



HAL
open science

How useful are no-arbitrage restrictions for forecasting the term structure of interest rates?

Andrea Carriero, Raffaella Giacomini

► To cite this version:

Andrea Carriero, Raffaella Giacomini. How useful are no-arbitrage restrictions for forecasting the term structure of interest rates?. *Econometrics*, 2011, 164 (1), pp.21. <10.1016/j.jeconom.2011.02.010>. <hal-00844809>

HAL Id: hal-00844809

<https://hal.science/hal-00844809v1>

Submitted on 16 Jul 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Accepted Manuscript

How useful are no-arbitrage restrictions for forecasting the term structure of interest rates?

Andrea Carriero, Raffaella Giacomini

PII: S0304-4076(11)00037-6

DOI: 10.1016/j.jeconom.2011.02.010

Reference: ECONOM 3456

To appear in: *Journal of Econometrics*



Please cite this article as: Carriero, A., Giacomini, R., How useful are no-arbitrage restrictions for forecasting the term structure of interest rates?. *Journal of Econometrics* (2011), doi:10.1016/j.jeconom.2011.02.010

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

How useful are no-arbitrage restrictions for forecasting the term structure of interest rates?

Andrea Carriero

Queen Mary, University of London

Raffaella Giacomini*

University College London and CEMMAP

July 2010

Abstract

We develop a general framework for analyzing the usefulness of imposing parameter restrictions on a forecasting model. We propose a measure of the usefulness of the restrictions that depends on the forecaster's loss function and that could be time varying. We show how to conduct inference about this measure. The application of our methodology to analyzing the usefulness of no-arbitrage restrictions for forecasting the term structure of interest rates reveals that: 1) the restrictions have become less useful over time; 2) when using a statistical measure of accuracy, the restrictions are a useful way to reduce parameter estimation uncertainty, but are dominated by restrictions that do the same without using any theory; 3) when using an economic measure of accuracy, the no-arbitrage restrictions are no longer dominated by atheoretical restrictions, but for this to be true it is important that the restrictions incorporate a time-varying risk premium.

Keywords: Forecast Combination; Encompassing; Loss functions; Instability; Affine Term Structure Models

Acknowledgments: We would like to thank Caio Almeida, Gianni Amisano, Albert Chun and seminar participants at the 2008 Forecasting in Rio conference, 2009 London and Oxbridge Time Series workshop and Queen Mary University for useful comments and discussions.

J.E.L. Codes: C52, C53, E43, E47

*Corresponding author, Department of Economics, University College London - Gower Street - WC1E 6BT London.
Tel: +44 0207679 5898 Fax: +44 02079162775 Email: r.giacomini@ucl.ac.uk.

1 Introduction

In recent years the finance literature has produced major advances in modeling the term structure of interest rates, building on the assumption of absence of arbitrage opportunities in bond markets. While the no-arbitrage approach has produced good results in terms of in-sample fit, see e.g. De Jong (2000) and Dai and Singleton (2000), the papers focusing on out-of sample forecasting have documented a mixed performance of these models. Duffee (2002) shows that beating a random walk with a traditional no-arbitrage affine term structure model is difficult. Ang and Piazzesi (2003) show that imposing no-arbitrage restrictions and an essentially affine specification of market prices of risk improves out-of-sample forecasts from a VAR(12), but the gains with respect to a random walk forecast are small. Carriero (2007) shows that the no-arbitrage restrictions provide better results if they are imposed on the data as prior information rather than as a set of restrictions. More encouraging results have been obtained by Almeida and Vicente (2008), Moench (2008), and Favero et al. (2009).

A drawback of the above conclusions is that they are based on informal comparisons of mean squared forecast errors computed over a particular out-of-sample period. In this paper, we develop a formal framework for investigating the usefulness of parameter restrictions in general - and no-arbitrage restrictions in particular - when a model is used for forecasting. We achieve several goals: 1) we propose a measure of the usefulness of the restrictions that is tailored to the forecaster's decision problem; 2) the measure can be time-varying; 3) we show how to perform inference about the proposed measure. Our framework can be used to answer questions such as "are no-arbitrage restrictions useful for forecasting the term structure of interest rates?", "are the restrictions useful for bond portfolio allocation?", and "have the restrictions become more or less useful over time?", which are not readily answered using conventional model evaluation and hypothesis testing tools.

Our main idea is to cast the problem in an out-of-sample forecast combination framework, in which there is only one forecast model, but the forecaster has the option of imposing some restrictions on its parameters or to forecast with the unrestricted model. We consider a forecast combination and estimate the optimal combination weight in an out-of-sample framework. We say that the restriction is "useful for forecasting" when the optimal weight is large, and we can formally test the hypothesis that the restrictions are useless by an out-of-sample encompassing test. Optimality of the weight is defined with respect to a general forecast loss function, but we show how to specialize the results to either the commonly used quadratic loss or to a loss based on (minus) the utility of a bond portfolio constructed using the model. The latter example of an economically meaningful loss has not been considered before for evaluating no-arbitrage models, and we show how its use can lead to substantially different conclusions than those based on conventional statistical measures of accuracy.

We further generalize the techniques to an environment with possible instability and provide a

test to assess whether the usefulness of the restrictions is time-varying. To gain some intuition for why the usefulness of restrictions could be time-varying, consider the case of a quadratic loss, for which it can be shown that the measure of usefulness captures the bias/variance tradeoff between a possibly misspecified restricted model and the unrestricted model. In this case, time variation could be due to the variance of the unrestricted model changing or the restrictions becoming more or less misspecified over time. We should point out that our methods do not allow one to disentangle the two sources of time variation.

We stress that our techniques are not only applicable to the comparison between an unrestricted and a restricted forecast, but they can be more generally used for measuring the usefulness of two alternative sets of restrictions imposed on the same forecasting model. For example, the random walk model that is often used as a benchmark in forecasting can also be viewed as a set of restrictions on a VAR, and one could ask whether the no-arbitrage restrictions are useful relative to the random walk restrictions. Finally, our framework can be used to compare and combine forecasts from nested models, which is similar to the problem considered by Clark and McCracken (2009) in a different asymptotic context.

From the perspective of forecast combination, our problem is non-standard because we do not combine forecasts from different models, but forecasts from the same model that are based on different estimators. This in principle poses challenges for the econometric methodology in that the unrestricted and restricted forecasts may be perfectly correlated in large samples if the restrictions are true. We overcome this problem by considering an out-of-sample environment with non-vanishing estimation uncertainty, as that considered by Giacomini and White (2006) in the different context of equal predictive ability tests, and use it to derive out-of-sample encompassing tests.

Encompassing tests are appealing in our context because in case of rejection of the null hypothesis they provide as a by-product a combination weight that can be naturally interpreted as a measure of usefulness of the restrictions, whereas a test of equal predictive ability would force one to choose either the restricted or the unrestricted forecast. This combination weight can further be used to produce combined forecasts that exploit the information contained in the economic restrictions in a way that is optimal for the loss function of interest. Finally, the encompassing approach naturally lends itself to extensions to unstable environments, since the combination weight can be postulated to be time-varying. Our contribution to the literature in this respect is to provide a valid asymptotic theory for testing hypotheses about the time-varying weight.

Note that our problem is also different from testing the restrictions in-sample, since we allow for the possibility that the restrictions are not true, but are still useful for out-of-sample forecasting for a given loss function.

2 A measure of the usefulness of economic restrictions

2.1 Set-up and notation

Let $y_t = (x_t, z_t)'$ indicate the vector of observables, which include the (scalar) variable of interest x_t and the vector of predictors z_t . We assume that the user has obtained two sequences of h -step ahead out-of-sample forecasts for x_t , by first estimating the model without imposing the restrictions (the "unrestricted forecast") and then re-estimating the model subject to the restrictions (the "restricted forecast"). If the interest is in comparing two alternative sets of restrictions, the unrestricted forecast will be replaced by the alternative restricted forecast, but for simplicity we will continue to refer to the forecasts as "restricted" and "unrestricted".

The forecasts are obtained by a rolling window estimation scheme, which entails estimating the model using data indexed $t - m + 1, \dots, t$ for each $t = m, \dots, T - h$ and using the estimated model at time t to produce a forecast for x_{t+h} . This gives two sequences of $n \equiv T - h - m + 1$ forecasts $\left\{ f_{t,h}^U \right\}_{t=m}^{T-h}$ and $\left\{ f_{t,h}^R \right\}_{t=m}^{T-h}$, denoting respectively the unrestricted and the restricted forecasts.

The asymptotic framework considers the in-sample size m fixed and lets the out-of-sample size n grow to infinity, so that all results are implicitly conditional on the choice of m , which is user-defined. The computation of the time-varying measure of usefulness further requires choosing a smoothing window of size d , which is a constant fraction π of the out-of-sample size n .

The user must finally choose a forecast loss function $L(x_{t+h}, f_{t,h})$. We consider in particular two types of loss functions, a quadratic loss and a portfolio utility loss. The quadratic loss is defined as

$$L(x_{t+h}, f_{t,h}) = (x_{t+h} - f_{t,h})^2, \quad (1)$$

where x_t in our application will be the yield on a zero coupon bond of maturity τ . The portfolio utility loss considers the asset allocation problem of an investor who is buying a portfolio of q assets in period t and then sells it in period $t+1$. In our application such assets will be q zero coupon bonds of maturities $\tau_1, \tau_2, \dots, \tau_q$. Defining x_t as the vector of returns on each asset: $x_t = (x_1, x_2, \dots, x_q)'$, and w^* as a vector of optimal weights, the return on such a portfolio is given by $w^{*'}x_t$. The portfolio utility loss is similar to that considered by West, Edison and Cho (1993), and is given by

$$L(x_{t+h}, f_{t,h}) = -w^{*'}(f_{t,h})'x_{t+h} + \frac{\gamma}{2}w^{*'}(f_{t,h})'\Sigma w^*(f_{t,h}), \quad (2)$$

where $w^*(f_{t,h})$ are the optimal portfolio weights for a quadratic utility, and are linear functions of the forecasts (the exact expression is given in (18) below). Note that in this case $f_{t,h}$ is a vector containing the forecasts of each element in x_t . The matrix Σ is the variance-covariance matrix of x_{t+h} , and γ is a user-defined parameter related to the coefficient of relative risk aversion δ by the relationship $\frac{\gamma}{1-\gamma} = \delta$. Our empirical results are obtained by setting $\delta = 1$, so that $\gamma = .5$.

2.2 Methodology for a general loss function

Consider a combination of the restricted and unrestricted forecast, $f_{t,h}^* = f_{t,h}^R + (1 - \lambda)(f_{t,h}^U - f_{t,h}^R)$, so that λ is the weight on the restricted forecast.

The optimal weight λ^* minimizes the expected out-of-sample loss of the combined forecast:

$$\begin{aligned} \lambda^* &= \arg \min_{\lambda \in \mathbb{R}} E \left[\frac{1}{n} \sum_{t=m}^{T-h} L(x_{t+h}, f_{t,h}^R + (1 - \lambda)(f_{t,h}^U - f_{t,h}^R)) \right] \\ &= \arg \min_{\lambda \in \mathbb{R}} E [Q_n(\lambda)], \end{aligned} \quad (3)$$

and is estimated by

$$\begin{aligned} \hat{\lambda} &= \arg \min_{\lambda \in \mathbb{R}} \frac{1}{n} \sum_{t=m}^{T-h} L(x_{t+h}, f_{t,h}^R + (1 - \lambda)(f_{t,h}^U - f_{t,h}^R)) \\ &= \arg \min_{\lambda \in \mathbb{R}} Q_n(\lambda). \end{aligned} \quad (4)$$

The estimated optimal weight $\hat{\lambda}$ is our measure of the usefulness of the economic restrictions for forecasting, for a given loss function $L(\cdot)$. A small $\hat{\lambda}$ indicates that the restrictions are not useful for forecasting, whereas a large $\hat{\lambda}$ suggests that the economic restrictions can be usefully imposed to obtain more accurate forecasts. $\hat{\lambda}$ in (4) can be computed for a general loss function using numerical methods, but we show how to derive simple analytical expressions for the special cases of a quadratic and portfolio loss functions in Section 3 below.

The asymptotic distribution of $\hat{\lambda}$ is obtained by recognizing that $\hat{\lambda}$ is an M-estimator, which minimizes the (typically well-behaved) objective function $Q_n(\lambda)$. A similar remark was made by Elliott and Timmermann (2004), in an environment where the forecasts are based on different models and are taken as given. The fact that in our context the forecasts are based on the same model and depend on in-sample data and estimated parameters introduces some complications, which we handle using a generalization of the key insight in Giacomini and White (2006). Specifically, we show that an asymptotic theory for $\hat{\lambda}$ can still be derived by relying on laws of large numbers, central limit theorems and functional central limit theorems for the objective function and its derivatives in spite of the fact that such functions depend in a complex nonlinear manner on the in-sample data through $f_{t,h}^R$ and $f_{t,h}^U$. This is because we assume that the in-sample estimation window is finite, so that the objective function and its derivatives become functions of the finite history of "short memory" (mixing) processes, and are thus themselves short memory and plausibly satisfy laws of large numbers and central limit theorems.

We rely on the asymptotic properties of $\hat{\lambda}$ to obtain formal methods for testing the usefulness of the restrictions, both in an environment where such usefulness is constant over time (Section 2.2.1) and in an environment with possibly time-varying usefulness (Section 2.2.2).

2.2.1 Testing the global usefulness of parameter restrictions

We first consider an environment in which λ^* is constant over time, and can thus be interpreted as a "global" measure of the usefulness of the restrictions.

Proposition 1 below shows how to construct formal tests for whether the unrestricted forecast is useless ($H_0^U : \lambda^* = 1$) or whether the restricted forecast is useless ($H_0^R : \lambda^* = 0$), which are essentially out-of-sample encompassing tests. The tests are derived under the following assumptions.

- Assumption A. (1) $E[Q_n(\lambda)]$ is uniquely minimized at $\lambda^* < \infty$;
- (2) $L(x_{t+h}, f_{t,h}^R + (1-\lambda)(f_{t,h}^U - f_{t,h}^R))$ is convex and twice continuously differentiable with respect to λ ;
- (3) $\{x_t\}$ is mixing with ϕ of size $-r/(r-1)$ or α of size $-2r/(r-2)$, $r > 2$;
- (4) $E|L(x_{t+h}, f_{t,h}^R + (1-\lambda)(f_{t,h}^U - f_{t,h}^R))|^{r/2} < \infty$ for all t and all λ ;
- (5) $E|\nabla_\lambda L(x_{t+h}, f_{t,h}^R + (1-\lambda^*)(f_{t,h}^U - f_{t,h}^R))|^{2r} < \infty$ for all t , where ∇_λ indicates the first derivative with respect to λ ;
- (6) $\Omega = E[(\sqrt{n}\nabla_\lambda Q_n(\lambda^*))^2] > 0$ for all n ;
- (7) $H = E[\nabla_{\lambda\lambda} Q_n(\lambda^*)] > 0$ for all n ;
- (8) $\sup_{\lambda \in \Lambda} \|\nabla_{\lambda\lambda} Q_n(\lambda) - E[\nabla_{\lambda\lambda} Q_n(\lambda)]\| \rightarrow^p 0$, where Λ indicates a neighborhood of λ^* and $\nabla_{\lambda\lambda}$ the second derivative with respect to λ ;
- (9) $m < \infty$, $h < \infty$, $n \rightarrow \infty$.

Assumption A(1) is satisfied by the quadratic and the portfolio utility loss functions considered in Section 3, which are both quadratic polynomials in λ . Assumption A(2) is stronger than necessary and is only imposed for convenience and because it is satisfied by the loss functions in Section 3. Following Newey and McFadden (1994), it is straightforward to extend the results to an environment with non-convex and non-differentiable objective functions. Assumptions A(3) to A(7) are the familiar primitive conditions guaranteeing applicability of laws of large numbers and central limit theorems for the objective function and its derivatives. Note that these conditions, while ruling out the presence of unit roots, allow the data to be heterogeneous and dependent. Assumption A(7) could be violated if the forecasts were perfectly correlated in large samples. To see why, consider for simplicity the quadratic loss case, where $H = E\left[2\left(f_{t,h}^R - f_{t,h}^U\right)^2\right]$. If the restrictions were true, the forecasts would become perfectly correlated as the estimation sample grows, making H converge to zero. This occurrence is however ruled out in our context by A(9), which assumes that the estimation sample is fixed, thus preventing estimation uncertainty from disappearing asymptotically. Assumption A(8) requires a uniform law of large numbers for the second derivatives of the objective function. Primitive conditions for A(8) could easily be found, but we do not specify them here because A(8) becomes considerably simpler for the loss functions considered in Section 3, since in both cases the second derivative of the objective function does not depend on λ . For these loss functions, A(8) can be replaced with the condition that $\nabla_{\lambda\lambda} Q_n$ has finite $r/2 - th$ moments, which,

together with A(3), guarantees that a law of large numbers can be invoked for $\nabla_{\lambda\lambda}Q_n$. Assumption (9) shows that the asymptotic distribution is obtained by letting the out-of-sample size n grow to infinity, whereas the in-sample size m and the forecast horizon h are finite.

Proposition 1 (Tests of global usefulness) *Suppose Assumption A holds. Let*

$$\begin{aligned} t^U &= \frac{\sqrt{n}(\hat{\lambda} - 1)}{\hat{\sigma}}; \\ t^R &= \frac{\sqrt{n}\hat{\lambda}}{\hat{\sigma}}, \end{aligned} \quad (5)$$

where $\hat{\sigma}$ is given by

$$\begin{aligned} \hat{\sigma} &= \sqrt{\hat{H}^{-1}\hat{\Omega}\hat{H}^{-1}}; \\ \hat{H} &= \nabla_{\lambda\lambda}Q_n(\hat{\lambda}); \\ \hat{\Omega} &= \sum_{j=-p_n+1}^{p_n-1} \left(1 - \left|\frac{j}{p_n}\right|\right) n^{-1} \sum_{t=m+j}^{T-h} s_t(\hat{\lambda}) s_{t-j}(\hat{\lambda}); \\ s_t(\hat{\lambda}) &= \nabla_{\lambda} \partial L(x_{t+h}, f_{t,h}^R + (1 - \hat{\lambda})(f_{t,h}^U - f_{t,h}^R)), \end{aligned} \quad (6)$$

where p_n is a bandwidth that increases with the sample size (Newey and West, 1987).

Then the hypotheses $H_0^U : \lambda^* = 1$ and $H_0^R : \lambda^* = 0$ are rejected at a significance level α respectively when $|t^U| > c_{\alpha/2}$ and $|t^R| > c_{\alpha/2}$, with $c_{\alpha/2}$ indicating the $1 - \alpha/2$ quantile of a $N(0, 1)$ distribution.

The bandwidth p_n used in the construction of the test statistic must be appropriately chosen to account for the possible serial correlation in the first derivatives of the loss function. In practice, the accuracy of the estimate of Ω can be an issue, particularly for long-horizon forecasts (see, e.g., Kim and Nelson, 1993 and Harvey, Leybourne and Newbold, 1998). In our application, we follow Kim and Nelson (1993)'s recommendation and set $p_n = 2(h - 1)$.

2.2.2 Testing the usefulness of parameter restrictions in the presence of instability

A question that may be of further interest to forecasters is whether the usefulness of the restrictions varies over time. To answer this question, we extend the previous analysis to the case of time-varying forecast combination weights. These time-varying weights can be interpreted as measuring the "local" usefulness of the restrictions, and solve the problem

$$\lambda_t^* = \arg \min_{\lambda_t \in \mathbb{R}} E [L(x_{t+h}, f_{t,h}^R + (1 - \lambda_t)(f_{t,h}^U - f_{t,h}^R))], \quad t = m, \dots, T - h. \quad (7)$$

A simple nonparametric estimator of (7) can be obtained by computing rolling average weights

over windows of size d :

$$\widehat{\lambda}_{t,d} = \arg \min_{\lambda_t \in \mathbb{R}} \sum_{j=t-d+1}^t [L(x_{j+h}, f_{j,h}^R + (1 - \lambda_t)(f_{j,h}^U - f_{j,h}^R))], \quad t = m + d - 1, \dots, T - h. \quad (8)$$

Instead of adopting a standard asymptotic approximation to conduct inference about (8), which would require the bandwidth d/n to go to zero as d and n grow to infinity, we follow a similar approach as Giacomini and Rossi (2010) and obtain a distribution theory for $\widehat{\lambda}_{t,d}$ that has better finite-sample properties by using a non-standard asymptotic approximation with fixed bandwidth. Note that in this fixed-bandwidth approximation, however, $\widehat{\lambda}_{t,d}$ is no longer a consistent estimator of λ_t^* , but it consistently estimates a "smoothed" version of λ_t^* :

$$\lambda_{t,d}^* = \arg \min_{\lambda_t \in \mathbb{R}} \sum_{j=t-d+1}^t E [L(x_{j+h}, f_{j,h}^R + (1 - \lambda_t)(f_{j,h}^U - f_{j,h}^R))], \quad t = m + d - 1, \dots, T - h. \quad (9)$$

Further note that, as a result of adopting a non-standard fixed bandwidth approximation, standard results for optimal bandwidth selection obtained in the nonparametric literature do not apply here. Instead, in our framework different choices of bandwidth result in a different null hypothesis being tested.

A plot of the sample path of $\{\widehat{\lambda}_{t,d}\}_{t=m+d-1}^{T-h}$ in (8) can uncover possible time-variation in the usefulness of the economic restrictions. Proposition 2 below further shows how to test the hypothesis that the unrestricted forecast was consistently useless ($H_0^U : \lambda_{t,d}^* = 1$ for all t) or that the restricted forecast was consistently useless ($H_0^R : \lambda_{t,d}^* = 0$ for all t) over time. We control the overall size of the procedure by deriving uniform confidence bands that have the desired coverage under the null hypothesis.

The proposition relies on the following set of assumptions.

Assumption B. Let $\tau \in [0, 1]$. Under the hypothesis that $\lambda_{t,d}^*$ is constant and equal to λ^* ,

- (1) $\left\{ n^{-1/2} \sum_{j=m}^{m+\lceil \tau n \rceil} \nabla_{\lambda} L(x_{j+h}, f_{j,h}^R + (1 - \lambda^*)(f_{j,h}^U - f_{j,h}^R)) \right\}$ obeys a Functional Central Limit Theorem with $\Omega = \lim_{n \rightarrow \infty} E(n^{-1/2} \sum_{j=m}^{T-h} \nabla_{\lambda} L(x_{j+h}, f_{j,h}^R + (1 - \lambda^*)(f_{j,h}^U - f_{j,h}^R))^2) > 0$;
- (2) $d/n \rightarrow \pi \in (0, \infty)$ as $d \rightarrow \infty, n \rightarrow \infty$. $m < \infty$ and $h < \infty$;
- (3) $\widehat{\sigma} \rightarrow^p \sigma$ and $\widehat{\lambda} \rightarrow^p \lambda^*$.

Primitive conditions for B(1) and B(3) analogous to those listed in Assumption A could be similarly specified here.

Proposition 2 (Tests of time variation in usefulness) *Suppose Assumption B holds. For a significance level α , first construct the bands:*

$$\left(\widehat{\lambda}_{t,d} - k_{\alpha,\pi} \frac{\widehat{\sigma}}{\sqrt{d}}, \widehat{\lambda}_{t,d} + k_{\alpha,\pi} \frac{\widehat{\sigma}}{\sqrt{d}} \right), \quad t = m + d - 1, \dots, T - h, \quad (10)$$

where $k_{\alpha,\pi}$ is tabulated in Table 1 for various values of $\pi = d/n$ and $\hat{\sigma}$ is as in Proposition 1.

The null hypotheses $H_0^U : \lambda_{t,d}^* = 1$ for all t and $H_0^R : \lambda_{t,d}^* = 0$ for all t can be rejected if there exists at least one t at which, respectively, 1 or 0 fall outside the bands.

3 Special cases: quadratic and portfolio utility loss

This section specializes the general methods described in Section 2.2 to the cases of a quadratic and a portfolio utility loss.

3.1 Quadratic loss

For a quadratic loss, the objective function in (3) is $Q_n(\lambda) = \frac{1}{n} \sum_{t=m}^{T-h} (x_{t+h} - f_{t,h}^R - (1-\lambda)(f_{t,h}^U - f_{t,h}^R))^2$, which is minimized by

$$\lambda^* = \frac{E \left[\sum_{t=m}^{T-h} (x_{t+h} - f_{t,h}^U) (f_{t,h}^R - f_{t,h}^U) \right]}{E \left[\sum_{t=m}^{T-h} (f_{t,h}^R - f_{t,h}^U)^2 \right]}. \quad (11)$$

A consistent estimator of λ^* is

$$\hat{\lambda} = \frac{\sum_{t=m}^{T-h} (x_{t+h} - f_{t,h}^U) (f_{t,h}^R - f_{t,h}^U)}{\sum_{t=m}^{T-h} (f_{t,h}^R - f_{t,h}^U)^2}, \quad (12)$$

or, equivalently, the OLS estimator of λ in the regression

$$x_{t+h} - f_{t,h}^U = \lambda(f_{t,h}^R - f_{t,h}^U) + \varepsilon_{t+h}, \quad t = m, \dots, T-h. \quad (13)$$

The estimator $\hat{\sigma}$ that is needed for constructing the tests in Proposition 1 and Proposition 2 is in this case given by

$$\hat{\sigma} = \left(\frac{1}{n} \sum_{t=m}^{T-h} (f_{t,h}^R - f_{t,h}^U)^2 \right)^{-1} \sqrt{\sum_{j=-p_n+1}^{p_n-1} \left(1 - \left| \frac{j}{p_n} \right| \right) n^{-1} \sum_{t=m+j}^{T-h} (f_{t,h}^R - f_{t,h}^U) \hat{\varepsilon}_{t+h} (f_{t-j,h}^R - f_{t-j,h}^U) \hat{\varepsilon}_{t+h-j}}, \quad (14)$$

where $\hat{\varepsilon}_{t+h}$ are regression residuals from (13) and p_n is a bandwidth that increases with the sample size (Newey and West, 1987).

In the presence of possible instability, a consistent estimator of the smoothed measure of usefulness $\lambda_{t,d}^*$ in (9) can be similarly obtained as

$$\hat{\lambda}_{t,d} = \frac{\sum_{j=t-d+1}^t (x_{j+h} - f_{j,h}^U) (f_{j,h}^R - f_{j,h}^U)}{\sum_{j=t-d+1}^t (f_{j,h}^R - f_{j,h}^U)^2}, \quad t = m+d-1, \dots, T-h, \quad (15)$$

or, equivalently, by estimating the OLS coefficient in the following regression over rolling samples of size d :

$$\begin{aligned} x_{j+h} - f_{j,h}^U &= \lambda_{t,d}(f_{j,h}^R - f_{j,h}^U) + \varepsilon_{j+h}; \\ j &= t - d + 1, \dots, t; \\ t &= m + d - 1, \dots, T - h. \end{aligned} \quad (16)$$

In the empirical application, the variable to forecast will be $x_t = y_t^{(\tau)}$, i.e. the yield of a bond of maturity τ .

3.2 Portfolio utility loss

Let $x_t = (x_1, x_2, \dots, x_q)'$ be a $q \times 1$ vector of risky assets and consider the portfolio $w'x_t$, with weights summing to 1. In analogy with our empirical application to no-arbitrage VARs, we suppose the forecaster has a model for x_{t+h} and has the option of estimating it unrestricted or by imposing restrictions that only affect the conditional mean parameters. We further assume that the model does not specify conditional variance dynamics, so that the conditional variance of x_{t+h} at time t simply equals the unconditional variance-covariance matrix of the q assets Σ , so that $Var_t[x_{t+h}] = Var[x_{t+h}] = \Sigma$.

We suppose that at each time $t = m, \dots, T - h$ the forecaster constructs a portfolio by choosing the weights that minimize a quadratic utility function:

$$w^* = \arg \min_w \left\{ w' E_t[x_{t+h}] - \frac{\gamma}{2} w' \Sigma w \right\}, \quad (17)$$

where $E_t[\cdot]$ denotes the conditional mean at time t . The classical solution (Markowitz, 1952) to this problem is given by

$$\begin{aligned} w^* &= a + B E_t[x_{t+h}]; \\ a &= \frac{\Sigma^{-1} \iota}{\iota' \Sigma^{-1} \iota}; \\ B &= \frac{1}{\gamma} \left(\Sigma^{-1} - \frac{\Sigma^{-1} \iota \iota' \Sigma^{-1}}{\iota' \Sigma^{-1} \iota} \right), \end{aligned} \quad (18)$$

where ι is a $q \times 1$ vector of ones.

When the economic restrictions only affect the conditional mean of the assets, as is the case for the no-arbitrage restrictions that we are interested in, the forecaster can construct two different portfolios, one by forecasting the conditional mean with the unrestricted model, so that $E_t[x_{t+h}] = f_{t,h}^U$, and one by imposing the restriction and letting $E_t[x_{t+h}] = f_{t,h}^R$.

We can similarly consider the portfolio whose optimal weights are a function of the combination

forecast, and our measure of usefulness is then obtained by minimizing the expected portfolio utility loss in (2) with respect to the forecast combination weight λ :

$$\lambda^* = \arg \min_{\lambda \in \mathbb{R}} E \left\{ \left(-a - B (f_{t,h}^R + (1 - \lambda)(f_{t,h}^U - f_{t,h}^R)) \right)' x_{t+h} + \frac{\gamma}{2} \left(a + B (f_{t,h}^R + (1 - \lambda)(f_{t,h}^U - f_{t,h}^R)) \right)' \Sigma \left(a + B (f_{t,h}^R + (1 - \lambda)(f_{t,h}^U - f_{t,h}^R)) \right) \right\}. \quad (19)$$

The closed-form solution for this problem is

$$\lambda^* = \frac{E \left[(f_{t,h}^R - f_{t,h}^U)' B' \left(x_{t+h} - \gamma \Sigma \left(a + B f_{t,h}^U \right) \right) \right]}{E \left[\gamma (f_{t,h}^R - f_{t,h}^U)' B' \Sigma B (f_{t,h}^R - f_{t,h}^U) \right]}. \quad (20)$$

A consistent estimator of λ^* is

$$\hat{\lambda} = \frac{\sum_{t=m}^{T-h} \left[(f_{t,h}^R - f_{t,h}^U)' \hat{B}'_t \left(x_{t+h} - \gamma \hat{\Sigma}_t \left(\hat{a}_t + \hat{B}_t f_{t,h}^U \right) \right) \right]}{\sum_{t=m}^{T-h} \left[\gamma (f_{t,h}^R - f_{t,h}^U)' \hat{B}'_t \hat{\Sigma}_t \hat{B}_t (f_{t,h}^R - f_{t,h}^U) \right]}, \quad (21)$$

where \hat{a}_t and \hat{B}_t are as defined in (18) with Σ substituted at each time t by an estimate computed over each rolling window of data up to time t :

$$\hat{\Sigma}_t = \frac{1}{m} \sum_{j=t-m+1}^t (x_j - \bar{x})(x_j - \bar{x})', \quad \text{with } \bar{x} = \frac{1}{m} \sum_{j=t-m+1}^t x_j. \quad (22)$$

The estimator of the asymptotic variance $\hat{\sigma}$ that is needed for constructing the test in Proposition 1 and the bands in Proposition 2 are obtained by setting

$$\begin{aligned} s_t(\hat{\lambda}) &= (f_{t,h}^R - f_{t,h}^U)' \hat{B}'_t \left[x_{t+h} - \gamma \hat{\Sigma}_t \left(\hat{a}_t + \hat{B}_t \left(f_{t,h}^R + (1 - \hat{\lambda})(f_{t,h}^U - f_{t,h}^R) \right) \right) \right] \quad \text{and} \\ \frac{\partial s_t(\hat{\lambda})}{\partial \lambda} &= \gamma (f_{t,h}^R - f_{t,h}^U)' \hat{B}'_t \hat{\Sigma}_t \hat{B}_t (f_{t,h}^R - f_{t,h}^U) \end{aligned} \quad (23)$$

in equation (6).

In the presence of time variation, a consistent estimator of the smoothed measure of usefulness (9) for a portfolio utility loss can be obtained as

$$\hat{\lambda}_{t,d} = \frac{\sum_{j=t-d+1}^t \left[(f_{j,h}^R - f_{j,h}^U)' \hat{B}'_j \left(x_{j+h} - \gamma \hat{\Sigma}_j \left(\hat{a}_j + \hat{B}_j f_{j,h}^U \right) \right) \right]}{\sum_{j=t-d+1}^t \left[\gamma (f_{j,h}^R - f_{j,h}^U)' \hat{B}'_j \hat{\Sigma}_j \hat{B}_j (f_{j,h}^R - f_{j,h}^U) \right]}, \quad t = m + d - 1, \dots, T - h. \quad (24)$$

In the empirical application, the variable to be forecasted will be $x_t = (r_t^{(\tau_1)}, r_t^{(\tau_2)}, \dots, r_t^{(\tau_q)})'$, i.e. a vector of returns on bonds of q different maturities.

4 Illustrative example and finite sample properties

To gain intuition about the determinants of our measure of usefulness λ^* and to assess the finite-sample properties of our tests, we consider a simple example of two competing sets of restrictions imposed on the parameters of a linear model and investigate the size properties of the global usefulness test in Proposition 1.

Suppose the data-generating process is:

$$\begin{aligned} x_t &= \beta_1 z_{1t} + \beta_2 z_{2t} + \varepsilon_t; \\ z_t &\sim iidN(0, I_2), \varepsilon_t \sim iidN(0, 1), \end{aligned} \quad (25)$$

and that the two models M_U and M_R impose the competing restrictions $\beta_1 = \beta_1^U$ or $\beta_1 = \beta_1^R$, while leaving β_2 unrestricted. This yields the one-step-ahead forecasts $f_t^U = \beta_1^U z_{1,t+1} + \hat{\beta}_2 z_{2,t+1}$ and $f_t^R = \beta_1^R z_{1,t+1} + \hat{\gamma}_2 z_{2,t+1}$ with $\hat{\beta}_2$ and $\hat{\gamma}_2$ OLS estimators.

The optimal weight (11) for a quadratic loss function in this example is¹:

$$\lambda^* = \frac{\beta_1 - \beta_1^U}{\beta_1^R - \beta_1^U}, \quad (26)$$

which reveals that the usefulness of the restrictions is determined by the relative amount of bias implied by the two models, so that λ^* equals 0 or 1 when either restriction is true. Notice that the weight does not necessarily fall between 0 and 1, but in principle could be any value on the real line. Intuitively, in this simple example the relative bias of the models is the sole determinant of the usefulness of the restrictions because the two forecasts imply the same amount of estimation uncertainty, but in more general cases there will be a bias-variance trade-off between different sets of restrictions.

We now proceed to illustrate the small sample properties of our tests by using a simple Monte Carlo simulation. We assume that the true DGP is given by (25) and we consider the following two restricted models. The first model imposes on (25) the restriction $\beta_1 = \beta_1^U = 1$:

$$M_U : x_t = z_{1t} + \beta_2 z_{2t} + \varepsilon_t. \quad (27)$$

The second model is imposing on (25) the restriction $\beta_1 = \beta_1^R = 0$.

$$M_R : x_t = \gamma_2 z_{2t} + \varepsilon_t. \quad (28)$$

¹To see why, note that the numerator in (11) is $(\beta_1 - \beta_1^U)(\beta_1^R - \beta_1^U) + E(\beta_2 - \hat{\beta}_2)(\hat{\gamma}_2 - \hat{\beta}_2)$ and the denominator is $(\beta_1^R - \beta_1^U)^2 + E(\hat{\gamma}_2 - \hat{\beta}_2)^2$, from which the result follows by noting that $E(\hat{\beta}_2) = E(\hat{\gamma}_2)$ and $E(\hat{\beta}_2 \hat{\gamma}_2) = E(\hat{\beta}_2^2) = E(\hat{\gamma}_2^2)$.

For this set of restrictions the expression in (26) simplifies to $\lambda^* = 1 - \beta_1$. We draw 5000 random samples from the DGP in (25) setting² $\beta_2 = 1$ and using in turn $\beta_1 = 1$ and $\beta_1 = 0$. The case $\beta_1 = 1$ yields by construction $\lambda^* = 0$ (M_U imposes the right restrictions and M_R is useless). The case $\beta_1 = 0$ yields by construction $\lambda^* = 1$ (M_R imposes the right restrictions and M_U is useless).

For each of the 5000 replications we first estimate the two restricted models and obtain the corresponding forecasts. Then we compute λ^* and $\hat{\sigma}$ using (13) and (14), and use them to obtain the test statistics in (5). Results of our simulation for different sizes of the estimation window m and forecast window n are reported in Table 2. Panel A displays the empirical size of the test in (5) for $H_0 : \lambda = 0$, Panel B reports the empirical size the test in (5) for $H_0 : \lambda = 1$. In both cases the nominal size of the tests is 0.05.

As is clear from the table, the tests in (5) are well sized, and for a given estimation window the size tends to improve as the size of the forecast window increases.

Before turning to the empirical application we want to briefly discuss the relation of our approach to Bayesian forecast combination. A Bayesian econometrician would attach a-priori probabilities $p(M_U) = 1 - p(M_R)$ and $p(M_R)$ to models M_U and M_R . Note that the two models exhaust all the possibilities. Then he would use Bayes' formula to derive the posterior probability of M_R , i.e., $p(M_R|x)$, where x denotes the available data. Under a flat prior this posterior probability is simply given by the (properly normalized) marginal likelihood of M_R . The optimal point forecast for a quadratic loss is then given by:

$$f_t^* = E[x_{t+1}|x] = f_t^U(1 - p(M_R|x)) + f_t^R p(M_R|x). \quad (29)$$

The expression above is conceptually different from that resulting from forecast combination in a classical framework. In (29) the optimal weights are probabilities, and as such they are constrained to be between 0 and 1, while in the classical framework the weights depend on the variances and correlation between the forecasts and they can in principle be negative or greater than 1. As the weights are between 0 and 1, the combined forecast produced by the Bayesian econometrician is a convex combination of the forecasts of the original models, which makes sense as he believes that the two competing models exhaust all the possibilities. On the other hand, for the classical econometrician both models can be misspecified, but their forecasts can be useful if combined in some optimal way. For a comparison of the two approaches with a discussion about the cases in which they can coincide see Palm and Zellner (1992).

There is a simple example in which the optimal combination weights computed in our framework can have an interpretation similar to the Bayesian one. Consider the DGP in (25), but further assume that $\beta_1 = \beta_1^U$ with probability $1 - \pi$ (in which case M_U is true) and $\beta_1 = \beta_1^R$ with probability π (in which case M_R is true). In this case, the optimal weight is³ $\lambda^* = \pi$, so the estimate $\hat{\lambda}$ can be

²As is clear from (26) the choice of the value of β_2 in the true DGP does not influence the results as both models will estimate it unbiasedly.

³To see why note β_1 is a discrete random variable independent from all the other variables in (25). It follows that

thought of as the frequentist equivalent of the posterior probability of M_R computed by a Bayesian econometrician with a flat prior.

5 Application: usefulness of the no-arbitrage restrictions for predicting the term structure of interest rates

In this section we apply our proposed framework to the problem of forecasting the yield curve using no-arbitrage restrictions. Our framework enables us to address several questions such as: "are no-arbitrage restrictions useful for forecasting the term structure of interest rates?", "are the restrictions useful for bond portfolio allocation?", "does time variation in the term premium help in forecasting?", and "have the restrictions become more or less useful over time?".

We will start with describing how the no-arbitrage restrictions can be imposed on a VAR model for the yields, and then turn to the forecasting exercise and provide the results for a quadratic and a portfolio loss function.

5.1 A benchmark no-arbitrage affine term structure model (ATSM)

We consider the affine term structure model (*ATSM*) proposed by Ang and Piazzesi (2003), which is a discrete-time version of the affine class introduced by Duffie and Kan (1996), where bond prices are exponential affine functions of underlying state variables. The assumption of no-arbitrage (Harrison and Kreps, 1979) guarantees the existence of a risk neutral measure Q such that the price at time t of an asset V_t that does not pay any dividends at time $t + 1$ satisfies $V_t = E_t^Q [\exp(-i_t)V_{t+1}]$, where the expectation is taken with respect to the measure Q and i_t is the short term rate. The assumption of no-arbitrage is equivalent to the assumption of the existence of the Radon-Nikodym derivative ξ_{t+1} , which allows to convert the risk neutral measure into the data generating measure: $E_t^Q [\exp(-i_t)V_{t+1}] = E_t [(\xi_{t+1}/\xi_t) \exp(-i_t)V_{t+1}]$. Assume ξ_{t+1} follows a log-normal process:

$$\xi_{t+1} = \xi_t \exp(-0.5\Lambda_t'\Lambda_t - \Lambda_t'\varepsilon_{t+1}). \quad (30)$$

Λ_t is called the market price of risk and is an affine function of a vector of k factors F_t :

$$\Lambda_t = \Lambda_0 + \Lambda_1 F_t, \quad (31)$$

where Λ_0 is a k -dimensional vector and Λ_1 a $k \times k$ matrix. The short term rate is also assumed to be an affine function of F_t :

$$i_t = \delta_0 + \delta_1' F_t, \quad (32)$$

the optimal λ is $\lambda^* = \frac{E[\beta_1] - \beta_1^U}{\beta_1^R - \beta_1^U}$. Substituting $E[\beta_1] = (1 - \pi)\beta_1^U + \pi\beta_1^R$ in the expression for λ^* provides the result.

where δ_0 is a scalar and δ_1 a k -dimensional vector. We assume that the factors follow a zero-mean stationary vector process:

$$F_t = \Psi F_{t-1} + \Omega \varepsilon_t, \quad (33)$$

where $\varepsilon_t \sim iidN(0, \Sigma_\varepsilon)$ with $\Sigma_\varepsilon = I$ with no loss of generality. The nominal pricing kernel is defined as:

$$m_{t+1} = \exp(-i_t) \xi_{t+1} / \xi_t = \exp(-\delta_0 - \delta_1' F_t - 0.5 \Lambda_t' \Lambda_t - \Lambda_t' \varepsilon_{t+1}), \quad (34)$$

where the second equality comes from (32) and (30). The nominal pricing kernel prices all assets in the economy, so, by letting $p_t^{(\tau)}$ denote the time t price of a τ -period zero coupon, we have:

$$p_t^{(\tau+1)} = E_t(m_{t+1} p_{t+1}^{(\tau)}). \quad (35)$$

Using the above equations, it is possible to show that bond prices are an affine function of the state variables:

$$p_t^{(\tau)} = \exp(\bar{A}_\tau + \bar{B}_\tau' F_t), \quad (36)$$

where \bar{A}_τ and \bar{B}_τ are a scalar and a k -dimensional vector obeying:

$$\begin{aligned} \bar{A}_{\tau+1} &= \bar{A}_\tau + \bar{B}_\tau' (-\Omega \Lambda_0) + 0.5 \bar{B}_\tau' \Omega \Omega' \bar{B}_\tau - \delta_0; \\ \bar{B}_{\tau+1} &= \bar{B}_\tau' (\Psi - \Omega \Lambda_1) - \delta_1', \end{aligned} \quad (37)$$

with $\bar{A}_1 = -\delta_0$ and $\bar{B}_1 = -\delta_1$. See Ang and Piazzesi (2003) for a formal derivation. The continuously compounded yield on a τ -period zero coupon bond is:

$$y_t^{(\tau)} = -\ln p_t^{(\tau)} / \tau = A_\tau + B_\tau' F_t, \quad (38)$$

with $A_\tau = -\bar{A}_\tau / \tau$ and $B_\tau = -\bar{B}_\tau / \tau$, so yields are also an affine function of the factors. Equations (33) and (38) define a state-space model:

$$\begin{aligned} F_t &= \Psi F_{t-1} + \Omega \varepsilon_t; \\ Y_t &= A + B F_t + v_t, \end{aligned} \quad (39)$$

where $Y_t = (y_t^{(\tau_1)}, y_t^{(\tau_2)}, \dots, y_t^{(\tau_q)})'$ is a q -dimensional vector process collecting all the yields at maturities $\tau_1, \tau_2, \dots, \tau_q$, $A = (A_{\tau_1}, A_{\tau_2}, \dots, A_{\tau_q})'$ and $B = (B_{\tau_1}, B_{\tau_2}, \dots, B_{\tau_q})'$ are functions of the structural coefficients of the model according to equation (37), and v_t is a vector of i.i.d. Gaussian measurement errors with variance Σ_v .

Following common practice, we use three factors, which can be interpreted as the level, slope and curvature of the term structure. Given that scaling, shifting, or rotation of the factors provides observational equivalence, a normalization is required. Following Dai and Singleton (2000) we identify

the factors by assuming factor mean equal to zero, a lower triangular structure for the matrix Ψ , and we set $\delta_1 = (1, 1, 0)'$. Given this identification scheme, the coefficient δ_0 equals the unconditional mean of the instantaneous rate, which can be approximated by the sample average of the 1-month yield. As for second order coefficients, we assume Ω and Σ_v to be diagonal, while we assume absence of correlation between the state and the measurement equation disturbances, i.e. $\Sigma_{\varepsilon v} = 0$.

We collect all the parameters to be estimated in the vector:

$$\theta = \{\Psi, \Omega, \Lambda_0, \Lambda_1, \Sigma_v\}. \quad (40)$$

We estimate θ with the EM algorithm, evaluating the likelihood at each iteration by means of the Kalman Filter. In our application we also consider a specification of the model with constant risk premium, which amounts to setting $\Lambda_1 = 0$ in equation (31).

5.2 VARs with no-arbitrage ATSM restrictions (ATSM-VAR)

Now consider a $VAR(p)$ representation of the q -dimensional vector collecting all the yields at hand:

$$Y_t = \Phi_0 + \Phi_1 Y_{t-1} + \dots + \Phi_p Y_{t-p} + u_t \quad (41)$$

where $Y_t = (y_t^{(\tau_1)}, y_t^{(\tau_2)}, \dots, y_t^{(\tau_q)})'$ and u_t is a vector of one-step-ahead forecast errors having a multivariate normal distribution with variance Σ_u . The VAR in (41) can be interpreted as an approximation of the Moving Average (MA) representation of Y_t . The approximation gets better as more dynamics are added to the system.

Importantly, as is clear from equation (39), the $ATSM$ features an MA representation. As the $ATSM$ depends on a vector of coefficients θ (eq. (40)) having much fewer elements than the coefficient matrices of the VAR , the $ATSM$ imposes a set of nonlinear cross-equation restrictions on the VAR in (41).

To impose such restrictions on the VAR we follow Del Negro and Schorfheide (2004), i.e. we first compute the moments of Y_t under the state-space in equation (39), and then impose them on the VAR in (41). To do so, rewrite the VAR in the data-matrix notation:

$$Y = X\Phi + U, \quad (42)$$

where Y is a $T \times q$ data-matrix with rows Y_t' , X is a $T \times k$ (where $k = 1 + qp$) data-matrix with rows $X_t = (1, Y_{t-1}', Y_{t-2}', \dots, Y_{t-p}')$, $\Phi = (\Phi_0, \Phi_1, \dots, \Phi_p)'$, and U is a $T \times q$ data-matrix with rows u_t' . Let E_θ denote the expectation under the $ATSM$ model and define the autocovariance matrices $\Gamma_{xx}^*(\theta) = E_\theta(X_t X_t')$ and $\Gamma_{XY}^*(\theta) = E_\theta(X_t Y_t')$, which can be computed using the state-space representation in (39) for a given θ . Then, under the $ATSM$, the relation between the $ATSM$

parameters and the *VAR* parameters is⁴ $\Phi^* = [\Gamma_{xx}^*(\theta)]^{-1}\Gamma_{xy}^*(\theta)$, where the star indicates that the *ATSM* restrictions hold. Defining $\hat{\theta}$ as the maximum likelihood estimator of θ , the maximum likelihood estimator for the *VAR* coefficients under the *ATSM* is:

$$\hat{\Phi}^* = [\Gamma_{xx}^*(\hat{\theta})]^{-1}\Gamma_{xy}^*(\hat{\theta}). \quad (43)$$

In the following we will refer to this model as the *ATSM – VAR*. We also consider a specification in which we impose on the *ATSM – VAR* the additional restriction of a constant risk premium, which is obtained simply by setting $\Lambda_1 = 0$ in equation (31). We label this case *CRP – VAR*. In the empirical application the structural coefficients θ and the corresponding *ATSM* moments $\Gamma_{xx}^*(\hat{\theta})$ and $\Gamma_{xy}^*(\hat{\theta})$ are re-estimated in pseudo-real-time as forecasting proceeds forward within our rolling estimation-forecasting scheme. The maximum likelihood estimator of the unrestricted VAR (*UVAR*) is simply $\hat{\Phi} = (X'X)^{-1}X'Y$.

5.3 Forecasting exercise

For our exercise we use monthly data on zero coupon bond yields of maturities 1-, 3-, 12-, 36-, and 60-month, from January 1964 to December 2003. The data are taken from the Fama CRSP zero coupon and Treasury Bill files.

We produce 1-step ahead forecasts using the *UVAR* (we label such forecasts f_t^U), the *ATSM – VAR* (f_t^{ATSM}), the *CRP – VAR* (f_t^{CRP}), and a simple random walk (*RW*) forecast (f_t^{RW}).

For each of the models at hand the sequences of forecasts are produced over the sample 1974:1 to 2003:12 using a rolling estimation window of 10 years. The procedure thus starts with estimating all the models using the estimation window 1964:1 to 1973:12, and producing the forecasts for the vector of yields in 1974:1. Then the estimation window is moved one period ahead, to 1964:2 to 1974:1, and the new estimates are used to produce the forecasts for the vector of yields in 1974:2.

In light of the forecasting focus of the paper, the structural coefficients θ and the corresponding *ATSM* moments $\Gamma_{xx}^*(\hat{\theta})$ and $\Gamma_{xy}^*(\hat{\theta})$ are re-estimated using the data in each of the rolling samples. We maximize the likelihood of the VAR with no-arbitrage *ATSM* restrictions using the Broyden, Fletcher, Goldfarb, and Shanno (BFGS) algorithm with Brent line search.⁵ The procedure is iterated until the last forecast (i.e., that for 2003:12) is obtained. For the VAR models, we use a specification

⁴As stressed above, the approximation is not exact because the state-space representation of the ATSM generates moving average terms

⁵In the first estimation window we initialize our algorithm as follows. First we compute a maximum, then we draw 100 alternative starting points by randomizing around this maximum (drawing from a normal with variance derived from the Hessian at the maximum), maximize again, and check that none of the random initial points leads to a point with higher likelihood (i.e. a new maximum). If this is not the case, we take the new maximum and repeat the randomization until no points with higher likelihood are found. Then, for all the remaining estimation windows, we use the optimum obtained in the previous period $t - 1$ as initial condition for the maximization performed in period t . The in-sample fit of the estimated models is extremely high throughout the sample (with the R^2 being around 0.998).

with 3 lags which provides well-behaved residuals.⁶

5.4 Results for a quadratic loss

We first consider the results for a quadratic loss function. In the case at hand the loss function in (1) specializes to:

$$L(y_{t+1}^{(\tau)}, \hat{y}_{t+1}^{(\tau)}) = \left(y_{t+1}^{(\tau)} - \hat{y}_{t+1}^{(\tau)} \right)^2, \quad (44)$$

where $y_{t+1}^{(\tau)}$ is the yield to maturity of a bond of maturity τ in period $t + 1$ and $\hat{y}_{t+1}^{(\tau)}$ is the 1-step ahead forecast of such variable. We provide results for bonds of five different maturities: 1-, 3-, 12-, 36-, 60- months.

We start with the results based on the global measure of usefulness. Results are displayed in Table 3, which is composed of four panels, each corresponding to a different combination: *ATSM* – *VAR* and *UVAR* (panel A), *CRP* – *VAR* and *UVAR* (panel B), *ATSM* – *VAR* and *RW* (panel C), *CRP* – *VAR* and *RW* (panel D). For each yield, column (1) in Table 3 contains the Root Mean Squared Forecast Error (*RMSFE*) (i.e., the realized loss) of the first model considered in the combination. Columns (2) to (4) report the percentage gains over the *RMSFE* of the model in column (1) obtained by using, respectively, the second model in the combination, a combination with equal weights ($f_t^{\lambda=1/2}$), and a combination based on the estimated optimal weight (f_t^*). Column (5) reports the value of the estimated optimal weight $\hat{\lambda}$ (as defined in equation (12)). Finally, columns (6) to (8) report the statistics for the encompassing tests of Proposition (1).

We included in the comparison the case of a combination with equal weights for reference, because there is a long literature documenting the fact that in practice equal weights usually yield more accurate forecasts than optimal combination weights. Of course this can happen in a real-time forecasting exercise, when the optimal weights are estimated using only past information, while in our case the optimal weights outperform the equal weights by construction. Still, it may be interesting to see whether the optimal weights are statistically different from 0.5, because this might explain the success of combinations with equal weights, and how large are the gains in using optimal rather than equal weights.

From Panel A we see that the estimated optimal weights $\hat{\lambda}$ range between 0.514 and 1.054, and in all cases the encompassing test rejects the null that the optimal weight is zero, i.e., the restricted model is useless, while it cannot reject (except for the 1-month yield) the null that the unrestricted forecast is useless. Therefore there is evidence that imposing the no-arbitrage *ATSM* restrictions on a VAR might help in forecasting, although the restrictions are not uniformly useful across yields, in particular they seem to work better for bonds with longer maturity. This is in line with Carriero (2007) who shows that the misspecification of the *ATSM* restrictions is more pronounced at the

⁶The Bayesian Information Criterion selects 1 lag, but the LM test statistic reported in Johansen (1995) rejects the null of no residual autocorrelation. The specification with 3 lags is the most parsimonious one which eliminates this problem. Our results are robust to specifications with 1 to 4 lags.

short-end of the yield curve while it is milder for yields of longer maturities. The pattern of the forecast gains is obviously related to that of the weights, with the gains from using the combination and using the restricted model being similar for longer maturities. With regards to the comparison with the combination with equal weights, in general the optimal weights lead to larger gains, as the optimal weights are well above 0.5 for all cases except the 1-month yield.

Panel B provides results for the case in which we impose on the *ATSM* – *VAR* the additional restriction of a constant risk premium, i.e. the *CRP* – *VAR* . The estimated optimal weights for the 3-, 36-, and 60- month yields do not change dramatically, while there is a sharp decrease in the usefulness of the restrictions for the 12-month and especially for the 1-month yield. The results in terms of significance of the optimal weights are entirely in line with those obtained with the specification with variation in the risk premium, with a slightly stronger evidence against the usefulness of the restrictions. Therefore, most of the forecasting gains do not seem to be strongly related to the presence of time varying rather than constant risk premia in the model, except at the short end of the yield curve. The combined forecasts based on the optimal weights lead to larger gains with respect to the case with equal weights.

Panel C and D provide results for the combination between the random walk and the VAR with no-arbitrage *ATSM* restrictions (with and without variation in the risk premium). Results for this case show that most of the gains documented in Panel A and Panel B seem to be related to the failure of the unrestricted VAR to provide a good forecast of the yield curve rather than to the merits of the no-arbitrage *ATSM* restrictions. In particular, for the combination of the *ATSM* – *VAR* with the *RW*, the estimated optimal weights $\hat{\lambda}$ range between -0.186 and 0.629 , while for the combination of *CRP* – *VAR* with the *RW* the weights $\hat{\lambda}$ range between 0.090 and 0.380 . These figures are much lower than those obtained when considering the combination of the VAR with no-arbitrage *ATSM* restrictions with the unrestricted VAR (see panels A and B). In particular, all the weights decrease quite dramatically, and the encompassing test does not reject the null that the no-arbitrage *ATSM* restrictions are useless, with the only exception of the weight on the 1-month yield. Indeed, the latter is the only case in which the random walk produces quite poor forecasts, worse than those produced by the unrestricted VAR. As a result the gains coming from the optimal combination, though positive, are relatively small for all maturities but the 1-month. The combination with equal weights does not work very well for this case, yielding small losses rather than gains in most of the cases. This happens because the equal weights are too high with respect to the optimal ones.

The difficulty in beating the random walk forecasts is not surprising, given the results usually obtained for the 1-step ahead case in the literature. For example, in the paper by Diebold and Li (2006) the gains in *RMSFE* of all the considered models at 1-step-ahead with respect to the random walk are quite low. Among all the models they consider, the best forecasts for the 3-month and 12-month yield provide gains of at most 4%, the best forecasts of the 24-month yield provide a gain of 1%, while for yields of longer maturity the gains are either zero or negative. The maximum gain

in *RMSFE* in using an *ATSM* against the random walk found by Almeida and Vicente (2008) for the 1-step ahead case is of 5% for the 2-year yield, while they found gains in the range of 1%-3.5% for intermediate maturities, and negative gains for yields of longer maturity. The *ATSM* estimated by Moench (2008) outperforms the random walk forecast only for the 6-month yield, with a gain of 3%, while it is outperformed for all the remaining maturities. Better results are obtained by Favero et al. (2009) and Moench (2008) by extending the approach with the inclusion of a broad macroeconomic information set.

We now turn to the results for the local measure of usefulness, which are summarized in Figure 1. The figure is composed of 20 panels displayed in 5 rows and 4 columns. The rows display results for different yields, while the columns represent different forecast combinations (respectively f_t^{ATSM} and f_t^U in the first column, f_t^{CRP} and f_t^U in the second column, f_t^{ATSM} and f_t^{RW} in the third column, and f_t^{CRP} and f_t^{RW} in the last column). Each panel contains a plot of the estimated smoothed weights, as defined in equation (15), together with the 95% bands described in Proposition 2. We set $\pi = 0.3$, which given that our out-of sample size $n = 360$ implies that the smoothed weights are estimated using a window of $d = 108$ observations.

Looking at the first two columns in Figure 1, it is clear that the forecasting gains from imposing the no-arbitrage *ATSM* restrictions onto the VAR are not constant over time. In particular, the optimal weight is not statistically different from one in the first part of the sample, but in more recent years the estimated optimal weight $\hat{\lambda}_t$ decreases, signalling that there are small or no gains in using the no-arbitrage *ATSM* restrictions from around 1994 to 2003.

Moreover, by comparing the first and the second column of Figure 1, it is apparent that the effect of including a time varying rather than constant risk premium is not strong, and is mostly limited to the short end of the yield curve, which confirms the results found using the global measure (see Table 2).

Finally, by looking at the last two columns of Figure 1, it is clear that when the VAR with no-arbitrage restrictions is combined with the random walk, one cannot reject the null that the no-arbitrage *ATSM* restrictions are useless throughout the sample.

5.5 Results for a portfolio utility loss

The portfolio utility loss considers the asset allocation problem of an investor who is buying a portfolio of bonds in period t and then sells it in period $t + 1$, and therefore earning/losing the change occurring in the value of the portfolio within t and $t + 1$. The holding period return on a yield of maturity τ is:

$$r_{t+1}^{(\tau+1)} = p_{t+1}^{(\tau)} - p_t^{(\tau+1)} = -\tau y_{t+1}^{(\tau)} + (\tau + 1)y_t^{(\tau+1)}. \quad (45)$$

Equation (45) shows that a forecast of the yield $\hat{y}_{t+1}^{(\tau)}$ provides a forecast of the holding period return $\hat{r}_{t+1}^{(\tau+1)}$ via a simple transformation⁷:

$$\hat{r}_{t+1}^{(\tau+1)} = -\tau\hat{y}_{t+1}^{(\tau)} + (\tau + 1)y_t^{(\tau+1)}. \quad (46)$$

Collecting all the returns under consideration in the vector $r_{t+1} = (r_{t+1}^{(\tau_1+1)}, r_{t+1}^{(\tau_2+1)}, \dots, r_{t+1}^{(\tau_q+1)})'$, and setting $x_{t+1} = r_{t+1}$ the loss function in (2) specializes to:

$$L(r_{t+1}, f_{t,1}) = -w^*(f_{t,1})'r_{t+1} + \frac{\gamma}{2}w^*(f_{t,1})'\Sigma w^*(f_{t,1}), \quad (47)$$

where $f_{t,1}$ is a vector of forecasts of r_{t+1} and can be derived from the forecasts of the yields by using (46).

Results for the global measure of usefulness are in Table 4. Each panel in the table corresponds to a different forecast combination: Panels A and B contain results from the combination of f_t^{ATSM} and f_t^{CRP} with the unrestricted VAR forecasts f_t^U , while panels C and D contain results of the combination of f_t^{ATSM} and f_t^{CRP} with the random walk forecasts f_t^{RW} . Column (1) contains the portfolio utility loss of the first model considered in the combination. Columns (2) to (4) report the percentage gains over the utility loss of the model in column (1) obtained by using, respectively, the second model in the combination, a combination with equal weights ($f_t^{\lambda=1/2}$), and a combination based on the estimated optimal weight (f_t^*). Column (5) reports the value of the estimated optimal weight $\hat{\lambda}$ (as defined in equation (12)). Finally, columns (6) to (8) report the statistics for the encompassing tests of Proposition (1).

Two main results emerge from Table 4. First, by looking at the t-statistics for the encompassing tests of Proposition 1 it appears that the random walk restrictions no longer dominate the no-arbitrage *ATSM* restrictions when considering a portfolio utility loss, but both forecasts are useful. As a result, the gains from combining these two models are high (depending on the assumption on the risk premium, they are respectively 18.82% and 14.24%).

Second, it appears that the choice between a constant or time varying risk premium can be important. The losses occurring when the forecasts are produced with the no-arbitrage restricted VAR are indeed much higher in the case of constant risk premium than in the case of time varying risk premium. This can be interpreted as evidence that keeping fixed the risk premium worsens the forecasts because it increases model misspecification. When the risk premium is time varying the optimal weights are not statistically different from the equal weights (see column 8), while when the risk premium is fixed the optimal weight decreases.

We now turn on the results for the local measure of usefulness, which are summarized in Figure

⁷It also follows that the forecast error made in forecasting the holding period return is proportional to that made in forecasting the yield of a given bond: $\hat{r}_{t+1}^{(\tau+1)} - r_{t+1}^{(\tau+1)} = -\tau(\hat{y}_{t+1}^{(\tau)} - \tau y_{t+1}^{(\tau)})$. This also implies that using the holding period return rather than the yields in the *quadratic* loss function would not change the optimal weight $\hat{\lambda}$ in that case.

2. The Figure is composed of four panels, each corresponding to a different forecast combination. Each panel contains a plot of the estimated smoothed weight, as defined in equation (24), together with the 95% bands described in Proposition 2.

Similarly to the case of a quadratic loss, we observe a clear pattern of decreasing usefulness of the no-arbitrage *ATSM* restrictions over time.

Another interesting result, which is in stark contrast with the quadratic loss case, is that assuming a constant risk premium clearly worsens the performance of the no-arbitrage forecasts. As is clear from the Figure, the weights are uniformly lower in panels B and D with respect to panels A and C. Moreover, as is clear in panels B and D, in the second part of the sample it is not possible to reject the null that the no-arbitrage VAR with constant risk premium is useless. This suggests that the incorporation of a time-varying risk premium in the no-arbitrage *ATSM* restrictions may not be important from a statistical point of view, but it is essential when evaluating the forecasts in terms of their usefulness for constructing bond portfolios.

The results displayed in Figure 2 also confirm the fact that, differently from the quadratic loss case, the random walk restrictions no longer dominate the no-arbitrage *ATSM* restrictions when considering a portfolio utility loss. Even though the usefulness of the no-arbitrage *ATSM* restrictions relative to a random walk has decreased over time, both restrictions appear to be useful, and the optimal forecast combination exploits information from both.

5.6 Additional results

In this subsection we address two additional issues. First we provide results for a 12-step-ahead forecast horizon. Second, we provide results for the case in which the optimal λ is chosen ex-ante within a pseudo-real-time forecasting exercise.

Table 5 displays results based on a 12-step-ahead forecast horizon for both the quadratic loss and the portfolio utility loss. The multi-step forecasts are obtained by iteration. Results are in line with those obtained for the 1-step-ahead case. For the quadratic loss imposing the no-arbitrage *ATSM* restrictions improves the forecasting performance of a VAR in the yields, but the restrictions are less useful when used in combination with a random walk forecast. For the portfolio utility loss the no-arbitrage *ATSM* restrictions are always useful, providing very large forecasting gains. For both loss functions the forecasts based on optimal weights largely outperform the simple equal combination weights forecast. The only major difference with respect to the 1-step-ahead case is that the assumption of constant term premium does not seem to play an important role in the case of the portfolio utility loss.

Table 6 displays results for a forecasting exercise in which rather than evaluating the usefulness of the restrictions we try to understand whether forecast combination using estimated optimal weights is a viable and efficient way to produce forecasts in real time. Note that to implement our scheme in real time two estimation windows are needed. The first window is needed to estimate the model's

coefficients. Then a second window is needed to estimate λ using past (from the point of view of the simulation) out of sample forecasts. Once λ is obtained, forecasts for the future can be computed. As a result the sample on which the figures displayed in Table 6 are based is different from that used for Tables 3, 4, and 5. For analogy with our analysis for the estimation of the smoothed λ we select the estimation window to be equal to $d = 108$ observations. This implies that the first forecast used for evaluation is that of the period 1983:2, and the estimated optimal weights used at time t in this exercise are the lagged values (i.e., the values at time $t - 1$) of the optimal weights depicted in Figure 1 and Figure 2. For both loss functions the *ATSM - VAR* works well when combined with an unrestricted VAR, while when combined with a random walk the no-arbitrage *ATSM* restrictions are useful only under the portfolio utility loss, and not under the quadratic loss (except at the short end of the curve). The results obtained with equal weights are similar to those obtained with the optimal weights for the quadratic loss, while they are worse for the utility loss. The latter result is not surprising, as Figure 2 shows that the optimal weights for the utility loss are statistically different from 0.5 in most of the sample.

6 Conclusions

In this paper we have developed a general framework for analyzing the usefulness of imposing parameter restrictions on a forecasting model. We have proposed a measure of usefulness based on the weight that a set of restrictions receives within an optimal forecast combination. Importantly, the proposed measure can vary over time and depends on the forecaster's loss function. We have shown how to estimate the measure of usefulness out-of-sample and perform inference about it, both in a stable framework and in a framework with possible instability.

We have applied our methodology to the problem of analyzing the usefulness of no-arbitrage *ATSM* restrictions for forecasting the term structure of interest rates.

Our results reveal that: 1) the restrictions have become less useful over time; 2) using a statistical measure of accuracy, the restrictions are a useful way to reduce parameter estimation uncertainty, but are dominated by restrictions that do the same without using any theory; 3) using an economic measure of accuracy, the no-arbitrage *ATSM* restrictions are no longer dominated by atheoretical restrictions, but for this to be true it is important that they incorporate a time-varying risk premium.

We want to stress the fact that these conclusions do not necessarily imply that bond markets have become more or less efficient over time, because first of all our method can only reveal time variation in the usefulness of the restrictions, but does not allow us to determine its source. Secondly, the results are conditional on the particular specification of no-arbitrage restrictions, and for the *ATSM* in particular this involves a number of additional assumptions that are not necessarily grounded in economic theory.

References

- [1] Almeida, C. and J. Vicente, 2008 The role of no-arbitrage on forecasting: Lessons from a parametric term structure model. *Journal of Banking & Finance* 32(12) 2695-2705.
- [2] Ang, A. and M. Piazzesi, 2003, A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables. *Journal of Monetary Economics* 50, 4, 745-787.
- [3] Carriero, A., 2007, Forecasting the Yield Curve using Priors from No Arbitrage Affine Term Structure Models. Queen Mary Working Paper no. 612.
- [4] Clark, T.E. and M. W. McCracken, 2007, Combining forecasts from nested models, *Oxford Bulletin of Economics and Statistics* 71(3), 303-329.
- [5] Clark, T. and M. McCracken, 2001, Tests of Equal Forecast Accuracy and Encompassing for Nested Models. *Journal of Econometrics* 105(1), 85-110.
- [6] Dai, Q. and K. Singleton, 2000, Specification analysis of affine term structure models. *Journal of Finance* 55, 1943–1978.
- [7] Del Negro, M. and F. Schorfheide, 2004, Priors from General Equilibrium Models for VARs. *International Economic Review* 45, 643-673
- [8] De Jong, F., 2000, Time series and cross section information in affine term structure models. *Journal of Business and Economic Statistics* 18, 300–314.
- [9] Diebold, F. X. and C. Li, 2006, Forecasting the term structure of government bond yields. *Journal of Econometrics* 130, 337–364
- [10] Diebold, F. X and R. Mariano, 1995, Comparing Predictive Accuracy. *Journal of Business & Economic Statistics* 13(3), 253-63.
- [11] Duffee, G., 2002, Term premia and interest rate forecasts in affine models. *Journal of Finance* 57, 405–443.
- [12] Duffie, D. and R. Kan, 1996, A yield-factor model of interest rates. *Mathematical Finance* 6, 379–406.
- [13] Elliott, G. and A. Timmermann, 2004, Optimal forecast combinations under general loss functions and forecast error distributions. *Journal of Econometrics* 122, 47-79
- [14] Favero, C.A., Niu L., and L. Sala, 2007, Term Structure Forecasting: No-Arbitrage Restrictions vs Large Information Set. CEPR Discussion Papers 6206
- [15] Giacomini, R. and B. Rossi, 2010, Forecast Comparisons in Unstable Environments. *Journal of Applied Econometrics* 25, 595-620.
- [16] Giacomini, R. and H. White, 2006, Tests of Conditional Predictive Ability. *Econometrica* 74, 1545-1578.
- [17] Harrison, J.M. and D.M. Kreps, 1979, Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory* 2, 381–408.
- [18] Harvey, D. I., S. J. Leybourne and P. Newbold, 1998, Tests for Forecast Encompassing. *Journal of Business & Economic Statistics* 16, 254-259.
- [19] Kim M.J. and C.R. Nelson, 1993, Predictable Stock Returns: The Role of Small Sample Bias. *The Journal of Finance* 48(2), 641-661.

- [20] Johansen S., 1995, Likelihood-based Inference in Cointegrated Vector Auto-regressive Models, Oxford University Press, Oxford.
- [21] Markowitz, H. 1952, Portfolio Selection. The Journal of Finance 7(1), 77-91.
- [22] Moench, E., 2008, Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented VAR approach. Journal of Econometrics 146(1) 26-43.
- [23] Newey, W. K. and D. McFadden, 1994, Large sample estimation and hypothesis testing, in: R.F. Engle and D.L. McFadden, (eds), Handbook of Econometrics, Vol. 4, North-Holland, Amsterdam, pp. 2112-2245.
- [24] Newey, W. and K. West, 1987, A Simple, Positive Semi-Definite, Heteroskedasticity and Auto-correlation Consistent Covariance Matrix. Econometrica 55, 703-708.
- [25] Palm, F.C. and A. Zellner, 1992, To Combine or not to Combine? Issues of Combining Forecasts. Journal of Forecasting 11, 687-701.
- [26] West, K. D., 1996, Asymptotic Inference about Predictive Ability. Econometrica 64, 1067-1084.
- [27] West, K. D., H. J. Edison, and D. Cho, 1993, A Utility-Based Comparison of Some Models of Exchange Rate Volatility. Journal of International Economics 35, 23-45.
- [28] White, H., 2001, Asymptotic theory for econometricians, Revised Edition, Academic Press, New York.
- [29] Wooldridge, J.M. and H. White, 1988, Some invariance principles and central limit theorems for dependent heterogeneous processes. Econometric Theory 4, 210-230.

Proofs

Proof of Proposition 1. We first show that, under Assumption A, $\widehat{\lambda}$ is asymptotically normal, so that $\sigma^{-1}\sqrt{n}(\widehat{\lambda} - \lambda^*) \rightarrow^d N(0, 1)$, where $\sigma^2 = H^{-1}\Omega H^{-1}$. The results in the Proposition then follow from showing consistency of $\widehat{\sigma}^2$ for σ^2 . Asymptotic normality of $\widehat{\lambda}$ is obtained by verifying the assumptions of Theorem 3.1 of Newey and McFadden (1994), since $\widehat{\lambda}$ can be viewed as an extremum estimator obtained by maximizing the objective function $-Q_n(\lambda)$ over \mathbb{R} . First, we show that assumptions (i)-(iii) of Theorem 2.7 of Newey and McFadden (1994) are satisfied, so that $\widehat{\lambda} \rightarrow^p \lambda^*$. Assumption (i) of Theorem 2.7 is equivalent to A(1). Assumption (ii) of Theorem 2.7 requires concavity of $-Q_n(\lambda)$, which is implied by A(2). Assumption (iii) requires that $Q_n(\lambda) - E[Q_n(\lambda)] \rightarrow^p 0$ for all λ . Since any measurable function of the finite history of y_t is mixing of the same size as y_t , $f_{t,h}^U$ and $f_{t,h}^R$ are mixing of the same size as y_t , because they are functions of a window of in-sample data m that is finite by A(9). This implies that $L(x_{t+h}, f_{t,h}^R + (1-\lambda)(f_{t,h}^U - f_{t,h}^R))$ is also mixing with ϕ of size $-r/(2r-1)$ or α of size $-r/(r-1)$, which, together with A(4), implies that the conditions of Corollary 3.48 of White (2001) are satisfied and thus $Q_n(\lambda) - E[Q_n(\lambda)] \rightarrow^p 0$ for all λ . We next verify conditions (i)-(v) of Theorem 3.1 of Newey and McFadden (1994). Conditions (i) and (ii) are implied by A(1) and A(2). Condition (iii) requires that $\Omega^{-1/2}\sqrt{n}\nabla_\lambda Q_n(\lambda^*) \rightarrow^d N(0, 1)$. Ω is finite by A(5) and it is positive by A(6). By arguments similar to those used above, one can show that A(3) implies that $Z_t \equiv \Omega^{-1/2}\nabla_\lambda L(x_{t+h}, f_{t,h}^R + (1-\lambda^*)(f_{t,h}^U - f_{t,h}^R))$ is mixing with ϕ of size $-r/(2r-2)$ or α of size $-r/(r-2)$. This, together with A(5), implies that the sequence $\{Z_t\}$ satisfies the conditions of Corollary 3.1 of Wooldridge and White's (1988), and thus condition (iii) is satisfied. Conditions (iv) and (v) of Newey and McFadden (1994) coincide with A(7) and A(8). Finally, A(3), A(5) and A(9) imply that the conditions of Theorem 6.20 of White (2001) are satisfied and thus $\widehat{\Omega}$ is a consistent estimator of Ω . This, in turn, implies that the conditions of Theorem 4.1 of Newey and McFadden (1994) are satisfied and thus $\widehat{\sigma}^2 \rightarrow^p \sigma^2$, which completes the proof. ■

Proof of Proposition 2. Let $L_{t+h}(\lambda^*) = L(x_{t+h}, f_{t,h}^R + (1-\lambda^*)(f_{t,h}^U - f_{t,h}^R))$ and, for ease of notation, henceforth drop the subscript d from $\widehat{\lambda}_{t,d}$. For $t = m + d - 1, \dots, T - h$ we have

$$\begin{aligned} \widehat{\sigma}^{-1}\sqrt{d}(\widehat{\lambda}_t - \lambda^*) &= \widehat{H}^{-1}(\bar{\lambda})\widehat{\sigma}^{-1}d^{-1/2} \sum_{j=t-d+1}^t \nabla_\lambda L_{j+h}(\lambda^*) \\ &= (d/n)^{-1/2} \widehat{H}^{-1}(\bar{\lambda})\widehat{\sigma}^{-1}\Omega^{1/2} \left(\Omega^{-1/2}n^{-1/2} \sum_{j=m}^t \nabla_\lambda L_{j+h}(\lambda^*) \right. \\ &\quad \left. - \Omega^{-1/2}n^{-1/2} \sum_{j=m}^{t-d} \nabla_\lambda L_{j+h}(\lambda^*) \right) \end{aligned}$$

where $\bar{\lambda}$ lies between $\widehat{\lambda}_t$ and λ^* . By B(1), we have

$$\Omega^{-1/2}n^{-1/2} \left(\sum_{j=m}^t \nabla_\lambda L_{j+h}(\lambda^*) - \sum_{j=m}^{t-d} \nabla_\lambda L_{j+h}(\lambda^*) \right) \implies [\mathcal{B}(\tau) - \mathcal{B}(\tau - \pi)],$$

where \mathcal{B} is a standard univariate Brownian motion. By B(2), $(d/n)^{-1/2} \rightarrow \pi^{-1/2}$. By B(3), $\widehat{H}^{-1}(\bar{\lambda})\widehat{\sigma}^{-1}\Omega^{1/2} \rightarrow^p H^{-1}\sigma^{-1}\Omega^{1/2} = 1$ and thus $\widehat{\sigma}^{-1}\sqrt{d}(\widehat{\lambda}_t - \lambda^*) \implies [\mathcal{B}(\tau) - \mathcal{B}(\tau - \pi)]/\sqrt{\pi}$ under the null hypothesis. Let $k_{\alpha,\pi}$ solve $\Pr\{\sup_\tau |[\mathcal{B}(\tau) - \mathcal{B}(\tau - \pi)]/\sqrt{\pi}| > k_{\alpha,\pi}\} = \alpha$. Then, under either H_0^U or H_0^R , $(1-\alpha)\%$ of the time λ^* is contained within $(\widehat{\lambda}_t - k_{\alpha,\pi}\frac{\widehat{\sigma}}{\sqrt{d}}, \widehat{\lambda}_t + k_{\alpha,\pi}\frac{\widehat{\sigma}}{\sqrt{d}})$ for all $t = m + d - 1, \dots, T - h$. The values of $k_{\alpha,\pi}$ in Table 1 are obtained by Monte Carlo simulation. ■

Tables and Figures

Table 1. Critical values $k_{\alpha,\pi}$ for the confidence bands in Proposition 2

α	$\pi = d/n$								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.05	3.393	3.179	3.012	2.890	2.779	2.634	2.560	2.433	2.248
0.10	3.170	2.948	2.766	2.626	2.500	2.356	2.252	2.130	1.950

Table 2: Empirical Size of the tests in (5).**Panel A: Case $\beta_1 = 1$ ($\lambda^* = 0$)**

Percentage of rejections of $H_0 : \lambda = 0$							
$m \setminus n \rightarrow$	50	100	150	250	500	1000	
↓ 50		0.070	0.068	0.053	0.056	0.054	0.052
100		0.067	0.059	0.056	0.053	0.051	0.051
150		0.074	0.064	0.056	0.052	0.047	0.048
250		0.073	0.060	0.058	0.052	0.049	0.048

Panel B: Case $\beta_1 = 0$ ($\lambda^* = 1$)

Percentage of rejections of $H_0 : \lambda = 1$							
$m \setminus n \rightarrow$	50	100	150	250	500	1000	
↓ 50		0.068	0.069	0.052	0.058	0.053	0.052
100		0.067	0.057	0.059	0.053	0.051	0.049
150		0.074	0.064	0.057	0.052	0.046	0.047
250		0.070	0.060	0.058	0.052	0.049	0.051

The table contains the percentage of rejections for the tests in (5) with a nominal size of 0.05

Table 3. Results with Quadratic Loss.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: ATSM-VAR and UVAR								
	RMSFE	%gain						
Yields	(f_t^U)	(f_t^{ATSM})	$(f_t^{\lambda=1/2})$	(f_t^*)	$\hat{\lambda}$	$t^{\lambda=1}$	$t^{\lambda=0}$	$t^{\lambda=1/2}$
1-month	0.808	0.51	4.84	4.85	0.514	-4.301***	4.553***	0.126
3-month	0.709	6.42	6.62	7.41	0.736	-1.486	4.147***	1.331
12-month	0.682	11.93	8.50	11.96	1.054	-0.231	4.525***	2.378**
36-month	0.526	8.19	6.78	8.37	0.875	-0.619	4.331***	1.856*
60-month	0.460	6.70	5.79	6.97	0.840	-0.872	4.572***	1.850*
Panel B: CRP-VAR and UVAR								
	RMSFE	%gain						
Yields	(f_t^U)	(f_t^{CRP})	$(f_t^{\lambda=1/2})$	(f_t^*)	$\hat{\lambda}$	$t^{\lambda=1}$	$t^{\lambda=0}$	$t^{\lambda=1/2}$
1-month	0.808	-13.91	1.04	2.70	0.280	-8.833***	3.444***	-2.695***
3-month	0.709	6.20	6.83	7.51	0.709	-1.792*	4.367***	1.288
12-month	0.682	9.46	8.97	10.35	0.777	-1.071	3.473***	1.336
36-month	0.526	9.22	7.50	9.37	0.889	-0.567	4.545***	1.989**
60-month	0.460	6.05	6.18	6.93	0.740	-1.753*	4.995***	1.621
Panel C: ATSM-VAR and RW								
	RMSFE	%gain						
Yields	(f_t^{RW})	(f_t^{ATSM})	$(f_t^{\lambda=1/2})$	(f_t^*)	$\hat{\lambda}$	$t^{\lambda=1}$	$t^{\lambda=0}$	$t^{\lambda=1/2}$
1-month	0.837	3.86	5.73	5.99	0.629	-2.695***	4.562***	0.933
3-month	0.607	-9.33	-3.06	0.25	-0.186	-5.528***	-0.870	-3.199***
12-month	0.596	-0.88	-0.09	0.05	0.188	-2.135**	0.496	-0.820
36-month	0.473	-2.13	-0.35	0.05	0.128	-3.649***	0.536	-1.557
60-month	0.424	-1.11	-0.002	0.14	0.249	-2.530***	0.839	-0.845
Panel D: CRP-VAR and RW								
	RMSFE	%gain						
Yields	(f_t^{RW})	(f_t^{CRP})	$(f_t^{\lambda=1/2})$	(f_t^*)	$\hat{\lambda}$	$t^{\lambda=1}$	$t^{\lambda=0}$	$t^{\lambda=1/2}$
1-month	0.837	-10.08	4.00	4.80	0.356	-7.607***	4.212***	-1.697**
3-month	0.607	-9.59	-1.94	0.10	0.090	-8.043***	0.800	-3.621***
12-month	0.596	-3.71	-0.42	0.81	0.295	-2.945***	1.234	-0.855
36-month	0.473	-0.99	0.55	0.61	0.380	-2.645***	1.626	-0.509
60-month	0.424	-1.83	-0.27	0.06	0.148	-3.948***	0.685	-1.632

Each of the four panels displays the results from the combination of two models. The considered combinations are: VAR with no-arbitrage ATSM restrictions (f_t^{ATSM}) and unrestricted VAR (f_t^U) (panel A), VAR with no-arbitrage ATSM restrictions and constant risk premium (f_t^{CRP}) with unrestricted VAR (f_t^U) (panel B), VAR with no-arbitrage ATSM restrictions (f_t^{ATSM}) and random walk (f_t^{RW}) (panel C), VAR with no-arbitrage ATSM restrictions and constant risk premium (f_t^{CRP}) and random walk (f_t^{RW}) (panel D).

For each panel, results are reported for each of the five yields at hand. Column (1) contains the Root Mean Squared Forecast Error (RMSFE) (i.e. the realized loss) of the first model considered in the combination. Columns (2) to (4) report the percentage gains over the RMSFE of the model in column (1) obtained by using, respectively, the second model in the combination, a combination with equal weights ($f_t^{\lambda=1/2}$), and a combination based on the estimated optimal weight (f_t^*). Column (5) reports the value of the estimated optimal weight $\hat{\lambda}$ (as defined in equation (12)). Finally, columns (6) to (8) report the statistics for the encompassing tests of Proposition (1). The stars *, **, ***, indicate rejection of the null at 10%, 5%, and 1% level. The statistic $t^{\lambda=1}$ is used to test the null that the unrestricted forecast is useless. The statistic $t^{\lambda=0}$ is used to test the null that the restricted forecast is useless. The statistic $t^{\lambda=1/2}$ is used to test the null that the optimal weight is 0.5.

Table 4. Results with Portfolio Utility Loss.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: ATSM-VAR and UVAR							
Utility Loss	%gain			$\hat{\lambda}$	$t^{\lambda=1}$	$t^{\lambda=0}$	$t^{\lambda=1/2}$
(f_t^U)	(f_t^{ATSM})	$(f_t^{\lambda=1/2})$	(f_t^*)				
-1.35	7.96	23.55	23.75	0.551	-6.082***	7.460***	0.689
Panel B: CRP-VAR and UVAR							
Utility Loss	%gain			$\hat{\lambda}$	$t^{\lambda=1}$	$t^{\lambda=0}$	$t^{\lambda=1/2}$
(f_t^U)	(f_t^{CRP})	$(f_t^{\lambda=1/2})$	(f_t^*)				
-1.35	-66.12	9.09	15.57	0.304	-13.527***	5.906***	-3.811***
Panel C. ATSM-VAR and RW							
Utility Loss	%gain			$\hat{\lambda}$	$t^{\lambda=1}$	$t^{\lambda=0}$	$t^{\lambda=1/2}$
(f_t^{RW})	(f_t^{ATSM})	$(f_t^{\lambda=1/2})$	(f_t^*)				
-1.60	-8.66	18.61	18.82	0.453	-9.267***	7.668***	-0.799
Panel D. CRP-VAR and RW							
Utility Loss	%gain			$\hat{\lambda}$	$t^{\lambda=1}$	$t^{\lambda=0}$	$t^{\lambda=1/2}$
(f_t^{RW})	(f_t^{CRP})	$(f_t^{\lambda=1/2})$	(f_t^*)				
-1.60	-71.34	6.74	14.24	0.289	-17.48***	7.131***	-5.176***

Each of the four panels displays the results from the combination of two models. The considered combinations are: VAR with no-arbitrage ATSM restrictions (f_t^{ATSM}) and unrestricted VAR (f_t^U) (panel A), VAR with no-arbitrage ATSM restrictions and constant risk premium (f_t^{CRP}) with unrestricted VAR (f_t^U) (panel B), VAR with no-arbitrage ATSM restrictions (f_t^{ATSM}) and random walk (f_t^{RW}) (panel C), VAR with no-arbitrage ATSM restrictions and constant risk premium (f_t^{CRP}) and random walk (f_t^{RW}) (panel D).

Column (1) contains the portfolio utility loss of the first model considered in the combination. Columns (2) to (4) report the percentage gains over the utility loss of the model in column (1) obtained by using, respectively, the second model in the combination, a combination with equal weights ($f_t^{\lambda=1/2}$), and a combination based on the estimated optimal weight (f_t^*). Column (5) reports the value of the estimated optimal weight $\hat{\lambda}$ (as defined in equation (21)). Finally, columns (6) to (8) report the statistics for the encompassing tests of Proposition (1). The stars *, **, ***, indicate rejection of the null at 10%, 5%, and 1% level. The statistic $t^{\lambda=1}$ is used to test the null that the unrestricted forecast is useless. The statistic $t^{\lambda=0}$ is used to test the null that the restricted forecast is useless. The statistic $t^{\lambda=1/2}$ is used to test the null that the optimal weight is 0.5.

Table 5. Quadratic and Portfolio Utility Loss: 12-step-ahead results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: ATSM-VAR and UVAR								
	Loss	%gain						
	(f_t^U)	(f_t^{ATSM})	$(f_t^{\lambda=1/2})$	(f_t^*)	$\hat{\lambda}$	$t^{\lambda=1}$	$t^{\lambda=0}$	$t^{\lambda=1/2}$
1-month	1.632	41.37	28.75	41.37	0.999	-0.007	18.728***	9.361***
3-month	1.711	50.29	34.14	50.29	0.996	-0.145	35.379***	17.617***
12-month	1.570	51.51	33.50	51.67	1.046	0.952	21.834***	11.393***
36-month	1.242	46.39	32.88	46.50	0.961	-0.474	11.677***	5.601***
60-month	1.119	41.02	31.03	41.55	0.911	-0.923	9.471***	4.274***
Portfolio	4.258	61.88	60.57	68.65	0.761	-4.137***	13.174***	4.518***
Panel B: CRP-VAR and UVAR								
	Loss	%gain						
	(f_t^U)	(f_t^{CRP})	$(f_t^{\lambda=1/2})$	(f_t^*)	$\hat{\lambda}$	$t^{\lambda=1}$	$t^{\lambda=0}$	$t^{\lambda=1/2}$
1-month	1.632	20.77	21.63	24.45	0.733	-2.715***	7.455***	2.370***
3-month	1.711	37.29	30.29	38.62	0.860	-2.668***	16.406***	6.869***
12-month	1.571	42.60	31.98	43.11	0.915	-1.551	16.653***	7.542***
36-month	1.242	51.37	36.03	51.53	0.956	-0.966	21.234***	10.134***
60-month	1.119	56.28	39.21	56.57	0.947	-1.447	26.029***	12.291***
Portfolio	4.258	87.90	83.31	95.58	0.779	-3.710***	13.089***	4.670***
Panel C: ATSM-VAR and RW								
	Loss	%gain						
	(f_t^{RW})	(f_t^{ATSM})	$(f_t^{\lambda=1/2})$	(f_t^*)	$\hat{\lambda}$	$t^{\lambda=1}$	$t^{\lambda=0}$	$t^{\lambda=1/2}$
1-month	0.837	-14.38	1.27	2.96	0.285	-6.480***	2.585***	-1.947***
3-month	0.607	-40.09	-13.24	0.44	-0.104	-14.538***	-1.376	-7.957***
12-month	0.596	-27.87	-8.92	0.29	-0.105	-12.260***	-1.168	-6.714***
36-month	0.473	-40.80	-13.03	0.24	-0.075	-15.014***	-1.047	-8.031***
60-month	0.424	-55.50	-17.60	0.13	-0.044	-23.389***	-0.983*	-12.186***
Portfolio	-1.601	-201.36	-26.67	7.56	0.160	-37.034***	7.048***	-14.993***
Panel D: CRP-VAR and RW								
	Loss	%gain						
	(f_t^{RW})	(f_t^{CRP})	$(f_t^{\lambda=1/2})$	(f_t^*)	$\hat{\lambda}$	$t^{\lambda=1}$	$t^{\lambda=0}$	$t^{\lambda=1/2}$
1-month	0.837	-54.55	-8.57	2.79	0.163	-15.086***	2.944***	-6.071***
3-month	0.607	-76.70	-21.15	0.33	0.053	-34.100***	1.904**	-16.098***
12-month	0.596	-51.39	-12.97	0.30	0.063	-17.912***	1.275	-8.811***
36-month	0.473	-27.73	-6.93	0.06	0.042	-13.531***	0.590	-6.471***
60-month	0.424	-15.24	-4.14	0.003	-0.015	-10.685***	-0.157	-5.421***
Portfolio	-1.601	-132.15	-5.12	12.78	0.229	-18.337***	5.446***	-6.445***

The table summarizes the results for the 12-step ahead forecast horizon, using the quadratic utility loss and the portfolio utility loss. Each panel considers a different forecast combination. For each panel, the first 5 rows display the results for the quadratic loss. Therefore, the first five entries of column (1) in each panel are the RMSFE in forecasting the corresponding yield 12-step-ahead. The last row (Portfolio) displays results for the portfolio utility loss. Therefore, the last entry in column (1) is the utility loss associated with the portfolio. Columns (2)-(8) have the same structure as in Tables 2 and 3. In particular columns (2) to (4) report the percentage gains over the RMSFE or Utility Loss of the model in column (1) obtained by using, respectively, the second model in the combination, a combination with equal weights ($f_t^{\lambda=1/2}$), and a combination based on the estimated optimal weight (f_t^*). Column (5) reports the value of the estimated optimal weight $\hat{\lambda}$ (as defined in equation (12) for the quadratic loss and in equation (21) for the utility loss). Finally, columns (6) to (8) report the statistics for the encompassing tests of Proposition (1). The stars *, **, ***, indicate rejection of the null at 10%, 5%, and 1% level. The statistic $t^{\lambda=1}$ is used to test the null that the unrestricted forecast is useless. The statistic $t^{\lambda=0}$ is used to test the null that the restricted forecast is useless. The statistic $t^{\lambda=1/2}$ is used to test the null that the optimal weight is 0.5. The used bandwidth is $2(h-1)$.

Table 6: Pseudo-real-time simulation

Panel A: ATSM-VAR and UVAR				
	Loss	%gain		
	(f_t^U)	(f_t^{ATSM})	$(f_t^{\lambda=1/2})$	(f_t^*)
1-month	0.561	1.66	6.05	5.55
3-month	0.341	11.63	12.87	13.47
12-month	0.376	7.31	7.04	7.73
36-month	0.380	3.78	4.77	4.42
60-month	0.371	2.75	3.69	3.20
Portfolio	-1.218	-37.19	-4.17	1.00
Panel B: CRP-VAR and UVAR				
	Loss	%gain		
	(f_t^U)	(f_t^{CRP})	$(f_t^{\lambda=1/2})$	(f_t^*)
1-month	0.561	-30.71	-3.58	1.02
3-month	0.341	-3.66	7.20	8.56
12-month	0.376	-1.54	3.91	2.21
36-month	0.380	4.06	4.67	4.20
60-month	0.371	3.57	5.32	5.39
Portfolio	-1.218	-167.69	-34.70	-0.60
Panel C: ATSM-VAR and RW				
	Loss	%gain		
	(f_t^{RW})	(f_t^{ATSM})	$(f_t^{\lambda=1/2})$	(f_t^*)
1-month	0.594	7.14	11.38	11.03
3-month	0.294	-2.31	0.46	1.30
12-month	0.341	-1.99	-0.50	-0.66
36-month	0.352	-3.94	-0.88	-0.25
60-month	0.349	-3.62	-1.04	-1.11
Portfolio	-1.231	-38.07	5.18	8.20
Panel D: CRP-VAR and RW				
	Loss	%gain		
	(f_t^{RW})	(f_t^{CRP})	$(f_t^{\lambda=1/2})$	(f_t^*)
1-month	0.594	-23.42	6.26	8.13
3-month	0.294	-20.02	0.16	-0.22
12-month	0.341	-11.73	-2.69	-1.97
36-month	0.352	-3.64	-0.55	-1.48
60-month	0.349	-2.75	-0.06	-1.50
Portfolio	-1.231	-166.99	-22.17	5.02

The table summarizes the results for the pseudo-real-time implementation of our forecasting exercise, at 1-step-ahead horizon. The sample used for forecast evaluation is 1983:2 to 2003:12, and it differs from the evaluation sample used in Table 3, Table 4 and Table 5. This happens because at each point in time we need a training sample for which past out of sample forecasts are available in order to estimate the optimal λ . For analogy with the computation of the smoothed λ we use a training sample of 108 observations. Each panel considers a different forecast combination. For each panel, the first 5 rows display the results for the quadratic loss. Therefore, the first five entries of column (1) in each panel are the RMSFE in forecasting the corresponding yield 12-step-ahead. The last row (Portfolio) displays results for the portfolio utility loss. Therefore, the last entry in column (1) is the utility loss associated with the portfolio. Columns (2)-(8) have the same structure as in Tables 2 and 3. In particular columns (2) to (4) report the percentage gains over the RMSFE or utility loss of the model in column (1) obtained by using, respectively, the second model in the combination, a combination with equal weights ($f_t^{\lambda=1/2}$), and a combination based on the estimated optimal weight (f_t^*).

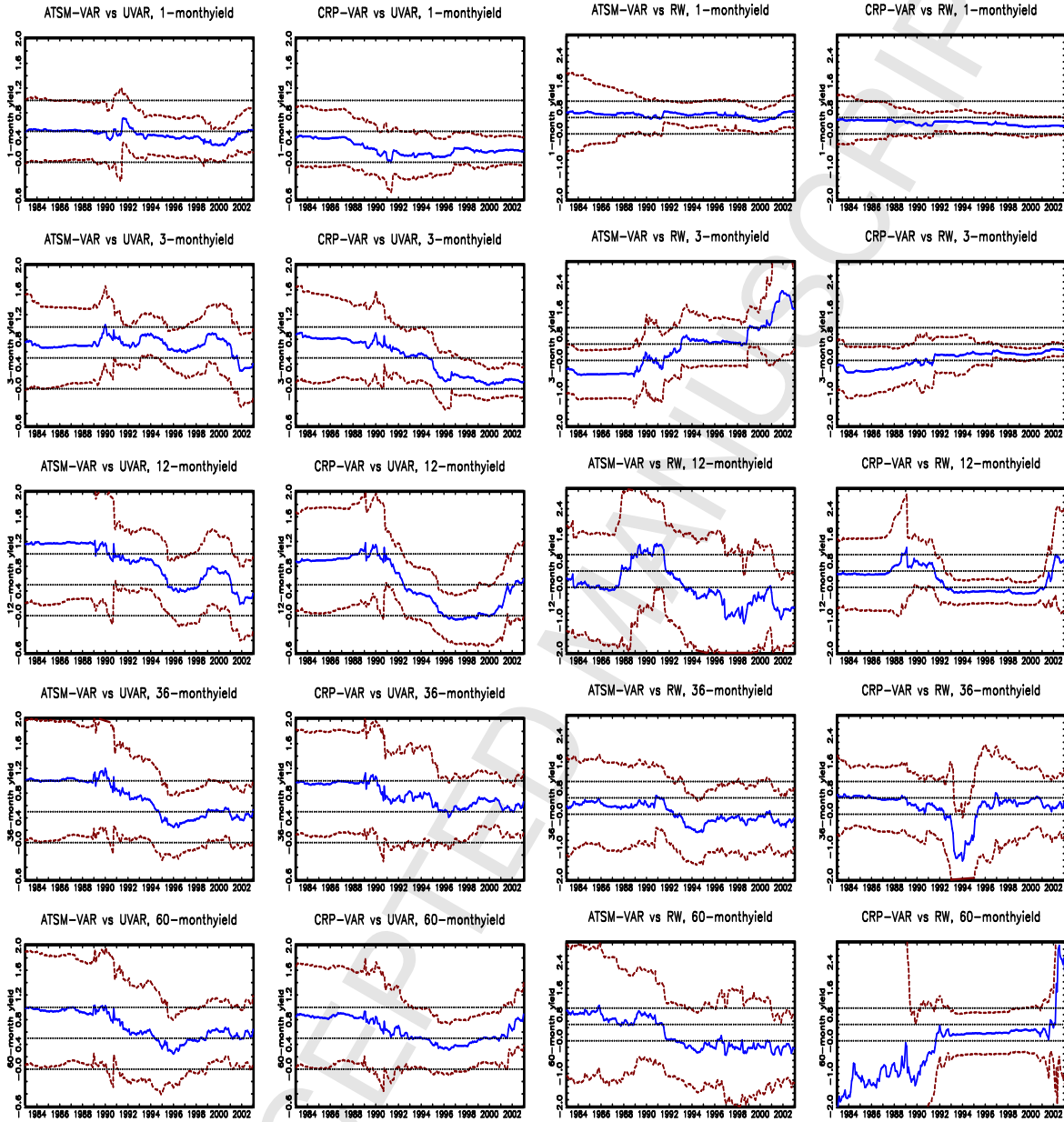


Figure 1: Results for Quadratic Loss. The blue solid line is the estimated optimal weight $\hat{\lambda}_t$ in equation (15). The red dashed lines are the 95% bands in equation (10).

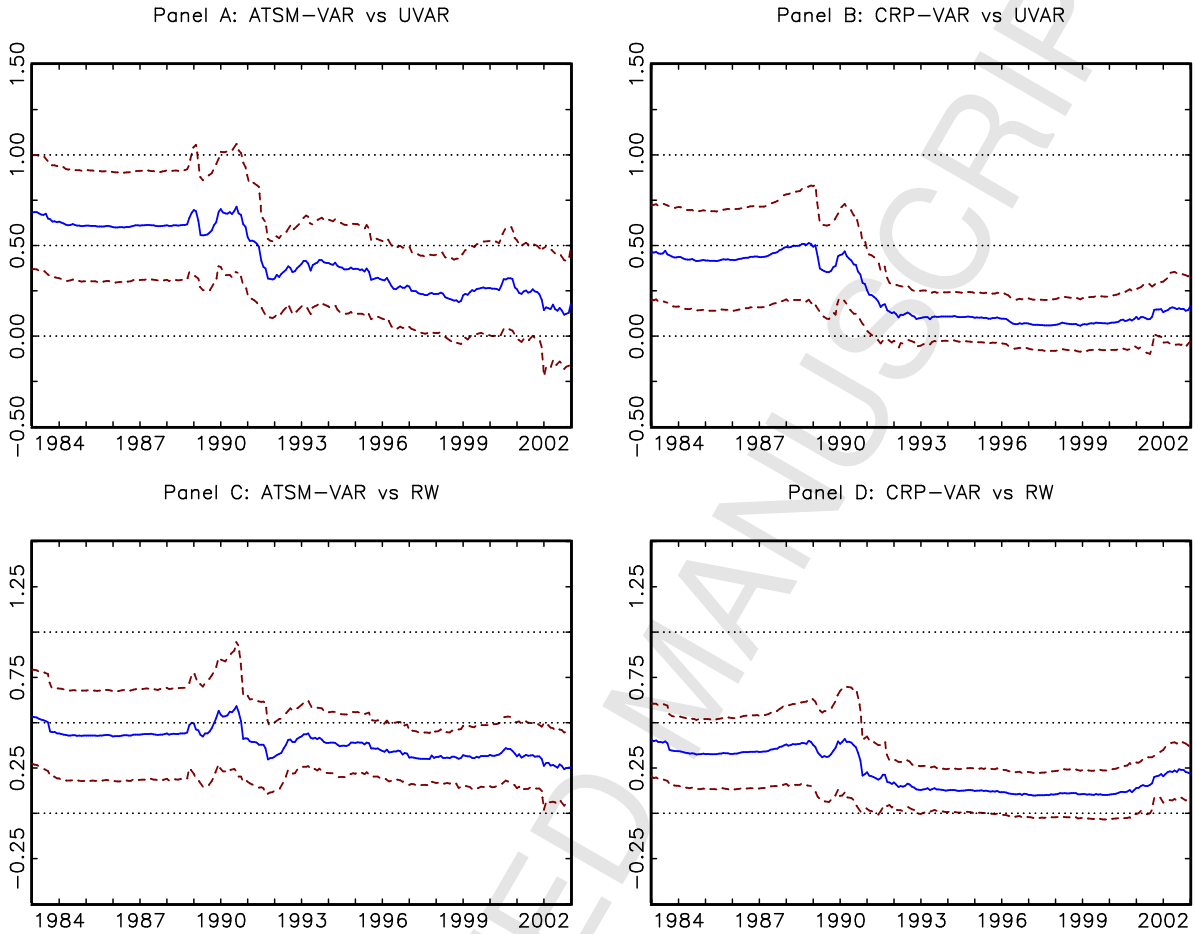


Figure 2: Results for Portfolio Utility Loss. The blue solid line is the estimated optimal weight $\hat{\lambda}_t$ in equation (24). The red dashed lines are the 95% bands in equation (10).