Estimating Loss Function Parameters*

Graham Elliott
University of California San Diego

Ivana Komunjer
Caltech

Allan Timmermann
University of California San Diego

October 4, 2002

Abstract

We examine estimation and inference on loss function parameters within classes of loss functions. In situations where forecasts or implied forecasts are observed, a common strategy is to examine 'rationality' given a loss function. Rejection of the hypothesis is conditional on the loss function chosen. We examine this from a different perspective - supposing that we have a family of loss functions indexed by a vector of parameters, then given the forecasts can we back out the loss function parameters consistent with the forecasts being rational. We establish identification and provide estimation methods and asymptotic distributional results for classes of loss functions. The methods are applied in an empirical analysis of IMF and OECD forecasts of budget deficits for the G7 countries. We find that allowing for asymmetric loss can significantly change the outcome of empirical tests of forecast rationality.

*Graham Elliott and Allan Timmermann are grateful to the NSF for financial assistance under grant SES 0111238. Carlos Capistrano provided excellent research assistance.

1 Introduction

That agents are rational when they construct forecasts of economic variables is an important assumption maintained throughout much of economics and finance. Much effort has been devoted to empirically testing the validity of this proposition in areas such as efficient market models of stock prices (Dokko and Edelstein (1989), Lakonishok (1980)), models of the term structure of interest rates (Cargill and Meyer (1980), De Bondt and Bange (1992), Fama (1975)), models of currency rates (Hansen and Hodrick (1980)), inflation forecasting (Figlewski and Wachtel (1981), Keane and Runkle (1990), Mishkin (1981), Pesando (1975), Schroeter and Smith (1986)) and tests of the Fisher equation (Gultekin (1983)).

Invariably the empirical literature has tested rationality of forecasts in conjunction with the assumption that mean squared error (MSE) loss adequately represents the forecaster's objectives.[1] Under this loss function forecasts are easy to compute through least squares methods and they also have well established properties such as unbiasedness and lack of serial correlation at the single-period horizon, c.f. Diebold and Lopez (1996). This makes inference about optimality of a particular forecast series an easy exercise.

Symmetry of the loss function, albeit a widely used assumption, is, however often difficult to justify on economic grounds and is certainly not universally accepted. Granger and Newbold (1986, page 125), for example, argue that "An assumption of symmetry for the cost function is much less acceptable [than an assumption of a symmetric forecast error density]." It is easy to understand their argument. There is, for example, no reason why the consequences of underpredicting the demand for some product (los of potential sales, customers and reputation) should be identical to the costs of overpredicting it (added costs of production and storage). As a second example, central banks are likely to have asymmetric preferences, c.f. Peel and Nobay (1998).

Consequently, in economics and finance forecasting performance is increasingly evaluated under more general loss functions that account for asymmetries as witnessed by recent studies such as Christofferson and Diebold (1997), Granger and Newbold (1986), Granger

[1] In addition to the studies cited in the first paragraph, Hafer and Hein (1985) and Zarnovitz (1979) use mean squared error loss or mean absolute error loss as a metric for measuring forecast accuracy.

and Pesaran (2000) and West, Edison and Choi (1996). Frequently used loss functions include lin-lin, linex and quad-quad loss which allow for asymmetries through a single shape parameter. Under these more general loss functions, the forecast error no longer retains the optimality properties that are typically tested in empirical work, c.f. Granger (1999). This raises the possibility that the many rejections of forecast optimality reported in the empirical literature may simply be driven by the assumption of MSE loss rather than by absence of forecast rationality as such.

This paper develops new methods for testing forecast optimality under general classes of loss functions that include Mean absolute deviations (MAD) or MSE loss as a special case. This allows us to separate the question of forecast rationality from that of whether MAD or MSE loss accurately represents the decision maker's objectives. Instead our results allow us to test the joint hypothesis that the loss function belongs to a more flexible family and that the forecast is optimal. This situation is very different from MSE loss where the properties of the observed forecast error are independent of the parameters of the loss function. This may be the reason why the empirical literature often overlooks that tests of forecast rationality relying on properties such as unbiasedness and lack of serial correlation in forecast errors are really joint tests.

In each case the family of loss functions is indexed by a single unknown parameter. We establish conditions under which this parameter is exactly identified. Since first order conditions for optimality of the forecast take the form of moment conditions, exact identification corresponds to the situation where the number of moment conditions equals the number of parameters of the loss function. When there are more moments than parameters, the model is overidentified and the null hypothesis of rationality can be tested through a J-test.

Our approach therefore reverses the usual procedure - which conditions on a maintained loss function and tests rationality of the forecast - and instead asks what sort of parameters of the loss function would be most consistent with forecast rationality. We treat the loss function parameters as unknowns that have to be estimated and effectively 'back out' the parameters of the loss function from the observed time-series of forecast errors.

The plan of the paper is as follows. Section 2 outlines the conditions for optimality of forecasts under a general class of loss functions, including ones that are non-differentiable at a finite number of points. Section 3 develops the theory for identification and estimation of loss function parameters and also derives tests for forecast optimality in overidentified models. Section 4 explores the small sample performance of our methods in a Monte Carlo simulation experiment, while Section 5 provides an application to two international organizations' forecasts of government budget deficits. Section 6 concludes. Technical details are provided in appendices at the end of the paper.

2 Asymmetric Loss and Optimal Properties of Forecasts

It is common in the literature to test for 'rationality' using data on forecasts. Optimal properties (or properties of rational forecasts) can only be established jointly with, or in the context of, a maintained loss function. Typically this is taken to be squared loss, where loss is assumed to be symmetric in the losses. This choice is useful in practice for a number of reasons - it provides simple optimal properties and relates directly to least squares regression on forecast errors, so fits directly within the standard econometric toolbox. In this section we review the optimal properties of forecasts for more general loss functions. We then 'turn the problem around' and motivate the idea of estimating loss functions from observed forecasts.

Consider a stochastic process $X \equiv \{X_t : \Omega \longrightarrow \mathbb{R}^{m+1}, m \in \mathbb{N}, t = 1, \ldots, T\}$ defined on a complete probability space $(\Omega, \mathcal{F}, \mathcal{P})$ where $\mathcal{F} = \{\mathcal{F}_t, t = 1, \ldots, T\}$ and $\mathcal{F}_t$ is the σ-field $\mathcal{F}_t \equiv \sigma\{X_s, s \leqslant t\}$. In what follows we partition the observed vector X_t as $X_t \equiv (Y_t, Z_t')'$, where $Y_t \in \mathbb{R}$ is the variable of interest and $Z_t \in \mathbb{R}^m$ is a vector of exogenous variables. The random variable Y_t is further assumed to be continuous.[2] We denote by y_t and z_t the observations of the variables Y_t and Z_t respectively.[3] The forecasting problem considered

[2]In standard notation the subscript t under the distribution function $P(\cdot)$ of Y_{t+1}, its density $p(\cdot)$, and the expectation $E[\cdot]$ denotes conditioning on the information set $\mathcal{F}_t$.

[3]As a general rule, we hereafter use upper case letters for random variables, i.e. Y_t and Z_t, and lower case letters for their realizations, i.e. y_t and z_t.

here involves forecasting the variable Y_{t+h}, where h is the prediction horizon of interest, $h \geqslant 1$. In what follows, we set $h = 1$ and examine the one-step-ahead predictions of the realization y_{t+1}, knowing that all results developed in this case can readily be generalized to any $h \geqslant 1$.

The setup used here is fairly standard in the forecasting literature: we let f_{t+1} be the forecast of Y_{t+1} conditional on the information set $\mathcal{F}_t$. In what follows we restrict ourselves to the class of linear forecasts, $f_{t+1} \equiv \theta' W_t$, in which θ is an unknown k-vector of parameters, $\theta \in \Theta$, Θ compact in $\mathbb{R}^k$, and W_t is an h-vector of variables that are $\mathcal{F}_t$-measurable. It is important to note that both the model $M \equiv \{f_{t+1}\}$ and the vector W_t are specified by the agent producing the forecast (e.g., IMF, OECD, EC) and they need not be known by the forecast user. As a general rule, W_t should include variables that are observed at time t and which are thought to help forecast Y_{t+1} (e.g., a subset of the m-vector of exogenous variables Z_t, lags of Y_t, and/or different functions of the above). Should W_t fail to incorporate all the relevant information, we say that the model $M = \{f_{t+1}\}$ is wrongly specified. Misspecification will equally occur if the form of f_{t+1}, linear here, is wrongly specified by the forecaster, or if the original forecasts were manipulated in order to satisfy some institutional criterion. Keeping in mind this possibility, we do *not* assume that $M = \{f_{t+1}\}$ is correctly specified, i.e. we allow for model misspecification in the construction of the optimal forecasts.

When constructing optimal forecasts, we assume that, given Y_{t+1} and W_t, the forecaster has in mind a generalized loss function $\mathcal{L}$ defined by

$$\mathcal{L}(p, \alpha, \theta) \equiv [\alpha + (1 - 2\alpha) \cdot 1(Y_{t+1} - \theta' W_t < 0)] \cdot |Y_{t+1} - \theta' W_t|^p, \tag{1}$$

where $p \in \mathbb{N}^*$, $\alpha \in (0, 1)$, $\theta \in \Theta$ and $Y_{t+1} - \theta' W_t$ corresponds to the forecast error ε_{t+1}.[4] We let α_0 and p_0 be the unknown true values of α and p used by the forecaster. Hence, the loss function in (1) is a function of not only the realization of Y_{t+1} and the forecast $\theta' W_t$, but also of the shape parameters α and p of $\mathcal{L}$. Special cases of $\mathcal{L}$ include: (i) square loss function $\mathcal{L}(2, 1/2, \theta) = (Y_{t+1} - \theta' W_t)^2$, (ii) absolute deviation loss function $\mathcal{L}(1, 1/2, \theta) = |Y_{t+1} - f_{t+1}|$,

[4]Note that the function $\mathcal{L}(p, \alpha, \theta)$ is $\mathcal{F}_{t+1}$-measurable. In order to simplify the notations, however, we drop the reference to the time t and use the notation $\mathcal{L}(p, \alpha, \theta)$ instead of $\mathcal{L}_{t+1}(p, \alpha, \theta)$.

as well as their asymmetrical counterparts obtained when $\alpha \neq 1/2$, i.e. (iii) quad-quad loss, $\mathcal{L}(2, \alpha, \theta)$, and (iv) lin-lin loss, $\mathcal{L}(1, \alpha, \theta)$.

Given p_0 and α_0, the forecaster is thus assumed to constructs the optimal one-step-ahead forecast of Y_{t+1}, $f^*_{t+1} \equiv \theta^{*\prime} W_t$, by solving

$$\min_{\theta \in \Theta} E[\mathcal{L}(p_0, \alpha_0, \theta)]. \tag{2}$$

We let ε^*_{t+1} be the optimal forecast error, $\varepsilon^*_{t+1} \equiv y_{t+1} - f^*_{t+1} = y_{t+1} - \theta^{*\prime} w_t$, which depends on the unknown true values p_0 and α_0. Optimal forecasts have properties that follow directly from the construction of the forecasts. In the general case, the relevant condition is the one given in the following Proposition.

Proposition 1 (Necessary Optimality Condition) *Under Assumption (A0), given* $(p_0, \alpha_0) \in \mathbb{N}^* \times (0, 1)$*, if* θ^* *is the minimum of* $E[\mathcal{L}(p_0, \alpha_0, \theta)]$*, then* θ^* *satisfies the first order condition*

$$E[W_t \cdot (1(Y_{t+1} - \theta^{*\prime} W_t < 0) - \alpha_0) \cdot |Y_{t+1} - \theta^{*\prime} W_t|^{p_0 - 1}] = 0. \tag{3}$$

In other words, if the optimal forecast f^*_{t+1} is such that θ^* is an interior point of Θ (Assumption (A0)), the sequence of optimal forecast errors ε^*_{t+1} will satisfy the moment condition $E[W_t \cdot (1(\varepsilon^*_{t+1t} < 0) - \alpha_0) \cdot |\varepsilon^*_{t+1}|^{p_0-1}] = 0$.

This result implies that the first derivative of the loss function evaluated at the forecast errors is a martingale difference sequence with respect to all information available at the time of the forecast. When the forecasts are 'optimal', then any information must be correctly included in f^*_{t+1}, which is orthogonal to the transformed forecast errors. This enables the researcher to get around the problem that although they observe only the forecasts f^*_{t+1} rather than the components that made up these forecasts (i.e. the form of f_{t+1} which would be needed to determine W_t) since rationality implies that any variable that is useful should be in this model, we can use variables that could have been used to construct the forecasts.[5]

[5]This means we are concerned with partial rationality, i.e. the efficient use of a particular subset of information as opposed to full rationality that requires efficient utilization of all relevant information at the time the forecast is produced, c.f. Brown and Maital (1981).

Assuming that the forecast user observes a $d-$vector of variables V_t that would have been available to the forecast producer. In the case of rational forecasts V_t is a subvector of W_t. Given values for (α_0, p_0), the hypothesis of rational forecasts can be tested by testing that the moment conditions

$$E[V_t \cdot (1(Y_{t+1} - f^*_{t+1} < 0) - \alpha_0) \cdot |Y_{t+1} - f^*_{t+1}|^{p_0 - 1}] = 0. \tag{4}$$

hold.

In the typical case of mean square error loss, the parameter choice is $(0.5, 2)$. This choice simplifies the expression as $-(1(Y_{t+1} - f^*_{t+1} < 0) - \alpha_0) \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}_{t+1}) = \varepsilon_{t+1}$ and after dividing by -1 we have the property that the forecast errors themselves are a martingale difference sequence with respect to all t-dated information. For this special loss function one need work only with the forecast errors themselves, which is one of the simplification properties mentioned above that makes this a popular assumption for the loss function. It is this result that is typically tested in practice with data (Campbell and Ghysels (1995), Keane and Runkle (1990), Zarnowitz (1985)). These tests on the forecast errors directly have been split up into tests known as 'unbiasedness' and 'orthogonality' tests. Unbiasedness tests set $V_t = 1$, and hence test that $E[\varepsilon^*_{t+1}] = 0$. This can be undertaken by either directly testing the mean from an observed sequence of forecast errors is zero, i.e. having observed a $T \times 1$ time series of forecast errors $\left\{\varepsilon^*_{t+1}\right\}_1^T$ the regression

$$\varepsilon^*_{t+1} = \beta_0 + u_t$$

is run and the test is of the hypothesis $H_0 : \beta_0 = 0$ versus the alternative $H_A : \beta_0 \neq 0$. Alternatively, this idea is extended noting that $\varepsilon^*_{t+1} \equiv y_{t+1} - f^*_{t+1}$ and the regression

$$y_{t+1} = \beta_0 + \beta_1 f^*_{t+1} + u_{t+1}$$

is run and we can consider the joint test $H_0 : \beta_0 = 0, \beta_1 = 1$ versus the alternative that one or both coefficients differ from their null values. Finally, the general idea of 'orthogonality' extends these ideas to other more general specifications for V_t. In the typical linear regression we estimate

$$\varepsilon_{t+1} = \beta' V_t + u_{t+1}$$

and test the hypothesis $H_0 : \beta = 0$. This final idea of 'orthogonality' regressions thus includes as special cases the 'unbiasedness' regression. This is the Mincer-Zarnowitz (1969) regression.

However, these tests rely on the specification of the loss function as mean square error to justify the test procedure. In a more general loss function setting it is not the forecast errors themselves that are orthogonal to time t dated information but a transformation of these forecast errors (i.e. the first difference of the loss function). Hence, as noted in passing by most papers which undertake these tests, any rejection could stem from the joint nature of the testing procedure - jointly testing rationality and the form of the loss function. It is completely unclear when there is a rejection of any of these hypotheses as to the economic interpretation of the rejection. It may be that power is quite high for even small deviations from squared loss functions, resulting in rejections of rationality when all that is actually going on is that the forecaster had a slightly different loss function to squared loss. Thus it is reasonable to try and extend the class of loss functions for which the tests are valid.

It is this point that motivates the approach of this paper. Rather than assume an explicit loss function, we will generalize this idea to a class of loss functions indexed by the parameter set (α, p). We will then show that given observed forecasts and outcomes we can estimate the parameter of the loss function (α) within the families we examine (choice of p). Further, we are with the additional time t dated information V_t we are able to jointly test rationality and the class of loss functions (rather than imposing a particular loss function). The test will simply be a test of overidentification in a GMM estimation procedure.

3 Estimating Loss Function Parameters

In order to recover the shape parameters of the loss function $\mathcal{L}$ used by the forecaster in the minimization problem (2) we propose to use the first order condition (3). The main idea behind our approach is fairly simple: if for given shape parameters p_0 and α_0 the forecaster uses (3) to determine θ^*, then for a given θ^* we can reverse the problem and use the same moment condition (3) to recover p_0 and α_0. It is important to note that our approach is valid only if knowing a solution to (3) allows the forecast user to identify p_0 and α_0. The identification requirement is not easy to meet in general and we now turn to the construction of a setup where the estimation of the loss function parameters is possible.

First, note that the first order condition (3) is merely a necessary condition for θ^* to be optimal, i.e. not every value θ^* solving (3) is going to be the minimum of $E[\mathcal{L}(p_0, \alpha_0, \theta)]$ on $\mathring{\Theta}$ (the interior of Θ). The following result gives a set of sufficient conditions for a solution of (3) to be a strict local minimum of $E[\mathcal{L}(p_0, \alpha_0, \theta)]$.

Proposition 2 (First Order Condition) *Under Assumptions (A0)-(A2), and given* $(p_0, \alpha_0) \in \mathbb{N}^* \times (0, 1)$*, if* $\theta^* \in \mathring{\Theta}$ *is a solution to the first order condition* (3) *then* θ^* *is a strict local minimum of* $\mathcal{L}(p_0, \alpha_0, \theta)$ *on* $\mathring{\Theta}$*, i.e. given* $(p_0, \alpha_0) \in \mathbb{N}^* \times (0, 1)$ *there exists a neighborhood* $\mathcal{V}$ *of* θ^* *such that for any* $\theta \neq \theta^*$ *in* $\mathcal{V}$ *we have* $E[\mathcal{L}(p_0, \alpha_0, \theta)] < E[\mathcal{L}(p_0, \alpha_0, \theta^*)]$.

This is a sufficient condition for an interior point of Θ to be a local minimum of $\mathcal{L}$. Note that the first order condition does not necessarily hold if θ^* is on the boundary of Θ, i.e. if $\theta \in \Theta \backslash \mathring{\Theta}$. Also, note that the condition in Proposition 2 is slightly stronger than a necessary condition for $\theta^* \in \mathring{\Theta}$ to be a local minimum of $\mathcal{L}$. Indeed $\theta^* \in \mathring{\Theta}$ local minimum of $\mathcal{L}$ implies that the first order condition (3) holds, and that the Hessian matrix of second derivatives of $\mathcal{L}$ with respect to θ, evaluated at θ^*, is positive semidefinite. In Proposition 2, however, Assumptions (A1)-(A2) imply that the last one is positive definite so that θ^* is a strict local minimum of $\mathcal{L}$ on $\mathring{\Theta}$.

In order to identify and estimate α_0 we need to further limit the class of loss functions in (1), so that the loss function $\mathcal{L}$ is identified up to the parameter $\alpha_0 \in (0, 1)$. In what follows we consider two popular sets of loss functions: (i) the lin-lin loss function, obtained when $p_0 = 1$, and (ii) the quad-quad loss function obtained when $p_0 = 2$. The lin-lin loss function has been employed in the literature to allow for asymmetry. The quad-quad loss function is based on the same idea however with quadratic loss. When this loss function is symmetric it is identical to mean squared loss, hence it is a direct generalization of the typical loss function assumed in the forecast evaluation literature.

Having fixed the parameter p_0 of the loss function $\mathcal{L}$ in (1), we now consider the following problem: for a given $\alpha_0 \in (0, 1)$, is the optimal value θ^*, obtained as a solution to the first order condition (3), unique? Recall the result from Proposition 2: any $\theta^* \in \mathring{\Theta}$ solution to (3) is a strict *local* minimum of $\mathcal{L}$ in $\mathring{\Theta}$. In other words, for a given $\alpha_0 \in (0,1)$, we may have two or more local minima θ_i^* of $\mathcal{L}$ in $\mathring{\Theta}$ only one of them being the *absolute* minimum θ^* of $\mathcal{L}$ as defined by (2). If, given a solution θ_i^* to (3) we want to identify α_0 used in the minimization problem (2), we need to make sure that θ_i^* is the absolute minimum of $\mathcal{L}$. One way of solving this identification problem is to make sure that there is only one strict local minimum of $\mathcal{L}$ in $\mathring{\Theta}$. Indeed, if a solution to (3) - a local strict minimum θ^* - is *unique* in $\mathring{\Theta}$ then we know that θ^* is the *absolute* minimum of $\mathcal{L}$. Hence, we need to have the uniqueness of the solution θ^* to (3) (at least in some neighborhood of α_0) if, by reversing the problem, we want to identify α_0 given p_0 and θ^*.

As an illustrative example, let us first consider the case where the forecaster's model $M = \{f_{t+1}\}$ is correctly specified. In that case, the h-vector W_t contains all the relevant information from $\mathcal{F}_t$, so that the first order condition (3) is equivalent to

$$E_t[(1(Y_{t+1} - \theta^{*\prime} W_t < 0) - \alpha_0) \cdot |Y_{t+1} - \theta^{*\prime} W_t|^{p_0 - 1} | \mathcal{F}_t] = 0. \tag{5}$$

Note that for $p_0 = 1, 2$, and conditional on $\mathcal{F}_t$, the term $|Y_{t+1} - \theta^{*\prime} W_t|^{p_0-1}$ is strictly positive $a.s. - \mathcal{P}$ so that the condition (5) can only be satisfied if $1(Y_{t+1} - \theta^{*\prime} W_t < 0) - \alpha_0 = 0$, $a.s. - \mathcal{P}$. In other words, if the forecasting model $M = \{f_{t+1}\}$ is correctly specified, then $E_t[1(Y_{t+1} - \theta^{*\prime} W_t < 0] = \alpha_0$, so that the conditional α_0-quantile of the optimal forecast error $\varepsilon_{t+1}^* \equiv Y_{t+1} - \theta^{*\prime} W_t$ is exactly equal to zero. Hence, the optimal value θ^* is unique: $\theta^{*\prime} W_t = \overset{-1}{P_t}(\alpha_0)$, where $\overset{-1}{P_t}$ is the inverse of the conditional distribution function of Y_{t+1}. The uniqueness property in particular allows us to compute the value of α_0 used in the construction of the cost function $\mathcal{L}$, by inverting the preceding equation. Thus, we obtain

$$\alpha_0 = P_t(\theta^{*\prime} W_t). \tag{6}$$

Let us now turn to a more realistic case where the forecaster's model $M = \{f_{t+1}\}$ may be misspecified. The misspecification typically occurs when Y_{t+1} depends on some set of $\mathcal{F}_t$-measurable variables which are not contained in W_t. In this case, the first order condition (3) is weaker than (5) and the aforementioned property of the optimal forecast errors ε_{t+1}^* is no longer true. Hence, in presence of forecaster's model misspecification we cannot deduct from

(3) that the conditional α_0-quantile of the optimal forecast error ε^*_{t+1} is zero. In particular, this implies that the unicity of θ^* is not trivially verified, which makes the true value of the probability level α_0 more difficult to recover. Fortunately, by using the implicit function theorem we can show that, given $p_0 \in \mathbb{N}^*$, there exists an open subset F of $\mathring{\Theta}$ such that, for any $\alpha_0 \in (0,1)$, the equation (3) has a *unique* solution θ^* in F and that this solution is implicitly defined as a function $\theta_{p_0}(\alpha_0)$ of α_0. This result is established in the following Proposition.

Proposition 3 (Unicity) *Under Assumptions (A0)-(A2), given $p_0 \in \mathbb{N}^*$, there exist an open set F, $F \subseteq \mathring{\Theta}$, such that, for any $\alpha_0 \in (0,1)$, the equation* (3)*,*

$$E[W_t \cdot (1(Y_{t+1} - \theta^{*\prime} W_t < 0) - \alpha_0) \cdot |Y_{t+1} - \theta^{*\prime} W_t|^{p_0 - 1}] = 0,$$

has a unique solution θ^* in F. Moreover, the function $\theta^* = \theta_{p_0}(\alpha_0)$ defined implicitly by (3) *is bijective and continuously differentiable from* $(0,1)$ *to F with*

$$\theta'_{p_0}(\alpha) = \begin{cases} \{E[W_t W_t' \cdot p_t(\theta_{p_0}(\alpha)' W_t)]\}^{-1} \cdot E[W_t], & \textit{if } p_0 = 1, \\ \{E[W_t W_t' \cdot (\alpha + (1-2\alpha) \cdot P_t(\theta_{p_0}(\alpha)' W_t))]\}^{-1} \cdot E[W_t \cdot |Y_{t+1} - \theta_{p_0}(\alpha)' W_t|], & \textit{if } p_0 = 2, \\ \{(p_0 - 1) E[W_t W_t' \cdot (\alpha + (1-2\alpha) \cdot 1(Y_{t+1} - \theta_{p_0}(\alpha)' W_t < 0)) & \\ \cdot |Y_{t+1} - \theta_{p_0}(\alpha)' W_t|^{p_0 - 2}]\}^{-1} \cdot E[W_t \cdot |Y_{t+1} - \theta_{p_0}(\alpha)' W_t|^{p_0 - 1}], & \textit{if } p_0 > 2, \end{cases}$$

where P_t and p_t are the distribution function and the density of Y_{t+1} conditional on $\mathcal{F}_t$.

We now turn to the problem of estimating the true value α_0 used in the loss function $\mathcal{L}$ minimization problem (2). As previously, we are interested in recovering α_0 by assuming that the value of p_0 is already known by the forecast user. Recall that the forecast user need not know the forecasting model $M = \{f_{t+1}\}$ used to construct the forecasts. In other words, the components of the h-vector W_t need not be known and/or available in their entirety. We assume here that the forecast user knows and observes a sub-vector of W_t, whose dimension d is less than h and which we denote by V_t. As noted earlier, V_t being a sub-vector of W_t, the moment conditions (3) in particular imply that

$$E[V_t \cdot (1(Y_{t+1} - f^*_{t+1} < 0) - \alpha_0) \cdot |Y_{t+1} - f^*_{t+1}|^{p_0 - 1}] = 0. \tag{7}$$

The following lemma will be useful in the construction of an estimator for α_0.

Lemma 4 *Under Assumptions (A0)-(A3), given* $p_0 \in \mathbb{N}^*$ *and given* $f^*_{t+1} = \theta^{*\prime} W_t$ *where* θ^* *is the solution to* (3)*, the true value* $\alpha_0 \in (0,1)$ *is the unique minimum of a quadratic form*

$$\begin{aligned} Q_0(\alpha) \ \equiv \ & E[V_t \cdot (1(Y_{t+1} - f^*_{t+1} < 0) - \alpha) \cdot |Y_{t+1} - f^*_{t+1}|^{p_0 - 1}]' \cdot \\ & W \cdot E[V_t \cdot (1(Y_{t+1} - f^*_{t+1} < 0) - \alpha) \cdot |Y_{t+1} - f^*_{t+1}|^{p_0 - 1}] \ , \end{aligned}$$

i.e.

$$\alpha_0 = \frac{E[V_t \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}]' \cdot W \cdot E[V_t \cdot (1(Y_{t+1} - f^*_{t+1} < 0) - \alpha) \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}]}{E[V_t \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}]' \cdot W \cdot E[V_t \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}]} \tag{8}$$

where V_t *is a sub-vector of* W_t *and* W *is a positive definite weighting matrix.*

If we observe the sequence of optimal one-step-ahead forecasts $f^*_{t+1} \equiv \theta^{*\prime} W_t$ provided by the forecaster, we can estimate α_0 directly from equation (8). In practice however, we only observe the sequence $\{\hat{f}_{t+1}\}$ where $\hat{f}_{t+1} \equiv \hat{\theta}'_t w_t$ where $\hat{\theta}'_t$ is an estimate of θ^* obtained by using the data up to the date t. Let $n+1$ be the total number of periods available. There are $n - \tau + 1 \equiv T$ forecasts available, starting at $t = \tau + 1$ and ending at $n + 1 = T + \tau$. The first one-step-ahead forecast $\hat{f}_{\tau+1}$ of the random variable $Y_{\tau+1}$ is constructed as follows: the data from $s = 1$ to $s = \tau$, i.e. $(y_2, w'_1, \ldots, y_\tau, w'_{\tau-1})'$, is used to compute an estimate $\hat{\theta}_\tau$ of θ^*. The corresponding forecast of $y_{\tau+1}$ is then given by $\hat{f}_{\tau+1} = \hat{\theta}'_\tau w_\tau$. The second forecast $\hat{f}_{\tau+2}$ is obtained by computing $\hat{\theta}_{\tau+1}$ using the data available from $s = 1$ to $s = \tau + 1$, i.e. $(y_2, w'_1, \ldots, y_{\tau+1}, w'_\tau)$, and then forming $\hat{f}_{\tau+2} = \hat{\theta}'_{\tau+1} w_{\tau+1}$. By repeating the same procedure, for $t = n$, an estimate $\hat{\theta}'_n$ of θ^* is obtained by using the data $(y_2, w'_1, \ldots, y_n, w'_{n-1})'$, and the corresponding one-step-ahead forecast of y_{n+1} is given by $\hat{f}_{n+1} = \hat{\theta}'_n w_n$. To recap, the forecaster provide a sequence of $T = n - \tau + 1$ forecasts, $\{\hat{f}_{t+1}\}_{\tau \leqslant t < T+\tau}$, where for each t, $\tau \leqslant t < T + \tau$, the forecasts are constructed as $\hat{f}_{t+1} = \hat{\theta}'_t w_t$ and where $\hat{\theta}_t$ is an estimate of θ^* that relies on the data from period t and earlier, i.e. $(y_2, w'_1, \ldots, y_t, w'_{t-1})'$.

Having observed the sequence $\{\hat{f}_{t+1}\}_{\tau \leqslant t < T+\tau}$ provided by the forecaster, we now construct an estimator for α_0 based on equation (8). Given the T observations $(v'_\tau, \ldots, v'_{T+\tau-1})'$ of the

Figure 1: Description of the available data

d-vector V_t, we consider a linear Instrumental Variable (IV) estimator of α_0, $\hat{\alpha}_T$, defined as

$$\hat{\alpha}_T \equiv \frac{[T^{-1}\sum_{t=\tau}^{T+\tau-1} v_t|y_{t+1}-\hat{f}_{t+1}|^{p_0-1}]'\cdot\hat{S}^{-1}\cdot[T^{-1}\sum_{t=\tau}^{T+\tau-1} v_t(1(y_{t+1}-\hat{f}_{t+1}<0)|y_{t+1}-\hat{f}_{t+1}|^{p_0-1}]}{[T^{-1}\sum_{t=\tau}^{T+\tau-1} v_t|y_{t+1}-\hat{f}_{t+1}|^{p_0-1}]'\cdot\hat{S}^{-1}\cdot[T^{-1}\sum_{t=\tau}^{T+\tau-1} v_t|y_{t+1}-\hat{f}_{t+1}|^{p_0-1}]}, \tag{9}$$

where $\hat{S}$ is some consistent estimate of $S \equiv E[V_t V_t' \cdot (1(Y_{t+1}-\hat{f}_{t+1}<0)-\alpha)^2 \cdot |Y_{t+1}-\hat{f}_{t+1}|^{2p_0-2}]$. The consistency result for $\hat{\alpha}_T$ is as follows.

Proposition 5 (Consistency) *Given $p_0 = 1, 2$, let $\hat{\alpha}_T$ be the linear IV estimator defined in (9). Under Assumptions (A0)-(A6), $\hat{\alpha}_T$ exists with probability approaching one and $\hat{\alpha}_T \xrightarrow{p} \alpha_0$.*

In other words, even with the domain of α_0 not being compact, the linear IV estimator exists with probability approaching one and is moreover consistent for the true value α_0. Note that this result is particularly interesting since the construction of $\hat{\alpha}_T$ does not require the full knowledge of the h-vector W_t used by the forecaster. Indeed, by considering some publicly available sub-vector V_t of W_t, the forecast users can still consistently estimate the true value α_0 used in the loss function minimization problem (2).

Proposition 6 (Asymptotic Normality) *Given* $p_0 = 1, 2$*, let* $\hat{\alpha}_T$ *be the linear IV estimator defined in* (9)*. Under Assumptions (A0)-(A4), (A5') and (A6),* $\hat{\alpha}_T$ *exists with probability approaching one and*

$$T^{1/2}(\hat{\alpha}_T - \alpha_0) \xrightarrow{d} \mathcal{N}(0, (h_0' \cdot S^{-1} \cdot h_0)^{-1}),$$

where $S \equiv E[V_t V_t' \cdot (1(Y_{t+1} - \hat{f}_{t+1} < 0) - \alpha)^2 \cdot |Y_{t+1} - \hat{f}_{t+1}|^{2p_0-2}]$ *and* $h_0 \equiv E[V_t \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}]$.

In other words, the linear IV estimator $\hat{\alpha}_T$ is asymptotically normal, with asymptotic variance which does not depend on neither W_t nor θ^*, which are a priori unknown to the forecast user. Indeed, the asymptotic variance of $\hat{\alpha}_T$ is identical to the one obtained with a standard GMM-type estimator. It is interesting to note that this result stems from the fact that $\hat{\theta}_t$, for $\tau \leqslant t < T$, and $\hat{\alpha}_T$ are obtained as solutions to the same first order condition (7), which moreover is linear in α. Were $\hat{\theta}_t$, $\tau \leqslant t < T$, and $\hat{\alpha}_T$ obtained with different loss functions, they would no longer satisfy the same first order condition. Hence, the uncertainty of parameter estimates $\hat{\theta}_t$, $\tau \leqslant t < T$, would in that case affect the asymptotic variance of $\hat{\alpha}_T$ and make it substantially more complicated.[6]

In practice, the computation of the linear IV estimator $\hat{\alpha}_T$ is done iteratively. According to equation (9), $\hat{\alpha}_T$ depends on a consistent estimator of S^{-1}. For example, S can be consistently estimated by replacing the population moment by a sample average and the true parameter by its estimated value, so that $\hat{S}(\bar{\alpha}_T) = T^{-1} \sum_{t=\tau}^{T+\tau-1} v_t v_t' (1(y_{t+1} - \hat{f}_{t+1} < 0) - \bar{\alpha}_T)^2 |y_{t+1} - \hat{f}_{t+1}|^{2p_0-2}$, where $\bar{\alpha}_T$ is some consistent initial estimate of α_0, or by using some heteroskedasticity and autocorrelation robust estimator, such as Newey and West's (1987) estimator. The computation of $\hat{\alpha}_T$ is then carried out by first choosing a $d \times d$ identity weight matrix $S = I_{d \times d}$ and using (9) to compute the corresponding $\hat{\alpha}_{T,1}$. The resulting new weight matrix $\hat{S}^{-1}(\hat{\alpha}_{T,1})$ is more efficient than the previous one, which when plugged into (9) leads to a new estimator $\hat{\alpha}_{T,2}$. The last two steps can then be repeated until $\hat{\alpha}_{T,j}$ equals its previous value $\hat{\alpha}_{T,j-1}$.

[6]For general results on asymptotic inference in presence of parameter uncertainty see, e.g., West, 1996, West and McCracken, 1998. Note that the assumptions placed on the forecast method differ from those examined here.

Finally, we can use the moment conditions to test the hypothesis that the forecasts are rational with respect to available information within the class of loss functions (i.e. without specifying a value for α). We note that if indeed the forecasts are rational, then V_t is a subvector of W_t. Thus all moment conditions must hold (which is how we obtain estimates for α above). A different question that can be asked is that given V_t, does there exist some value for α for which the forecasts are rational? The usual test for overidentification of the overidentified IV estimation (so long as d>1) tests this proposition. One degree of freedom is used in the estimation of the loss parameter, so the resultant test statistic, i.e.

$$\begin{aligned} J &= T^{-1}\Bigg([\sum_{t=\tau}^{T+\tau-1} v_t(1(y_{t+1}-\hat{f}_{t+1}<0-\hat{\alpha}_T)|y_{t+1}-\hat{f}_{t+1}|^{p_0-1}]'\hat{S}^{-1} \\ & [\sum_{t=\tau}^{T+\tau-1} v_t(1(y_{t+1}-\hat{f}_{t+1}<0-\hat{\alpha}_T)|y_{t+1}-\hat{f}_{t+1}|^{p_0-1}] \Bigg) \end{aligned}$$

The test, from the results of Proposition 6, is distributed χ^2_{d-1} and rejects for large values as usual. The typical testing assumption of assuming mean square error loss then is closely related to this test when p_0 is chosen to be equal to 2. The difference is that if indeed $\alpha_0 = 0.5$, the typical test imposes this whereas here the test uses a consistent estimate. However, if α_0 were different from 0.5 then the typical test would have power in this direction - here the use of the consistent test avoids this problem, size will be controlled if the forecasters have chosen a different value for α. Asymptotically there is no loss from relaxing this parameter value, however there is clearly a gain in terms of directing power in the desired direction.

4 Simulation Results

We examine aspects of the methods derived above in a Monte Carlo experiment. Psuedo data were generatedby the model given by

$$y_{t+1} = \theta' w_t + u_t$$

where the vector $w_t = [1, w_{1t}, w_{2t}]$ where $w_{1t}\tilde{}N(1,1), w_{2t}\tilde{}N(-1,1), \theta = [1, 0.5, 0.5]$ and $u_t\tilde{}N(0, 0.5)$. Experiments were undertaken for differnent numbers of initial values available for estimating θ recursively (such data are available to the forecaster before the initial recorded forecast is made), denoted n_0, and for different numbers of data available for estimation of α and testing, denoted n_f (for number of forecasts). For $p = 1$ recursive forecasts are made using quantile regression methods and for $p = 2$ the nonlinear least squares method of estimation in Newey and Powell (1987) is used to estimate θ recursively.

Table 1 examines, for various samples sizes and values of α, the size of t-tests testing $\hat{\alpha} = \alpha_0$ (i.e. the true value) against two sided alternatives with size 5%. Results for $p = 1$ (Lin-Lin) are in the first panel and for $p = 2$ (Quad-Quad) in the second panel. In all cases we have not used any instruments other than the constant term, i.e. $V_t = 1$. Size is fairly well controlled overall, less so when α is far from one half (on average). The reason for this is straightforward, for the asymmetric models the 'errors' are less well balanced above and below the true value, hence we obtain asymmetric small sample distributions and require a larger n for the central limit theorem to provide a good approximation (this is identical to the usual result in applying the central limit theorem to the bernoulli distributed outcomes). Size is less well controlled for the Quad-Quad loss function. More observations (either in or out of sample) tend to help with size control.

Table 2 repeats the exercise in Table 1 but now employs the two available instruments, i.e. $V_t = w_t$. The use of the extra instruments results in larger size distortions across the board. The probelm is again more of an issue for the Quad-Quad case than the Lin-Lin loss function. As we expect, this is less of a problem when there are both more in sample and out of sample observations available. As before, more out of sample observations help more than more in sample observations in controlling size. Problems are again greater the further is α from one half.

We examine the tests for overidentification in Table 3. These tests examine whether or not the moment conditions are compatible with some α, left unspecified in the testing procedure. We report size of the tests under the null for the same cases as the previous two tables. For these tests, size is well controlled. The tests tend to be undersized rather than oversized, again the size departing from nominal size (of 5%) more when α is further from one half. When α is at one half, the sizes are very close to nominal size for all sample sizes.

Increasing the sample size helps, with adding more out of sample observations appearing to be more useful.

5 Government Deficit Forecasts

In this section we apply our estimation methods and tools for inference to forecasts of government budget deficits for the G7 countries produced by two international organizations, namely the IMF and the OECD. This data set is well suited to demonstrate our methods since, as pointed out by Artis and Marcellino (2001) fiscal forecast errors are likely to be particularly sensitive to political pressures and "the political context in which fiscal deficit forecasts emerge may well be one in which the costs of forecast misses are not symmetric." (Artis and Marcellino, page 20). This point is echoed by Campbell and Ghysels (1995) in the context of an analysis of federal budget projections.

The data that we use is identical to that considered in Artis and Marcellino (2001) and comprises budget deficit forecasts, reported as a percentage of GDP, for the G7 countries.[7] The data is reported as budget surpluses so that a budget deficit takes a negative value. Following standard practice forecast errors are defined as realizations minus predicted values. Since almost all realizations and predictions are negative, a positive forecast error corresponds to a larger predicted deficit than actually occurred. We refer to this as an overprediction of the budget deficit (underprediction of the budget surplus).

For the OECD, sufficient data was only available on four of these countries, namely France, Germany, Italy and the UK. In both cases the data comprises current year (published in May each year) and year-ahead forecasts (published in October of year t for year $t+1$). The IMF sample goes from 1976 to 1995 and thus has 20 observations. The OECD data contains between 18 and 21 data points. Clearly these are not large samples, so caution should be exercised in the interpretation of the statistical results.

[7] We are grateful to Massimiliano Marcellino for providing this data. The data source is the IMF's World Economic Outlook and the OECD's Economic Outlook. Artis and Marcellino (2001) also consider forecasts from the European Commission, but this data set is very short (14 observations) so we decided not to include it here.

Table 4 presents summary statistics for the forecast errors. The mean forecast error suggests that overpredictions of budget deficits is the typical situation, although there are also some countries, most notably Canada, with underpredictions. For some countries there are strong imbalances between the number of positive ($n+$) and negative ($n-$) forecast errors, particularly in the case of current-year IMF forecasts for Italy, Japan, UK and the US, where 15 or 17 of 20 forecast errors are positive. While most countries produce a majority of positive forecast errors, Canada is the exception with a majority of negative forecast errors. The current year RMSE values vary from 0.53 (France) to 2.14 (Italy), while the range goes from 0.83 to 2.33 for the 1-year ahead forecasts. For all countries the RMSE values are higher at the 1-year ahead horizon than at the current-year horizon.

The previous sections show how to test forecast optimality in conjunction with the assumption that the loss function belongs to a particular class. In our empirical tests we adopt the strategy of first assuming that the loss function is lin-lin. Authors such as Granger and Newbold (1986) have argued that lin-lin loss approximate other classes of asymmetric loss functions: "The linear cost function may well provide a good approximation to nonsymmetric cost functions that arise in practice" (Granger and Newbold (1986), page 126). However, as a robustness check of our results we subsequently conduct empirical tests under the assumption of quad-quad loss.

5.1 Evidence of Asymmetric Loss

Under the assumption that the loss function is piecewise linear (lin-lin), Table 5 presents the estimated asymmetry parameter ($\hat{\alpha}$) along with its standard error and p-values for tests of the null hypothesis of symmetric loss, i.e. $\alpha = 0.5$. The parameter estimates and test results are of separate economic interest since they may suggest the type of objectives the forecaster was operating under.

To explore the robustness of our results with respect to the number and type of instruments, we report results for four separate sets of instruments: (i) a constant; (ii) a constant and the lagged (real time) forecast error; (iii) a constant and the lagged budget deficit; (iv) a constant, the lagged forecast error and the lagged budget deficit. Given the small sample size, we do not consider more than three instruments.

First consider the current-year IMF results when the model is exactly identified and a constant is the only instrument. Five of seven countries generate α-estimates below one-half, one country (France) has an estimate of exactly one-half and another country (Canada) has an alpha estimate of 0.65. The null of symmetry is very strongly rejected for Italy, Japan, UK and the US. Similar results are obtained for the year-ahead IMF predictions, where the α-estimates are statistically significantly different from one-half for Japan, UK and the US.

In the overidentified models with two or three instruments the current-year results tend to be even stronger since the standard errors of $\hat{\alpha}$ tend to decline. Hence, the null of symmetric loss is rejected with p-values less than 0.01 for Italy, Japan, the UK and the US. In each case the point estimates for these four countries are at or below 0.25, thus suggesting economically strong evidence of asymmetry. At the year-ahead horizon the null of symmetric loss continues to be rejected at or below the 5% level for Japan, UK and the US. The null is marginally rejected for Italy for two of the four sets of instruments.

Turning to the OECD data, for the current year predictions, three out of four countries generate estimates of α below one half. Irrespective of the set of instruments used, the null of symmetric loss is rejected at the five percent significance level for Germany and Italy, while it is rejected at the 10 percent significance level for France. The evidence of asymmetric loss is somewhat weaker at the 1-year-ahead horizon, where the null is only rejected for Germany.

These results suggest that international organizations such as the IMF and the OECD systematically overpredict government losses and forecast budget deficits larger than they turn out to be. This is consistent with a loss function that penalizes underpredictions more heavily than overpredictions and our parameter estimates quantify the extent to which this is the case. Indeed, the point estimates of α suggest very strong asymmetries in the loss function both from an economic and a statistical point of view.

5.2 Tests of Forecast Rationality

Ultimately we are interested in testing whether the IMF and OECD forecasts are rational. We first conduct our tests under the assumption of symmetric loss. This is the null hypothesis that has been maintained throughout the literature, so it seems a natural starting point for our analysis. In our context we can test this hypothesis by imposing $\alpha = 1/2$ and examining the J-test which follows a χ^2_d-distribution under this restriction.

The outcome of the joint tests of rationality and $\alpha = 1/2$ is reported in Table 6. Overall, there are 36 cases where the null hypothesis is rejected at or below the 5% level. In the IMF data there is very strong evidence against the composite null hypotheses for Italy, Japan, the UK and the US, while the OECD data leads to rejections of the null for Germany. Notice that this is the same list of countries for which we found overwhelming evidence of asymmetric loss in Table 2.

Since the rejection of symmetric loss and forecast rationality may well be due to the symmetry assumption, we next test whether forecast rationality gets rejected once we allow for asymmetric loss. To investigate this possibility, Table 7 reports outcomes from adopting the J-test to our forecast data when α is not constrained to take a particular value.

The results are very interesting and in complete contrast to those found in Table 6. There is only very weak evidence against the null hypothesis of forecast rationality. Overall there are only seven cases where the null gets rejected at the 5% significance level. None of the $J-$tests associated with the current-year forecasts are statistically significant at the five percent level. At the 1-year horizon the IMF results for the UK, US and France indicate absence of optimality but it appears that the rejections of the composite null hypothesis of symmetric loss and optimal forecasts can be ascribed to asymmetric loss in the case of Italy and Japan.

5.3 Robustness to the Shape of the Loss Function

To check the robustness of our findings with respect to the assumed shape of the loss function and to consider a family of loss functions that embeds MSE loss, Tables 8-10 report results for the quad-quad loss function.

First consider the evidence of asymmetries in the quad-quad loss function. In the current year IMF forecasts there continues to be strong evidence against symmetric loss for Italy, Japan the UK and the US, all of which produce estimates of α below one-half. Interestingly, there is now also significant evidence of asymmetric loss for Canada, albeit with an α-estimate above one-half. At the 1-year forecast horizon Canada, France, Italy and Japan all produce strong evidence of asymmetric loss. In the OECD data there is strong evidence of asymmetric loss for Germany and, in the case of current-year predictions, also for Italy.

Turning next to the question of forecast rationality, Table 9 shows the outcome of tests of the joint hypothesis of MSE loss and rationality. In the current-year IMF forecasts this null hypothesis is strongly rejected for Canada, Japan and the UK and for Germany and Italy in the case of the OECD forecasts. There is some evidence in both data sets of loss asymmetry also for Italy. At the 1-year horizon the evidence against the null hypothesis is even stronger and the null gets rejected in the IMF data for Canada, France, Italy, Japan and the UK and, in the OECD data, for Germany and (for two out of four tests) the UK. Overall, the null is rejected at the 5 percent level in 40 cases.

Allowing for asymmetric quadratic loss, as we do in Table 10, the evidence against rationality is far weaker. In no case does the null get rejected at the 5% level for the current-year data. At the 1-year horizon, the null is strongly rejected only by the IMF predictions for the UK and, to some extent, France. Overall the null is only rejected at the 5% level in seven cases in Table 10.

Overall our conclusions thus appear to be very robust with respect to the assumed class of loss functions. This is fortunate since, in the absence of a more detailed analysis of the political pressures facing the international organizations, it is difficult to choose one over the other. Consistent with our findings under lin-lin loss, the tests for forecast rationality are significantly changed once we allow for asymmetric loss. While the joint null hypothesis of MSE loss and forecast rationality is strongly rejected in a large number of cases, there is far weaker evidence against this null once asymmetric quadratic loss is considered.

Artis and Marcellino (2001) perform a related exercise in which they first back out the asymmetry parameter of a quad-quad loss function and then, conditional on this estimate test whether the resulting forecast errors are serially correlated and uncorrelated with the forecast. Conditional on their first-stage parameter estimate they reject the null that the IMF forecasts are efficient for France, Germany, Italy, Japan and the US. Unfortunately the standard errors used in the second step of their analysis do not account for the first-stage estimation of the asymmetry parameter which introduces a generated regressor problem. In a sample as small as that considered here, this could be a major concern. In fact, our results are very different and shows that it is only really for the UK and France that there is evidence against forecast rationality and quad-quad loss.

6 Conclusion

This paper provided theory for identification and estimation of parameters of loss functions applicable to situations where time-series data on forecasts is available. We also provided test statistics that can be used when testing whether information is used efficiently in the computation of the forecasts. In applications our methods suggest that there is systematic evidence that international organizations such as the IMF and the OECD have asymmetric loss and that the composite hypothesis of symmetric loss and forecast rationality is rejected for many countries in our sample. Once we allow for asymmetric loss, there is far weaker evidence against forecast rationality.

Although our theory was derived in the context of commonly used loss functions, it can be adopted in many areas where data is observed on variables that are the outcome of an optimizing agent's actions. One example is investors' portfolio weights in asset classes such as stocks and, bonds and other asset classes. It is commonly assumed that investors' utility belongs to a certain family such as mean-variance or power utility. For each utility specification, one would of course have to verify that the primitive assumptions in our paper are satisfied. Provided that this could be done, it should be possible to back out the investor's risk aversion parameters from the observed portfolio weights and time-series on the underlying asset returns and any instruments used by the investors to model the conditional return distribution.

7 Appendix A: Assumptions

(A0) θ^* is interior to Θ, i.e. $\theta^* \in \mathring{\Theta}$;

(A1) the h-vector W_t (with the first component 1) is such that, given $p_0 \in \mathbb{N}^*$, for all $\theta \in \mathring{\Theta}$, $E[W_t \cdot |Y_{t+1} - \theta' W_t|^{p_0-1}] \neq 0$, $E[W_t \cdot 1(Y_{t+1} - \theta' W_t < 0) \cdot |Y_{t+1} - \theta' W_t|^{p_0-1}] \neq 0$, and $E[W_t W_t']$ exists and is positive definite;

(A2) the density of Y_{t+1} conditional on $\mathcal{F}_t$ is strictly positive, i.e. $p_t(y) > 0$, for every $y \in \mathbb{R}$;

(A3) the d-vector V_t is a sub-vector of the h-vector W_t ($d \leqslant h$) with the first component 1 and there exists a constant $m > 0$ such that $||W_t||^2 = W_t' W_t \leqslant m$, $a.s. - \mathcal{P}$;

(A4) for every t, $\tau \leqslant t < T + \tau$, $\hat{\theta}_t$ is a consistent estimate of θ^* and $\theta^* \in F$;

(A5) the stochastic processes Y_t and W_t are stationary and α-mixing with mixing coefficient α of size $-r/(r-2)$, $r > 2$, and, given $p_0 \in \mathbb{N}^*$, there exist some $\delta_f > 0$ and $\Delta_f > 0$ such that $E[(Y_{t+1} - \hat{f}_{t+1})^{(2+\delta_f)(p_0-1)}] \leqslant \Delta_f < \infty$ and some $\delta_W > 0$ and $\Delta_W > 0$ such that $E[||W_t||^{2+\delta_W}] \leqslant \Delta_W < \infty$;

(A5') the stochastic processes Y_t and W_t are stationary and α-mixing with mixing coefficient α of size $-r/(r-2)$, $r > 2$, and, given $p_0 \in \mathbb{N}^*$, there exist some $\delta_f > 0$ and $\Delta_f > 0$ such that $E[(Y_{t+1} - \hat{f}_{t+1})^{(4+\delta_f)(p_0-1)}] \leqslant \Delta_f < \infty$ and some $\delta_W > 0$ and $\Delta_W > 0$ such that $E[||W_t||^{4+\delta_W}] \leqslant \Delta_W < \infty$;

(A6) the density of Y_{t+1} conditional on $\mathcal{F}_t$ is bounded, i.e. there exist some $M > 0$ such that, $\sup_{y \in \mathbb{R}} p_t(y) \leqslant M < \infty$;

8 Appendix B: Proofs

Proof of Proposition 1. We know that if θ^* is the minimum of $\mathcal{L}$ in $\mathring{\Theta}$, i.e. if θ^* is the solution to the minimization problem

$$\min_{\theta \in \Theta} E\{[\alpha_0 + (1 - 2\alpha_0) \cdot 1(Y_{t+1} - \theta' W_t < 0)] \cdot |Y_{t+1} - \theta' W_t|^{p_0}\} \equiv \Sigma(\theta),$$

with $\Sigma(\theta)$ continuously differentiable on Θ, and $\theta^* \in \mathring{\Theta}$ (Assumption (A0)), then θ^* satisfies the first order condition $\nabla_\theta \Sigma(\theta^*) = 0$ (see, e.g., Theorem 3.7.13 in Schwartz, 1997, vol 2, p 168). Let $\Sigma_{t+1}(\theta) \equiv [\alpha_0 + (1-2\alpha_0) \cdot 1(Y_{t+1} - \theta' W_t < 0)] \cdot |Y_{t+1} - \theta' W_t|^{p_0}$. The function $\Sigma_{t+1}(\theta)$ is continuously differentiable on $\Theta \backslash A_{t+1}$ where $A_{t+1} \equiv \{\theta \in \Theta : Y_{t+1} = \theta' W_t\}$. Let $\nabla_\theta \Sigma_{t+1}(\theta)$ be the gradient of $\Sigma_{t+1}(\theta)$ on $\Theta \backslash A_{t+1}$. We have

$$\Sigma(\theta) = E[\Sigma_{t+1}(\theta)] = E\{E_{t+1}[\Sigma_{t+1}(\theta)]\},$$

so that

$$\begin{aligned}
\nabla_\theta \Sigma(\theta) &= E\{\nabla_\theta E_{t+1}[\Sigma_{t+1}(\theta)]\} \\
&= E\{\nabla_\theta E_{t+1}[\Sigma_{t+1}(\theta) \cdot 1(\theta \in A^c_{t+1})]\} + E\{\nabla_\theta E_{t+1}[\Sigma_{t+1}(\theta) \cdot 1(\theta \in A_{t+1})]\} \\
&= E\{\nabla_\theta \Sigma_{t+1}(\theta) \cdot E_{t+1}[1(\theta \in A^c_{t+1})]\} + E\{\nabla_\theta \Sigma_{t+1}(\theta) \cdot E_{t+1}[1(\theta \in A_{t+1})]\},
\end{aligned}$$

where $E_{t+1}[1(\theta \in A^c_{t+1})] = \mathcal{P}(A^c_\theta)$ with $A^c_\theta \equiv \Omega \backslash A_\theta$ and $A_\theta \equiv \{\omega \in \Omega : Y_{t+1}(\omega) - \theta' W_t(\omega)\}$. Hence, $E_{t+1}[1(\theta \in A^c_{t+1})] = 1$ and $E_{t+1}[1(\theta \in A_{t+1})] = 0$. $\Sigma(\theta)$ is therefore continuously differentiable on Θ and we have

$$\begin{aligned}
\nabla_\theta \Sigma(\theta) = \; & -\alpha_0 E\{W_t \cdot [1 - 2 \cdot 1(Y_{t+1} - \theta' W_t < 0)] \cdot |Y_{t+1} - \theta' W_t|^{p_0 - 1}\} \\
& +(1-2\alpha_0) E[\nabla_\theta 1(Y_{t+1} - \theta' W_t < 0) \cdot |Y_{t+1} - \theta' W_t|^{p_0}] \\
& -(1-2\alpha_0) E\{1(Y_{t+1} - \theta' W_t < 0) \cdot W_t \cdot [1 - 2 \cdot 1(Y_{t+1} - \theta' W_t < 0)] \cdot |Y_{t+1} - \theta' W_t|^{p_0 - 1}\},
\end{aligned}$$

so that

$$\begin{aligned}
\nabla_\theta \Sigma(\theta) = \; & E[W_t \cdot (1(Y_{t+1} - \theta' W_t < 0) - \alpha_0) \cdot |Y_{t+1} - \theta' W_t|^{p_0 - 1}] \\
& +(1-2\alpha_0) E[\nabla_\theta 1(Y_{t+1} - \theta' W_t < 0) \cdot |Y_{t+1} - \theta' W_t|^{p_0}].
\end{aligned}$$

Note that

$$\nabla_\theta 1(Y_{t+1} - \theta' W_t < 0) = W_t \cdot \delta(\theta' W_t - Y_{t+1})$$

where δ represents the Dirac function, i.e. for all $x \in \mathbb{R}^*$, $\delta(x) = 0$ and $\int_{\mathbb{R}} \delta(x) dx = 1$. Knowing that for any real function $\varphi : \mathbb{R} \to \mathbb{R}$ we have $\int_{\mathbb{R}} \varphi(x) \delta(x) dx = \varphi(0)$, we obtain

$$E[W_t \cdot \delta(\theta' W_t - Y_{t+1}) \cdot |Y_{t+1} - \theta' W_t|^{p_0}] = 0,$$

since $p_0 \in \mathbb{N}^*$. Thus,

$$\nabla_\theta \Sigma(\theta) = E[W_t \cdot (1(Y_{t+1} - \theta' W_t < 0) - \alpha_0) \cdot |Y_{t+1} - \theta' W_t|^{p_0 - 1}].$$

Given the values of p_0 and α_0 if θ^* is the minimum of $\Sigma(\theta)$ then θ^* is a solution to $\nabla_\theta \Sigma(\theta^*) = 0$, i.e. we have

$$E[W_t \cdot (1(Y_{t+1} - \theta' W_t < 0) - \alpha_0) \cdot |Y_{t+1} - \theta' W_t|^{p_0 - 1}] = 0,$$

which completes the proof of Proposition 1. ■

Proof of Proposition 2. We derive the set of sufficient conditions for $\theta^* \in \mathring{\Theta}$ to be a solution to the minimization problem

$$\min_{\theta \in \Theta} E\{[\alpha_0 + (1 - 2\alpha_0) \cdot 1(Y_{t+1} - \theta' W_t < 0)] \cdot |Y_{t+1} - \theta' W_t|^{p_0}\} \equiv \Sigma(\theta).$$

We know that θ^* is a strict local minimum of $\Sigma(\theta)$ on $\mathring{\Theta}$ if $\nabla_\theta \Sigma(\theta^*) = 0$ and $\Delta_{\theta\theta} \Sigma(\theta^*)$ positive definite (see, e.g., Theorem 3.7.13 in Schwartz, 1997, vol 2, p 169). Recall that,

$$\nabla_\theta \Sigma(\theta) = E[W_t \cdot (1(Y_{t+1} - \theta' W_t < 0) - \alpha_0) \cdot |Y_{t+1} - \theta' W_t|^{p_0 - 1}],$$

so if θ^* satisfies the moment condition (3) then $\nabla_\theta \Sigma(\theta^*) = 0$. Note that by Assumption (A1), we know that $E[W_t \cdot |Y_{t+1} - \theta' W_t|^{p_0 - 1}] \neq 0$ and $E[W_t \cdot 1(Y_{t+1} - \theta' W_t < 0) \cdot |Y_{t+1} - \theta' W_t|^{p_0 - 1}] \neq 0$ so that $\nabla_\theta \Sigma(\theta)$ is not identically equal to zero for all $\theta \in \mathring{\Theta}$. We now need to show that θ^* is a strict minimum of $\Sigma(\theta)$. Note that we have

$$\begin{aligned}
\Delta_{\theta\theta} \Sigma(\theta) = \; & E[W_t W_t' \cdot \delta(\theta' W_t - Y_{t+1}) \cdot |Y_{t+1} - \theta' W_t|^{p_0 - 1}] \\
& -(p_0 - 1) \cdot E\left\{W_t W_t' \cdot [1(Y_{t+1} - \theta' W_t < 0) - \alpha_0]\right. \\
& \left. \cdot [1 - 2 \cdot 1(Y_{t+1} - \theta' W_t < 0)] \cdot |Y_{t+1} - \theta' W_t|^{p_0 - 2}\right\} \\
= \; & E[W_t W_t' \cdot \delta(\theta' W_t - Y_{t+1}) \cdot |Y_{t+1} - \theta' W_t|^{p_0 - 1}] \\
& +(p_0 - 1) \cdot E\{W_t W_t' \cdot [\alpha_0 + (1 - 2\alpha_0) \cdot 1(Y_{t+1} - \theta' W_t < 0)] \cdot |Y_{t+1} - \theta' W_t|^{p_0 - 2}\}
\end{aligned}$$

where $\delta : \mathbb{R} \to \mathbb{R}$ is the Dirac function i.e. $\delta(x) = 0$ if $x \neq 0$ and $\int_{\mathbb{R}} \delta(x)dx = 1$. Thus

$$\begin{aligned}
\Delta_{\theta\theta}\Sigma(\theta) = \;& E[W_t W_t{}' \cdot \delta(\theta' W_t - Y_{t+1}) \cdot |Y_{t+1} - \theta' W_t|^{p_0 - 1}] \\
& +(p_0 - 1)E[W_t W_t{}' \cdot (\alpha_0 + (1 - 2\alpha_0) \cdot 1(Y_{t+1} - \theta' W_t < 0)) \cdot |Y_{t+1} - \theta' W_t|^{p_0 - 2}] \\
= \;& 1(p_0 = 1) \cdot E[W_t W_t{}' \cdot \delta(\theta' W_t - Y_{t+1})] \\
& +1(p_0 > 1) \cdot \{E[W_t W_t{}' \cdot \delta(\theta' W_t - Y_{t+1}) \cdot |Y_{t+1} - \theta' W_t|^{p_0 - 1}] \\
& +(p_0 - 1)E[W_t W_t{}' \cdot (\alpha_0 + (1 - 2\alpha_0) \cdot 1(Y_{t+1} - \theta' W_t < 0)) \cdot |Y_{t+1} - \theta' W_t|^{p_0 - 2}]\} \\
= \;& 1(p_0 = 1) \cdot E\{W_t W_t{}' \cdot E_t[\delta(\theta^{*\prime} W_t - Y_{t+1})]\} + 1(p_0 > 1) \cdot \\
& (p_0 - 1)E\{W_t W_t{}' \cdot E_t[(\alpha_0 + (1 - 2\alpha_0) \cdot 1(Y_{t+1} - \theta' W_t < 0)) \cdot |Y_{t+1} - \theta' W_t|^{p_0 - 2}]\} \\
= \;& 1(p_0 = 1) \cdot E[W_t W_t{}' \cdot p_t(\theta^{*\prime} W_t)] \\
& +1(p_0 = 2) \cdot E[W_t W_t{}' \cdot (\alpha_0 + (1 - 2\alpha_0) \cdot P_t(\theta' W_t))] \\
& +1(p_0 > 2) \cdot (p_0 - 1)E[W_t W_t{}' \cdot (\alpha_0 + (1 - 2\alpha_0) \cdot 1(Y_{t+1} - \theta' W_t < 0)) \cdot |Y_{t+1} - \theta' W_t|^{p_0 - 2}],
\end{aligned}$$

where P_t and p_t are the distribution function and the density of Y_{t+1} conditional on $\mathcal{F}_t$. We need to ensure that $\Delta_{\theta\theta}\Sigma(\theta^*)$ is positive definite. Consider the following two cases separately: $p_0 = 1$ and $p_0 > 1$.

CASE $p_0 = 1$: note that in that case for any $\varphi \in \mathbb{R}^k$ we have

$$\varphi' \Delta_{\theta\theta}\Sigma(\theta)\varphi = E[\varphi' W_t W_t{}' \varphi \cdot p_t(\theta' W_t)],$$

where p_t is the density of Y_{t+1} conditional on $\mathcal{F}_t$. By imposing the strict positivity of the conditional density p_t (Assumption (A2)), we have that

$$\varphi' \Delta_{\theta\theta}\Sigma(\theta)\varphi = 0 \Rightarrow \varphi' W_t W_t{}' \varphi = 0, a.s. - \mathcal{P} \Rightarrow \varphi' E[W_t W_t{}']\varphi = 0,$$

which, by Assumption (A1), in turn implies $\varphi = 0$. Hence, for any $\theta \in \mathring{\Theta}$ the matrix $\Delta_{\theta\theta}\Sigma(\theta)$ is positive definite, therefore it is positive definite at θ^* which is then a strict local minimum of $\Sigma(\theta)$ on $\mathring{\Theta}$.

CASE $p_0 > 1$: note that in that case we have

$$\varphi' \Delta_{\theta\theta}\Sigma(\theta)\varphi = (p_0 - 1)E\{\varphi' W_t W_t{}' \varphi \cdot E_t[(\alpha_0 + (1 - 2\alpha_0) \cdot 1(Y_{t+1} - \theta' W_t < 0)) \cdot |Y_{t+1} - \theta' W_t|^{p_0 - 2})]\},$$

and, conditional on $\mathcal{F}_t$, the conditional expectation on the right hand side of the previous equality is strictly positive, $a.s. - \mathcal{P}$, for any $(\alpha_0, \theta) \in (0, 1) \times \Theta$. Therefore, we again have

$$\varphi' \Delta_{\theta\theta}\Sigma(\theta)\varphi = 0 \Rightarrow \varphi' W_t W_t{}' \varphi = 0, a.s. - \mathcal{P} \Rightarrow \varphi' E[W_t W_t{}']\varphi = 0,$$

so that by Assumption (A1) $\varphi = 0$. Hence, for every $\theta \in \mathring{\Theta}$ the matrix $\Delta_{\theta\theta}\Sigma(\theta)$ is positive definite, and θ^* is a strict local local minimum of $\Sigma(\theta)$ on $\mathring{\Theta}$. This completes the proof of Proposition 2. ■

Proof of Proposition 3. Given $p_0 = 1, 2$, let the function $h_{p_0} : (0,1) \times \Theta \to \mathbb{R}^k$ be defined as

$$h_{p_0}(\alpha, \theta) \equiv E[W_t \cdot (1(Y_{t+1} - \theta' W_t < 0) - \alpha) \cdot |Y_{t+1} - \theta' W_t|^{p_0 - 1}], \tag{10}$$

so that the first order condition (3) is equivalent to $h_{p_0}(\alpha_0, \theta^*) = 0$. In order to show that the results from Proposition 3 hold, we use the implicit function theorem (see, e.g., Theorem 3.8.5. in Schwartz, 1997, vol 2, p 185). In order to do so, we need to show that (i) the function $h_{p_0} : (0,1) \times \Theta \to \mathbb{R}^k$ is continuously differentiable on $(0,1) \times \Theta$, and (ii) for every $\alpha_0 \in (0,1)$, the $\mathbb{R}^k \times \mathbb{R}^k$-matrix $\partial h_{p_0}/\partial\theta(\alpha_0, \theta^*)$ is nonsingular, i.e. $[\partial h_{p_0}/\partial\theta(\alpha_0, \theta^*)]^{-1}$ exists. According to equation (10) the function h_{p_0} is linear in α and we have

$$h_{p_0}(\alpha, \theta) = E[W_t \cdot 1(Y_{t+1} - \theta' W_t < 0) \cdot |Y_{t+1} - \theta' W_t|^{p_0 - 1}] - \alpha E[W_t \cdot |Y_{t+1} - \theta' W_t|^{p_0 - 1}].$$

The differentiability of $h_{p_0}(\cdot, \theta) : (0,1) \to \mathbb{R}^k$ is therefore trivially verified and, for every $\theta \in \Theta$, we have

$$\frac{\partial h}{\partial \alpha}(\alpha, \theta) = -E[W_t \cdot |Y_{t+1} - \theta' W_t|^{p_0 - 1}],$$

which is independent of α. Therefore, the function $\partial h_{p_0}/\partial\alpha(\cdot, \theta) : (0,1) \to \mathbb{R}^k$ is continuous on $(0,1)$. We now turn to the study of $h_{p_0}(\alpha, \cdot) : \Theta \to \mathbb{R}^k$. Note that $\dfrac{\partial h_{p_0}}{\partial \theta}(\alpha, \theta) = \Delta_{\theta\theta}\Sigma(\theta)$ where $\Sigma(\theta)$ is defined as in (**??**), so that

$$\begin{aligned}
\frac{\partial h_{p_0}}{\partial \theta}(\alpha, \theta) = \quad & 1(p_0 = 1) \cdot E[W_t W_t{}' \cdot p_t(\theta^{*\prime} W_t)] \\
& +1(p_0 = 2) \cdot E[W_t W_t{}' \cdot (\alpha + (1 - 2\alpha) \cdot P_t(\theta' W_t))] \\
& +1(p_0 > 2) \cdot (p_0 - 1) \\
& \quad \cdot E[W_t W_t{}' \cdot (\alpha + (1 - 2\alpha) \cdot 1(Y_{t+1} - \theta' W_t < 0)) \cdot |Y_{t+1} - \theta' W_t|^{p_0 - 2}],
\end{aligned}$$

where P_t and p_t are the distribution function and the density of Y_{t+1} conditional on $\mathcal{F}_t$. The function $\partial h_{p_0}/\partial\theta(\alpha, \cdot) : \Theta \to \mathbb{R}^k \times \mathbb{R}^k$ being an integral it is clearly continuous on Θ. We have therefore shown that (i) is verified, i.e. $h : (0,1) \times \Theta \to \mathbb{R}^k$ is continuously differentiable on $(0,1)\times\Theta$.

We know, from the previous proof that $\Sigma(\theta)$ is positive definite (by Assumptions (A1)-(A2)), therefore nonsingular for every $(p_0,\alpha_0,\theta)\in\mathbb{N}^*\times(0,1)\times\Theta$.

Hence, for any $p_0\in\mathbb{N}^*$, we conclude that $[\partial h_{p_0}/\partial\theta(\alpha_0,\theta^*)]^{-1}$ exists for every $\alpha_0\in(0,1)$, which verifies the condition (ii). We can now apply the implicit function theorem (Theorem 3.8.5. in Schwartz, 1997, vol 2, p 185) to show that for every $\alpha_0\in(0,1)$ there exist an open interval E_0 containing α_0 and an open set F_0 containing θ^*, $F_0\equiv\{\theta\in\mathring{\Theta}:||\theta-\theta^*||<\delta_0\}$ with $\delta_0>0$, such that for every $\alpha\in E_0$, the equation $h_{p_0}(\alpha,\theta)=0$ has a unique solution θ in F_0, and that the function $\theta=\theta_{p_0}(\alpha)$ defined implicitly by $h_{p_0}(\alpha,\theta_{p_0}(\alpha))=0$ is continuously differentiable from E_0 to F_0 with

$$\theta'_{p_0}(\alpha)=-[\frac{\partial h_{p_0}}{\partial\theta}(\alpha,\theta_{p_0}(\alpha))]^{-1}\cdot\frac{\partial h_{p_0}}{\partial\alpha}(\alpha,\theta_{p_0}(\alpha)),$$

i.e.

$$\theta'_{p_0}(\alpha)=\begin{cases}\{E[W_tW_t'\cdot p_t(\theta_{p_0}(\alpha)'W_t)]\}^{-1}\cdot E[W_t], & \text{if } p_0=1,\\ \{E[W_tW_t'\cdot(\alpha+(1-2\alpha)\cdot P_t(\theta_{p_0}(\alpha)'W_t))]\}^{-1}\cdot E[W_t\cdot|Y_{t+1}-\theta_{p_0}(\alpha)'W_t|], & \text{if } p_0=2,\\ \{(p_0-1)E[W_tW_t'\cdot(\alpha+(1-2\alpha)\cdot 1(Y_{t+1}-\theta_{p_0}(\alpha)'W_t<0)) & \\ \quad\cdot|Y_{t+1}-\theta_{p_0}(\alpha)'W_t|^{p_0-2}]\}^{-1}\cdot E[W_t\cdot|Y_{t+1}-\theta_{p_0}(\alpha)'W_t|^{p_0-1}], & \text{if } p_0>2.\end{cases}\tag{11}$$

It is important to note that we can extend the previous implicit function argument by continuity to the entire open interval $(0,1)$. Let $F\equiv\bigcup_{\alpha_0\in(0,1)}F_0$. F being a union of open sets, F is an open subset of $\mathring{\Theta}$. Hence, we have shown that given $p_0\in\mathbb{N}^*$, for every $\alpha_0\in(0,1)$, the equation $h_{p_0}(\alpha_0,\theta)=0$ has a unique solution θ^* in F and the implicit function $\theta^*=\theta_{p_0}(\alpha_0)$ is continuously differentiable from $(0,1)$ to F with $\theta'_{p_0}(\alpha)$ as given in (11). We now show that $\theta_{p_0}(\alpha)$ is bijective from $(0,1)$ to F. It is surjective by construction, so we only need to show that it is injective on $(0,1)$, i.e. $\alpha_1\neq\alpha_2$ implies $\theta_{p_0}(\alpha_1)\neq\theta_{p_0}(\alpha_2)$. This last implication is equivalent to: $\theta_{p_0}(\alpha_1)=\theta_{p_0}(\alpha_2)$ implies $\alpha_1=\alpha_2$. If $\theta_{p_0}(\alpha_1)=\theta_{p_0}(\alpha_2)$ then

$$\begin{aligned}0=\;& E[W_t\cdot(1(Y_{t+1}-\theta_{p_0}(\alpha_1)'W_t<0)-\alpha_1)\cdot|Y_{t+1}-\theta_{p_0}(\alpha_1)'W_t|^{p_0-1}]\\ &-E[W_t\cdot(1(Y_{t+1}-\theta_{p_0}(\alpha_2)'W_t<0)-\alpha_2)\cdot|Y_{t+1}-\theta_{p_0}(\alpha_2)'W_t|^{p_0-1}]\\ =\;&(\alpha_2-\alpha_1)E[W_t\cdot|Y_{t+1}-\theta_{p_0}(\alpha_2)'W_t|^{p_0-1}],\end{aligned}$$

which, by Assumption (A1), implies $\alpha_1 = \alpha_2$. Hence for a given $\theta^* \in F$ there is a unique $\alpha_0 \in (0,1)$ such that $\theta^* = \theta_{p_0}(\alpha_0)$. This completes the proof of Proposition 3. ■

Proof of Lemma 4. We first show that α_0 is a minimum of the quadratic form

$$Q_0(\alpha) \equiv E[V_t \cdot (1(Y_{t+1} - f^*_{t+1} < 0) - \alpha) \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}]' \cdot W \cdot E[V_t \cdot (1(Y_{t+1} - f^*_{t+1} < 0) - \alpha) \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}],$$

where V_t is a sub-vector of W_t and W is a positive definite weighting matrix. Note that $Q_0(\alpha) = c - 2b\alpha + a\alpha^2$ with

$$\begin{aligned}
a &\equiv E[V_t \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}]' \cdot W \cdot E[V_t \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}], \\
b &\equiv E[V_t \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}]' \cdot W \cdot E[V_t \cdot 1(Y_{t+1} - f^*_{t+1} < 0) \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}], \\
c &\equiv E[V_t \cdot 1(Y_{t+1} - f^*_{t+1} < 0) \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}]' \cdot W \cdot E[V_t \cdot 1(Y_{t+1} - f^*_{t+1} < 0) \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}].
\end{aligned}$$

The weighting matrix W being positive definite, we know that $a > 0$ so that $Q_0(\alpha)$ is a concave function of α. Any solution to the first order condition

$$\begin{aligned}
0 &= b - \alpha a \\
&= E[V_t \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}]' \cdot W \cdot E[V_t \cdot (1(Y_{t+1} - f^*_{t+1} < 0) - \alpha) \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}] \quad (12)
\end{aligned}$$

is therefore a minimum of $Q_0(\alpha)$. We know that if V_t is a sub-vector of W_t (Assumption (A3)) then $E[V_t \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}] \neq 0$ (Assumption (A1)). Moreover, W is nonsingular, so that the equality (12) implies $E[V_t \cdot (1(Y_{t+1} - f^*_{t+1} < 0) - \alpha) \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}] = 0$. Hence, any solution to the moment condition $E[V_t \cdot (1(Y_{t+1} - f^*_{t+1} < 0) - \alpha) \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}] = 0$ is a minimum of $Q_0(\alpha)$. We know from the condition (3) that this is the case for α_0, so α_0 is a minimum of $Q_0(\alpha)$.

We now need to show that $Q_0(\alpha)$ has a unique minimum and that this minimum is in $(0,1)$. We know that $a > 0$ so that $Q_0'(\alpha)$ is not identically equal to zero on $(0,1)$, i.e. $Q_0(\alpha)$ is not a constant nor a linear function of α. It therefore has a unique minimum at $\alpha^* = -b/a$,

$$\alpha^* = \frac{E[V_t \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}]' \cdot W \cdot E[V_t \cdot 1(Y_{t+1} - f^*_{t+1} < 0) \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}]}{E[V_t \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}]' \cdot W \cdot E[V_t \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}]}. \quad (13)$$

We now need to verify that α^* defined in (13) is in $(0,1)$. First, we show that $\alpha^* \in (0,1)$ holds if all the elements of the d-vector V_t are strictly positive, i.e. $V_t > 0_d$, $a.s. - \mathcal{P}$, where 0_d is a d-vector of zeros. In that case we have

$$0 \leqslant V_t \cdot 1(Y_{t+1} - f^*_{t+1} < 0) \cdot |Y_{t+1} - f^*_{t+1}|^{p_0 - 1} \leqslant V_t \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}, \; a.s. - \mathcal{P},$$

so that

$$0 \leqslant E[V_t \cdot 1(Y_{t+1} - f^*_{t+1} < 0) \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}] \leqslant E[V_t \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}].$$

Using Assumption (A1) we know that $0 < E[V_t \cdot 1(Y_{t+1} - f^*_{t+1} < 0) \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}]$ since V_t is a sub-vector of W_t. Knowing that W is positive definite, we have

$$\begin{aligned}
0 < \; & E[V_t \cdot 1(Y_{t+1} - f^*_{t+1} < 0) \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}]' \cdot W \cdot E[V_t \cdot 1(Y_{t+1} - f^*_{t+1} < 0) \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}] \\
\leqslant \; & E[V_t \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}]' \cdot W \cdot E[V_t \cdot (1(Y_{t+1} - f^*_{t+1} < 0) - \alpha) \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}] \\
\leqslant \; & E[V_t \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}]' \cdot W \cdot E[V_t \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}],
\end{aligned}$$

i.e. $0 < c \leqslant b \leqslant a$. Hence $\alpha^* > 0$. We also know that for all $\alpha \in (0,1)$ $Q_0(\alpha) > 0$ so that the reduced discriminant $\Delta = b^2 - ac < 0$. Hence, $b < \sqrt{ac} \leqslant a$ so that $\alpha^* < 1$. So, if $V_t > 0_d$, $a.s. - \mathcal{P}$ then $\alpha^* \in (0,1)$. Now consider a case where the first element of V_t is a constant 1 and that there exists some constant $c > 0$ such that $V_t > -c \cdot 1_d$, $a.s. - \mathcal{P}$, where 1_d is a d-vector of ones. Note that this inequality is implied by the Assumption (A3), which ensures that $||V_t|| \leqslant ||W_t|| \leqslant m$ so that the components of V_t are necessarily bounded by some constant c. Now, consider the rotation of the d-vector V_t,

$$\bar{V}_t = KV_t = \begin{pmatrix} 1 & 0 \\ c & I_{d-1} \end{pmatrix} V_t,$$

where now $\bar{V}_t = KV_t > 0$, $a.s. - \mathcal{P}$ (I_{d-1} is a $(d-1) \times (d-1)$ identity matrix). Notice that K is positive definite and that $(K^{-1})' \cdot W \cdot K^{-1}$ is positive definite if W is positive definite. Now, note that

$$\begin{aligned}
\alpha^* &= \frac{E[V_t \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}]' \cdot W \cdot E[V_t \cdot 1(Y_{t+1} - f^*_{t+1} < 0) \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}]}{E[V_t \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}]' \cdot W \cdot E[V_t \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}]} \\
&= \frac{E[\bar{V}_t \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}]' \cdot (K^{-1})'WK^{-1} \cdot E[\bar{V}_t \cdot 1(Y_{t+1} - f^*_{t+1} < 0) \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}]}{E[\bar{V}_t \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}]' \cdot (K^{-1})'WK^{-1} \cdot E[\bar{V}_t \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}]},
\end{aligned}$$

so that if α^* is the minimum of $Q_0(\alpha)$ then α^* is also a minimum of the quadratic form $\bar{Q}(\alpha)$, with

$$\bar{Q}(\alpha) \equiv E[\bar{V}_t \cdot (1(Y_{t+1} - f^*_{t+1} < 0) - \alpha) \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}]' \\ \cdot K^{-1} W (K^{-1})' \cdot E[\bar{V}_t \cdot (1(Y_{t+1} - f^*_{t+1} < 0) - \alpha) \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}].$$

¿From the results above we then know that $\alpha^* \in (0,1)$ since $\bar{V}_t > 0$, $a.s. - \mathcal{P}$. Hence, under Assumptions (A0)-(A3), we know that $Q_0(\alpha)$ is uniquely minimized at α^* defined in (13) and that $\alpha^* \in (0,1)$. Therefore, we conclude that $\alpha_0 = \alpha^*$, which completes the proof of Lemma 4. ■

Proof of Proposition 5. We impose the following conditions, in addition to Assumptions (A0)-(A3):

(A4) for every t, $\tau \leqslant t < T + \tau$, $\hat{\theta}_t$ is a consistent estimate of θ^* and $\theta^* \in F$;

(A5) the stochastic processes Y_t and W_t are stationary and α-mixing with mixing coefficient α of size $-r/(r-2)$, $r > 2$, and, given $p_0 \in \mathbb{N}^*$, there exist some $\delta_f > 0$ and $\Delta_f > 0$ such that $E[(Y_{t+1} - \hat{f}_{t+1})^{(2+\delta_f)(p_0-1)}] \leqslant \Delta_f < \infty$ and some $\delta_W > 0$ and $\Delta_W > 0$ such that $E[||W_t||^{2+\delta_W}] \leqslant \Delta_W < \infty$;

(A6) the density of Y_{t+1} conditional on $\mathcal{F}_t$ is bounded, i.e. there exist some $M > 0$ such that, $\sup_{y \in \mathbb{R}} p_t(y) \leqslant M < \infty$;

Recall that form (9) we have

$$\hat{\alpha}_T \equiv \frac{[T^{-1} \sum_{t=\tau}^{T+\tau-1} v_t |y_{t+1} - \hat{f}_{t+1}|^{p_0-1}]' \cdot \hat{S}^{-1} \cdot [T^{-1} \sum_{t=\tau}^{T+\tau-1} v_t 1(y_{t+1} - \hat{f}_{t+1} < 0) |y_{t+1} - \hat{f}_{t+1}|^{p_0-1}]}{[T^{-1} \sum_{t=\tau}^{T+\tau-1} v_t |y_{t+1} - \hat{f}_{t+1}|^{p_0-1}]' \cdot \hat{S}^{-1} \cdot [T^{-1} \sum_{t=\tau}^{T+\tau-1} v_t |y_{t+1} - \hat{f}_{t+1}|^{p_0-1}]}.$$

In order to show that $\hat{\alpha}_T \xrightarrow{p} \alpha_0$ it is sufficient to show that: *(i)* $T^{-1} \sum_{t=\tau}^{T+\tau-1} v_t |y_{t+1} - \hat{f}_{t+1}|^{p_0-1} \xrightarrow{p} E[V_t \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}]$, *(ii)* $T^{-1} \sum_{t=\tau}^{T+\tau-1} v_t 1(y_{t+1} - \hat{f}_{t+1} < 0) |y_{t+1} - \hat{f}_{t+1}|^{p_0-1} \xrightarrow{p} E[V_t \cdot 1(Y_{t+1} - f^*_{t+1} < 0) \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}]$, and *(iii)* S positive definite. Then, by using Lemma 4, the consistency of $\hat{S}$, the positive definiteness of S (thus of S^{-1}), the Assumptions (A1) and (A3) which ensure that $E[V_t \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}] \neq 0$ and $E[V_t \cdot 1(Y_{t+1} - f^*_{t+1} < 0) \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}] \neq 0$, and the continuity of the inverse function (away form zero), we will have that $\hat{\alpha}_T \xrightarrow{p} \alpha_0$. Note that an alternative way to prove the same result would be to work with the quadratic form $Q_0(\alpha)$ and then use the results of Theorem 2.7 in Newey and McFadden (1994, p 2133), for example. Here however, we use the fact that we know the exact analytic form of α_0 which considerably simplifies the consistency proof.

First, let us show that S is positive definite. Recall that, given $p_0 \in \mathbb{N}^*$, we have

$$S \equiv E[V_t V_t' \cdot (1(Y_{t+1} - \hat{f}_{t+1} < 0) - \alpha)^2 \cdot |Y_{t+1} - \hat{f}_{t+1}|^{2p_0-2}],$$

so that for every $\varphi \in \mathbb{R}^d$ we have $\varphi' S \varphi = E[\varphi' V_t V_t' \varphi \cdot (1(Y_{t+1} - \hat{f}_{t+1} < 0) - \alpha)^2 \cdot |Y_{t+1} - \hat{f}_{t+1}|^{2p_0-2}]$. Note that $(1(Y_{t+1} - \hat{f}_{t+1} < 0) - \alpha)^2 \cdot |Y_{t+1} - \hat{f}_{t+1}|^{2p_0-2} > 0$, $a.s. - \mathcal{P}$, so that

$$\varphi' S \varphi = 0 \Rightarrow \varphi' V_t V_t' \varphi = 0, a.s. - \mathcal{P} \Rightarrow \varphi' E[V_t V_t'] \varphi = 0.$$

Now, note that the positive definiteness of $E[W_t W_t']$ implies that all upper-left submatrices of $E[W_t W_t']$ have strictly positive determinant. By rearranging (if necessary) the elements of W_t, we can easily show that $E[V_t V_t']$ is an upper-left $d \times d$ submatrix of $E[W_t W_t']$. Therefore $\det E[V_t V_t'] > 0$. Together with the fact that $E[V_t V_t']$ is positive semi-definite (for every $\varphi \in \mathbb{R}^d$, we have $\varphi' E[V_t V_t'] \varphi = E[\varphi' V_t V_t' \varphi] = E[(\varphi' V_t)^2] \geqslant 0$), this implies that $E[V_t V_t']$ is positive definite. Therefore $\varphi' E[V_t V_t'] \varphi = 0$ implies $\varphi = 0$, which shows that S is positive definite. We can therefore use the results of Lemma 4.

We now show that conditions *(i)* and *(ii)* hold. Given $p_0 \in \mathbb{N}^*$ and for every t, $\tau \leqslant t < T + \tau$, let

$$\begin{aligned}
g_t &\equiv v_t 1(y_{t+1} - \hat{f}_{t+1} < 0)|y_{t+1} - \hat{f}_{t+1}|^{p_0-1}, \\
\hat{g}_T &\equiv T^{-1} \sum_{t=\tau}^{T+\tau-1} g_t, \\
g_0 &\equiv E[V_t \cdot 1(Y_{t+1} - f^*_{t+1} < 0) \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}], \\
g^* &\equiv E[V_t \cdot 1(Y_{t+1} - \hat{f}_{t+1} < 0) \cdot |Y_{t+1} - \hat{f}_{t+1}|^{p_0-1}],
\end{aligned}$$

and let

$$\begin{aligned}
h_t &\equiv v_t |y_{t+1} - \hat{f}_{t+1}|^{p_0-1}, \\
\hat{h}_T &\equiv T^{-1} \sum_{t=\tau}^{T+\tau-1} h_t, \\
h_0 &\equiv E[V_t \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}], \\
h^* &\equiv E[V_t \cdot |Y_{t+1} - \hat{f}_{t+1}|^{p_0-1}].
\end{aligned}$$

We now check that the conditions *(i)-(ii)* hold: by triangle inequality we have $||\hat{g}_T - g_0|| \leqslant ||\hat{g}_T - g^*|| + ||g^* - g_0||$ and $||\hat{h}_T - h_0|| \leqslant ||\hat{h}_T - h^*|| + ||h^* - h_0||$. We first show that $||\hat{g}_T - g^*|| \xrightarrow{p} 0$ and $||\hat{h}_T - h^*|| \xrightarrow{p} 0$ by using a law of large numbers (LLN) for stationary and α-mixing sequences (e.g., Corollary 3.48 in White 2001). By Assumption (A5), the Cauchy-Schwartz inequality, and using $E[||V_t||^2] \leqslant E[||W_t||^2]$, we know that, for $\delta = \min(\delta_f, \delta_W)/2 > 0$,

$$\begin{aligned} E[||h_t||^{1+\delta}] &\leqslant (E[||V_t||^{2+2\delta}])^{1/2} \cdot (E[(Y_{t+1} - \hat{f}_{t+1})^{2(1+\delta)(p_0-1)}])^{1/2} \\ &\leqslant \max(1, \Delta_W^{1/2}) \cdot \max(1, \Delta_f^{1/2}) < \infty. \end{aligned}$$

Similarly,

$$E[||g_t||^{1+\delta}] \leqslant (E[||V_t \cdot 1(Y_{t+1} - \hat{f}_{t+1} < 0)||^{2+2\delta}])^{1/2} \cdot (E[(Y_{t+1} - \hat{f}_{t+1})^{2(1+\delta)(p_0-1)}])^{1/2},$$

and, since $V_t'V_t \cdot 1(Y_{t+1} - \hat{f}_{t+1} < 0) \leqslant V_t'V_t$, $a.s. - \mathcal{P}$, so that $E[||V_t \cdot 1(Y_{t+1} - \hat{f}_{t+1} < 0)||^2] \leqslant E[||V_t||^2]$, by same reasoning as previously, we get $E[||g_t||^{1+\delta}] < \infty$. Hence, both $\hat{g}_T$ and $\hat{h}_T$ converge in probability to their expected values. Next we need to show that the same holds for $||g^* - g_0|| \xrightarrow{p} 0$ and $||h^* - h_0|| \xrightarrow{p} 0$. We treat the two cases $p_0 = 1$ and $p_0 = 2$ separately. CASE: $p_0 = 1$: note that in that case $h^* = h_0$ and that by triangular and Cauchy-Schwartz inequalities, we have

$$\begin{aligned} ||g^* - g_0||^2 &\leqslant ||E[V_t \cdot (1(Y_{t+1} - \hat{f}_{t+1} < 0) - 1(Y_{t+1} - f^*_{t+1} < 0))]||^2 \\ &\leqslant E[||V_t||^2] \cdot E[(1(Y_{t+1} - \hat{f}_{t+1} < 0) - 1(Y_{t+1} - f^*_{t+1} < 0))^2]. \end{aligned}$$

Note that for every t, $\tau \leqslant t < T + \tau$, we have

$$\begin{aligned} E\{[1(Y_{t+1} - \hat{f}_{t+1} &< 0) - 1(Y_{t+1} - f^*_{t+1} < 0)]^2\} \\ &= E\{[1(f^*_{t+1} \leqslant Y_{t+1} < \hat{f}_{t+1}) - 1(\hat{f}_{t+1} \leqslant Y_{t+1} < f^*_{t+1})]^2\} \\ &= E[1(f^*_{t+1} \leqslant Y_{t+1} < \hat{f}_{t+1}) + 1(\hat{f}_{t+1} \leqslant Y_{t+1} < f^*_{t+1})] \\ &= E\{E_t[1(f^*_{t+1} \leqslant Y_{t+1} < \hat{f}_{t+1}) + 1(\hat{f}_{t+1} \leqslant Y_{t+1} < f^*_{t+1})]\}, \end{aligned}$$

where

$$\begin{aligned} E_t[1(f^*_{t+1} \leqslant Y_{t+1} < \hat{f}_{t+1}) + 1(\hat{f}_{t+1} \leqslant Y_{t+1} < f^*_{t+1})] &= |\textstyle\int_{\theta^{*\prime} W_t}^{\hat{\theta}_t' W_t} p_t(y) dy| \\ &\leqslant |\hat{\theta}_t' W_t - \theta^{*\prime} W_t| \cdot \sup_{y \in \mathbb{R}} p_t(y) \\ &\leqslant ||\hat{\theta}_t - \theta^*|| \cdot ||W_t|| \cdot \sup_{y \in \mathbb{R}} p_t(y). \end{aligned}$$

Hence, by Assumption (A3) and (A6) we have

$$||g^* - g_0||^2 \leqslant E[||V_t||^2] \cdot E[||\hat{\theta}_t - \theta^*||] \cdot m \cdot M.$$

Therefore, by using Assumptions (A3) and (A5), $||g^* - g_0||^2 \leqslant \max(1, \Delta_W) \cdot m \cdot M \cdot E[||\hat{\theta}_t - \theta^*||]$ which shows that when $\hat{\theta}_t$ is a consistent estimate of θ^* (Assumption (A4)), $||g^* - g_0|| \xrightarrow{p} 0$. Hence, when $p_0 = 1$, we have shown that $\hat{\alpha}_T \xrightarrow{p} \alpha_0$.

CASE $p_0 = 2$: we now have

$$\begin{aligned}
||h^* - h_0|| &\leqslant ||E[V_t \cdot (|Y_{t+1} - \hat{f}_{t+1}| - |Y_{t+1} - f^*_{t+1}|)]|| \\
&\leqslant ||E[V_t \cdot |f^*_{t+1} - \hat{f}_{t+1}|]|| \\
&\leqslant E[||V_t|| \cdot ||W_t|| \cdot ||\hat{\theta}_t - \theta^*||] \\
&\leqslant m^2 \cdot E[||\hat{\theta}_t - \theta^*||],
\end{aligned}$$

so that by same argument as previously, $||h^* - h_0|| \xrightarrow{p} 0$. Finally, note that

$$\begin{aligned}
||g^* - g_0||^2 &\leqslant ||E[V_t \cdot (1(Y_{t+1} - \hat{f}_{t+1} < 0) \cdot |Y_{t+1} - \hat{f}_{t+1}| - 1(Y_{t+1} - f^*_{t+1} < 0) \cdot |Y_{t+1} - f^*_{t+1}|)]||^2 \\
&\leqslant ||E[V_t \cdot (1(f^*_{t+1} \leqslant Y_{t+1} < \hat{f}_{t+1}) \cdot |Y_{t+1} - \hat{f}_{t+1}| - 1(\hat{f}_{t+1} \leqslant Y_{t+1} < f^*_{t+1}) \cdot |Y_{t+1} - f^*_{t+1}|)]||^2 \\
&\leqslant E[||V_t||^2] \cdot E[(1(f^*_{t+1} \leqslant Y_{t+1} < \hat{f}_{t+1}) \cdot |Y_{t+1} - \hat{f}_{t+1}| - 1(\hat{f}_{t+1} \leqslant Y_{t+1} < f^*_{t+1}) \cdot |Y_{t+1} - f^*_{t+1}
\end{aligned}$$

As previously,

$$\begin{aligned}
E[(1(f^*_{t+1} &\leqslant Y_{t+1} < \hat{f}_{t+1}) \cdot |Y_{t+1} - \hat{f}_{t+1}| - 1(\hat{f}_{t+1} \leqslant Y_{t+1} < f^*_{t+1}) \cdot |Y_{t+1} - f^*_{t+1}|)^2] \\
&= E[1(f^*_{t+1} \leqslant Y_{t+1} < \hat{f}_{t+1}) \cdot (Y_{t+1} - \hat{f}_{t+1})^2] + E[1(\hat{f}_{t+1} \leqslant Y_{t+1} < f^*_{t+1}) \cdot (Y_{t+1} - f^*_{t+1})^2],
\end{aligned}$$

so that by Assumptions (A3), (A5) and (A6) we have $||g^* - g_0||^2 \leqslant \max(1, \Delta_W^{5/2}) \cdot \max(1, \Delta_f) \cdot M \cdot m \cdot E||\hat{\theta}_t - \theta^*||]$ and so $||g^* - g_0|| \xrightarrow{p} 0$ when $\hat{\theta}_t \xrightarrow{p} \theta^*$ (Assumption (A4)). Hence, for $p_0 = 2$ we have $\hat{\alpha}_T \xrightarrow{p} \alpha_0$, which completes the proof of Proposition 5. ■

Proof of Proposition 6. In addition to Assumptions (A0)-(A4) and (A6) we assume the following:

(A5') the stochastic processes Y_t and W_t are stationary and α-mixing with mixing coefficient α of size $-r/(r-2)$, $r > 2$, and, given $p_0 \in \mathbb{N}^*$, there exist some $\delta_f > 0$ and $\Delta_f > 0$

such that $E[(Y_{t+1} - \hat{f}_{t+1})^{(4+\delta_f)(p_0-1)}] \leqslant \Delta_f < \infty$ and some $\delta_W > 0$ and $\Delta_W > 0$ such that $E[||W_t||^{4+\delta_W}] \leqslant \Delta_W < \infty$;

Note that the Assumption (A5') is identical to the Assumption (A5), except for the moment conditions which are stronger than previously. We now show that $T^{1/2}(\hat{\alpha}_T - \alpha_0)$ is asymptotically normal:

using the previous notation we know that $\hat{\alpha}_T$ solves the first order condition

$$\hat{h}'_T \cdot \hat{S}^{-1} \cdot (\hat{g}_T - \hat{h}_T \cdot \hat{\alpha}_T) = 0.$$

Using the fact that the previous equation is linear in α we can easily expand it around α_0 so that

$$0 = \hat{h}'_T \cdot \hat{S}^{-1} \cdot (\hat{g}_T - \hat{h}_T \cdot \alpha_0) - \hat{h}'_T \cdot \hat{S}^{-1} \cdot \hat{h}_T \cdot (\hat{\alpha}_T - \alpha_0). \tag{14}$$

The idea then is to show that

$$\hat{h}'_T \cdot \hat{S}^{-1} \cdot (\hat{g}_T - \hat{h}_T \cdot \alpha_0) = \bar{h}'_T \cdot \hat{S}^{-1} \cdot (\bar{g}_T - \bar{h}_T \cdot \alpha_0) + o_p(1), \tag{15}$$

where we use the following notation:

$$\begin{aligned} \bar{g}_t &\equiv v_t 1(y_{t+1} - f^*_{t+1} < 0)|y_{t+1} - f^*_{t+1}|^{p_0-1}, \\ \bar{g}_T &\equiv T^{-1} \sum_{t=\tau}^{T+\tau-1} \bar{g}_t, \end{aligned}$$

and

$$\begin{aligned} \bar{h}_t &\equiv v_t |y_{t+1} - f^*_{t+1}|^{p_0-1}, \\ \bar{h}_T &\equiv T^{-1} \sum_{t=\tau}^{T+\tau-1} \bar{h}_t. \end{aligned}$$

We now show that the development (15) holds: note that by the triangle inequality we have

$$\begin{aligned} ||\hat{h}'_T \cdot \hat{S}^{-1} \cdot (\hat{g}_T - \hat{h}_T \cdot \alpha_0) - \bar{h}'_T \cdot \hat{S}^{-1} \cdot (\bar{g}_T - \bar{h}_T \cdot \alpha_0)|| &\leqslant ||\hat{h}'_T \cdot \hat{S}^{-1} \cdot (\hat{g}_T - \bar{g}_T)|| \\ &\quad + ||(\hat{h}_T - \bar{h}_T)' \cdot \hat{S}^{-1} \cdot \bar{g}_T|| \\ &\quad + ||(\hat{h}_T - \bar{h}_T)' \cdot \hat{S}^{-1} \cdot \hat{h}_T \cdot \alpha_0|| \\ &\quad + ||\bar{h}'_T \cdot \hat{S}^{-1} \cdot (\hat{h}_T - \bar{h}_T) \cdot \alpha_0|| \end{aligned}$$

so that

$$||\hat{h}'_T \cdot \hat{S}^{-1} \cdot (\hat{g}_T - \hat{h}_T \cdot \alpha_0) - \bar{h}'_T \cdot \hat{S}^{-1} \cdot (\bar{g}_T - \bar{h}_T \cdot \alpha_0)|| \leqslant ||\hat{h}'_T \cdot \hat{S}^{-1}|| \cdot ||\hat{g}_T - \bar{g}_T|| + (||\hat{S}^{-1} \cdot \bar{g}_T|| + \alpha_0 ||\hat{S}^{-1} \cdot \hat{h}_T|| + \alpha_0 ||\bar{h}'_T \cdot \hat{S}^{-1}||) \cdot ||\hat{h}_T - \bar{h}_T||.$$

Now note that $||\hat{g}_T - \bar{g}_T|| \leqslant ||\hat{g}_T - g_0|| + ||\bar{g}_T - g_0||$. From the previous proof we have that $||\hat{g}_T - g_0|| \xrightarrow{p} 0$ if Assumption (A4) holds. Moreover, by the LLN we have $||\bar{g}_T - g_0|| \xrightarrow{p} 0$ so that $||\hat{g}_T - \bar{g}_T|| \xrightarrow{p} 0$ if $\hat{\theta}_t$ is consistent (Assumption (A4)). By the same type of argument, we show that $||\hat{h}_T - \bar{h}_T|| \xrightarrow{p} 0$. Therefore, by using the fact that $||\hat{S}^{-1} \cdot \hat{h}_T|| < \infty$, $||\hat{S}^{-1} \cdot \bar{h}_T|| < \infty$ and $||\hat{S}^{-1} \cdot \bar{g}_T|| < \infty$ we show that (15) holds.

Now we use central limit theorem (CLT) for α-mixing sequences (e.g., Theorem 5.20 in White 2001) to show that $T^{1/2}(\bar{g}_T - \bar{h}_T \cdot \alpha_0) \xrightarrow{d} \mathcal{N}(0, S)$: the Cauchy-Schwartz inequality and Assumption (A5') imply that

$$\begin{aligned} E[||V_t \cdot (1(Y_{t+1} - f^*_{t+1} &< 0) - \alpha_0) \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}||^{2+\delta}] \leqslant E[||V_t||^{2+\delta} \cdot (Y_{t+1} - f^*_{t+1})^{(2+\delta)(p_0-1)}] \\ &\leqslant E[||V_t||^{4+2\delta}]^{1/2} \cdot E[(Y_{t+1} - f^*_{t+1})^{(4+2\delta)(p_0-1)}]^{1/2} \\ &\leqslant \max(1, \Delta_W^{1/2}) \cdot \max(1, \Delta_f^{1/2}) < \infty \end{aligned}$$

for $\delta = \min(\delta_W, \delta_f)/2 > 0$. The CLT then ensures $T^{1/2}(\bar{g}_T - \bar{h}_T \cdot \alpha_0) \xrightarrow{d} \mathcal{N}(0, S)$ so that $T^{1/2}[\bar{h}'_T \cdot \hat{S}^{-1} \cdot (\bar{g}_T - \bar{h}_T \cdot \alpha_0)] \xrightarrow{d} \mathcal{N}(0, h'_0 \cdot S^{-1} \cdot h_0)$. Together with (15) this implies (by Slutzky theorem)

$$T^{1/2}[\hat{h}_T \cdot \hat{S}^{-1} \cdot (\hat{g}_T - \hat{h}_T \cdot \alpha_0)] \xrightarrow{d} \mathcal{N}(0, h'_0 \cdot S^{-1} \cdot h_0). \tag{16}$$

The remainder of the asymptotic normality proof is similar to the standard case: the positive definiteness of S^{-1}, $\hat{S} \xrightarrow{p} S$ and $\hat{h}_T \xrightarrow{p} h_0$, together with Assumptions (A1) and (A3), ensure that $h'_0 \cdot S^{-1} \cdot h_0 \neq 0$ and $\hat{h}'_T \cdot \hat{S}^{-1} \cdot \hat{h}_T \neq 0$ with probability one, so that the expansion (14) is equivalent to $T^{1/2}(\hat{\alpha}_T - \alpha_0) = [\hat{h}'_T \cdot \hat{S}^{-1} \cdot \hat{h}_T]^{-1} T^{1/2}[\hat{h}'_T \cdot \hat{S}^{-1} \cdot (\hat{g}_T - \hat{h}_T \cdot \alpha_0)]$. We then use the limit result in (16) and the Slutzky theorem to show that

$$T^{1/2}(\hat{\alpha}_T - \alpha_0) \xrightarrow{d} \mathcal{N}(0, (h'_0 \cdot S^{-1} \cdot h_0)^{-1}),$$

which completes the proof of Proposition 6. ■

References

[1] Artis, M. and M. Marcellino, 2001, Fiscal forecasting: The track record of the IMF, OECD and EC. Econometrics Journal 4, S20-S36.

[2] Bonham, C. and R. Cohen, 1995, Testing the rationality of price forecasts: Comment. American Economic Review 85, 284-289.

[3] Brown, B.Y. and S. Maital, 1981, What do economists know? An empirical study of experts' expectations. Econometrica 49, 491-504.

[4] Campbell, B. and E. Ghysels, 1995, Federal budget projections: A nonparametric assessment of bias and efficiency. Review of Economics and Statistics, 17-31.

[5] Cargill, T.F. and R.A. Meyer, 1980, The Term Structure of Inflationary Expectations and Market Efficiency. Journal of Finance 35, 57-70.

[6] De Bondt, W.F.M. and M.M. Bange, 1992, Inflation Forecast Errors and Time Variation in Term Premia. Journal of Financial and Quantitative Analysis 27, 479-496.

[7] Christoffersen, P.F. and F.X. Diebold, 1997, Optimal prediction under asymmetric loss. Econometric Theory 13, 808-817.

[8] Diebold, F.X. and J.A. Lopez, 1996, Forecast Evaluation and Combination. Ch. 8 in G.S. Maddala and C.R. Rao, eds., Handbook of Statistics, Vol. 14.

[9] Dokko, Y. and R. H. Edelstein, 1989, How Well do Economists Forecast Stock Market Prices? A study of the Livingston Surveys. American Economic Review 79, 865-871.

[10] Fama, E.F., 1975, Short-Term Interest Rates as Predictors of Inflation. American Economic Review 65, 269-82.

[11] Figlewski, S. and P. Wachtel, 1981, The Formation of Inflationary Expectations. Review of Economics and Statistics 63, 1-10.

[12] Granger, C.W.J., 1999, Outline of Forecast Theory Using Generalized Cost Functions. Spanish Economic Review 1, 161-173.

[13] Granger, C.W.J., and P. Newbold, 1986, Forecasting Economic Time Series, Second Edition. Academic Press.

[14] Granger, C.W.J. and M.H. Pesaran, 2000, Economic and Statistical Measures of Forecast Accuracy. Journal of Forecasting 19, 537-560.

[15] Gultekin, N.B., 1983, Stock Market Returns and Inflation Forecasts. Journal of Finance 38, 663-673.

[16] Hafer, R.W. and S.E. Hein, 1985, On the Accuracy of Time-series, Interest Rate, and Survey Forecasts of Inflation. Journal of Business 58, 377-398.

[17] Hansen, L.P. and R.J. Hodrick, 1980, Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Investigaion. Journal of Political Economy 88, 829-853.

[18] Keane, M.P. and D.E. Runkle, 1990, Testing the Rationality of Price Forecasts: New Evidence from Panel Data. American Economic Review 80, 714-735.

[19] Lakonishok, J., 1980, Stock Market Return Expectations: Some General Properties. Journal of Finance 35, 921-931.

[20] Mincer, J. and V. Zarnowitz, 1969, The Evaluation of Economic Forecasts. In J. Mincer, ed., Economic Forecasts and Expectations. National Bureau of Economic Research, New York.

[21] Mishkin, F.S., 1981, Are Markets Forecasts Rational? American Economic Review 71, 295-306.

[22] Newey, W. and D. McFadden, 1994, Large Sample Estimation and Hypothesis Testing. In R.F.Engel and D.L.McFadden eds. Handbook of Econometrics, volume 4, Elsevier: Amsterdam.

[23] Newey, W. and West, K., 1987, A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. Econometrica, 55, 703-708.

[24] Peel, D.A. and A.R. Nobay, 1998, Optimal Monetary Policy in a Model of Asymmetric Central Bank Preferences. FMG discussion paper 0306.

[25] Pesando, J.E., 1975, A Note on the Rationality of the Livingston Price Expectations. Journal of Political Economy 83, 849-858.

[26] Schroeter, J.R. and S.L. Smith, 1986, A Reexamination of the Livingston Price Expectations. Journal of Money, Credit and Banking 18, 239-246.

[27] Schwartz, L., 1997, Analyse. Hermann: Paris.

[28] West, K.D., 1996, Asymptotic Inference about Predictive Ability. Econometrica 64, 1067-84

[29] West, K.D. and M.W. McCracken, 1998, "Regression-Based Tests of Predictive Ability", International Economic Review 39, 817-840.

[30] 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, 23-46.

[31] White, H., 2001, Asymptotic Theory for Econometricians, 2nd edition, Academic Press, San Diego: California.

[32] Zarnowitz, V., 1979, An Analysis of Annual and Multiperiod Quarterly Forecasts of Aggregate Income, Output, and the Price Level. Journal of Business 52, 1-33.

[33] Zarnowitz, V., 1985, Rational Expecations and Macroeconomic Forecasts. Journal of Business and Economic Statistics 3, 293-311.