambda = \sigma^{-2}\Omega_{u\varepsilon}'\Omega_{uu}^{-1}\Omega_{u\varepsilon}$.

The proofs of the theorems are collected in Appendix 1. Theorem 1 implies that both estimators are consistent, regardless of the correlation between the elements of w_t and ε_t—even if w_t is worthless augmenting the regression does not affect consistency. If w_t and ε_t are correlated, however, then $\tilde{\beta}$ is asymptotically more efficient than $\hat{\beta}$. The relative efficiency ($\frac{Var(\tilde{\beta})}{Var(\hat{\beta})}$) is $1-\lambda$ where λ is the population R^2 in a regression of ε_t on w_t. In other words, the more variation of ε_t that is explained by w_t, the greater is the reduction in the asymptotic variance of the estimator.[2]

Now consider forecast accuracy. The researcher observes x_T and wishes to forecast y_{T+1}. The standard forecast in this work is $\hat{\beta}'x_T$. Our proposed alternative substitutes $\tilde{\beta}$ for $\hat{\beta}$, giving the forecast $\tilde{\beta}'x_T$. The unconditional mean square prediction error (MSPE) for either of these forecasts is

$$\sigma^2 + T^{-1}tr(V(.)E(x_T x_T')) + o(T^{-1})$$

as $T \rightarrow \infty$, where $V(.)$ is the asymptotic variance-covariance matrix of the β estimate in question.[3]

Note that the augmenting regressors are not used directly in forming the alternative forecast; they are used only at the estimation stage to improve the precision

[2]It is tempting to regress w_t on x_{t-1} and use these residuals instead as the augmenting variables. However, this would be equivalent to standard SUR estimation of (1)-(3), which would yield the OLS estimate of β and destroy the efficiency gains.

[3]This is the unconditional MSPE as opposed to the conditional MSPE (see Phillips (1979) for a discussion of the distinction). It could equivalently be thought of as the average conditional MSPE where the parameter is estimated over the first T periods and that fixed parameter estimate is then used for forecasting over a subsequent large sample.

of the estimate of the projection coefficient of y_t on x_{t-1}. Thus, the MSPE in both cases involves the variance of ε_t. The advantage of the augmented estimator comes in the second term, which is that due to error in estimating β. This advantage of the augmented estimator diminishes asymptotically at the rate T.

Any advantage of the procedure comes from reduced variance in the estimate of β, reducing forecast error loss for the conventional mean squared-error loss function. The gains will not necessarily carry over to other loss functions, especially nonsymmetric loss functions. Indeed there is little basis for focussing on either the ordinary or augmented OLS estimates of β under more general forecast error loss functions.

The basic idea of our procedure is that the augmented regression soaks up a component of the error term thereby reducing the error variance and giving more precise parameter estimates and, hence, better forecasts. This idea is related to the recent work of Campbell and Yogo (2006), who consider a system consisting of equations (1) and (2) alone, with a scalar x_t. They note that if A is known, then v_t (the innovations to x_t) are observed and that one can then obtain a more powerful test of the hypothesis $\beta = \beta_0$ by subtracting the component of ε_t that is correlated with v_t. More precisely, if $\beta_0 = 0$ and the errors are Gaussian with known variance-covariance matrix, they show that the optimal test reduces to the conventional t-statistic in a regression of $y_t - \frac{\Omega_{v\varepsilon}}{\Omega_{vv}} v_t$ on x_{t-1}.[4] When the error variance-covariance matrix is not known, this suggests augmenting the regression of y_t on x_{t-1} with the additional

[4] This test is optimal in the sense that it is the uniformly most powerful (UMP) test in the system given by equations (1) and (2), conditional on the ancillary statistic $\Sigma_{t=1}^{T} x_{t-1}^2$. Campbell and Yogo (2006) derive the conditional UMP test for the general hypothesis $\beta = \beta_0$. It only reduces to this t-statistic when $\beta_0 = 0$.

regressor v_t . Of course the parameter A is generally not known, but Campbell and Yogo model it as being local-to-unity allowing superconsistent estimation. They show how to use Bonferroni methods in conjunction with a superconsistent estimate of A to improve inference on β.

Our setup is different. The A parameter need only be $T^{1/2}$-consistently estimable. In this case, the system estimation of (1) and (2) as in Campbell and Yogo, would reduce to OLS. We posit an additional measured variable in (3), which allows us to form a more efficient system estimator.

The key requirements for the proposed method to give an improvement in one-step-ahead mean-squared prediction error in this baseline model are that the additional variable, w_t, is uncorrelated with the predictor x_{t-1} but correlated with the error in the predictive regression. In this simplified case, we have also assumed that the errors in the predictive regression are homoskedastic and serially uncorrelated.

2.2 The General Model

The baseline model is quite stylized. The general regression that we consider in this paper is the h-period-ahead forecasting regression

$$y_{h,t} = \beta'_h x_{t-h} + \varepsilon_{h,t} \tag{6}$$

where $y_{h,t}$ is a return from time $t-h$ to time t, x_{t-h} is a px1 vector of predictors and $w_{h,t}$ is a kx1 vector of augmenting variables. We make the following assumptions:

(i) $w_{h,t} = \Gamma_h x_{t-h} + u_{h,t}$, where $\Gamma_h = 0$ and $E(u_{h,t}|x_{t-h}, x_{t-h-1}, ...) = 0$

(ii) $\varepsilon_{h,t} = \phi'_h w_{h,t} + \xi_{h,t}$, where $E(\xi_{h,t}|w_{h,t}, x_{t-h}, x_{t-h-1}, ...) = 0$

(iii) $E(\zeta_{h,t}|x_{t-h}, x_{t-h-1}, ...) = 0$, where $\zeta_{h,t} = (\xi_{h,t}, w'_{h,t})'$

(iv) $T^{-1}\Sigma_{t=1}^{T} x_t x'_t \rightarrow_p \Sigma_{xx}$, a full-rank matrix

(v) $T^{-1/2}\Sigma_{t=1}^{T} x_{t-h} \otimes \zeta_{h,t} \rightarrow_d N(0, \Omega_h)$ where

$$\Omega_h = \lim_{T\rightarrow\infty} T^{-1} E(x_{t-h}x'_{t-h} \otimes \zeta_{h,t}\zeta'_{h,t})$$

which can be partitioned conformably with $\zeta_{h,t} = (\xi_{h,t}, w'_{h,t})'$ as

$$\Omega_h = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega'_{12} & \Omega_{22} \end{pmatrix}.$$

In this model, we allow for heteroskedasticity and serial correlation in the errors, with the latter being the norm in the case of overlapping forecasts, $h > 1$. Let $\hat{\beta}_h$ denote the OLS estimator of β_h in (6), which is the standard estimator of a predictive regression. Let $\tilde{\beta}_h$ denote the OLS estimator of β_h in the augmented regression

$$y_{h,t} = \beta'_h x_{t-h} + \phi'_h w_{h,t} + \xi_{h,t} \tag{7}$$

Both $\hat{\beta}_h$ and $\tilde{\beta}_h$ will be consistent; we derive their asymptotic distributions in the next Theorem:

Theorem 2. Under assumptions (i)-(v) of the general model, $T^{1/2}(\hat{\beta}_h - \beta_h) \rightarrow_d N(0, \Sigma_{xx}^{-1} U_1 \Sigma_{xx}^{-1})$ and $T^{1/2}(\tilde{\beta}_h - \beta_h) \rightarrow_d N(0, \Sigma_{xx}^{-1} U_2 \Sigma_{xx}^{-1})$ where

$$U_1 = \omega_{11} + (I \otimes \phi'_h)\Omega_{22}(I \otimes \phi_h) + 2\omega'_{12}(I \otimes \phi'_h)$$
$$U_2 = \omega_{11}.$$

From Theorem 2, a sufficient (but not necessary) condition for the augmented estimator, $\tilde{\beta}_h$, to be more efficient than the OLS estimator, $\hat{\beta}_h$ is that $\omega_{12} = 0$. In the one-step case, as in the baseline model, $\omega_{12} = 0$, so the baseline results carryover fairly directly. In Appendix 2, we provide other conditions implying $\omega_{12} = 0$ (these are that the xs are strictly exogenous or that there is in fact no predictability, i.e. $\beta_h = 0$) and in many contexts one might expect that ω_{12} would be small. A deeper finite-sample rationale for advocating $\tilde{\beta}_h$, however, can be seen as follows.

We can see our estimators as GMM exploiting two moment conditions: $E(x_{t-h}\varepsilon_{h,t}) = 0$ and $E(x_{t-h} \otimes w_{h,t}) = 0$. The OLS estimator, $\hat{\beta}_h$ uses the identity weight matrix[5], and $\tilde{\beta}_h$ uses as the weight matrix the inverse of

$$H_{AUG} = T^{-1}\Sigma_t x_{t-h}x'_{t-h} \otimes T^{-1}\Sigma_t l_t l'_t$$

where $l_t = (\varepsilon_{h,t}, w'_{h,t})'$, iterating until convegence.

Efficient GMM instead would use the inverse of a consistent estimate of

$$H_{EGMM} = \lim_{T\to\infty} ET^{-1}\Sigma_s\Sigma_t(x_t x'_s \otimes l_t l'_s)$$

Call the efficient GMM estimator β^*_h. Under our assumptions, its limiting distribution is

$$T^{1/2}(\beta^*_h - \beta_h) \to_d N(0, \Sigma_{xx}^{-1}U_3\Sigma_{xx}^{-1})$$

where $U_3 = \omega_{11} - \omega_{12}\Omega_{22}^{-1}\omega'_{12}$.

Obviously, β^*_h is at least as efficient as both the other two estimators. If $\omega_{12} = 0$, $\tilde{\beta}_h$ is equivalent to efficient GMM and more efficient than OLS.

[5] Indeed GMM using any suitably block diagonal positive definite matrix will deliver the OLS estimator given the stated moment conditions.

Note that H_{AUG} can be seen as a version of H_{EGMM} in which we have shrunken to zero all terms involving serial correlation in the moment condition errors. Thus, our progression from β_h^* to $\tilde{\beta}_h$ to $\hat{\beta}_h$ can be seen as GMM imposing stronger and stronger shrinking restrictions to the weight matrix.

In our Monte-Carlo simulations below, we compare the small-sample properties of $\hat{\beta}_h$, $\tilde{\beta}_h$ and β_h^* in models calibrated to match broad features of our return prediction applications. We find that the augmented estimator $\tilde{\beta}_h$ generally gives the most accurate estimate of β_h. Consistent with much earlier evidence, the small-sample performance of the efficient GMM estimator, β_h^*, is quite erratic–in finite samples, a more constrained choice of the weight matrix may result in better performance of GMM (e.g. Altonji and Segal (1996)).

Our advocacy of a less-than-fully-efficient estimation on finite-sample grounds—parameter shrinkage—is, of course, completely consistent with the spirit of the return prediction literature and the forecasting literature more generally. Also, as is standard in this area, the assumptions that we have made in the general model do not completely specify the DGP. If we did specify the full DGP, then we could use the system estimator of β_h which would exploit a host of moment conditions involving the ws and xs. However, as has been widely discussed (e.g. West, Wong and Anatolyev (2009)) small-sample issues and/or misspecification of the original system may make the full system estimator badly behaved. Indeed it is for this reason that OLS on the simple forecasting regression has been a preferred approach. We are advocating modestly increasing the exploitation of information; the finite sample merits are to be found in simulations and out-of-sample prediction exercises.

2.3 Many augmenting variables: the Large k Case

There are two clear ways one could go awry in applying our approach: using a large number of augmenting variables that are of limited value, and using augmenting variables that are, in fact, predictable. In this and the next subsection, we provide some results about these two problems. For this discussion, we revert to the baseline model.

Asymptotically, adding augmenting variables cannot cause an efficiency loss, even if they are uncorrelated with ε_t. However, in a finite sample, choosing a number of augmenting variables, k, that is large relative to T will reduce efficiency. To illustrate this intuition formally, we consider an alternative asymptotic nesting of our baseline model in which the number of extra regressors goes to infinity at the same rate as the sample size so that k/T approaches a fixed, positive limit.

As usual with such asymptotic nestings, it is not that we are taking a position about what we would do if someone gave us arbitrarily large samples. Rather, we hope this alternative asymptotic theory will give a better guide as to the finite-sample distributions of $\hat{\beta}$ and $\tilde{\beta}$ in moderate sample sizes when k is fairly large relative to T.

We find that there still may be efficiency gains, but there may be losses, and the trade-off turns on how strongly the extra regressors are correlated with ε_t.

Theorem 3. Take the assumptions of the baseline model with the alterations that $k = \alpha T$ and ζ_t is Gaussian such that $\Omega_{v\varepsilon} = 0$ and $\Omega_{uv} = 0$. Then

$$T^{1/2}(\hat{\beta} - \beta) \rightarrow_d N(0, \sigma^2 \Sigma_{xx}^{-1})$$

and

$$T^{1/2}(\tilde{\beta} - \beta) \rightarrow_d N(0, \sigma^2 \tfrac{1-\lambda}{1-\alpha} \Sigma_{xx}^{-1})$$

as $T \rightarrow \infty$.

Under this asymptotic formulation, the efficiency of $\tilde{\beta}$ relative to $\hat{\beta}$ falls as k/T rises. As before, efficiency rises with λ.

2.4 Small Violations of the Identifying Assumption

The assumption that w_t and x_{t-1} are uncorrelated ($\Gamma = 0$) is central: the gains in efficiency come exclusively from imposing what we call our *identifying assumption.* If this assumption is not satisfied, then $\tilde{\beta}$ will not be consistent. Of course, it will often be difficult in practice to rule out small violations of the assumption. To illustrate the implications of such violations we consider a variation on the baseline model in which Γ is not exactly zero, but instead is local-to-zero: $\Gamma = T^{-1/2}G$. The limiting distributions of $\hat{\beta}$ and $\tilde{\beta}$ are provided in,

Theorem 4. Take the assumptions of the baseline model with the alteration that $\Gamma = GT^{-1/2}$.

$T^{1/2}(\hat{\beta} - \beta) \rightarrow_d N(0, \sigma^2 \Sigma_{xx}^{-1})$

$T^{1/2}(\tilde{\beta} - \beta) \rightarrow_d N(-G'\phi, \sigma^2(1-\lambda)\Sigma_{xx}^{-1})$

$E((y_{T+1} - \hat{y}_{T+1|T})^2) = \sigma^2 + T^{-1}\sigma^2 p + o(T^{-1})$

and

$E((y_{T+1} - \tilde{y}_{T+1|T})^2) = \sigma^2 + T^{-1} tr\{[G'\phi\phi'G + \sigma^2(1-\lambda)\Sigma_{xx}^{-1}]\Sigma_{xx}\} + o(T^{-1})$

as $T \rightarrow \infty$.

When $G \neq 0$, $\tilde{\beta}$ is biased, which will tend to degrade forecasting performance. It still may be true, however, that the mean square error of $\tilde{\beta}$ is smaller than that of

$\hat{\beta}$. Thus, we have the sort of bias-variance trade-off that often arises in forecasting. Theorem 4 suggests that the balance can go either way.

Both theorems 3 and 4 set out trade-offs, and one might seek to optimize these trade-offs to choose a number of augmenting variables, k, taking account of modest violations of the identifying assumption, $G \neq 0$. Following up this idea would require a much richer framework, and we do not pursue these lines. In the end, we believe that the case for using the augmented regression would be hard to make if we do not have strong reason to believe that w_t and x_{t-1} are very nearly uncorrelated and so we henceforth revert to our baseline model in which it is assumed that $\Gamma = 0$.

2.5 A Monte Carlo Simulation

We illustrate the potential gains or losses from using augmenting variables using a Monte Carlo simulation. The design of the experiment is

$$y_t = \beta x_{t-1} + \varepsilon_t$$
$$x_t = a x_{t-1} + v_t$$
$$\varepsilon_t = \phi' w_t + \xi_t$$

where β is normalized to zero (without loss of generality), $a = 0.9$ (to capture the high degree of persistence that is common in the explanatory variables in the returns prediction exercises we take up below), and the sample size is T. The shocks v_t and ε_t are iid standard normal random variables with correlation δ. We set ϕ to be a constant times a kx1 vector of ones such that the population R^2 in the regression of ε_t on w_t is λ, while $(w_t', \xi_t)'$ is iid N$(0, \frac{1}{\phi'\phi+1} I_{k+1})$.

Let $y_{h,t} = y_t + y_{t-1}.. + y_{t-h+1}$ and define $\varepsilon_{h,t}$ and $w_{h,t}$ in the same way. In the regression

$$y_{h,t} = \beta_h' x_{t-h} + \varepsilon_{h,t}$$

we can then consider the three estimators of β_h described in subsection 2.2 (OLS, augmented OLS and efficient GMM). The true value of β_h is zero in this setting; if $h = 1$, then this reduces to a special case of the baseline model. Table 1 shows the simulated mean square errors of $\tilde{\beta}_h$ and β_h^* relative to those of the OLS estimator $\hat{\beta}_h$ for various values of k, δ, λ and T, and for $h = 1, 4$. Entries less than one mean that the augmented OLS or efficient GMM estimators, $\tilde{\beta}_h$ or β_h^*, respectively, are doing better than the OLS estimator in a mean-square error sense.

As expected, the higher is λ (the population R^2 in a regression of ε_t on w_t), the better the augmented estimator, $\tilde{\beta}_h$, fares, consistent with Theorem 1. Meanwhile, other things being equal, the advantage of using $\tilde{\beta}_h$ rather than the OLS estimator declines as k increases relative to T, consistent with Theorem 3. Indeed $\tilde{\beta}_h$ can even give a larger mean square prediction error than OLS if k is very large relative to T, or if λ very small. But even with λ as small as 10 percent, $\tilde{\beta}_h$ is better than OLS in all the cases considered.

It is likewise true that β_h^* gives an improvement in mean-square error provided that λ is not very small and k is not too large relative to T. However, the mean-square error of β_h^* exceeds that of $\tilde{\beta}_h$ in nearly all cases, and there are indeed many cases in which β_h^* has a higher mean-square error than OLS. This estimator appears to have a more erratic performance, which must owe to difficulty in estimating the weight matrix. From this point on, the estimators of predictive regressions that we consider in this paper are the OLS estimator, $\hat{\beta}_h$, and the augmented estimator, $\tilde{\beta}_h$. Also, from this point on, we omit the h subscript denoting the horizon of the forecasting regression.

3. Our specific inference procedures

There are three natural hypotheses of interest: 1) Is the identifying assumption that the ws are unpredictable satisfied ($H_1 : \Gamma = 0$)? 2) Are the augmenting variables irrelevant ($H_2 : \phi = 0$)? 3) Is the recursive out-of-sample root mean square prediction error of the augmented approach relative to the baseline OLS approach ($RRMSPE$) smaller than one ($H_3 : RRMSPE = 1$ against the one-sided alternative $RRMSPE < 1$)? As we define it, RRMSPEs less than one indicate that the augmented approach has smaller errors. In the best case for the approach, we fail to reject the first hypothesis, and reject the latter two. Unfortunately, all of these inferences are likely to be complicated by well-known problems arising from persistent variables.

For example, the natural way to test the identifying assumption that $\Gamma = 0$ and that the augmenting variables are worthless ($\phi = 0$) is to simply use the relevant regression-based Wald test that the parameters of interest are zero. In the present case, however, conventional asymptotics for evaluating marginal significance are known to provide a very poor approximation to behavior in relevant sample sizes. The main culprit here is that in our applications—as in most predictive regressions in finance—the predictors, the xs, are highly persistent. Meanwhile, the overlapping nature of the long-horizon returns we will be predicting implies that the appropriate augmenting variables will also be overlapping and, hence, persistent as well. Regressing persistent variables on each other poses severe challenges to small-sample inference: the relationship between $\{w_t\}$ and $\{x_{t-h}\}$ is akin to a spurious regression (see, for example, Hodrick (1992), Goetzmann and Jorion (1993) Elliott and Stock (1994) and Stambaugh (1999)). Thus, we expect that the conventional asymptotic p-values would be misleading. Various conventional bootstraps do not resolve the problem because the test statistics have non-pivotal distributions in the presence of roots that are local to unity (Basawa et al. (1991)). In some simple models, such as an AR(1), a grid bootstrap provides a workable alternative (Hansen (1999)); but in more complicated models with local-to-unit roots there is no widely accepted practical solution.

Kilian's (1998) bias-adjusted bootstrap provides a pragmatic alternative that—while not asymptotically valid in the local-to-unit root case—has been shown to have relatively small size and coverage distortions in the presence of near-unit roots. We adapt this approach to our problem and report a Monte-Carlo simulation providing evidence that the approach has reasonable small-sample properties in the current context.

For our bootstrap, we first use the bias-adjusted bootstrap of Kilian (1998) to fit a VAR(1) to 1, 2, 3, 4 and 5 year zero-coupon interest rates. In each bootstrap sample, we draw interest rates from this VAR and then compute excess returns $\{y_t\}$ and predictors $\{x_{t-h}\}$. Each replication of the bootstrap separately re-samples from the ws, making them uncorrelated with both excess returns and the predictors. The details of how we re-sample the ws depend on the character of the ws and are described below. This resampling scheme ensures that $\Gamma = 0$ and $\phi = 0$ by construction under the null, while preserving the persistence properties of the variables.

We use this resampling scheme to assess the statistical significance of tests of the hypotheses that $\Gamma = 0$ and $\phi = 0$, by comparing the relevant Wald statistic to the bootstrap distribution of the test statistic.[6] We also use this method to test the hypothesis that the RRMSPE is equal to 1, by comparing the Diebold-Mariano statistic (Diebold and Mariano (1995)) to the bootstrap distribution, in a one-sided test.[7]

We use a Monte Carlo experiment summarized here and reported in detail in Appendix 3 to assess whether our concerns with conventional asymptotics are warranted and whether our bootstrap approach overcomes these problems . This experiment varies the largest autoregressive root of interest rates in the range from 0.8 to 0.99. Using the 5 percent critical value from the asymptotic χ^2 distribution, we find that the actual sizes of the Wald tests of $\phi = 0$ and $\Gamma = 0$ are about 20 percent and 50 percent, respectively. The well-known problems with inference in persistent data manifest themselves quite dramatically in this case. Meanwhile, the actual size of the nominal 5 percent bootstrap tests are between 2 and 8 percent. For the test of statistical significance of the RRMSPE, the empirical size of a nominal 5 percent bootstrap test is close to 5 percent. Thus, we conclude our inference approach–although not theoretically justified in the local-to-unit root case–is nonetheless fairly

[6] The Wald statistics are computed using Newey-West heteroskedasticity and autocorrelation consistent variance-covariance matrices with a truncation lag of h.

[7] The hypothesis that we are effectively testing in this inference procedure is that the augmenting variables are strictly exogenous with respect to $\{y_t\}$ and $\{x_t\}$. That is sufficient, but not necessary, for the augmenting variables to be irrelevant for forecasting. If there is feedback from w_t to future values of x_t, then w_t is not strictly exogenous, but is not necessarily of any help in prediction.

well calibrated, even with roots that are very close to the unit circle.

4. Predicting Excess Bond Returns

As noted in the introduction, regression-based excess return prediction is a natural application for our method. We expect these regressions to have very modest predictive power (if any) for excess returns, which are very volatile. These facts suggest that the relevant β may be difficult to estimate precisely. Further, the facts suggest that there will be a great deal of variance in the forecast error that could potentially be *soaked up* if we can find augmenting variables. These intuitions are examined in two widely studied areas: excess bond and equity return prediction.

There are many regressions predicting excess bond returns, but we take as our baseline the recent and influential work of Cochrane and Piazzesi (2005). Their predictions are based on a regression of excess bond returns on the term structure of forward rates.

To describe the regressions, define $P_{n,t}$ to be the price of an n-month zero-coupon bond in month t; the yield on this bond is $z_{n,t} = -\frac{1}{n}\log(P_{n,t})$, and the 12-month forward rate ending n months hence is $f_{n,t} = \log(P_{n-12,t}) - \log(P_{n,t})$. The return from buying an n-month bond in month $t-12$ and selling it as an $n-12$-month bond in month t is $\log(P_{n-12,t}) - \log(P_{n,t-12})$ and the excess return from holding an n-month bond for 12 months over holding a 12-month bond for that same holding period is

$$rx^n_{t-12,t} = \log(P_{n-12,t}) - \log(P_{n,t-12}) - z_{1,t-12}$$

Cochrane and Piazzesi consider the regression of excess returns on forward rates at the first five annual horizons:

$$rx^n_{t-12,t} = \beta_{0,n} + \beta_{1,n} z_{12,t-12} + \beta_{2,n} f_{24,t-12} + \beta_{3,n} f_{36,t-12} + \beta_{4,n} f_{48,t-12} + \beta_{5,n} f_{60,t-12} + \varepsilon_{t-12,t} \quad (8)$$

for $n = 24, 36, 48, 60$. They also estimate a restricted version of this model in which the coefficients on the forward rates are the same, up to a scaling factor, for each maturity n, i.e. $\beta_{j,n} = \gamma_n \beta_j$. To estimate this restricted model, they first run the regression

$$rx^*_{t-12,t} = \beta_0 + \beta_1 z_{12,t-12} + \beta_2 f_{24,t-12} + \beta_3 f_{36,t-12} + \beta_4 f_{48,t-12} + \beta_5 f_{60,t-12} + \varepsilon^*_{t-12,t} \quad (9)$$

where $rx^*_{t-12,t} = \frac{1}{4}\Sigma^5_{j=2} rx^{12j}_{t-12,t}$, and then regress the excess returns on the fitted values from (9). We focus on improving the precision of the estimates in equations (8) and (9). Our baseline regressions consist of (8) and (9) estimated using the CRSP Fama-Bliss dataset of monthly zero-coupon bond prices.

We consider several different sets of augmenting variables, each constructed as a sort of *ex post* forecast error from some forecast related to the macroeconomy:

A1. *Ex post* errors from the Survey of Professional Forecasters. Each quarter, in the middle of February, May, August and November, the Survey of Professional Forecasters (SPF) reports analysts' predictions for several variables over the next four quarters. For each SPF back to the beginning of the survey in 1968Q4 we take the median forecast of nominal GDP growth, GDP deflator inflation and the unemployment rate four quarters hence and then take the differences between these forecasts and the actual realized values to form the augmenting variables $\{w_t\}$.[8] Because the

[8]For GDP growth and GDP deflator inflation, this is the annualized growth rate from the quarter

survey data are available only at the quarterly frequency, the regressions (8) and (9) are run only using data from each January, April, July and October (i.e. just before the survey) for a total of four observations per year. For example, we use the *ex-post* forecast error from the forecast made in February as the augmenting variable for predicting 12-month excess bond returns starting in January. The timing ensures that only what forecasters learned *after* time $t-12$ goes into the augmenting variables[9].

A2. Expanded SPF errors. Starting with the 1981Q3 survey, the SPF expanded the set of variables being predicted to include the CPI and some interest rates. Accordingly, for each SPF back to 1981Q3, we take the median predictions of the variables considered in A1 plus CPI inflation, short-term Treasury bill yields and long-term Treasury bond yields and again construct realized forecast errors as described above. We then take the first 3 principal components of these 6 realized forecast errors to form the augmenting variables $\{w_t\}$. As with A1, only four observations are used per year.

A3. A News Index of Macro Announcement Surprises. We take the following monthly macroeconomic news announcements: CPI, durable goods orders, housing starts, industrial production, index of leading indicators, nonfarm payrolls, PPI, retail sales and unemployment. For each month and each of these announcement types, we construct the difference between the actual released value and the expected value as found in the MMS survey taken the previous Friday. We form news index as a

---

before the survey to four quarters later. As the actual realized values of the series, we use the first released values from the Federal Reserve Bank of Philadelphia's realtime dataset.

[9]The survey deadline date is a few days before the SPF publication date, but is always in the second month of each quarter.

weighted average of these surprises, giving each type of release a weight equal to the slope coefficient in a regression of the intraday changes in the fourth Eurodollar futures contract[10] from 5 minutes before the announcement until 15 minutes afterwards on the surprise component of the news announcement.[11] This is designed to weight each type of announcement by its market impact. We then cumulate the resulting index over all months from $t-h+1$ to t, inclusive, to form the augmenting variable w_t where h denotes the horizon of the regression. Our data for these announcements and, hence, the news index spans 1985:02 to 2006:12.

A4. Expanded News Index. As in A3, except adding the following announcements as well: capacity utilization, core CPI, factory orders, the advance release of GDP[12], new home sales, personal consumption expenditures, core PPI, retail sales excluding autos. Our data for this larger set of announcements go back to 1989:09. The extra variables are not available earlier.

A5. Further Expanded News Index. This is as in A3, except adding the following

[10] This is the fourth contract in the quarterly cycle and settles to the three-month interest rate about one year hence.

[11] Some announcements come out concurrently. In A3, A4 and A5, the slope coefficients were obtained from a single regression of the intraday change in the fourth Eurodollar futures rate on the surprise components of all of the following announcements: capacity utilization, consumer confidence, CPI (total and core), durable goods orders, the employment cost index, factory orders, the advance release of GDP, hourly earnings, housing starts, initial jobless claims, industrial production, the index of leading indicators, the Michigan survey, NAPM, nonfarm payrolls, new homes sales, personal consumption expenditures, PMI, PPI (total and core), retail sales (total and ex autos) and unemployment. Each surprise was set to zero whenever that particular announcement type did not come out or was missing from our dataset. The regression was run over the period 1982 to 2006.

[12] This is the one quarterly release that we consider. The monthly advance GDP surprise series is set to zero in all months for which there was no advance GDP announcement.

announcements as well: consumer confidence, initial jobless claims[13], the NAPM index. Our data for this largest set of announcements go back only to 1991:07.

A6. Alternative Aggregation of News Index. We take the first three principal components of the monthly surprises that go into the construction of the index in A5. We then cumulate these principal components over all months from $t-h+1$ to t, inclusive, to form the augmenting variables w_t.

The cumulative economic news indexes in A3, A4 and A5 are plotted in Figure 1. If the SPF and MMS forecasts are predictions of upcoming releases and quarterly macroeconomic data, respectively, made by a rational forecaster who minimizes quadratic loss, then these forecasts must be efficient (i.e. equal to the conditional expectation) and all the variables A1–A6 must be orthogonal to everything in the information set at the time the forecasts were made, including x_{t-h}. Evidence on the efficiency of SPF survey forecasts is mixed. Froot (1989) and Romer and Romer (2000) report evidence against the efficiency of survey forecasts, but Thomas (1999), Mehra (2002) and Ang, Bekaert and Wei (2007) report more favorable evidence. Much of the discrepancy appears to relate to the sample period considered. In the 1970s and early 1980s the surveys appear to have had poor success in forecasting some variables, notably inflation, but have been more successful subsequently. Evidence on the efficiency of MMS survey forecasts (forecasts for a specific news release taken the previous Friday) is more uniformly favorable (see, for example, Balduzzi, Elton and Green (2001)).

In any case, forecast efficiency is a sufficient but not necessary condition for

[13] This is the one weekly release that we consider. All surprises within a given month are cumulated to form the monthly claims surprise series.

our identifying assumption. If the expectations are efficient but for an expectational error that is orthogonal to x_{t-h}, then w_t and x_{t-h} will still be uncorrelated and the augmented estimator will be consistent for the population projection coefficient of y_t onto x_{t-h}. Our identifying assumption will only fail if the expectational error is correlated with x_{t-h}, which seems unlikely to us, but which we test below.

With the sets of augmenting variables defined, we can now complete our description of the method for re-sampling from the augmenting variables for the bootstrap (owing to differences in the SPF versus news index variables, the re-sampling methods are slightly different in these two case). The SPF data are for overlapping forecast periods and to preserve this structure insofar as possible, we draw blocks of four-quarter-ahead SPF forecast errors with a block length of 3 years. For the news index, we have underlying monthly surprises, which are not overlapping and are arguably uncorrelated. Thus, we re-sample randomly from the monthly surprise indexes, and then, as with the actual data, we aggregate these monthly surprise data between $t-h$ and t to form augmenting variables in each bootstrap sample. As a result, the augmenting variables in the bootstrap sample are by construction of no value in increasing efficiency ($\phi = 0$ in (7)) and are uncorrelated with the predictors ($\Gamma = 0$).

4.1 Results

The results from estimating equation (8) using the regressors A1-A6 in the augmented estimator, as well as the results of the baseline OLS regression over the same sample periods for $n = 24, 36, 48$ and 60 are shown in Table 2. The corresponding results for the estimation of equation (9) are also shown in Table 2. The Table reports estimates of β and asymptotic standard errors from the regressions, both baseline and augmented.

Consistent with the theory, the standard errors on the elements of β in the augmented regressions are typically—though not always—lower than in the baseline regressions. Often they are often substantially lower. This is a preliminary indication that the inclusion of additional regressors may be improving efficiency.

Cochrane and Piazzesi emphasized a "tent" shape in the coefficients on the forward rates whereby the shortest and longest term forward rates have negative coefficients while the intermediate term forward rates have positive coefficients. In some cases, the inclusion of the additional regressors indicates a more pronounced tent shape than is found in the OLS regressions.

Table 2 also gives the Wald test statistics testing the hypothesis that the coefficients on the augmenting regressors are jointly equal to zero ($\phi = 0$) which would imply that the augmenting variables are not correlated with the forecast error. For augmenting variables A1–A5, this hypothesis is rejected at conventional significance levels, typically at levels between 0.001 and 0.05. For variables A6 (panel 2-6), the hypothesis is not rejected at the 10 percent level.

Finally Table 2 gives p-values for testing our identifying restriction $\Gamma = 0$. This assumption is never rejected at conventional significance levels using the bootstrap p-values. It is true, however, that for variables A1, the result is borderline with a p-value of about 0.12.

4.2 Pseudo-Out-of-Sample Forecasting

Perhaps the most stringent test of the practical usefulness of the proposed approach to prediction can be obtained in a standard recursive out-of-sample prediction exercise. Starting half-way through each of the respective estimation periods, we estimate equation (8) using all data back to the beginning of the sample period that would have been available in that month, and then construct predictions of excess bond returns over the subsequent year using benchmark and augmented regressions; we repeat this estimation and prediction exercise in each subsequent month, in each case using all data back to the start of the sample period. We compute the out-of-sample root mean square prediction error for each of the augmented regressions, relative to the out-of-sample root mean square prediction error from the corresponding baseline regression model. The results are reported in Table 3 for both the unrestricted, (8), and restricted (9), models. A relative root mean square prediction error (RRMSPE) below one means that the augmented regression is giving better out-of-sample predictions of excess bond returns than the baseline.

As can be seen from the Table, the relative root mean square prediction errors are mostly between 0.7 and 0.95, implying about a 5-30 percent reduction in root mean square prediction error. The best results obtain with the SPF regressors (additional regressors A1 and A2) and the news index that combines the largest number of announcements (additional regressors A5). The weakest results are found when using the three principal components of the surprises as additional regressors (A6).

The improvement in root mean square prediction error is significant at the 5 percent level for additional regressors A1, A2 and A5 and the p-values are between 5 and 20 percent for A3 and A4. Using the principal components of the surprises as additional variables (A6), the improvement in root mean square prediction error is not close to being significant at any conventional level.

Thus, our new method leads to reductions in the estimated standard error of the estimator of β and improvements in out-of-sample RMSPE. One may well wonder, however, whether the improvements in RMSPE are plausibly due to the mechanism we describe. To shed some light on this, we did a "back-of-the-envelope" calculation. The expected relative RMSPE is approximately $\sqrt{\frac{\sigma^2+(1-\lambda)tr(V_{\hat{\beta}}\Sigma_x)}{\sigma^2+tr(V_{\hat{\beta}}\Sigma_x)}}$ where σ^2 denotes the error variance, $\Sigma_x = E(x_t x_t')$, λ is the population R-squared in the regression of the errors on the augmenting variables, $V_{\hat{\beta}}$ is the estimated variance-covariance matrix of the parameter estimates (which is $O_p(T^{-1})$) and x_t is the vector of predictors. We can form a crude estimate of the expected improvement by plugging in estimates of these parameters generated from our application. When we do this, we get estimated mean improvements in relative RMSPEs in the range of 0.92 to 0.95.[14] While our improvements are generally a bit larger than our estimate of the expected improvement, these crude estimates of expected improvement are broadly consistent with what we report for additional regressors A1-A4 and A6.

5. Predicting Excess Stock Returns

The second predictive regression that we consider is the prediction of excess stock returns, following authors such as Fama and French (1988), Campbell and Shiller (1988) and Ang and Bekaert (2007). We use the following notation: the return on the CRSP value-weighted portfolio from month $t-1$ to month t is $R_{t-1,t} = \log(\frac{P_t+D_t}{P_{t-1}})$

[14] We plug in the standard regression estimates of σ^2, Σ_x and λ. In the current context, the standard Newey-West estimate of $V_{\hat{\beta}}$ is well known to work poorly in small samples, so we fit a VAR(1) to yields and forward rates over the period since January 1985 and used this as the data-generating process to form a Monte-Carlo estimate of $V_{\hat{\beta}}$ in samples of the average sample size for the in-sample period in our forecasting exercise.

where P_t denotes the price and D_t denotes the dividend in month t. We define three-month excess stock returns as $rx^S_{t-3,t} = \Sigma^2_{j=0} R_{t-3+j,t-3+j+1} - z_{1,t-3}$, where $z_{1,t-3}$ is the three-month Fama-Bliss risk-free rate. Define the log dividend-price ratio in month t as $dp_t = \log(\Sigma^{11}_{j=0} D_{t-j}/P_t)$. Summing the dividends over the past year deals with the seasonality in dividends. Finally define the stochastically detrended short-term interest rate as $\tilde{r}^{RF}_t = r^{RF}_t - \frac{1}{4}(r^{RF}_t + r^{RF}_{t-3} + r^{RF}_{t-6} + r^{RF}_{t-9})$ where r^{RF}_t is the three-month Fama-Bliss riskfree rate. We consider predicting excess stock returns with the following predictors: dp_t alone, $\tilde{r}^{RF}_t$ alone, and dp_t and $\tilde{r}^{RF}_t$ together in the following three regressions:

$$rx^S_{t-3,t} = \alpha_{1,0} + \alpha_{1,1} dp_{t-3} + \varepsilon_{1t} \tag{10}$$

$$rx^S_{t-3,t} = \alpha_{2,0} + \alpha_{2,1} \tilde{r}^{RF}_{t-3} + \varepsilon_{2t} \tag{11}$$

and

$$rx^S_{t-3,t} = \alpha_{3,0} + \alpha_{3,1} dp_{t-3} + \alpha_{3,2} \tilde{r}^{RF}_{t-3} + \varepsilon_{3t} \tag{12}$$

As in section 3, we take these regressions as the baseline prediction equations for excess stock returns, and our focus is on improving the precision of the estimates in equations (10), (11) and (12). The horizon here is $h = 3$. We estimate these baseline regressions and augmented regressions including the augmenting regressors A1-A6, where these news surprise measures are now cumulated over 3-month rather than 12-month horizons.

Table 4 shows the coefficient estimates and Wald tests from the estimation of (10), (11) and (12) both with and without the additional regressors. Exactly the same information is shown as for the corresponding excess bond return prediction equations. With additional regressors A1 and A2 (SPF forecast errors), the standard errors in the augmented regressions are a bit lower than those in the baseline regressions. The Wald test testing the hypothesis that the additional variables are relevant is significant in (10), (11) or (12), using the bootstrap p-values. On the other hand, using additional regressors A3-A6 (announcement surprise measures), the standard errors in the augmented regressions are little changed from those in the baseline regressions and the Wald test for the joint relevance of these extra regressors is not significant. This is quite different from what we found for bond returns, but is perhaps not surprising since many authors have found that a considerably greater fraction of bond price movements can be explained by macroeconomic news announcements than is the case for stock prices (see e.g. Andersen, Bollerslev, Diebold and Vega (2007)).

For each of the augmented regressions in Table 4, we recursively compute the out-of-sample mean square prediction error relative to that from the baseline regression starting half-way through the sample period. The results are reported in Table 5. Again, the results are different depending on whether one uses the SPF forecast errors or announcement surprise measures as additional variables. With the SPF forecast errors, the RRMSPE is a bit below 1, showing some improvement in root mean square prediction error. Using our bootstrap test, the improvement is statistically significant, at least at the 10 percent level, in all cases. The improvement is more modest than was obtained in forecasting excess bond returns with additional regressors A1 and A2. Meanwhile, with additional regressors A3-A6, the RRMSPE is around 1 in all cases and there are no cases in which the augmented regression gives a significant improvement in out-of-sample forecasting performance, as one might expect given that the announcement surprise measures do not have a significant association with quarterly excess stock returns.

6. Predicting Returns on Orange Juice Futures

The final predictive regression that we consider in this paper is quite different, namely the prediction of returns on frozen orange juice concentrate futures over the course of the winter, using the weather as an augmenting variable. It is meant as an illustration of other contexts in which the methodology proposed in this paper may be useful. Let $R_{t-1,t}$ denote the returns to buying a March orange juice futures contract on the last day of the previous November and selling it on the last day of February. We consider regressions of these returns on two predictors that are commonly used in forecast commodity futures returns (see, for example, Bessembinder (1992)). These are (i) the slope of the futures curve at the start of the holding period (the log difference between the prices of nine-month-ahead and front futures contracts at the end of November) and (ii) the net long positions of speculators in the orange juice futures market at the end of November from the Commitment of Traders reports of the Commodity and Futures Trading Commission. For the augmenting variable, w_t, we use the total number of freezing degree days at Orlando airport during December, January and February. This is calculated as the sum of the number of degrees that the minimum temperature fell below 32^0F, summed across all days in these three months that the temperature fell below freezing.

Our identifying assumption is that w_t is unforecastable at the end of November. The augmenting variable will help with efficiency as long as the weather is correlated with orange juice futures returns, which seems very likely as frost is very damaging to orange trees and central Florida is a major orange-growing area.

The results of the baseline and augmented regressions are shown in Table 6. For both regressions, including the augmenting variable reduces the standard error on the predictor. Using the slope of the futures curve as the predictor, we do not reject the null of no predictability in either case. Using the net long speculative positions as the predictor, the t-statistic on the slope coefficient goes from 1.94 in the baseline regression to 2.71 in the augmented regression, and so controlling for the weather allows us to reject the null of no predictability.

7. Conclusions

Researchers using a regressor x_{t-h} to forecast excess returns, y_t, conventionally regress the excess returns on x_{t-h} and use the resulting coefficients for forecasting. However, if there exists a variable w_t that is correlated with the regression error, but not with x_{t-h}, then a more efficient approach to estimating the coefficient on x_{t-h} in the forecasting regression is to augment the regression with w_t. This may in turn enable better forecasts to be constructed, because the coefficient on x_{t-h} is more precisely estimated, even though w_t is not observed at the time the forecast is made, and so cannot be directly used in prediction.

In this paper, we demonstrate the merits of augmenting the estimation model for predictive regressions with *ex post* measures of any unpredictable component of variable being forecasted. This method is most likely to yield advantages in cases such as forecasting excess returns where there is a large unforecastable component and precision of the coefficient estimates is likely to be a major issue. In the excess returns context, we argue that *ex-post* SPF survey forecast errors and the surprise components of macroeconomic news announcements may satisfy the required conditions for augmenting variables.

We demonstrate the merits of the approach using canonical predictive regressions for excess bond and equity returns. The gains are quite pronounced in our extension of the Cochrane and Piazzesi (2005) study of excess bond returns. We find little, if any, gains in conventional equity returns regressions. Our goal in the empirical work was to show the benefits in well-known cases. We suspect further gains may be found in other cases and using other augmenting variables. Indeed, we present a small illustration for agricultural commodity future returns using the weather as an augmenting variable. Other possible augmenting variables include data revisions (Koenig et al. (2003)) and oil supply shocks (as identified, for example, by Kilian (2006)), though, at least for the oil supply shocks, we think that the case that these are uncorrelated with the predictors is relatively weak.

The approach could of course easily be misused, say, by searching over a large set of potential augmenting variables for those that give the greatest reduction in (in sample) standard errors. While out-of-sample tests can provide some protection against the sort of false inference this could promote, we believe that, in practice, our approach should only be entertained if the researcher has a strong belief that the identifying assumption is satisfied. In any case, we think our results suggest that the approach deserves serious consideration.

Appendix 1: Proof of Theorems

Proof of Theorem 1.

This follows from the usual formula for the asymptotic distribution of OLS given that $Var(\varepsilon_t) = \sigma^2$ and

$$Var(\xi_t) = Var(\varepsilon_t - \phi' w_t) = \Omega_{\varepsilon e} + \phi'\Omega_{uu}\phi - 2\phi'\Omega_{u\varepsilon} = \Omega_{\varepsilon e} - \Omega'_{u\varepsilon}\Omega_{uu}^{-1}\Omega_{u\varepsilon} =$$
$$\sigma^2 - \Omega'_{u\varepsilon}\Omega_{uu}^{-1}\Omega_{u\varepsilon} = \sigma^2(1-\lambda)$$

Proof of Theorem 2.

Since $\hat{\beta}_h - \beta_h = (\Sigma_{t=1}^T x_t x_t')^{-1}\Sigma_{t=1}^T x_t \varepsilon_{h,t}$ and $\tilde{\beta}_h - \beta_h = (\Sigma_{t=1}^T x_t x_t')^{-1}\Sigma_{t=1}^T x_t \xi_{h,t}$, the result follows from the facts that $T^{-1}\Sigma_{t=1}^T x_t x_t' \rightarrow_p \Sigma_{xx}$, $T^{-1/2}\Sigma_{t=1}^T x_t \varepsilon_{h,t} \rightarrow_d N(0, U_1)$ and $T^{-1/2}\Sigma_{t=1}^T x_t \xi_{h,t} \rightarrow_d N(0, U_2)$.

Proof of Theorem 3.

We begin with a lemma proved by Bekker (1994) (lemma 2 in that paper).

Lemma. Let U be any $n \times m$ matrix with rows $\{u_i\}_{i=1}^n$ that are iid Gaussian with mean zero and variance-covariance matrix Φ, c be any fixed $m \times 1$ vector, and let P be any nonstochastic $n \times n$ projection matrix of rank r. If $r/n \rightarrow \psi$ as $n \rightarrow \infty$, then

$$\frac{1}{n^{1/2}}(U'PUc - E(U'PUc)) \rightarrow_d N(0, \psi c'\Phi c\Phi + \psi\Phi cc'\Phi) \qquad \text{(A1)}$$

and

$$n^{-1}U'PUc \rightarrow_p \psi\Phi c \qquad \text{(A2)}$$

The result for $\hat{\beta}$ follows from the usual formula for the asymptotic distribution of OLS with a fixed number of regressors. To prove the result for $\tilde{\beta}$, write the model in matrix form as $y = X\beta + W\phi + \xi$ where W is a $T \times k$ matrix, the t^{th} row of which is w_t'. Under the stated assumptions, the regressors in this equation are strictly exogenous and so we may condition on them. We can then write

$$\tilde{\beta} = \beta + (X'P_W X)^{-1} X' P_W \xi \tag{A3}$$

where $P_W = I - W(W'W)^{-1}W'$.

Now P_W is a projection matrix of rank $T - k$. In the notation of the lemma, let $U = [X \quad \xi]$, $n = T$ and

$$\Phi = \begin{pmatrix} \Sigma_{xx} & 0 \\ 0 & Var(\xi_t) \end{pmatrix}$$

From (A2), letting c be each of the first p standard basis vectors in turn, conditional on W,

$$T^{-1} X' P_W X \rightarrow_p (1 - \alpha)\Sigma_{xx}$$

From (A1), letting c be the $p + 1$st standard basis vector, conditional on W,

$$T^{-1/2} X' P_W \xi \rightarrow_d N(0, Var(\xi_t)(1 - \alpha)\Sigma_{xx})$$

and so, from (A3), again conditional on W,

$$T^{1/2}(\tilde{\beta} - \beta) \rightarrow_d N(0, \tfrac{1}{1-\alpha} Var(\xi_t)\Sigma_{xx}^{-1})$$

As this limiting distribution does not depend on W, it holds unconditionally as well. As $Var(\xi_t) = \sigma^2(1 - \lambda)$, it follows that

$$T^{1/2}(\tilde{\beta}-\beta) \to_d N(0, \tfrac{\sigma^2(1-\lambda)}{1-\alpha}\Sigma_{xx}^{-1})$$

as required.

Proof of Theorem 4.

The proof for $\hat{\beta}$ is as in Theorem 1. Turning to $\tilde{\beta}$, let $\phi = \Omega_{uu}^{-1}\Omega_{u\varepsilon}$ and $\xi_t = \varepsilon_t - \phi' u_t$. From (1),

$$y_t = \beta' x_{t-1} + \varepsilon_t = \beta' x_{t-1} + \phi' u_t + \xi_t = \beta' x_{t-1} + \phi'(w_t - \Gamma x_{t-1}) + \xi_t = \beta^{*\prime} x_{t-1} - \phi^{'} w_t + \xi_t$$

where $\beta^* = \beta - \Gamma'\phi$ and $E(\xi_t | x_{t-1}, w_t) = 0$. The estimator $\tilde{\beta}$ is the OLS estimate of the coefficient on x_{t-1} in this regression. Since $\Gamma = GT^{-1/2}$, $T^{1/2}(\beta - \beta^*) = G'\phi$. From the usual formula for the asymptotic distribution of OLS, given that $Var(\xi_t | x_{t-1}, w_{t-1}) = \Omega_{\varepsilon e} - \Omega_{u\varepsilon}' \Omega_{uu}^{-1} \Omega_{u\varepsilon} = \sigma^2(1-\lambda)$, and

$$T^{-1}\Sigma x_{t-1} w_t^{'} = T^{-1}\Sigma x_{t-1}(x_{t-1}^{'}\Gamma' + u_t^{'}) = T^{-3/2}\Sigma x_{t-1} x_{t-1}^{'} G' + T^{-1}\Sigma x_{t-1} u_t^{'} \to_p 0,$$

it follows that

$$T^{1/2}(\tilde{\beta} - \beta^*) \to_d N(0, \sigma^2(1-\lambda)\Sigma_{xx}^{-1})$$

Finally, since $T^{1/2}(\tilde{\beta} - \beta) = T^{1/2}(\tilde{\beta} - \beta^*) - T^{1/2}(\beta - \beta^*)$ we have $T^{1/2}(\tilde{\beta} - \beta) \to_d N(-G'\phi, \sigma^2(1-\lambda)\Sigma_{xx}^{-1})$, as required. The formulas for the unconditional mean square prediction errors follow from simple algebra.

Appendix 2: Cases in which $\omega_{12} = 0$

This appendix gives two other cases in which $\omega_{12} = 0$ in the notation of the general model in subsection 2.2 (besides the case of the one-period forecasting horizon discussed in the text). We assume in all cases that the DGP for one-period returns, y_t, is

$$y_t = \beta' x_{t-1} + \varepsilon_t$$
$$x_t = A x_{t-1} + v_t$$
$$\varepsilon_t = \phi w_t + \xi_t$$

where $(\xi_t, w_t', v_t')'$ is iid $N(0, \bar{\Omega})$ and is independent of $\{x_{t-1}, x_{t-2}, ...\}$ and the roots of A lie outside the unit circle.

Aggregating this to the corresponding model for $h-$period returns gives

$$y_{h,t} = \beta_h' x_{t-h} + \varepsilon_{h,t}$$
$$\varepsilon_{h,t} = \phi_h w_{h,t} + \xi_{h,t}$$

where $y_{h,t} = \Sigma_{i=0}^{h-1} y_{t-i}$ are h-period returns, $\beta_h = \beta \Sigma_{i=0}^{h-1} A^i$, $\varepsilon_{h,t} = \Sigma_{i=0}^{h-1} \varepsilon_{t-i} + \beta \Sigma_{j=1}^{h-1} [\Sigma_{k=1}^{j} A^{k-1}] v_{t-j}$, $w_{h,t} = \Sigma_{i=0}^{h-1} w_{t-i}$ is the h-period sum of the augmenting variables, $\phi_h = E(w_{h,t} w_{h,t}')^{-1} E(w_{h,t} \varepsilon_{h,t})$ and $\xi_{h,t} = \varepsilon_{h,t} - \phi_h w_{h,t}$. Note that $E(\xi_{h,t} | w_{h,t}, x_{t-h}, x_{t-h-1}, ...) = E(\xi_{h,t} | w_{h,t}) = 0$.

The model for $h-$period returns satisfies assumptions (i)-(v) of the general model in subsection 2.2. And there are two cases in which $w_{12} = 0$, where w_{12} denotes the cross-spectral density at zero frequency between $x_{t-h} \xi_t$ and $x_{t-h} \otimes w_t$. These cases are:

Case 1: x_t is strictly exogenous meaning that $\{v_t\}$ is independent of $\{(\xi_t, w_t')'\}$.

Case 2: $\beta = 0$.

To show this, note that in either case 1 or case 2, $\phi_h = E(w_t w_t')^{-1} E(w_t \varepsilon_t) = \phi$. Also, in either case,

$$E(\varepsilon_{h,t} w_{h,t+L}) = (h - |L|) E(w_t w_t') \phi$$

if $|L| \leq h-1$ and 0 otherwise. Accordingly, for all L, we have, $E(\xi_{h,t} w_{h,t+L}) = 0$. Now

$$E(\xi_{h,t} x_{t-h} (x_{t-h} \otimes w_{h,t})') = E(E(\xi_{h,t} | w_{h,t}, x_{t-h}) x_{t-h} (x_{t-h} \otimes w_{h,t})') = 0$$

For $L < 0$

$$E(\xi_{h,t} x_{t-h} (x_{t-h+L} \otimes w_{h,t+L})') = E(A^L x_{t-h+L} (x_{t-h+L} \otimes \xi_{h,t} w_{h,t+L})') =$$
$$E(A^L x_{t-h+L} (x_{t-h+L} \otimes E(\xi_{h,t} w_{h,t+L} | x_{t-h+L}))') = 0$$

because $E(\xi_{h,t} w_{h,t+L} | x_{t-h+L}) = E(\xi_{h,t} w_{h,t+L}) = 0$. Similarly, for $L > 0$

$$E(\xi_{h,t} x_{t-h} (x_{t-h+L} \otimes w_{h,t+L})') = E(x_{t-h} (A^L x_{t-h} \otimes \xi_{h,t} w_{h,t+L})') =$$
$$E(x_{t-h} (A^L x_{t-h} \otimes E(\xi_{h,t} w_{h,t+L} | x_{t-h}))') = 0$$

Combining these, we have that $\omega_{12} = 0$.

Appendix 3: Monte Carlo Evaluation of Inference Procedures

We rely mainly on bootstrap p-values to assess statistical significance of our results. To investigate the properties of our bootstrap procedure, and of the alternative based on conventional asymptotics, we run a Monte-Carlo experiment. Let $i_t(n)$ denote the continuously compounded interest rate on an n-month zero-coupon bond. In each simulation, we generated artificial data on interest rates from the dynamic Nelson-Siegel model of Diebold and Li (2006):

$$i_t(n) = \beta_{0t} + \beta_{1t}(\tfrac{1-\exp(-\lambda n)}{\lambda n}) + \beta_{2t}(\tfrac{1-\exp(-\lambda n)}{\lambda n} - \exp(-\lambda n)) + u_{nt}$$

where

$$\beta_{jt} = \tilde{\mu} + \rho\beta_{jt-1} + v_{jt} \text{ for } j = 0, 1, 2$$

and the errors $\{u_{nt}\}$ and $\{v_{jt}\}$ are all iid normally distributed with mean zero and variances $\{\sigma^2_{n,u}\}$ and $\{\sigma^2_{j,v}\}$, respectively. Diebold and Li found that this simple model provided good out-of-sample interest rate forecasts; for our purposes it suffices that it gives a reasonable approximation to the dynamics of interest rates. From this model, we can then compute the excess returns $\{y_t\}$ and predictors $\{x_{t-h}\}$ in each simulation. The parameter ρ controls the degree of persistence of the variables. We set this to 0.8, 0.9, 0.97 and 0.99. Following Diebold and Li, we set λ to 0.0609 (in months). For each value of ρ, we picked $\tilde{\mu}$ and the $\sigma^2_{j,v}$s so that the mean and variance of β_{jt} were equal to their estimated values in fitting the model to the Fama-Bliss data from 1985:01-2006:12. We also set the $\sigma^2_{n,u}$s to their estimated values in this sample. Finally, to draw augmenting variables to mimic key features of the economic news index, we draw news index values from a normal distribution with mean zero and standard deviation 8 (this is roughly the standard deviation of each of these economic news indexes) and then cumulated from time $t-12$ to time t to form the augmenting variable, making this a Gaussian MA(12) process. The sample size is 240, which corresponds to 20 years of monthly data (roughly the sample size for our applications involving the economic news index).

The asymptotic and bootstrap p-values for the Wald tests of $\phi = 0$ and $\Gamma = 0$ are then computed. We report the simulated percentage of samples for which these p-values are less than 5 percent. Obviously, a well-calibrated test will give numbers near 5 percent; numbers greater than 5 percent indicate that the true size is greater than the nominal size.

The choices of ρ we consider include values near the estimates in our sample, about 0.8, and higher values to assess how the size distortions change as one gets close to the unit circle.

As can be seen from Table A1, the effective size of the asymptotic test of the hypothesis that $\phi = 0$ is elevated, at around 25 percent, while the asymptotic test of the hypothesis that $\Gamma = 0$ shows massive size distortions, with an actual rejection rate of over 60 percent. In contrast, the bootstrap test is roughly correctly sized with the effective size of the test being between 2 and 5 percent. The third block in the table, labelled RRMSPE=1, gives the percentage of times that the bootstrap p-value for testing the hypothesis that RRMSPE=1 in the out-of-sample forecasting experiment is less than 5 percent, using the unrestricted regression. We see that this test has a size that is very close to the nominal level. Note that the Diebold-Mariano test does not in this case satisfy the regularity conditions for the asymptotic distributions discussed in West (2006) to apply.

References

Altonji, J.G. and L.M. Segal (1996): Small-Sample Bias in GMM Estimation of Covariance Structures, *Journal of Business and Economic Statistics*, 14, pp.353-366.

Andersen, T., T. Bollerslev, F.X. Diebold and C. Vega (2007): Real-Time Price Discovery in Stock, Bond and Foreign Exchange Markets, *Journal of International Economics*, 73, pp.251-277.

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

Ang, A., G. Bekaert and M. Wei (2007): Do Macro Variables, Asset Markets, or Surveys Forecast Inflation Better?, *Journal of Monetary Economics*, 54, pp.1163-1212.

Balduzzi, P. Elton, E.J. and T.C. Green (2001): Economic News and Bond Prices: Evidence from the U.S. Treasury Market, *Journal of Financial and Quantitative Analysis*, 36, pp.523-543.

Basawa, I.V., A.K. Mallik, W.P. McCormick, J.H. Reeves and R.L. Taylor (1991). Bootstrapping Unstable First-order Autoregressive Processes, *Annals of Statistics*, 19, pp.1098-1101.

Bekker, P.A. (1994): Alternative Approximations to the Distributions of Instrumental Variables Estimators, *Econometrica*, 62, pp.657-681.

Bessembinder, P.A. (1992): Systematic risk, Hedging Pressure and Risk Premiums in Futures Markets, *Review of Financial Studies*, 5, pp.637-667.

Campbell, J.Y. and R.J. Shiller (1988): The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors, *Review of Financial Studies*, 1, pp.195-228.

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

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

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

Diebold, F.X. and C. Li (2006) : Forecasting the Term Structure of Government Bond Yields, *Journal of Econometrics*, 130, pp.337-364.

Diebold, F.X. and R.S. Mariano (1995): Comparing Predictive Accuracy, *Journal of Business and Economic Statistics*, 13, pp.253-265.

Eichenbaum, M.S., L.P. Hansen and K.J. Singleton (1988): A Time-Series Analysis of Representative Agent Models of Consumption and Leisure Choice under Uncertainty, *Quarterly Journal of Economics*, 103, pp.51-78.

Elliott, G. and J.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, E. and R.R. Bliss (1987): The Information in Long-Maturity Forward Rates, *American Economic Review*, 77, pp.680-692.

Fama, E. and K. French (1988): Dividend Yields and Expected Stock Returns, *Journal of Financial Economics*, 22, pp.3-26.

Froot, K.A. (1989): New Hope for the Expectations Hypothesis of the Term Structure of Interest Rates, *Journal of Finance*, 44, pp.283-305.

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

Goldberger, A. (1970): Estimation of a Regression Coefficient Matrix Containing a Block of Zeros, Working Paper, University of Wisconsin.

Hansen, B.E. (1995): Rethinking the Univariate Approach to Unit Root Testing: Using Covariates to Increase Power, *Econometric Theory*, 11, pp.1148-1171.

Hansen, B.E. (1999): The Grid Bootstrap and the Autoregressive Model, *Review of Economics and Statistics*, 81, pp.594-607.

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

Kilian, L. (1998): Small-Sample Confidence Intervals for Impulse Response Functions, *Review of Economics and Statistics*, 80, pp.218-230.

Kilian, L. (2006): Exogenous Oil Price Shocks: How Big are They and How Much do They Matter for the U.S. Economy, Working Paper, University of Michigan.

Koenig, E., S. Dolmas, and J. Piger (2003): The Use and Abuse of 'Real-Time' Data in Economic Forecasting, *Review of Economics and Statistics*, August, 85(3), pp. 618–28.

Mehra, Y.P. (2002): Survey Measures of Expected Inflation: Revisiting the Issues of Predictive Content and Rationality, *Federal Reserve Bank of Richmond Economic Quarterly*, 88, pp.17-36.

Phillips, P.C.B. (1979): The Sampling Distribution of Forecasts from a First Order Autoregression, *Journal of Econometrics*, 9, pp.241-261.

Romer, C.D. and D.H. Romer (2000): Federal Reserve Information and the Behavior of Interest Rates, *American Economic Review*, 90, pp.429-457.

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

Thomas, L.B. (1999): Survey Measures of Expected U.S. Inflation, *Journal of Economic Perspectives*, 13, pp.125-144.

West K.D., K-F. Wong and S. Anatolyev (2009): Instrumental Variables Estimation of Heteroskedastic Linear Models Using All Lags of Instruments, *Econometric Reviews*, forthcoming.

West K.D. (2006): "Forecast Evaluation", in Handbook of Economic Forecasting, Volume 1, G. Elliott, C.W.J. Granger and A. Timmermann (eds.), Elsevier.

Table 1: Simulated mean square errors of $\tilde{\beta}_h$ and β_h^* relative to $\hat{\beta}_h$

Panel A: Case h=1

λ	δ	k=1	k=5	k=10	k=1	k=5	k=10	k=1	k=5	k=10
			T=100			T=500			T=1000	
		MSE of Augmented estimator ($\tilde{\beta}_h$) relative to OLS								
0	0	1.00	1.02	1.06	1.00	1.01	1.01	1.00	1.00	1.01
0.01	0	0.99	1.02	1.05	0.99	1.00	1.01	1.00	0.99	1.00
0.1	0	0.95	0.97	1.00	0.94	0.96	0.96	0.96	0.95	0.96
0.3	0	0.84	0.86	0.88	0.83	0.85	0.85	0.85	0.84	0.85
0.5	0	0.71	0.74	0.75	0.70	0.73	0.72	0.71	0.72	0.72
0	0.5	1.00	1.01	1.03	1.00	1.01	1.01	1.00	1.00	1.01
0.01	0.5	0.99	1.01	1.02	0.99	1.00	1.00	1.00	0.99	1.00
0.1	0.5	0.94	0.95	0.96	0.95	0.96	0.96	0.95	0.94	0.96
0.3	0.5	0.82	0.83	0.83	0.83	0.85	0.85	0.84	0.83	0.85
0.5	0.5	0.67	0.70	0.69	0.70	0.72	0.72	0.71	0.70	0.72
0	0.9	1.00	1.00	1.02	1.00	1.01	1.00	1.00	1.00	1.01
0.01	0.9	0.99	0.99	1.01	1.00	1.00	1.00	0.99	0.99	1.00
0.1	0.9	0.93	0.94	0.95	0.95	0.95	0.95	0.93	0.94	0.95
0.3	0.9	0.79	0.81	0.80	0.83	0.82	0.83	0.80	0.81	0.83
0.5	0.9	0.64	0.66	0.65	0.69	0.68	0.70	0.67	0.68	0.69
		MSE of EGMM estimator (β_h^*) relative to OLS								
0	0	1.01	1.03	1.08	1.01	1.03	1.05	1.00	1.01	1.03
0.01	0	1.01	1.02	1.07	1.00	1.02	1.05	1.00	1.01	1.02
0.1	0	0.97	0.98	1.03	0.95	0.98	1.00	0.96	0.96	0.98
0.3	0	0.86	0.89	0.95	0.84	0.87	0.89	0.85	0.85	0.87
0.5	0	0.74	0.79	0.85	0.71	0.74	0.76	0.71	0.73	0.74
0	0.5	1.02	1.06	1.08	1.00	1.02	1.05	1.01	1.02	1.03
0.01	0.5	1.01	1.05	1.08	1.00	1.02	1.04	1.00	1.01	1.03
0.1	0.5	0.96	1.00	1.04	0.95	0.98	1.00	0.96	0.96	0.98
0.3	0.5	0.85	0.90	0.94	0.84	0.87	0.89	0.84	0.85	0.86
0.5	0.5	0.71	0.78	0.85	0.70	0.74	0.76	0.71	0.71	0.73
0	0.9	1.03	1.07	1.12	1.00	1.04	1.07	1.01	1.03	1.06
0.01	0.9	1.02	1.07	1.12	1.00	1.04	1.07	1.00	1.03	1.06
0.1	0.9	0.96	1.01	1.08	0.96	0.98	1.02	0.93	0.97	1.00
0.3	0.9	0.83	0.90	0.97	0.84	0.85	0.90	0.81	0.84	0.87
0.5	0.9	0.69	0.77	0.86	0.70	0.71	0.76	0.68	0.69	0.72

Table 1 (continued): Simulated mean square errors of $\tilde{\beta}_h$ and β_h^* relative to $\hat{\beta}_h$

Panel B: Case h=4

λ	δ	k=1	k=5	k=10	k=1	k=5	k=10	k=1	k=5	k=10
			T=100			T=500			T=1000	
		MSE of Augmented estimator ($\tilde{\beta}_h$) relative to OLS								
0	0	1.01	1.02	1.09	1.00	1.02	1.04	1.00	1.00	1.01
0.01	0	1.00	1.01	1.09	0.99	1.01	1.03	1.00	1.00	1.01
0.1	0	0.95	0.97	1.04	0.94	0.97	0.98	0.95	0.95	0.96
0.3	0	0.84	0.87	0.92	0.83	0.86	0.87	0.83	0.84	0.85
0.5	0	0.72	0.74	0.78	0.71	0.73	0.74	0.70	0.72	0.72
0	0.5	1.01	1.05	1.14	1.00	1.02	1.06	1.00	1.01	1.02
0.01	0.5	1.00	1.05	1.13	0.99	1.02	1.05	1.00	1.00	1.02
0.1	0.5	0.96	0.99	1.08	0.94	0.97	1.01	0.94	0.95	0.97
0.3	0.5	0.84	0.87	0.93	0.82	0.86	0.89	0.83	0.83	0.85
0.5	0.5	0.71	0.72	0.76	0.69	0.72	0.75	0.70	0.70	0.71
0	0.9	1.03	1.11	1.23	1.00	1.03	1.09	1.01	1.02	1.05
0.01	0.9	1.02	1.11	1.23	1.00	1.03	1.08	1.00	1.01	1.04
0.1	0.9	0.97	1.05	1.16	0.95	0.97	1.03	0.94	0.95	0.98
0.3	0.9	0.84	0.92	0.99	0.83	0.84	0.89	0.81	0.82	0.84
0.5	0.9	0.69	0.75	0.80	0.68	0.69	0.74	0.68	0.68	0.70
		MSE of EGMM estimator (β_h^*) relative to OLS								
0	0	1.03	1.01	1.00	1.01	1.07	1.09	1.01	1.04	1.05
0.01	0	1.03	1.01	1.00	1.01	1.06	1.08	1.00	1.03	1.05
0.1	0	0.99	0.99	0.99	0.95	1.02	1.04	0.95	0.99	1.01
0.3	0	0.91	0.95	0.98	0.84	0.91	0.94	0.84	0.88	0.90
0.5	0	0.81	0.90	0.96	0.72	0.79	0.83	0.71	0.75	0.77
0	0.5	1.03	1.05	1.01	1.01	1.06	1.11	1.01	1.05	1.08
0.01	0.5	1.03	1.05	1.01	1.00	1.05	1.10	1.01	1.04	1.08
0.1	0.5	0.99	1.02	1.00	0.95	1.01	1.06	0.95	0.99	1.03
0.3	0.5	0.91	0.97	0.98	0.83	0.90	0.96	0.84	0.87	0.91
0.5	0.5	0.81	0.91	0.96	0.70	0.78	0.84	0.71	0.74	0.77
0	0.9	1.05	1.06	1.03	1.02	1.09	1.15	1.02	1.08	1.13
0.01	0.9	1.05	1.07	1.03	1.02	1.08	1.15	1.02	1.07	1.12
0.1	0.9	1.01	1.05	1.02	0.97	1.03	1.10	0.96	1.01	1.06
0.3	0.9	0.92	1.00	0.99	0.84	0.90	0.98	0.83	0.86	0.92
0.5	0.9	0.81	0.94	0.97	0.70	0.76	0.84	0.69	0.72	0.77

Notes. Mean square errors of $\tilde{\beta}_h$ and β_h^* divided by those of $\hat{\beta}_h$, obtained by Monte-Carlo simulations, as described in the text. In each experiment, 1000 replications were conducted.

Table 2-1: Regressions predicting excess bond returns using forward rates and additional regressors A1

	$n = 24$		$n = 36$		$n = 48$		$n = 60$		Average	
Regression:	Baseline	Aug.	Baseline	Aug.	Baseline	Aug.	Baseline	Aug.	Baseline	Aug.
y_{12}	-0.75	-1.34	-1.41	-2.44	-2.13	-3.47	-2.73	-4.37	-1.76	-2.91
	(0.22)	(0.20)	(0.38)	(0.33)	(0.50)	(0.44)	(0.61)	(0.52)	(0.43)	(0.37)
f_{24}	0.12	1.04	-0.25	1.35	0.05	2.10	0.25	2.75	0.05	1.81
	(0.45)	(0.36)	(0.77)	(0.64)	(1.01)	(0.85)	(1.26)	(1.07)	(0.86)	(0.72)
f_{36}	1.04	1.18	2.63	2.85	2.74	3.01	2.99	3.31	2.35	2.59
	(0.38)	(0.32)	(0.64)	(0.58)	(0.86)	(0.80)	(1.10)	(1.02)	(0.74)	(0.68)
f_{48}	0.80	0.08	1.48	0.24	2.83	1.25	3.43	1.51	2.14	0.77
	(0.29)	(0.25)	(0.52)	(0.48)	(0.67)	(0.65)	(0.84)	(0.80)	(0.58)	(0.54)
f_{60}	-0.94	-0.88	-2.02	-1.89	-2.92	-2.75	-3.23	-3.00	-2.28	-2.13
	(0.37)	(0.28)	(0.63)	(0.48)	(0.84)	(0.66)	(1.07)	(0.84)	(0.72)	(0.56)
Wald ϕ		64.33		47.05		43.33		43.37		46.79
p-val (boot)		0.000		0.000		0.000		0.000		0.000
p-val (asy)		0.000		0.000		0.000		0.000		0.000
Wald Γ		120.88								
p-val (boot)		0.126								
p-val (asy)		0.000								

Notes. The baseline regressions show the estimated coefficients in regressions of excess 24-, 36-, 48- and 60-month bond returns on the term structure of forward rates and of the average of these four excess returns on the term structure of forward rates. The regressions are run on data from the first month of each quarter from 1968Q4 to 2006Q4. Asymptotic standard errors are shown in parentheses. All asymptotic standard errors and p-values are Newey-West with a lag length of 4. The augmented regressions control for the additional variable A1. The estimated coefficients on this additional variable are not reported, but the row Wald ϕ denotes the Wald statistic testing the hypothesis that it is equal to zero along with the associated asymptotic and bootstrap p-values, constructed as described in the text. The row Wald Γ denotes the Wald statistic testing the hypothesis that $\Gamma = 0$, again along with the associated asymptotic and bootstrap p-values. These are the same for each regression in the table because they depend only on $\{x_t\}$ and $\{w_t\}$.

Table 2-2: Regressions predicting excess bond returns using forward rates and additional regressors A2

	$n=24$		$n=36$		$n=48$		$n=60$		Average	
Regression:	Baseline	Aug.	Baseline	Aug.	Baseline	Aug.	Baseline	Aug.	Baseline	Aug.
y_{12}	-0.09	-0.56	-0.36	-1.26	-0.83	-2.02	1.07	-2.52	-0.59	-1.59
	(0.39)	(0.14)	(0.73)	(0.27)	(0.97)	(0.40)	(1.22)	(0.51)	(0.82)	(0.31)
f_{24}	-0.29	0.67	-0.81	1.04	-0.60	1.87	-0.57	2.48	-0.57	1.52
	(0.60)	(0.26)	(1.10)	(0.50)	(1.45)	(0.75)	(1.84)	(1.06)	(1.23)	(0.62)
f_{36}	0.42	0.36	1.56	1.43	1.38	1.16	1.19	0.88	1.14	0.96
	(0.40)	(0.15)	(0.79)	(0.33)	(1.13)	(0.56)	(1.39)	(0.78)	(0.92)	(0.44)
f_{48}	0.81	-0.33	1.50	-0.69	2.81	-0.09	3.45	-0.11	2.14	-0.30
	(0.23)	(0.12)	(0.44)	(0.22)	(0.60)	(0.35)	(0.75)	(0.45)	(0.50)	(0.27)
f_{60}	-0.49	-0.01	-1.31	-0.38	-2.03	-0.76	-2.08	-0.51	-1.48	-0.41
	(0.27)	(0.09)	(0.57)	(0.20)	(0.80)	(0.34)	(1.01)	(0.49)	(0.66)	(0.27)
Wald ϕ		419.60		520.30		383.26		321.19		415.27
p-val (boot)		0.000		0.000		0.000		0.000		0.000
p-val (asy)		0.000		0.000		0.000		0.000		0.000
Wald Γ		121.16								
p-val (boot)		0.329								
p-val (asy)		0.000								

Notes. As in Table 2-1, except using the additional variable A2 and the sample period 1981Q3 to 2006Q4.

Table 2-3: Regressions predicting excess bond returns using forward rates and additional regressors A3

	$n=24$		$n=36$		$n=48$		$n=60$		Average	
Regression:	Baseline	Aug.	Baseline	Aug.	Baseline	Aug.	Baseline	Aug.	Baseline	Aug.
y_{12}	-0.28	-0.60	-0.84	-1.45	-1.60	-2.43	-2.43	-3.41	-1.29	-1.97
	(0.44)	(0.33)	(0.83)	(0.61)	(1.09)	(0.80)	(1.29)	(0.96)	(0.91)	(0.67)
f_{24}	-0.55	-0.15	-0.99	-0.22	-0.63	0.40	0.19	1.41	-0.50	0.36
	(0.77)	(0.54)	(1.37)	(0.95)	(1.81)	(1.24)	(2.17)	(1.50)	(1.53)	(1.05)
f_{36}	1.32	1.46	3.12	3.38	3.56	3.91	3.74	4.15	2.94	3.23
	(0.93)	(0.62)	(1.01)	(1.11)	(2.31)	(1.51)	(2.76)	(1.84)	(1.92)	(1.26)
f_{48}	0.71	0.53	0.87	0.66	1.90	1.43	1.74	1.19	1.34	0.95
	(0.45)	(0.30)	(0.59)	(0.62)	(1.22)	(0.92)	(1.54)	(1.25)	(1.00)	(0.76)
f_{60}	-0.92	-0.88	-1.80	-1.72	-2.50	-2.39	-2.35	-2.22	-1.89	-1.80
	(0.44)	(0.33)	(0.82)	(0.63)	(1.13)	(0.89)	(1.39)	(1.13)	(0.94)	(0.74)
Wald ϕ		31.52		29.27		26.45		20.89		26.00
p-val (boot)		0.000		0.002		0.004		0.007		0.004
p-val (asy)		0.000		0.000		0.000		0.000		0.000
Wald Γ		13.41								
p-val (boot)		0.556								
p-val (asy)		0.037								

Notes. The baseline regressions show the estimated coefficients in regressions of excess 24-, 36-, 48- and 60-month bond returns on the term structure of forward rates and of the average of these four excess returns on the term structure of forward rates. The regressions are run on monthly data from 1985:02 through 2006:12. Asymptotic standard errors are shown in parentheses. All asymptotic standard errors and p-values are Newey-West with a lag length of 18. The augmented regressions control for the additional variable A3. The estimated coefficients on this additional variable are not reported, but the row Wald ϕ denotes the Wald statistic testing the hypothesis that it is equal to zero along with the associated asymptotic and bootstrap p-values, constructed as described in the text. The row Wald Γ denotes the Wald statistic testing the hypothesis that $\Gamma = 0$, again along with the associated asymptotic and bootstrap p-values. These are the same for each regression in the table because they depend only on $\{x_t\}$ and $\{w_t\}$.

Table 2-4: Regressions predicting excess bond returns using forward rates and additional regressors A4

	$n = 24$		$n = 36$		$n = 48$		$n = 60$		Average	
Regression:	Baseline	Aug.	Baseline	Aug.	Baseline	Aug.	Baseline	Aug.	Baseline	Aug.
y_{12}	-0.17	-0.39	-0.74	-1.17	-1.57	-2.13	-2.54	-3.19	-1.25	-1.72
	(0.48)	(0.39)	(0.88)	(0.71)	(1.14)	(0.91)	(1.34)	(1.06)	(0.96)	(0.76)
f_{24}	-1.18	-0.57	-1.99	-0.83	-1.87	-0.34	-1.03	0.74	-1.52	-0.25
	(0.82)	(0.62)	(1.43)	(1.07)	(1.85)	(1.38)	(2.24)	(1.67)	(1.58)	(1.18)
f_{36}	2.46	1.96	5.34	4.39	6.74	5.49	7.63	6.17	5.54	4.50
	(0.96)	(0.65)	(1.74)	(1.15)	(2.26)	(1.49)	(2.70)	(1.79)	(1.90)	(1.26)
f_{48}	0.01	0.24	-0.68	–0.25	-0.78	-0.21	-1.90	-1.24	-0.84	-0.37
	(0.54)	(0.41)	(1.01)	(0.74)	(1.32)	(0.98)	(1.58)	(1.22)	(1.10)	(0.82)
f_{60}	-0.55	-0.85	-0.95	-1.51	-1.21	-1.95	-0.65	-1.51	-0.84	-1.45
	(0.46)	(0.44)	(0.96)	(0.90)	(1.36)	(1.29)	(1.74)	(1.67)	(1.12)	(1.07)
Wald ϕ		15.77		13.37		11.08		8.81		11.23
p-val (boot)		0.027		0.040		0.048		0.079		0.047
p-val (asy)		0.000		0.000		0.001		0.003		0.001
Wald Γ		26.042								
p-val (boot)		0.474								
p-val (asy)		0.000								

Notes. As in Table 2-3, except using the additional variable A4 and the sample period 1989:09 to 2006:12.

Table 2-5: Regressions predicting excess bond returns using forward rates and additional regressors A5

	$n = 24$		$n = 36$		$n = 48$		$n = 60$		Average	
Regression:	Baseline	Aug.	Baseline	Aug.	Baseline	Aug.	Baseline	Aug.	Baseline	Aug.
y_{12}	-0.11	-0.20	-0.64	-0.82	-1.44	-1.68	-2.22	-2.51	-1.10	-1.30
	(0.63)	(0.37)	(1.17)	(0.67)	(1.50)	(0.84)	(1.77)	(0.98)	(1.26)	(0.71)
f_{24}	-1.38	-0.89	-2.27	-1.30	-2.20	-0.90	-1.57	0.00	-1.86	-0.77
	(0.92)	(0.60)	(1.62)	(1.02)	(2.06)	(1.32)	(2.46)	(1.59)	(1.76)	(1.12)
f_{36}	2.87	2.31	5.93	4.84	7.43	5.96	8.34	6.57	6.14	4.92
	(0.96)	(0.63)	(1.78)	(1.13)	(2.36)	(1.49)	(2.86)	(1.82)	(1.98)	(1.25)
f_{48}	-0.28	0.00	-1.13	-0.58	-1.33	-0.59	-2.56	-1.67	-1.32	-0.71
	(0.58)	(0.35)	(1.08)	(0.62)	(1.42)	(0.82)	(1.70)	(1.04)	(1.19)	(0.69)
f_{60}	-0.53	-0.89	-0.88	-1.59	-1.10	-2.06	-0.31	-1.46	-0.71	-1.50
	(0.58)	(0.44)	(1.25)	(0.93)	(1.76)	(1.34)	(2.24)	(1.75)	(1.45)	(1.10)
Wald ϕ		25.39		26.28		23.64		20.33		23.60
p-val (boot)		0.013		0.010		0.012		0.017		0.011
p-val (asy)		0.000		0.000		0.000		0.000		0.000
Wald Γ		16.756								
p-val (boot)		0.659								
p-val (asy)		0.010								

Notes. As in Table 2-3, except using the additional variable A5 and the sample period 1991:07 to 2006:12.

Table 2-6: Regressions predicting excess bond returns using forward rates and additional regressors A6

	$n = 24$		$n = 36$		$n = 48$		$n = 60$		Average	
Regression:	Baseline	Aug.	Baseline	Aug.	Baseline	Aug.	Baseline	Aug.	Baseline	Aug.
y_{12}	-0.11	-0.22	-0.64	-0.86	-1.44	-1.75	-2.22	-2.61	-1.10	-1.36
	(0.63)	(0.37)	(1.17)	(0.70)	(1.50)	(0.91)	(1.77)	(1.09)	(1.26)	(0.76)
f_{24}	-1.38	-0.88	-2.27	-1.28	-2.20	-0.84	-1.57	0.07	-1.86	-0.73
	(0.92)	(0.63)	(1.62)	(1.12)	(2.06)	(1.46)	(2.46)	(1.77)	(1.76)	(1.24)
f_{36}	2.87	2.38	5.93	4.99	7.43	6.17	8.34	6.83	6.14	5.09
	(0.96)	(0.61)	(1.78)	(1.05)	(2.36)	(1.37)	(2.86)	(1.65)	(1.98)	(1.15)
f_{48}	-0.28	0.26	-1.13	-0.15	-1.33	-0.13	-2.56	-1.20	-1.32	-0.30
	(0.58)	(0.36)	(1.08)	(0.58)	(1.42)	(0.72)	(1.70)	(0.84)	(1.19)	(0.60)
f_{60}	-0.53	-1.00	-0.88	-1.74	-1.10	-2.12	-0.31	-1.45	-0.71	-1.58
	(0.58)	(0.36)	(1.25)	(0.75)	(1.76)	(1.04)	(2.24)	(1.29)	(1.45)	(0.85)
Wald ϕ		20.30		23.57		23.08		22.50		23.12
p-val (boot)		0.208		0.163		0.162		0.160		0.163
p-val (asy)		0.000		0.000		0.000		0.000		0.000
Wald Γ		429.29								
p-val (boot)		0.174								
p-val (asy)		0.000								

Notes. As in Table 2-3, except using the additional variables A4 and the sample period 1991:7 to 2006:12. In this case, the row Wald ϕ gives the Wald statistic testing the joint hypothesis that all the elements of ϕ are equal to zero.

Table 3: Out-of-Sample RRMSPE for excess bond returns from augmented regression relative to baseline

Additional Regressors		Bond Maturity			
		$n = 24$	$n = 36$	$n = 48$	$n = 60$
A1	Unrestricted	0.913	0.910	0.914	0.902
		0.030	*0.026*	*0.027*	*0.017*
	Restricted	0.948	0.925	0.900	0.892
		0.076	*0.036*	*0.020*	*0.012*
A2	Unrestricted	0.907	0.871	0.848	0.777
		0.063	*0.040*	*0.028*	*0.008*
	Restricted	0.815	0.815	0.826	0.851
		0.017	*0.020*	*0.025*	*0.026*
A3	Unrestricted	0.910	0.925	0.940	0.957
		0.071	*0.099*	*0.130*	*0.177*
	Restricted	0.960	0.948	0.938	0.931
		0.181	*0.148*	*0.131*	*0.107*
A4	Unrestricted	0.893	0.891	0.897	0.904
		0.111	*0.108*	*0.115*	*0.119*
	Restricted	0.908	0.903	0.894	0.895
		0.144	*0.127*	*0.107*	*0.104*
A5	Unrestricted	0.685	0.690	0.707	0.723
		0.018	*0.017*	*0.019*	*0.025*
	Restricted	0.730	0.719	0.699	0.698
		0.027	*0.021*	*0.020*	*0.015*
A6	Unrestricted	0.911	0.931	0.967	0.991
		0.238	*0.267*	*0.330*	*0.385*
	Restricted	0.973	0.965	0.958	0.956
		0.353	*0.322*	*0.311*	*0.300*

Notes. This table shows the root mean square prediction error for excess bond returns from the augmented model in which the augmenting regressors are predicted to be zero divided by the root mean square prediction error from the baseline model. Bootstrap p-values for one-sided tests testing the hypothesis of equality in root mean square prediction errors are shown in italics. The models are as described in the notes to Table 2. Models are either restricted or unrestricted: Restricted predictions impose that there is a single return forecasting factor with a different loading for each bond maturity n. Predictions are pseudo-out-of-sample with the first prediction made half-way through the estimation period.

Table 4-1: Regressions predicting excess stock returns using log dividend-price ratio and/or stochastically detrended short-term interest rate and additional regressors A1

Regression:	Baseline	Aug.	Baseline	Aug.	Baseline	Aug.
dp_t	0.013	0.020			0.013	0.020
	(0.015)	(0.014)			(0.015)	(0.014)
$\tilde{r}_t^{RF}$			-1.336	-0.894	-1.343	-0.897
			(0.668)	(0.615)	(0.646)	(0.596)
Wald ϕ		23.77		16.68		18.66
p-val (boot)		0.000		0.007		0.003
p-val (asy)		0.000		0.001		0.000
Wald Γ		12.78		24.49		34.85
p-val (boot)		0.573		0.000		0.139
p-val (asy)		0.005		0.000		0.000

Notes. The baseline regressions show the estimated coefficients in regressions of 3 month excess stock returns (CRSP value-weighted returns less the Fama-Bliss riskfree rate) on the corresponding log dividend-price ratio and/or the stochastically detrended short term interest rate. The regressions are run on data from the first month of each quarter from 1968Q4 through 2006Q4. Asymptotic standard errors are shown in parentheses. The augmented regressions control for the additional variable A1. The estimated coefficients on this additional variable are not reported, but the row Wald ϕ denotes the Wald statistic testing the hypothesis that it is equal to zero along with the associated asymptotic and bootstrap p-values, constructed as described in the text. The row Wald Γ denotes the Wald statistic testing the hypothesis that $\Gamma = 0$, again along with the associated asymptotic and bootstrap p-values. These are the same for each regression in the table because they depend only on $\{x_t\}$ and $\{w_t\}$.

Table 4-2: Regressions predicting excess stock returns using log dividend-price ratio and/or stochastically detrended short-term interest rate and additional regressors A2

Regression:	Baseline	Aug.	Baseline	Aug.	Baseline	Aug.
dp_t	0.018	0.031			0.016	0.031
	(0.018)	(0.016)			(0.019)	(0.016)
$\tilde{r}_t^{RF}$			-0.749	-0.357	-0.550	-0.103
			(1.037)	(0.920)	(1.036)	(0.903)
Wald ϕ		38.104		39.676		39.352
p-val (boot)		0.000		0.000		0.000
p-val (asy)		0.000		0.000		0.000
Wald Γ		1.863		7.323		8.063
p-val (boot)		0.739		0.196		0.553
p-val (asy)		0.601		0.062		0.234

Notes. As in Table 2-1, except using the additional variable A2 and the sample period 1981Q3 to 2006Q4.Notes.

Table 4-3: Regressions predicting excess stock returns using log dividend-price ratio and/or stochastically detrended short-term interest rate and additional regressors A3

Regression:	Baseline	Aug.	Baseline	Aug.	Baseline	Aug.
dp_t	0.034	0.036			0.034	0.037
	(0.018)	(0.019)			(0.019)	(0.019)
$\tilde{r}_t^{RF}$			0.221	0.270	0.386	0.476
			(1.005)	(0.987)	(0.983)	(0.956)
Wald ϕ		0.608		0.270		0.681
p-val (boot)		0.494		0.649		0.476
p-val (asy)		0.436		0.603		0.409
Wald Γ		0.125		1.237		1.540
p-val (boot)		0.771		0.385		0.642
p-val (asy)		0.724		0.266		0.463

Notes. The baseline regressions show the estimated coefficients in regressions of 3 month excess stock returns (CRSP value-weighted returns less the Fama-Bliss riskfree rate) on the corresponding log dividend-price ratio and/or the stochastically detrended short term interest rate. The regressions are run on monthly data from 1985:02 through 2006:12. Asymptotic standard errors are shown in parentheses. The augmented regressions control for the additional variable A3. The estimated coefficients on this additional variable are not reported, but the row Wald ϕ denotes the Wald statistic testing the hypothesis that it is equal to zero along with the associated asymptotic and bootstrap p-values, constructed as described in the text. The row Wald Γ denotes the Wald statistic testing the hypothesis that $\Gamma = 0$, again along with the associated asymptotic and bootstrap p-values. These are the same for each regression in the table because they depend only on $\{x_t\}$ and $\{w_t\}$.

Table 4-4: Regressions predicting excess stock returns using log dividend-price ratio and/or stochastically detrended short-term interest rate and additional regressors A4

Regression:	Baseline	Aug.	Baseline	Aug.	Baseline	Aug.
dp_t	0.035	0.033			0.038	0.036
	(0.025)	(0.025)			(0.025)	(0.025)
$\tilde{r}_t^{RF}$			0.752	0.845	1.006	1.077
			(1.126)	(1.119)	(1.072)	(1.069)
Wald ϕ		1.943		2.783		2.132
p-val (boot)		0.216		0.159		0.208
p-val (asy)		0.163		0.095		0.144
Wald Γ		4.575		0.986		4.765
p-val (boot)		0.250		0.512		0.427
p-val (asy)		0.032		0.321		0.092

Notes. As in Table 4-3, except using the additional variable A4 and the sample period 1989:09 to 2006:12.

Table 4-5: Regressions predicting excess stock returns using log dividend-price ratio and/or stochastically detrended short-term interest rate and additional regressors A5

Regression:	Baseline	Aug.	Baseline	Aug.	Baseline	Aug.
dp_t	0.066	0.065			0.065	0.065
	(0.025)	(0.025)			(0.024)	(0.025)
$\tilde{r}_t^{RF}$			1.060	1.076	0.981	0.990
			(1.206)	(1.206)	(1.113)	(1.111)
Wald ϕ		0.043		0.222		0.063
p-val (boot)		0.868		0.712		0.836
p-val (asy)		0.835		0.637		0.802
Wald Γ		2.001		0.193		2.067
p-val (boot)		0.384		0.728		0.608
p-val (asy)		0.157		0.660		0.356

Notes. As in Table 4-3, except using the additional variable A5 and the sample period 1991:07 to 2006:12.

Table 4-6: Regressions predicting excess stock returns using log dividend-price ratio and/or stochastically detrended short-term interest rate and additional regressors A6

Regression:	Baseline	Aug.	Baseline	Aug.	Baseline	Aug.
dp_t	0.066	0.068			0.065	0.068
	(0.025)	(0.024)			(0.024)	(0.023)
$\tilde{r}_t^{RF}$			1.060	1.356	0.981	1.276
			(1.206)	(1.176)	(1.113)	(1.057)
Wald ϕ		6.753		7.128		7.460
p-val (boot)		0.213		0.201		0.196
p-val (asy)		0.080		0.068		0.059
Wald Γ		0.348		2.293		2.905
p-val (boot)		0.968		0.698		0.944
p-val (asy)		0.951		0.514		0.821

Notes. As in Table 4-3, except using the additional variables A6 and the sample period 1991:7 to 2006:12. In this case, the row Wald ϕ gives the Wald statistic testing the joint hypothesis that all the elements of ϕ are equal to zero.

Table 5: Out-of-Sample RMSPE for excess stock returns from augmented regression relative to baseline

Additional Regressors	Predictors		
	dp_t	$\tilde{r}_t^{RF}$	dp_t & $\tilde{r}_t^{RF}$
A1	0.974	0.978	0.975
	0.030	*0.015*	*0.036*
A2	0.960	0.947	0.917
	0.018	*0.000*	*0.003*
A3	0.999	1.000	0.998
	0.332	*0.402*	*0.266*
A4	1.028	0.995	1.033
	0.929	*0.203*	*0.939*
A5	1.014	0.999	1.014
	0.842	*0.373*	*0.815*
A6	1.013	0.996	1.005
	0.675	*0.284*	*0.459*

Notes. This table shows the root mean square prediction error for excess stock returns from the augmented model in which the augmenting regressors are predicted to be zero divided by the root mean square prediction error from the baseline model. Bootstrap p-values for one-sided tests testing the hypothesis of equality in root mean square prediction errors are shown in italics. The models are as described in the notes to Table 4.

Table 6: Regressions predicting orange juice futures returns over the course of the winter with freezing degree days as additional regressors

Regression:	Baseline	Aug.	Baseline	Aug.
Slope_t	-0.044	0.054		
	(0.577)	(0.501)		
Net Long_t			1.002	1.378
			(0.518)	(0.508)

Notes. The baseline regressions show the estimated coefficients in regressions of returns on holding March frozen orange juice concentrate futures from the last day of the previous November to the last day of Febuary on either (i) the slope of the futures curve or (ii) the net long speculative positions, at the end of November. The augmented regression control for the number of freezing degree days in December, January and Feburary at Orlando airport. Heteroskedasticity-robust standard errors are shown in parentheses. There is one observation per year. The sample period is 1967/68-2007/08 for the regression on the slope of the futures curve and 1983/84-2007/08 for the regression on net long speculative positions, because of the absence of earlier positions data. This makes for a total of 41 and 25 observations in the two regressions, respectively. Futures data are from Norman's historical data and are closing quotes on the New York Board of Trade that was subsequently acquired by the Intercontinental Exchange . Positions data are from the Committment of Traders survey run by the Commodity and Futures Trading Commission. Freezing degree days were calculated from daily minimum temperatures at Orlando International Airport obtained from the National Climatic Data Center.

Table A1: Simulated Effective Size of Nominal 5 Percent Tests of $\phi = 0$, $\Gamma = 0$, and RRMSPE= 1

	$\phi = 0$					$\Gamma = 0$	RRMSPE=1			
	n=24	n=36	n=48	n=60	Average		n=24	n=36	n=48	n=60
$\rho = 0.8$										
Asymptotic	16.0	17.0	17.0	16.6	16.0	42.4				
Bootstrap	5.8	6.2	7.2	7.2	6.8	2.2	3.2	3.0	2.8	2.2
$\rho = 0.9$										
Asymptotic	16.0	17.2	17.6	16.8	16.0	47.4				
Bootstrap	3.0	2.0	2.6	3.0	2.4	2.2	2.6	2.8	2.6	2.6
$\rho = 0.97$										
Asymptotic	15.8	15.4	15.2	15.2	15.8	55.0				
Bootstrap	3.6	3.4	2.6	3.0	3.0	2.6	4.4	4.6	4.4	3.8
$\rho = 0.99$										
Asymptotic	20.2	19.6	18.8	17.8	20.2	50.8				
Bootstrap	5.6	5.0	5.0	4.8	5.0	4.2	5.0	4.8	4.8	4.2

Notes. In these simulations, samples of size 240 for $\{y_t\}$, $\{x_{t-h}\}$ and $\{w_t\}$ were generated from the design specified in Appendix 3. The Wald test statistics testing the hypothesis $\phi = 0$ (for the different excess return series) and $\Gamma = 0$ were then computed and compared to the asymptotic and bootstrap critical values. The table reports the fraction of simulations in which the test rejected using asymptotic and bootstrap critical values (nominal significance level: 5 percent). The table also reports the percentage of times that the bootstrap p-value for testing the hypothesis that RRMSPE=1 in the out-of-sample forecasting experiment is less than 5 percent, using the unrestricted regression. Note that the Diebold-Mariano test does not in this case satisfy the regularity conditions for the asymptotic distributions discussed in West (2006) to apply.

Fig. 1: 12-Month Moving Sum of Economic News Indexes Ending in Month Shown