

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

Andrea Carriero and Raffaella Giacomini

Queen Mary University of London and UCL

January 2009

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 such 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) 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 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; Shrinkage; Loss functions; Instability

Acknowledgments: We would like to thank Caio Almeida, Gianni Amisano, Albert Chun and seminar participants at the 2008 Forecasting in Rio conference for useful comments and discussions.

J.E.L. Codes: C22, C52, C53

1 Introduction

In the last 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 very 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 rather poor performance of such models. Duffee (2002) has shown 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.

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 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 that shrinks the unrestricted forecast towards the restricted forecast and estimate the optimal amount of shrinkage in an out-of-sample framework. We say that the restriction is "useful for forecasting" when the optimal amount of shrinkage towards the restricted model is large, and we can formally test the hypothesis that the restrictions are useless by an out-of-sample encompassing test. We then generalize the techniques to an environment where the usefulness of the restrictions is possibly time-varying, and provide a test to assess whether that is the case.

Optimality 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. This 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 stress that the 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 (2007) 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 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, which is the same asymptotic framework used by Giacomini and White (2006) in the different context of predictive ability testing.

Regarding the relationship between our approach and the out-of-sample predictive ability testing literature (e.g., West, 1996, Clark and McCracken, 2001, Giacomini and White, 2006), we first note that most of the existing literature focuses on comparing the performance of different models, and cannot therefore be applied to situations in which the forecasts are based on one model but are produced using two different estimators of its parameters. Exceptions are the test proposed by Giacomini and White (2006), and its extension to unstable environments (Giacomini and Rossi (in press)), which could in principle be used to test whether the restricted and unrestricted forecasts perform equally well, and then select the most accurate forecast. The difference between these approaches and our approach is akin to the difference between a pre-test estimator and a shrinkage estimator. The added advantage of our method is therefore that it does not force the user to choose either forecast, instead providing an "intermediate" forecast that optimally exploits the information contained in the economic restrictions.

Note that our problem is also different from standard hypothesis testing, since we allow for the possibility that the restrictions are not true, but are "close to the truth", in a way that makes them useful for forecasting. Note also that our definition of usefulness is loss-specific, which tailors the evaluation problem to the forecaster's decision problem, and may thus lead to different conclusions for different loss functions.

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 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 and "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 $L(x_{t+h}, f_{t,h})$. Here we consider 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 is a scalar process, which 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 (20) 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 convex 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 can be interpreted as the degree of "shrinkage" towards the restricted forecast. We consider convex combinations both because the results are more easily interpretable when restricting attention to a scalar measure λ , and in order to establish an analogy with James-Stein type estimators (the only difference is that our weight is non-random, whereas in a James-Stein estimator λ would be substituted by the random weight $\lambda / (f_{t,h}^U - f_{t,h}^R)^2$).

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

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

and is estimated by

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

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 4 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) $L\left(x_{t+h}, f_{t,h}^R + (1-\lambda)(f_{t,h}^U - f_{t,h}^R)\right)$ is convex and twice continuously differentiable with respect to λ ;
- (3) $\{y_t\}$ is mixing with ϕ of size $-r/(r-1)$ or α of size $-2r/(r-2)$, $r > 2$;
- (4) $E|L\left(x_{t+h}, f_{t,h}^R + (1-\lambda)(f_{t,h}^U - f_{t,h}^R)\right)|^{r/2} < \infty$ for all t and all λ ;
- (5) $E|\nabla_\lambda L\left(x_{t+h}, f_{t,h}^R + (1-\lambda^*)(f_{t,h}^U - f_{t,h}^R)\right)|^{2r} < \infty$ for all t , where ∇_λ indicates the first derivative with respect to λ ;
- (6) $\Omega = E\left[(\sqrt{n}\nabla_\lambda Q_n(\lambda^*))^2\right] > 0$ 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 4, 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 4. 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 4, 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*

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

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

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

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. Even though this serial correlation may be present for all forecast horizons, in analogy with the predictive ability testing literature (e.g., Diebold and Mariano, 1995), we recommend setting the bandwidth to $h - 1$ in practical applications, a choice that may improve the finite-sample properties of the test.

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

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

The solution to (7) in general cannot be estimated consistently if the expectation varies arbitrarily over time. Similarly to Giacomini and Rossi (in press), we overcome this problem by making our object of interest a "smoothed" version of λ_t^* , obtained by computing the expectations over rolling windows of size d :

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

which can be consistently estimated by solving the problem

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

A plot of the sample path of $\{\hat{\lambda}_{t,d}\}_{t=m+d-1}^{T-h}$ 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. 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) $\left\{ n^{-1/2} \sum_{j=m}^{m+[\tau n]} \nabla_{\lambda} L \left(x_{j+h}, f_{j,h}^R + (1 - \lambda^*)(f_{j,h}^U - f_{j,h}^R) \right) \right\}$ obeys a Functional Central Limit Theorem with $\Omega = \lim_{n \rightarrow \infty} E(n^{-1/2} \sum_{j=m}^{T-h} \nabla_{\lambda} L \left(x_{j+h}, f_{j,h}^R + (1 - \lambda^*)(f_{j,h}^U - f_{j,h}^R) \right))^2 > 0$;
- (2) $d/n \rightarrow \pi \in (0, \infty)$ as $d \rightarrow \infty, n \rightarrow \infty$. $m < \infty$ and $h < \infty$;
- (3) $\hat{\sigma} \xrightarrow{P} \sigma$ and $\hat{\lambda} \xrightarrow{P} \lambda^*$.

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:*

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

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 A simple illustrative example

To gain intuition about the determinants of the usefulness of economic restrictions for forecasting, consider the simple situation of a linear model with k regressors and iid data:

$$\begin{aligned} x_t &= \beta' z_t + \varepsilon_t; \\ z_t &\sim iid(0, \sigma_z^2 I_k), \varepsilon_t \sim iid(0, \sigma_\varepsilon^2). \end{aligned} \tag{11}$$

The unrestricted one-step-ahead forecast is $f_t^U = \hat{\beta}' z_{t+1}$, with $\hat{\beta}$ the OLS estimator, whereas the restricted forecast is $f_t^R = \tilde{\beta}' z_{t+1}$, with $\tilde{\beta}$ some estimator obtained by estimating the model subject to a set of (possibly non-linear) restrictions. For example, $\tilde{\beta}$ could be obtained by imposing zero restrictions on the coefficients of some components of z_t and estimating the remaining coefficients by OLS, which would correspond to the combination of forecasts from nested models considered by Clark and McCracken (2007) (who consider a different data-generating process and asymptotic environment).

The optimal weight (13) for a quadratic loss function is in this case $\lambda^* = \frac{E[(\beta - \hat{\beta})' z_{t+1} \cdot (\tilde{\beta} - \hat{\beta})' z_{t+1}]}{E[(\tilde{\beta} - \hat{\beta})' z_{t+1}]^2}$, which can be shown to be equivalent to

$$\lambda^* = \frac{tr(Var(\hat{\beta}) - cov(\hat{\beta}, \tilde{\beta}))}{tr\left(Var(\hat{\beta} - \tilde{\beta}) + (bias(\tilde{\beta}))^2\right)}. \tag{12}$$

Expression (12) corresponds to the optimal shrinkage when shrinking $\hat{\beta}$ towards $\tilde{\beta}$ using a mean squared error loss function, as derived by Kim and White (2000). That is, in this simple example our problem can be reframed as the problem of combining different estimators for β .

Expression (12) can be further simplified when the two estimators are independent, in which case $\lambda^* = \frac{tr(Var(\hat{\beta}))}{tr(Var(\hat{\beta}) + Var(\tilde{\beta}) + (bias(\tilde{\beta}))^2)}$, from which we see that a value of λ^* close to 0 ($\approx$ the restrictions are not useful) can be due to the restricted estimator having large bias, suggesting that the imposed restrictions are too far from the truth to be usefully exploited. Conversely, when the variance of the unrestricted estimator is much larger than the mean squared error of the restricted estimator, λ^* will be close to 1, indicating that the restrictions are useful for forecasting. Intermediate values of λ^* correspond to situations in which there is a bias-variance tradeoff that can be optimally exploited to obtain a combined forecast with that is more accurate than both original forecasts.

When the estimators are correlated, as will often be the case, one can use the intuition in Hausman (1978) to understand in which circumstances the optimal weight λ^* can equal 0 or 1. Suppose for example that both estimators are unbiased. The arguments in Hausman (1978) can then be used to show that, if the restricted estimator is efficient, then $Var(\hat{\beta} - \tilde{\beta}) = Var(\hat{\beta}) - Var(\tilde{\beta})$ and

$cov(\hat{\beta}, \tilde{\beta}) = Var(\tilde{\beta})$, so that expression (12) collapses to $\lambda^* = 1$. Similarly, when both estimators are unbiased and the unrestricted estimator is efficient, it can easily be shown that $\lambda^* = 0$.

Expression (12) can also provide some intuition on the possible reasons for observing time-variation in the usefulness of the economic restrictions. For simplicity, consider again the case of independent estimators and note that a reduction in λ^* could be due to a reduction in the variance of the unrestricted estimator (e.g., due to a reduction in the variance of the errors σ_ε^2) or to an increase in the bias of the restricted estimator, which would imply that the economic restrictions have become more misspecified over time. We stress that our method does not allow one to disentangle the possible causes of time variation in the usefulness of the economic restrictions, nor is it our objective to answer this question. Rather, the question of interest here is whether certain economic restrictions can be usefully imposed to improve forecast accuracy for a particular loss function, and whether the usefulness of the restrictions has changed over time.

4 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.

4.1 Quadratic loss

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

$$\lambda^* = \frac{E \left[\sum_{t=m}^{T-h} (x_{t+h} - f_{t,h}^U) (f_{t,h}^R - f_{t,h}^U) \right]}{E \left[\sum_{t=m}^{T-h} (f_{t,h}^R - f_{t,h}^U)^2 \right]}. \quad (13)$$

A consistent estimator of λ^* is

$$\hat{\lambda} = \frac{\sum_{t=m}^{T-h} (x_{t+h} - f_{t,h}^U) (f_{t,h}^R - f_{t,h}^U)}{\sum_{t=m}^{T-h} (f_{t,h}^R - f_{t,h}^U)^2}, \quad (14)$$

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

$$x_{t+h} - f_{t,h}^U = \lambda(f_{t,h}^R - f_{t,h}^U) + \varepsilon_{t+h}, \quad t = m, \dots, T-h. \quad (15)$$

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

in this case given by

$$\hat{\sigma} = \left(\frac{1}{n} \sum_{t=m}^{T-h} (f_{t,h}^R - f_{t,h}^U)^2 \right)^{-1} \sqrt{ \sum_{j=-p_n+1}^{p_n-1} \left(1 - \left| \frac{j}{p_n} \right| \right) n^{-1} \sum_{t=m+j}^{T-h} (f_{t,h}^R - f_{t,h}^U) \hat{\varepsilon}_{t+h} (f_{t-j,h}^R - f_{t-j,h}^U) \hat{\varepsilon}_{t+h-j} }, \quad (16)$$

where $\hat{\varepsilon}_{t+h}$ are regression residuals from (15) 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 (8) can be similarly obtained as

$$\hat{\lambda}_{t,d} = \frac{\sum_{j=t-d+1}^t (x_{j+h} - f_{j,h}^U) (f_{j,h}^R - f_{j,h}^U)}{\sum_{j=t-d+1}^t (f_{j,h}^R - f_{j,h}^U)^2}, \quad t = m + d - 1, \dots, T - h, \quad (17)$$

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

$$\begin{aligned} x_{j+h} - f_{j,h}^U &= \lambda_{t,d} (f_{j,h}^R - f_{j,h}^U) + \varepsilon_{j+h}; \\ j &= t - d + 1, \dots, t; \\ t &= m + d - 1, \dots, T - h. \end{aligned} \quad (18)$$

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 .

4.2 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 $\text{Var}_t[x_{t+h}] = \text{Var}[x_{t+h}] = \Sigma$.

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:

$$w^* = \arg \min_w \left\{ w' E_t[x_{t+h}] - \frac{\gamma}{2} w' \Sigma w \right\}, \quad (19)$$

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

problem is given by

$$\begin{aligned} w^* &= a + B E_t [x_{t+h}]; \\ a &= \frac{\Sigma^{-1} \iota}{\iota' \Sigma^{-1} \iota}; \\ B &= \frac{1}{\gamma} \left(\Sigma^{-1} - \frac{\Sigma^{-1} \iota \iota' \Sigma^{-1}}{\iota' \Sigma^{-1} \iota} \right), \end{aligned} \quad (20)$$

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 :

$$\begin{aligned} \lambda^* &= \arg \min_{\lambda \in R} E \left\{ \left(-a - B (f_{t,h}^R + (1 - \lambda)(f_{t,h}^U - f_{t,h}^R)) \right)' x_{t+h} + \right. \\ &\quad \left. \frac{\gamma}{2} (a + B (f_{t,h}^R + (1 - \lambda)(f_{t,h}^U - f_{t,h}^R)))' \Sigma (a + B (f_{t,h}^R + (1 - \lambda)(f_{t,h}^U - f_{t,h}^R))) \right\}. \end{aligned} \quad (21)$$

The closed-form solution for this problem is

$$\lambda^* = \frac{E \left[(f_{t,h}^R - f_{t,h}^U)' B' \left(x_{t+h} - \gamma \Sigma (a + B f_{t,h}^U) \right) \right]}{E \left[\gamma (f_{t,h}^R - f_{t,h}^U)' B' \Sigma B (f_{t,h}^R - f_{t,h}^U) \right]}. \quad (22)$$

A consistent estimator of λ^* is

$$\hat{\lambda} = \frac{\sum_{t=m}^{T-h} \left[(f_{t,h}^R - f_{t,h}^U)' \hat{B}_t' \left(x_{t+h} - \gamma \hat{\Sigma}_t (\hat{a}_t + \hat{B}_t f_{t,h}^U) \right) \right]}{\sum_{t=m}^{T-h} \left[\gamma (f_{t,h}^R - f_{t,h}^U)' \hat{B}_t' \hat{\Sigma}_t \hat{B}_t (f_{t,h}^R - f_{t,h}^U) \right]}, \quad (23)$$

where $\hat{a}_t$ and $\hat{B}_t$ are as defined in (20) with Σ substituted at each time t by an estimate computed over each rolling window of data up to time t :

$$\hat{\Sigma}_t = \frac{1}{m} \sum_{j=t-m+1}^t (x_j - \bar{x})(x_j - \bar{x})', \text{ with } \bar{x} = \frac{1}{m} \sum_{j=t-m+1}^t x_j. \quad (24)$$

The estimator of the asymptotic variance $\hat{\sigma}$ that is needed for constructing the test in Proposition

1 and the bands in Proposition 2 is obtained by setting

$$\begin{aligned} s_t(\hat{\lambda}) &= (f_{t,h}^R - f_{t,h}^U)' \hat{B}_t' \left[x_{t+h} - \gamma \hat{\Sigma}_t \left(\hat{a}_t + \hat{B}_t \left(f_{t,h}^R + (1 - \hat{\lambda})(f_{t,h}^U - f_{t,h}^R) \right) \right) \right] \text{ and} \\ \frac{\partial s_t(\hat{\lambda})}{\partial \lambda} &= \gamma (f_{t,h}^R - f_{t,h}^U)' \hat{B}_t' \hat{\Sigma}_t \hat{B}_t (f_{t,h}^R - f_{t,h}^U) \end{aligned} \quad (25)$$

in equation (6).

In the presence of time variation, a consistent estimator of the smoothed measure of usefulness (8) for a portfolio utility loss can be obtained as

$$\hat{\lambda}_{t,d} = \frac{\sum_{j=t-d+1}^t \left[(f_{j,h}^R - f_{j,h}^U)' \hat{B}_j' \left(x_{j+h} - \gamma \hat{\Sigma}_j \left(\hat{a}_j + \hat{B}_j f_{j,h}^U \right) \right) \right]}{\sum_{j=t-d+1}^t \left[\gamma (f_{j,h}^R - f_{j,h}^U)' \hat{B}_j' \hat{\Sigma}_j \hat{B}_j (f_{j,h}^R - f_{j,h}^U) \right]}, \quad t = m + d - 1, \dots, T - h. \quad (26)$$

In the empirical application, the variable to be forecast 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.

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

We consider the model 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:

$$\xi_{t+1} = \xi_t \exp(-0.5\Lambda_t'\Lambda_t - \Lambda_t'\varepsilon_{t+1}). \quad (27)$$

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

$$\Lambda_t = \Lambda_0 + \Lambda_1 F_t, \quad (28)$$

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 :

$$i_t = \delta_0 + \delta_1' F_t, \quad (29)$$

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

$$F_t = \Psi F_{t-1} + \Omega \varepsilon_t, \quad (30)$$

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

$$m_{t+1} = \exp(-i_t)\xi_{t+1}/\xi_t = \exp(-\delta_0 - \delta_1' F_t - 0.5\Lambda_t'\Lambda_t - \Lambda_t'\varepsilon_{t+1}), \quad (31)$$

where the second equality comes from (29) and (27). 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:

$$p_t^{(\tau+1)} = E_t(m_{t+1}p_{t+1}^{(\tau)}). \quad (32)$$

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

$$p_t^{(\tau)} = \exp(\bar{A}_\tau + \bar{B}_\tau' F_t), \quad (33)$$

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

$$\begin{aligned} \bar{A}_{\tau+1} &= \bar{A}_\tau + \bar{B}_\tau'(-\Omega\Lambda_0) + 0.5\bar{B}_\tau'\Omega\Omega'\bar{B}_\tau - \delta_0; \\ \bar{B}_{\tau+1}' &= \bar{B}_\tau'(\Psi - \Omega\Lambda_1) - \delta_1', \end{aligned} \quad (34)$$

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:

$$y_t^{(\tau)} = -\ln p_t^{(\tau)}/\tau = A_\tau + B_\tau' F_t, \quad (35)$$

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 (30) and (35) define a state-space model:

$$\begin{aligned} F_t &= \Psi F_{t-1} + \Omega \varepsilon_t; \\ Y_t &= A + B F_t + v_t, \end{aligned} \tag{36}$$

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 (34), and v_t is a vector of iid 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:

$$\theta = \{\Psi, \Omega, \Lambda_0, \Lambda_1, \Sigma_v\}. \tag{37}$$

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 (28).

5.2 VARs with no-arbitrage restrictions

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

$$Y_t = \Phi_0 + \Phi_1 Y_{t-1} + \dots + \Phi_p Y_{t-p} + u_t \tag{38}$$

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 (38) 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 (36), the $ATSM$ features an MA representation. As the $ATSM$ depends on a vector of coefficients θ (eq.(37)) having much fewer elements than the coefficient matrices of the VAR , the validity of the $ATSM$ imposes a set of nonlinear cross-equation

restrictions on the *VAR* in (38).

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 (36), and then impose them on the *VAR* in (38). To do so, rewrite the *VAR* in the data-matrix notation:

$$Y = X\Phi + U, \quad (39)$$

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 (36) for a given θ . Then, under the *ATSM*, the relation between the *ATSM* parameters and the *VAR* parameters is¹ $\Phi^* = [\Gamma_{xx}^*(\theta)]^{-1} \Gamma_{xy}^*(\theta)$, 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:

$$\hat{\Phi}^* = [\Gamma_{xx}^*(\hat{\theta})]^{-1} \Gamma_{xy}^*(\hat{\theta}). \quad (40)$$

The maximum likelihood estimator of the unrestricted *VAR* 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 Unrestricted *VAR* model (we label such forecasts f_t^U), the *VAR* with no-arbitrage restrictions (f_t^{NA}), the *VAR* with no-arbitrage restrictions and constant risk premium (f_t^{CRP}), and a simple Random Walk 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. The procedure goes on until the last forecast (i.e. that of 2003:12) is obtained.

We maximize the likelihood of the *VAR* with no-arbitrage restrictions using the Broyden, Fletcher, Goldfarb, and Shanno (BFGS) algorithm with Brent line search.² For the *VAR* models, we use a

¹As stressed in the text, the approximation is not exact because the state-space representation of the *ATSM* generates moving average terms

²In the first estimation window we initialize our algorithm as follows. First we compute a maximum, then we draw

specification with 3 lags which provides well-behaved residuals.³

5.4 Results for a quadratic loss

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

$$L(y_{t+1}^{(\tau)}, \hat{y}_{t+1}^{(\tau)}) = \left(y_{t+1}^{(\tau)} - \hat{y}_{t+1}^{(\tau)} \right)^2, \quad (41)$$

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 for the combination of the VAR with no-arbitrage restrictions and the Unrestricted VAR are displayed in Table 2. The table is composed of two panels: Panel A contains results from the combination of the forecasts from the VAR with no-arbitrage restrictions (f_t^{NA}) and forecasts from the Unrestricted VAR (f_t^U), i.e. $\hat{y}_{t+1}^{(\tau)} = f_{t,h}^{NA} + (1 - \lambda)(f_{t,h}^U - f_{t,h}^{NA})$. Panel B considers the combination of the no-arbitrage VAR with the additional restriction of constant risk premium (f_t^{CRP}) with the Unrestricted VAR (f_t^U) forecasts, i.e. $\hat{y}_{t+1}^{(\tau)} = f_{t,h}^{CRP} + (1 - \lambda)(f_{t,h}^U - f_{t,h}^{CRP})$. For each yield, Table 2 contains the Root Mean Squared Forecast Error (RMSFE) (i.e. the realized loss) of the two models used in the combination, the estimated optimal weight $\hat{\lambda}$ (as defined in equation (14)), and the RMSFE of the combined forecast (based on $\hat{\lambda}$). Finally, the last two columns in the table report the statistics for the encompassing tests of Proposition 1.

From Panel A of Table 2, 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 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 short-end of the yield curve while it is milder for yields of longer maturities.

Panel B of Table 2 provides results for the case in which we impose on the no-arbitrage VAR the additional restriction of a constant risk premium. The estimated optimal weights are similar

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).

³The 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.

for the 3-, 36-, and 60- month yields, while there is a 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. 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.

Then, we consider the combination of the no-arbitrage VAR forecasts (with and without variation in risk premia, i.e. f_t^{NA} and f_t^{CRP}) with the Random Walk forecasts (i.e. f_t^{RW}), which can be thought of as an alternative set of (atheoretical) restrictions on the baseline Unrestricted VAR model. Results for this case are displayed in Table 3, where Panel A contains results from the combination $\hat{y}_{t+1}^{(\tau)} = f_{t,h}^{NA} + (1-\lambda)(f_{t,h}^{RW} - f_{t,h}^{NA})$ and Panel B considers the combination $\hat{y}_{t+1}^{(\tau)} = f_{t,h}^{CRP} + (1-\lambda)(f_{t,h}^{RW} - f_{t,h}^{CRP})$.

As is clear from Table 3, most of the gains documented in Table 2 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 restricted VAR. In particular, once we combine the no-arbitrage VAR with the Random Walk, the estimated optimal weights $\hat{\lambda}$ range between -0.186 and 0.629 for the case with variation in risk premium and between 0.090 and 0.380 for the case of constant risk premium. These figures are much lower than those obtained when considering the combination of the no-arbitrage VAR with the Unrestricted VAR (see Table 2). In particular, all the weights decrease quite dramatically, and the encompassing test does not reject the null that the no-arbitrage 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.

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^{NA} and f_t^U in the first column, f_t^{CRP} and f_t^U in the second column, f_t^{NA} 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 (17), together with the 95% bands described in Proposition 2.

Looking at the first two columns in Figure 1, it is clear that the forecasting gains from imposing the no-arbitrage 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 and approaches zero, signalling that there are no gains in using the 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 restriction is combined with the Random Walk, one cannot reject the null that the no-arbitrage restrictions are useless.

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:

$$r_{t+1}^{(\tau+1)} = p_{t+1}^{(\tau)} - p_t^{(\tau+1)} = -\tau y_{t+1}^{(\tau)} + (\tau + 1)y_t^{(\tau+1)}. \quad (42)$$

Equation (42) 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 transformation⁴:

$$\hat{r}_{t+1}^{(\tau+1)} = -\tau \hat{y}_{t+1}^{(\tau)} + (\tau + 1)y_t^{(\tau+1)}. \quad (43)$$

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:

$$L(r_{t+1}, f_{t,1}) = -w^*(f_{t,1})'r_{t+1} + \frac{\gamma}{2}w^*(f_{t,1})'\Sigma w^*(f_{t,1}), \quad (44)$$

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 (43).

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^{NA} 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^{NA} and f_t^{CRP} with the Random Walk forecasts f_t^{RW} . In Table 4 the first two columns contain the portfolio utility loss of the two models considered in a given combination, the third column contains the portfolio utility loss for the combined forecasts computed using the estimated optimal weight $\hat{\lambda}$, the fourth column reports the value of the estimated optimal weight $\hat{\lambda}$ in (23), and the last two columns 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 restrictions when considering a portfolio utility loss. Second, by focusing on the losses occurring when the forecasts are produced with the no-arbitrage restricted VAR, it appears that

⁴It 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: $\hat{r}_{t+1}^{(\tau+1)} - r_{t+1}^{(\tau+1)} = -\tau(\hat{y}_{t+1}^{(\tau)} - y_{t+1}^{(\tau)})$. 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.

such losses are much higher in the case of constant risk premium (f_t^{CRP}) than in the case of time varying risk premium (f_t^{NA}).

We now turn on the results for the local measure of usefulness, which are summarized in Figure 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 (26), 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 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 restrictions may not be important from a statistical point of view, but it is essential when evaluating the forecasts in term 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 restrictions when considering a portfolio utility loss. Even though the usefulness of the no-arbitrage restrictions relative to a random walk has decreased over time, both restrictions appear to be useful, and the optimal forecast combination exploits information for both restrictions.

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 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 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.

References

- [1] 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.
- [2] Ang, A., M. Piazzesi, and M. Wei, 2004, What does the Yield Curve Tell us about GDP Growth? *Journal of Econometrics* 131, 359-403.
- [3] Carriero, A, 2007. Forecasting the Yield Curve using Priors from No Arbitrage Affine Term Structure Models, Queen Mary Working paper 612.
- [4] Clark, T.E. and M. W. McCracken, 2007. Combining forecasts from nested models, mimeo
- [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., Singleton, K., 2000. Specification analysis of affine term structure models. *Journal of Finance* 55, 1943-1978.
- [7] Del Negro, M., Schorfheide, F, 2004, Priors from General Equilibrium Models for VARs, *International Economic Review*, Vol. 45, 643-673.
- [8] Duffee, G., 2002. Term premia and interest rate forecasts in affine models. *Journal of Finance* 57, 405-443.
- [9] Duffie, D., Kan, R., 1996. A yield-factor model of interest rates. *Mathematical Finance* 6, 379-406.
- [10] Elliott, G. and A. Timmermann, 2004. Optimal forecast combinations under general loss functions and forecast error distributions, *Journal of Econometrics*, 122, 47-79
- [11] Giacomini, R. and B. Rossi, 2008. Forecast Comparisons in Unstable Environments, forthcoming *Journal of Applied Econometrics*
- [12] Giacomini, R. and H. White, 2006, Tests of Conditional Predictive Ability, *Econometrica*, 74, 1545-1578.
- [13] Harrison, J.M., Kreps, D.M., 1979. Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory* 2, 381-408.
- [14] Johansen S., 1995, *Likelihood-based Inference in Cointegrated Vector Auto-regressive Models*, Oxford University Press, Oxford.
- [15] Nelson, C.R. and A.F. Siegel, 1987. Parsimonious modeling of yield curve. *Journal of Business* 60, 473-489.
- [16] Newey, W. K. and D. McFadden, 1994. Large sample estimation and hypothesis testing, Handbook of Econometrics, Volume IV, edited by R.F. Engle and D.L. McFadden, Elsevier
- [17] Vasicek, O., 1977. An equilibrium characterization of the term structure. *Journal of Financial Economics* 5, 177-18
- [18] West, K. D., 1996. Asymptotic Inference about Predictive Ability, *Econometrica*, 64, 1067-1084.
- [19] White, H., 2001. Asymptotic theory for econometricians, Revised Edition, Academic Press
- [20] 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, $\hat{\lambda}$ is asymptotically normal, so that $\sigma^{-1}\sqrt{n}(\hat{\lambda} - \lambda^*) \rightarrow^d N(0, 1)$, where $\sigma^2 = H^{-1}\Omega H^{-1}$. The results in the Proposition then follow from showing consistency of $\hat{\sigma}^2$ for σ^2 . Asymptotic normality of $\hat{\lambda}$ is obtained by verifying the assumptions of Theorem 3.1 of Newey and McFadden (1994), since $\hat{\lambda}$ 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 $\hat{\lambda} \rightarrow^p \lambda^*$. 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 (1984) 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 $\hat{\Omega}$ 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 $\hat{\sigma}^2 \rightarrow^p \sigma^2$, 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 $\hat{\lambda}_{t,d}$. For $t = m + d - 1, \dots, T - h$ we have

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

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

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

where $\mathcal{B}$ is a standard univariate Brownian motion. By B.2, $(d/n)^{-1/2} \rightarrow \pi^{-1/2}$. By B.3, $\hat{H}^{-1}(\bar{\lambda}) \hat{\sigma}^{-1}\Omega^{1/2} \rightarrow^P H^{-1}\sigma^{-1}\Omega^{1/2} = 1$ and thus $\hat{\sigma}^{-1}\sqrt{d}(\hat{\lambda}_t - \lambda^*) \Rightarrow [\mathcal{B}(\tau) - \mathcal{B}(\tau - \pi)]/\sqrt{\pi}$ 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 the bands $(\hat{\lambda}_t - k_{\alpha,\pi}\frac{\hat{\sigma}}{\sqrt{d}}, \hat{\lambda}_t + k_{\alpha,\pi}\frac{\hat{\sigma}}{\sqrt{d}})$ 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. Quadratic Loss: no-arbitrage VAR and Unrestricted VAR

Panel A. No-arbitrage VAR and Unrestricted VAR						
Yields	$RMSFE(f_t^U)$	$RMSFE(f_t^{NA})$	$RMSFE(f_t^*)$	$\hat{\lambda}$	t^U	t^{NA}
1-month	0.808	0.804	0.769	0.514	-4.040*	4.276*
3-month	0.709	0.663	0.656	0.736	-1.470	4.103*
12-month	0.682	0.600	0.600	1.054	-0.301	5.901*
36-month	0.526	0.483	0.482	0.875	-0.733	5.129*
60-month	0.460	0.429	0.428	0.840	-1.018	5.338*
Panel B. No-arbitrage VAR with constant risk premium (CRP) and Unrestricted VAR						
Yields	$RMSFE(f_t^U)$	$RMSFE(f_t^{CRP})$	$RMSFE(f_t^*)$	$\hat{\lambda}$	t^U	t^{CRP}
1-month	0.808	0.921	0.786	0.280	-7.948*	3.099*
3-month	0.709	0.665	0.656	0.709	-1.816	4.428*
12-month	0.682	0.617	0.611	0.777	-1.290	4.510*
36-month	0.526	0.478	0.477	0.889	-0.686	5.502*
60-month	0.460	0.432	0.428	0.740	-1.912	5.448*

Panel A contains results from the combination of the forecasts made by the VAR with no-arbitrage restrictions (f_t^{NA}) with those of the Unrestricted VAR (f_t^U). Panel B considers the combination of the forecasts made by the VAR with no-arbitrage restrictions *and* constant risk premium (f_t^{CRP}) with those of the Unrestricted VAR (f_t^U). Results are reported for each of the yields at hand. The first two columns contain the Root Mean Squared Forecast Error (RMSFE) (i.e. the realized loss) of the two models considered. The third column contains the RMSFE for the combined forecasts f_t^* computed using the estimated optimal weight $\hat{\lambda}$. The fourth column reports the value of the estimated optimal weight $\hat{\lambda}$ (as defined in equation (14)). Finally, the last two columns in the table report the statistics for the encompassing tests of Proposition 1, where the unrestricted forecast is f_t^U and the restricted forecast is $f_t^R = f_t^{NA}$ (Panel A) or $f_t^R = f_t^{CRP}$ (Panel B). The statistic $t^U = \sqrt{n}(\hat{\lambda} - 1)/\hat{\sigma}$ is used to test the null that the unrestricted forecast is useless ($H_0^U : \lambda^* = 1$). The statistics $t^R = \sqrt{n}\hat{\lambda}/\hat{\sigma}$ is used to test the null that the restricted forecast is useless ($H_0^R : \lambda^* = 0$).

Table 3. Quadratic Loss : no-arbitrage VAR and Random Walk

Panel A. No-arbitrage VAR and Random Walk						
Yields	$RMSFE(f_t^{RW})$	$RMSFE(f_t^{NA})$	$RMSFE(f_t^*)$	$\hat{\lambda}$	t^{RW}	t^{NA}
1-month	0.837	0.804	0.787	0.629	-2.458*	4.162*
3-month	0.607	0.663	0.606	-0.186	-5.654*	-0.890
12-month	0.596	0.600	0.595	0.188	-2.186*	0.508
36-month	0.473	0.483	0.473	0.128	-3.577*	0.525
60-month	0.424	0.429	0.424	0.249	-2.418*	0.802

Panel B. No-arbitrage VAR with constant risk premium (CRP) and Random Walk						
Yields	$RMSFE(f_t^{RW})$	$RMSFE(f_t^{CRP})$	$RMSFE(f_t^*)$	$\hat{\lambda}$	t^{RW}	t^{CRP}
1-month	0.837	0.921	0.797	0.356	-7.213*	3.994*
3-month	0.607	0.665	0.607	0.090	-7.403*	0.736
12-month	0.596	0.617	0.590	0.295	-3.090*	1.295
36-month	0.473	0.478	0.470	0.380	-2.735*	1.681
60-month	0.424	0.432	0.424	0.148	-3.765*	0.653

Panel A contains results from the combination of the forecasts made by the VAR with no-arbitrage restrictions (f_t^{NA}) with those of the Random Walk (f_t^{RW}). Panel B considers the combination of the forecasts made by the VAR with no-arbitrage restrictions *and* constant risk premium (f_t^{CRP}) with those of the Random Walk (f_t^{RW}). Results are reported for each of the yields at hand. The first two columns contain the Root Mean Squared Forecast Error (RMSFE) (i.e. the realized loss) of the two models considered. The third column contains the RMSFE for the combined forecasts f_t^* computed using the estimated optimal weight $\hat{\lambda}$. The fourth column reports the value of the estimated optimal weight $\hat{\lambda}$ (as defined in equation (14)). Finally, the last two columns in the table report the statistics for the encompassing tests of Proposition 1, where $f_t^U = f_t^{RW}$ and $f_t^R = f_t^{NA}$ (Panel A) or $f_t^R = f_t^{CRP}$ (Panel B). The statistic $t^U = \sqrt{n}(\hat{\lambda} - 1)/\hat{\sigma}$ is used to test the null that the unrestricted forecast is useless ($H_0^U : \lambda^* = 1$). The statistics $t^R = \sqrt{n}\hat{\lambda}/\hat{\sigma}$ is used to test the null that the restricted forecast is useless ($H_0^R : \lambda^* = 0$).

Table 4. Portfolio Utility Loss

Panel A. No-arbitrage VAR and Unrestricted VAR					
$UL(f_t^U)$	$UL(f_t^{NA})$	$UL(f_t^*)$	$\hat{\lambda}$	t^U	t^{NA}
-1.360	-1.271	-1.585	0.458	-6.96	5.89

Panel B. No-arbitrage VAR (CRP) and Unrestricted VAR					
$UL(f_t^U)$	$UL(f_t^{CRP})$	$UL(f_t^*)$	$\hat{\lambda}$	t^U	t^{CRP}
-1.360	0.168	-1.494	0.221	-16.34	4.63

Panel C. No-arbitrage VAR and Random Walk					
$UL(f_t^{RW})$	$UL(f_t^{NA})$	$UL(f_t^*)$	$\hat{\lambda}$	t^{RW}	t^{NA}
-1.481	-1.271	-1.793	0.436	-10.35	8.00

Panel D. No-arbitrage VAR (CRP) and Random Walk					
$UL(f_t^{RW})$	$UL(f_t^{CRP})$	$UL(f_t^*)$	$\hat{\lambda}$	t^{RW}	t^{CRP}
-1.481	0.168	-1.709	0.258	-20.09	7.00

Panels A and B contain results from the combination of f_t^{NA} and f_t^{CRP} with the Unrestricted VAR forecasts f_t^U . Panels C and D contain results of the combination of f_t^{NA} and f_t^{CRP} with the Random Walk forecasts f_t^{RW} . The first two columns contain the Portfolio Utility Loss of the two models considered in a given combination. The third column contains the Portfolio Utility Loss for the combined forecasts f_t^* computed using the estimated optimal weight $\hat{\lambda}$. The fourth column reports the value of the estimated optimal weight $\hat{\lambda}$ (as defined in equation (23)). Finally, the last two columns in the table report the statistics for the encompassing tests of Proposition 1.

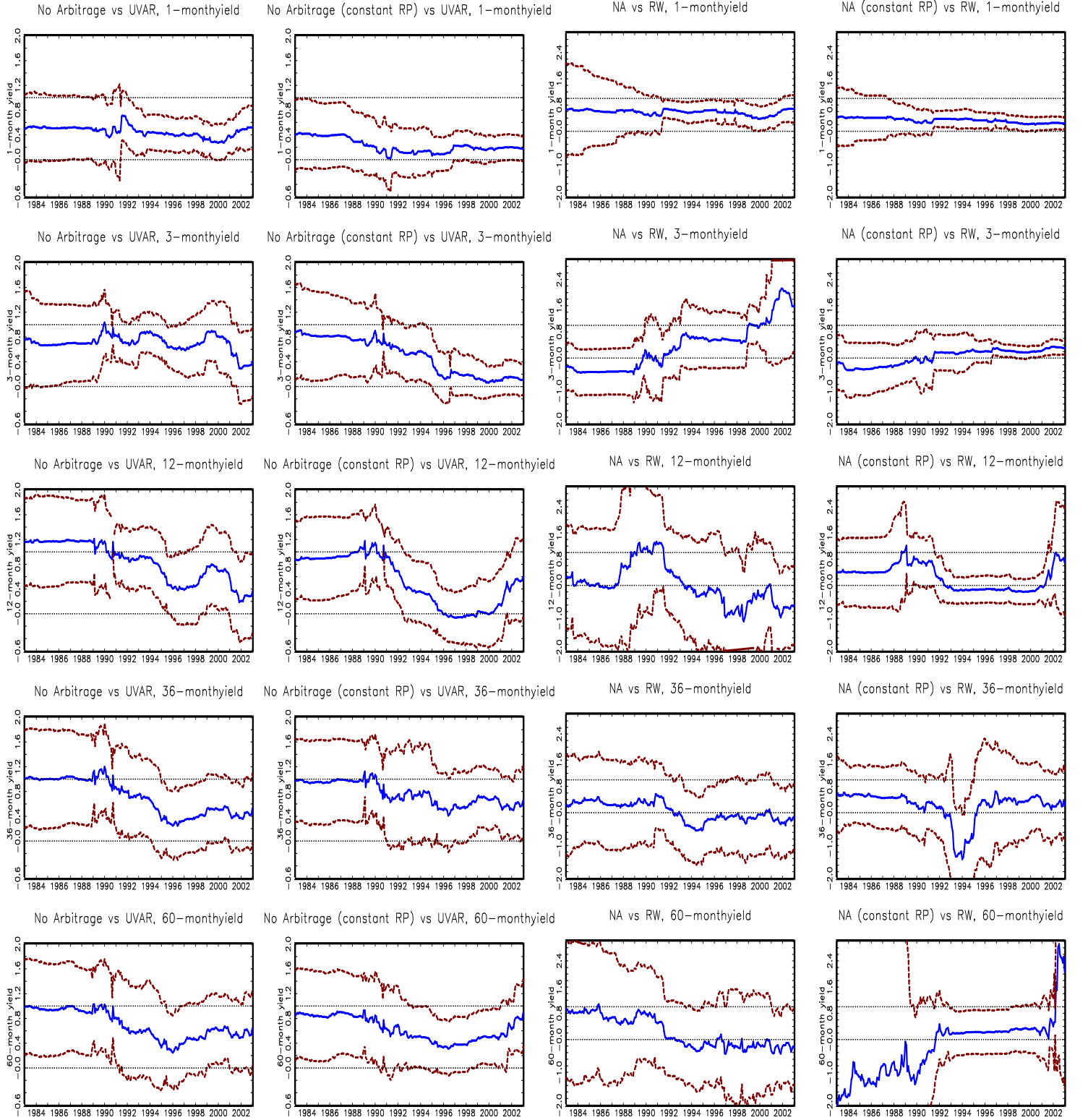


Figure 1: Results for Quadratic Loss. The blue solid line is the estimated optimal weight $\hat{\lambda}_t$ in equation (17). The red dashed lines are the 95% bands in equation (10).

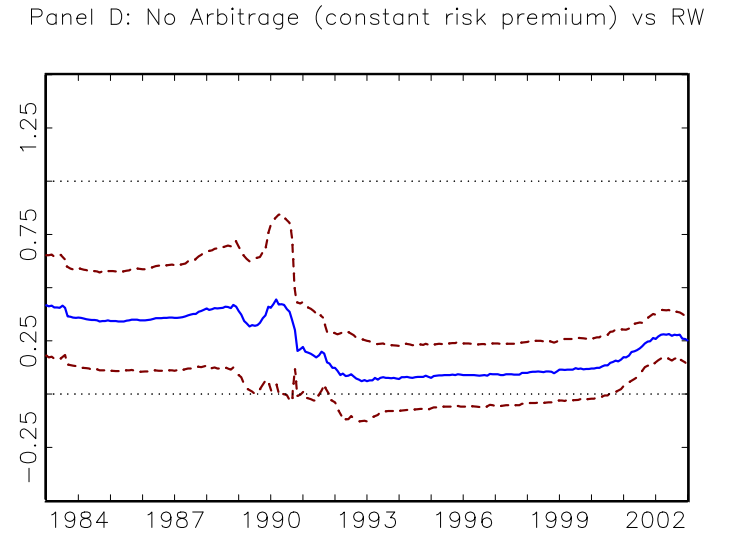
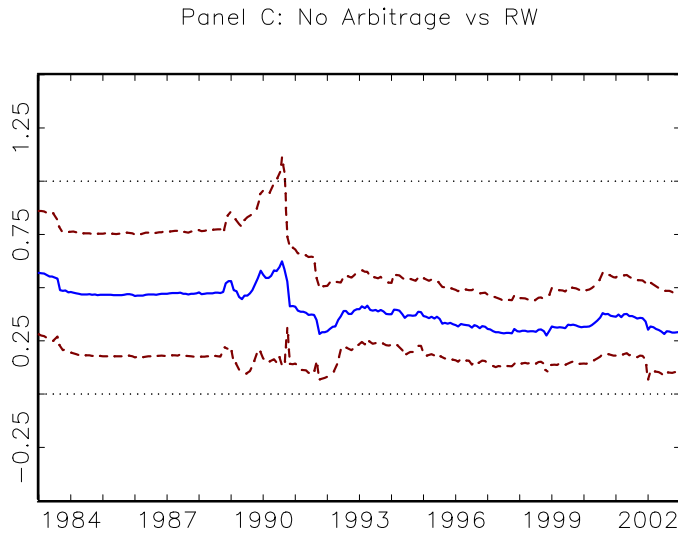
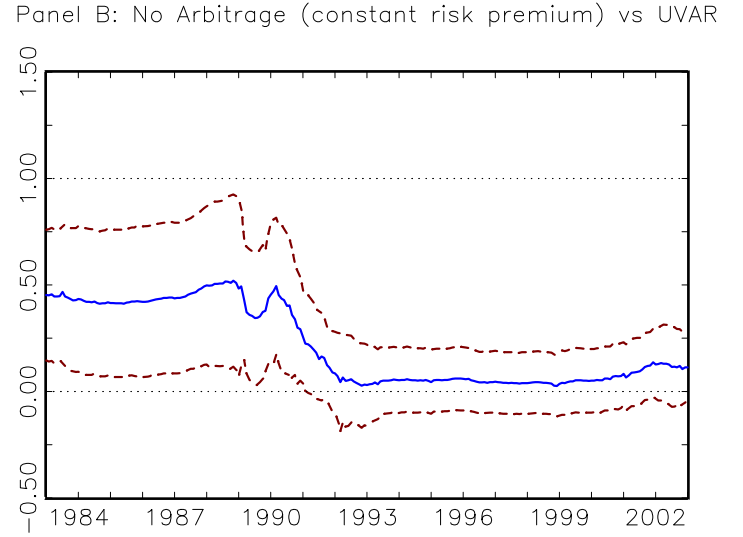
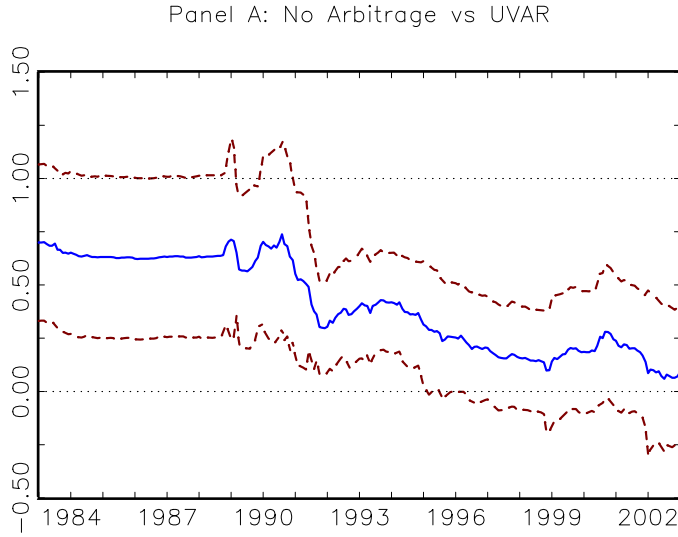


Figure 2: Results for Portfolio Utility Loss. The blue solid line is the estimated optimal weight $\hat{\lambda}_t$ in equation (26). The red dashed lines are the 95% bands in equation (10).