# Estimation and Testing of Forecast Rationality under Flexible Loss*

Graham Elliott — UCSD
Ivana Komunjer — Caltech
Allan Timmermann — UCSD

December 14, 2004

**Abstract**

In situations where a sequence of forecasts is observed, a common strategy is to examine 'rationality' conditional on a given loss function. We examine this from a different perspective - supposing that we have a family of loss functions indexed by unknown shape parameters, then given the forecasts can we back out the loss function parameters consistent with the forecasts being rational even when we do not observe the underlying forecasting model? We establish identification of the parameters of a general class of loss functions that nest popular loss functions as special cases and provide estimation methods and asymptotic distributional results for these parameters. This allows us to construct new tests of forecast rationality that allow for asymmetric loss. 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.

*We thank two anonymous referees, the editors, Hidehiko Ichimura and Bernard Salanié, as well as Max Auffhammer, Peter Bossaerts and seminar participants at Atlanta Fed, Bocconi, CentrA (Seville), Duke, St Louis Fed, Stanford, UCLA, UTS, UNSW, UT-Austin, Rice, Texas A&M, Caltech-UCLA-USC workshop on forecasting and the NBER-NSF-Penn conference on time-series in September 2002. 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. Considerable effort has been devoted to empirically testing the validity of this proposition using survey data on forecasts in areas such as efficient market models of stock prices, models of the term structure of interest or currency rates, inflation forecasting and tests of the Fisher equation.[1] Interpretation of this work is tempered by the fact that properties of rational forecasts can only be established jointly with a maintained loss function. Typically 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. Under this loss function forecasts are easy to compute through least squares methods and have well established properties such as unbiasedness and lack of serial correlation at the single-period horizon, c.f. Diebold and Lopez (1996). Inference about the optimality of a particular forecast series is easy and can be based directly on the observable forecast errors which do not depend on any unknown parameters of the forecasters's loss function.

Mean squared error loss, 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) argue that "An assumption of symmetry for the cost function is much less acceptable [than an assumption of a symmetric forecast error density]." Consequently, in economics and finance forecasting performance is increasingly evaluated under more general loss functions that account for asymmetries, c.f. Christoffersen and Diebold (1996, 1997), Granger and Newbold (1986), Granger and Pesaran (2000), West, Edison and Choi (1993) and Zellner (1986). Frequently used loss functions include lin-lin and linex 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. This raises the possibility that many of the rejections of forecast optimality reported in the empirical literature may simply be driven by the assumption of MSE loss rather than by the absence of forecast rationality per se. Indeed, if we are not sure that the loss function is of the MSE type, a key question then becomes what inference we can draw

[1] For references to numerous papers on forecast rationality see www.Phil.frb.org/econ/spf/spfbib.html.

from empirical inspection of a sequence of point forecasts.

This paper develops new methods for testing forecast optimality under general classes of loss functions that include mean absolute error (MAE) or MSE loss as a special case. This allows us to separate the question of forecast rationality from that of whether MAE or MSE loss accurately represents the decision maker's objectives. Instead our results let us test the joint hypothesis that the loss function belongs to a more flexible family and that the forecast is optimal.[2] In each case the family of loss functions is indexed by a single unknown parameter. We establish conditions under which this parameter is 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 essentially 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. These parameters are potentially of great economic interest as they provide information about the forecaster's objectives. For instance, if the mean forecast error is strongly negative, it could either be that the forecaster has MSE loss and is irrational or that loss is asymmetric and the forecaster rationally overpredicts due to higher costs associated with positive than with negative forecast errors.

The idea of backing out the parameter values that are most consistent with an optimizing agent's objective function has, in a different framework, been considered by Hansen and Singleton (1982). These authors study a representative investor with power utility and develop methods for estimating preference parameters from the investor's Euler equations.

[2] In general decision problems the forecasting and decision problem cannot be separated and an examination of the decision maker's action rule and full density forecast is required to test rationality, c.f. Diebold, Gunther and Tay (1998). Neither of these is, in general, observable and the vast majority of empirical data takes the form of point forecasts. Decision rules and utility functions giving rise to the loss function entertained in this paper can be established, however, c.f. Elliott, Komunjer and Timmermann (2004).

There is a major difference between this work and our approach, however, which has to do with the fact that Hansen and Singleton treat consumption and asset returns as observable state variables. When backing out the parameters of the forecaster's loss function from a sequence of point forecasts, this approach is less attractive, however. There is the real possibility that the forecasts are based on a misspecified model and this may well rule out identification of the parameters of the forecaster's loss function. Excluding this possibility requires carefully establishing conditions on the model used by the forecaster and the sense in which it may be misspecified. We develop new theoretical results that allow us to identify the source of rejection by establishing conditions on the decision maker's forecasting model under which the parameters of the loss function are identified and can be consistently estimated.

An area where asymmetric loss may play an important role is in the generation of government budget deficit forecasts by central banks and international organizations such as the IMF and OECD that are subject to political pressures from member countries but also play a role in imposing budgetary discipline. In an empirical analysis of forecasts generated by these organizations, we find evidence of systematic overpredictions of government budget deficits. This is inconsistent with forecast rationality and MSE loss. However, when we allow for asymmetric loss we can no longer reject forecast rationality. This suggests that unless it is known that forecast producers such as the IMF and OECD have symmetric loss, it is important to account for the possible effects of asymmetric loss. Furthermore, unless the forecast user happens to have the exact same loss function as the producer of the forecast, the raw forecasts cannot be used uncritically since they are only constructed to be optimal with respect to the forecast producer's loss. Knowing the direction of possible asymmetries in the loss function underlying the observed forecast - as can be obtained by estimating the loss function parameters - is thus important information to users of such forecasts.

The plan of the paper is as follows. Section 2 outlines the conditions for optimality of forecasts under a general class of loss functions. 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 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 Forecasts

In this section we examine families of loss functions which nest common ones as special cases. We study the forecasters' optimal problem and establish conditions under which we can identify the parameters describing the loss function from a sequence of observed rational forecasts.

Our setup is as follows: let $X \equiv \{X_t : \Omega \longrightarrow \mathbb{R}^{m+1}, m \in \mathbb{N}, t = 1, \ldots, n+1\}$ be a stochastic process defined on a complete probability space $(\Omega, \mathcal{F}, P)$ where $\mathcal{F} = \{\mathcal{F}_t, t = 1, \ldots, n+1\}$ and $\mathcal{F}_t$ is the $\sigma$-field $\mathcal{F}_t \equiv \sigma\{X_s, s \leqslant t\}$. Denote by $Y_t$ the component of interest of the observed vector $X_t$, $Y_t \in \mathbb{R}$, and interpret the remaining components as being an $m$-vector of other variables. We assume $Y_t$ is continuous. The distribution function $F(\cdot)$ of $Y_{t+1}$, its density $f(\cdot)$, and the expectation $E[\cdot]$ are subscripted by a $t$ to show that they are conditional on the information set $\mathcal{F}_t$.[3] The forecasting problem considered here involves forecasting the variable $Y_{t+s}$, where $s$ is the prediction horizon of interest, $s \geqslant 1$. In what follows, we set $s = 1$ and examine the one-step-ahead predictions of the realization $y_{t+1}$, knowing that all results can readily be generalized to any $s \geqslant 1$.

Let $f_{t+1} \equiv \theta' W_t$ be the forecast of $Y_{t+1}$ conditional on the information set $\mathcal{F}_t$ in which $\theta$ is an unknown $k$-vector of parameters, $\theta \in \Theta$, with $\Theta$ compact in $\mathbb{R}^k$, and $W_t$ is an $h$-vector of variables that are $\mathcal{F}_t$-measurable.[4] When constructing optimal forecasts we assume that, given $Y_{t+1}$ and $W_t$, the forecaster has in mind a generalized loss function $L$ defined by

$$L(p, \alpha, \theta) \equiv [\alpha + (1 - 2\alpha) \cdot 1(Y_{t+1} - f_{t+1} < 0)] |Y_{t+1} - f_{t+1}|^p, \tag{1}$$

where $p \in \mathbb{N}^*$, the set of positive integers, $\alpha \in (0, 1)$, $\theta \in \Theta$ and $Y_{t+1} - f_{t+1}$ corresponds to the forecast error $\varepsilon_{t+1}$. We let $\alpha_0$ and $p_0$ be the unknown true values of $\alpha$ and $p$ used

[3] Upper and lower case letters denote random variables and their realizations, respectively.

[4] Both the functional form of $f_{t+1}$ and the vector $W_t$ are specified by the agent producing the forecast. $W_t$ includes variables that are observed by the forecaster at time $t$ thought to help forecast $Y_{t+1}$ and which need not be known to the forecast user. If $W_t$ fails to incorporate all the relevant information in $\mathcal{F}_t$ or if the functional form of $f_{t+1}$ is misspecified, we say that the forecasting model is wrongly specified.

by the forecaster. Hence, the loss function in (1) is a function of not only the realization of $Y_{t+1}$ and the forecast $f_{t+1}$, but also of the shape parameters $\alpha$ and $p$ of $L$. Special cases of $L$ include: (i) squared loss function $L(2,1/2,\theta) = (Y_{t+1} - f_{t+1})^2$, (ii) absolute deviation loss function $L(1,1/2,\theta) = |Y_{t+1} - f_{t+1}|$, as well as their asymmetrical counterparts obtained when $\alpha \neq 1/2$, i.e. (iii) quad-quad loss, $L(2,\alpha,\theta)$, and (iv) lin-lin loss, $L(1,\alpha,\theta)$.[5] We shall take $p$ as given and focus on estimating $\alpha$.

Given $p_0$ and $\alpha_0$, the forecaster is assumed to construct the optimal one-step-ahead forecast of $Y_{t+1}$, $f^*_{t+1} \equiv \theta^{*\prime} W_t$, by solving

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

We let $\varepsilon^*_{t+1}$ 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 $\alpha_0$. Optimal forecasts have properties that follow directly from the construction of the forecasts. In the general case, the relevant optimality condition is the one given in the following Proposition. Assumptions referred to in the propositions are listed in Appendix A and proofs are provided in Appendix B.

**Proposition 1 (Optimality Condition)** *Under assumptions (A0)-(A2), and for given* $(p_0, \alpha_0) \in \mathbb{N}^* \times (0,1)$ *in* (2), *the forecast* $f^*_{t+1}$ *is optimal if and only if*

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

*Moreover, given* $p_0 \in \mathbb{N}^*$, *for any realization of* $W_t$, *the solution* $f^*_{t+1}$ *to the orthogonality condition* (3) *is unique, and the implicit function* $f^*_{t+1} = \theta_{p_0}(\alpha_0)' W_t$ *is a continuously differentiable one-to-one mapping from* $(0,1)$ *to* $\mathbb{R}$.

Proposition 1 shows that under fairly weak assumptions on $\theta^*$, $W_t$ and $Y_{t+1}$, the sequence of optimal forecast errors $\varepsilon^*_{t+1}$ satisfies the moment conditions $E[W_t(1(\varepsilon^*_{t+1t} < 0) - \alpha_0)|\varepsilon^*_{t+1}|^{p_0-1}] = 0$. When the forecasts are optimal, then any information must be correctly

[5] Linex loss is not a special case of (1). We chose not to focus on linex since the expected loss does not exist under linex loss for a wide class of distributions of the forecast error (e.g. student-t with finite degrees of freedom). Furthermore, linex loss only nests symmetric losss as a limiting case in the parameter space where loss is not defined. Obtaining symmetry only for a parameter on the boundary creates serious estimation problems and means that linex loss is not well-suited for our purpose.

included in $f^*_{t+1}$ which is orthogonal to the forecast errors and the quantity in (3) is a vector martingale difference sequence. If for given $p_0$ and $\alpha_0$ the forecaster uses (3) to determine $f^*_{t+1}$, then for a given $f^*_{t+1}$ we can back out $\alpha_0$ by using the same condition. However, this approach is valid only if knowing a solution to (3) allows the forecast user to identify $p_0$ and $\alpha_0$.

The second part of Proposition 1 shows that identification of $\alpha_0$ holds for fixed values of $p_0$. The result establishes a unique solution $f^*_{t+1}$ that in turn, knowing $p_0$ and $f^*_{t+1}$, yields a unique value for $\alpha$. Without this relationship we would not be able to identify $\alpha$.

In the case of a nonlinear forecasting model $f_{t+1} \equiv f(\theta, W_t)$, where the function $f : \Theta \times \mathbb{R}^h \to \mathbb{R}$ is continuously differentiable, the expression in (3) holds provided we replace $W_t$ with the gradient of $f$ with respect to $\theta$, evaluated at $(\theta^*, W_t)$. If in addition $f(\theta, W_t)$ is twice continuously differentiable and concave in the parameter $\theta$ on $\Theta$, for any realization of $W_t$, then $f^*_{t+1}$ is an optimal forecast. Provided the forecasting model is identifiable, so that $f(\theta_1, W_t) = f(\theta_2, W_t)$ for any realization of $W_t$ implies $\theta_1 = \theta_2$, we can replace $W_t$ by the gradient of $f$ with respect to $\theta$ in the assumptions (A0)-(A2) and show that $f^*_{t+1} = f(\theta_{p_0}(\alpha_0), W_t)$ is still a continuously differentiable one-to-one mapping from $(0,1)$ to $\mathbb{R}$.

Returning to the linear model, now suppose that the user of the forecast observes a $d$-vector of variables $V_t$ that were available to the forecast producer at the time $f^*_{t+1}$ was made. Assuming that the forecaster is rational this implies that $V_t$ is a subvector of $W_t$. For given values of $(\alpha_0, p_0)$ Proposition 1 then ensures that the following condition holds

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

Our next result shows that moment conditions (4) based on an observed subvector $V_t$ of $W_t$ are sufficient to identify $\alpha_0$.

**Lemma 2** *Under Assumptions (A0)-(A3), given $p_0 \in \mathbb{N}^*$ and given a solution $f^*_{t+1}$ to* (3), *the true value $\alpha_0 \in (0,1)$ is the unique minimum of a quadratic form*

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

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

An important implication of the result of Lemma 2 is that, in order to back out $\alpha_0$, the forecast user does not require the full vector of variables used by the forecaster $W_t$, but rather a subvector of these variables, $V_t$. This is a rather strong result which grants that with only a subvector of $W_t$ we can identify the loss function parameter $\alpha_0$, even though we cannot recover even a subset of $\theta^*$ or know the full forecasting model used to generate the forecast.

In practical applications, Lemma 2 is particularly relevant as we would generally expect that forecasters have access to not only publicly available information but also private information which is outside the information set of the forecast user. For example, it is a reasonable assumption that the IMF uses publicly available information provided by member governments in forecasting government budget deficits as well as private information gleaned from their country visits and discussions with finance ministers. However, even with only the public information available, the identification of $\alpha_0$ is still feasible.

It is this practical concern that limits our focus to linear models. The results established here continue to hold for nonlinear forecasting rules provided that $V_t$ is a subvector of the gradient of $f$ with respect to the parameter $\theta$, evaluated at $(\theta^*, W_t)$. In the linear case this gradient simplifies to $W_t$ and is therefore independent of $\theta^*$. In the nonlinear models, however, the gradient of $f$ potentially depends on both $W_t$ and the entire vector of parameter values $\theta^*$. To calculate $V_t$ we would therefore need to know the forecasting model $f$, its true parameters $\theta^*$ as well as the values of all the variables $W_t$ that were used to construct the forecast.

There are special cases (examples of nonlinear models) in which one can proceed in the same way as in Section 3 below. If the model is partially linear and $V_t$ is a subset of the linear terms, then the gradient of $f$ with respect to $\theta$ includes the vector $V_t$ and the orthogonality conditions (4) still hold. In other nonlinear models it is also possible that separability of the model would allow specification of $V_t$ with only partial knowledge of the model and variables. In these cases the results below would continue to hold with the appropriate redefinitions.

# 3 Estimating Loss Function Parameters

We now turn to the problem of recovering the true value $\alpha_0$ used in the loss minimization problem (2) assuming again that the value of $p_0$ is already known by the forecast user. If we observed the sequence $\{f^*_{t+1}\}$ of optimal one-step-ahead point forecasts $f^*_{t+1} \equiv \theta^{*\prime} w_t$ provided by the forecaster, $\alpha_0$ could be estimated directly from

$$\alpha_0 = \frac{E[V_t|Y_{t+1} - f^*_{t+1}|^{p_0-1}]' S^{-1} E[V_t 1(Y_{t+1} - f^*_{t+1} < 0)|Y_{t+1} - f^*_{t+1}|^{p_0-1}]}{E[V_t|Y_{t+1} - f^*_{t+1}|^{p_0-1}]' S^{-1} E[V_t|Y_{t+1} - f^*_{t+1}|^{p_0-1}]}, \tag{5}$$

where $S \equiv E[V_t V_t'(1(Y_{t+1} - f^*_{t+1} < 0) - \alpha_0)^2 |Y_{t+1} - f^*_{t+1}|^{2p_0-2}]$. In practice, however, we only observe the sequence $\{\hat{f}_{t+1}\}$ where $\hat{f}_{t+1} \equiv \hat{\theta}_t' w_t$ and $\hat{\theta}_t$ is an estimate of $\theta^*$ obtained by using the data up to time $t$. Let $n+1$ be the total number of periods available and assume that the first $\tau$ observations are used to produce the first one-step-ahead forecast $\hat{f}_{\tau+1}$. There are $n - \tau + 1 \equiv T$ forecasts available, starting at $t = \tau + 1$ and ending at $n + 1 = T + \tau$. These are assumed to be constructed recursively so that the parameter estimates use all information prior to the period covered by the forecast. In particular, the one-step-ahead forecast $\hat{f}_{\tau+i+1}$ of the random variable $Y_{\tau+i+1}$ is constructed using the data from $s = 1$ to $s = \tau + i$, i.e. $(y_2, w_1', \ldots, y_{\tau+i}, w_{\tau+i-1}')'$ to compute an estimate $\hat{\theta}_{\tau+i}$ of $\theta^*$. The forecast of $y_{\tau+i+1}$ is then given by $\hat{f}_{\tau+i+1} = \hat{\theta}_{\tau+i}' w_{\tau+i}$, $i = 1, ..., n-1$. Our approach allows for the possibility that the agent is recursively learning the parameters of the forecasting model. In many macroeconomic applications with small samples this is clearly more realistic than assuming that the agent's learning process has been completed.

Having observed the sequence of forecasts $\{\hat{f}_{t+1}\}_{\tau \leqslant t < T+\tau}$, we now construct an estimator for $\alpha_0$ based on equation (5). Given the $T$ observations $(v_\tau', \ldots, v_{T+\tau-1}')'$ of the $d$-vector $V_t$, we consider a linear Instrumental Variable (IV) estimator of $\alpha_0$, $\hat{\alpha}_T$, defined as

$$\hat{\alpha}_T \equiv \frac{[\frac{1}{T}\sum_{t=\tau}^{T+\tau-1} v_t|y_{t+1} - \hat{f}_{t+1}|^{p_0-1}]' \hat{S}^{-1} [\frac{1}{T}\sum_{t=\tau}^{T+\tau-1} v_t 1(y_{t+1} - \hat{f}_{t+1} < 0)|y_{t+1} - \hat{f}_{t+1}|^{p_0-1}]}{[\frac{1}{T}\sum_{t=\tau}^{T+\tau-1} v_t|y_{t+1} - \hat{f}_{t+1}|^{p_0-1}]' \hat{S}^{-1} [\frac{1}{T}\sum_{t=\tau}^{T+\tau-1} v_t|y_{t+1} - \hat{f}_{t+1}|^{p_0-1}]}, \tag{6}$$

where $\hat{S}$ is a consistent estimate of $S$. The consistency result for $\hat{\alpha}_T$ is as follows:

**Proposition 3 (Consistency)** *Given $p_0 = 1, 2$, let $\hat{\alpha}_T$ be the linear IV estimator defined in* (6). *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 $\alpha_0$ not being compact, the linear IV estimator is consistent for the true value $\alpha_0$. When the forecast rule is nonlinear, $V_t$ is now a subvector of the gradient of $f$ with respect to $\theta$, $\nabla_\theta f$, evaluated at $(\theta^*, W_t)$ where $\theta^*$ is unknown. Assuming one can observe a subvector $\hat{v}_t$ of $\nabla_\theta f$ evaluated at $(\hat{\theta}_t, W_t)$, where $\hat{\theta}_t$ is some consistent estimate of $\theta^*$, the results of Proposition 3 would still apply by replacing $v_t$ with $\hat{v}_t$ in the expression (6) for $\hat{\alpha}_T$.[6]

Results on the asymptotic distribution of $\hat{\alpha}_T$ can be established under a set of stronger mixing conditions:[7]

**Proposition 4 (Asymptotic Normality)** *Given $p_0 = 1, 2$, let $\hat{\alpha}_T$ be the linear IV estimator defined in* (6). *Under Assumptions (A0)-(A4), (A5') and (A6)-(A7), $\hat{\alpha}_T$ exists with probability approaching one and*

$$T^{1/2}(\hat{\alpha}_T - \alpha_0) \xrightarrow{d} \mathcal{N}(0, (h^{*\prime} S^{-1} h^*)^{-1}),$$

*where $S$ is defined after equation (5) and $h^* \equiv E[V_t \cdot |Y_{t+1} - f^*_{t+1}|^{p_0 - 1}]$.*

The linear IV estimator $\hat{\alpha}_T$ is asymptotically normal with asymptotic variance that does not depend on either $W_t$ or $\theta^*$, both of which are a priori unknown to the forecast user. Indeed, the asymptotic variance of $\hat{\alpha}_T$ is identical to that obtained with a standard GMM estimator. This stems from the slightly faster rate at which the forecaster's sample grows relative to the evaluator's sample. The result requires that the forecaster uses a consistent estimator, but not neccessarily an optimal one.

In practice, the computation of the linear IV estimator $\hat{\alpha}_T$ is done iteratively. Estimation of $\hat{\alpha}_T$ requires a consistent estimator of $S^{-1}$, which in turn depends on $\alpha_0$. $S$ can however be consistently estimated by replacing the population moment by a sample average and the true parameter by its estimated value, for example, $\hat{S}(\bar{\alpha}_T) \equiv T^{-1} \sum_{t=\tau}^{T+\tau-1} v_t v_t' (1(y_{t+1} - \hat{f}_{t+1} <$

[6] Note that the assumptions (A1), (A3) and (A5) used in the proof of Proposition 3 need to be appropriately modified.

[7] For general results on asymptotic inference in the presence of parameter uncertainty, see West (1996), West and McCracken (1998), McCracken (2000) and Corradi and Swanson (2002). Propositions 3 and 4 focus on the cases where $p_0 = 1, 2$. These are likely to be the cases most useful in empirical analysis as they nest MAE and MSE loss. The results are extendable to $p_0 > 2$ using the same approach.

$0) - \bar{\alpha}_T)^2|y_{t+1} - \hat{f}_{t+1}|^{2p_0-2}$, where $\bar{\alpha}_T$ is a consistent initial estimate of $\alpha_0$, or by using a robust estimator, such as Newey and West's (1987) estimator.[8] Computation of $\hat{\alpha}_T$ is then carried out by first choosing $S = I_{d\times d}$ and using (6) 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 (6) 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 $\hat{\alpha}_{T,j-1}$. Consistent estimates of the asymptotic variance of $\hat{\alpha}_{T,j}$ are obtained by replacing $S$ and $h^*$ in Proposition 4, with their consistent estimates $\hat{S}(\hat{\alpha}_{T,j})$ and $\hat{h}_T \equiv T^{-1}\sum_{t=\tau}^{T+\tau-1} v_t|y_{t+1} - \hat{f}_{t+1}|^{p_0-1}$, respectively.

In the single instrument case ($d = 1$), $\hat{\alpha}_T$ can be interpreted as justifying biased forecasts by adjusting the loss function to make them optimal.[9] However if indeed the forecasts are rational, then $V_t$ is a subvector of $W_t$ and all moment conditions must hold simultaneously. Thus a test for overidenfication when $d > 1$ provides a joint test of rationality of the forecasts and the more flexible loss function. One degree of freedom is used in the estimation of the loss parameter, $\hat{\alpha}_T$, so, from the results of Proposition 4, we have

**Corollary 5 (Rationality Test)** *Under the assumptions of Proposition 4, for a given value $p_0 = 1, 2$, a joint test of forecast rationality and the flexible loss function (1) can be conducted with $d > 1$ instruments through the test statistic*

$$J = \frac{1}{T}\left[\left(\sum_{t=\tau}^{T+\tau-1} v_t[1(y_{t+1} - \hat{f}_{t+1} < 0) - \hat{\alpha}_T]|y_{t+1} - \hat{f}_{t+1}|^{p_0-1}\right)' \hat{S}^{-1} \cdot \left(\sum_{t=\tau}^{T+\tau-1} v_t[1(y_{t+1} - \hat{f}_{t+1} < 0) - \hat{\alpha}_T]|y_{t+1} - \hat{f}_{t+1}|^{p_0-1}\right)\right] \sim \chi^2_{d-1}. \tag{7}$$

Tests based on an assumption of MSE loss are closely related to this test when $p_0$ is chosen to be equal to 2. The difference is that if indeed $\alpha_0 = 0.5$, tests based on MSE loss impose this restriction, whereas our test uses a consistent estimate of $\alpha$ which is treated as unknown. However, if $\alpha_0 \neq 1/2$ then standard tests would have power in this direction. Our use of a consistent test avoids this problem and controls for size if the forecaster's loss function reflects a different value of $\alpha_0$. Asymptotically there is no loss from relaxing the

[8] Consistency of $\hat{S}(\bar{\alpha}_T)$ can be shown by an argument analogous to the one in the proof of Proposition 3.

[9] When $d = 1$ the estimator is independent of $S$ and a closed form solution exists. For example, when $p_0 = 1$ and $V_t = 1$ the estimator is simply the proportion of negative forecast errors.

assumption that $\alpha_0 = 0.5$, but there is clearly a gain in terms of directing power in the desired direction.

Different choices of $V_t$ in the construction of our estimator $\hat{\alpha}_T$ result in different asymptotic variances, which naturally raises the question of how to optimally choose the instruments. It is possible to show that $\hat{\alpha}_T$ is asymptotically optimal - in the sense that its asymptotic variance is minimal - when $V_t = W_t$, i.e. when the forecast user has all the information used by the forecaster. Outside this situation, one could attempt to use data-based methods for selection of moment conditions using criteria such as those proposed by Donald and Newey (2001), replacing their MSE loss with our loss $L$ in (1) evaluated at $(p_0, \hat{\alpha}_T)$ where $\hat{\alpha}_T$ is a consistent estimate of $\alpha_0$.

# 4 Simulation Results

We briefly examine the behavior of the proposed estimator (6) and test (7) in a Monte Carlo experiment. Random data samples were generated by a linear forecasting model

$$Y_{t+1} = \theta' W_t + U_t$$

with the vector $W_t \equiv (1, W_{1t}, W_{2t})'$ where $W_{1t} \sim N(1,1), W_{2t} \sim N(-1,1), \theta = [1, 0.5, 0.5]$ and $U_t \sim N(0, 0.5)$. 5000 Monte Carlo simulation experiments were undertaken for different numbers of initial values available for estimating $\theta$ recursively (such data are available to the forecaster before the initial forecast is recorded), denoted by $n_0$, and for different numbers of data available for estimation of $\alpha_0$ and testing, denoted by $n_f$. For $p_0 = 1$ recursive forecasts were computed using quantile regression methods and for $p_0 = 2$ the nonlinear least squares estimation method of Newey and Powell (1987) was used to estimate $\theta$ recursively.

Panel A in Table 1 examines, for various sample sizes and values of $\alpha_0$, the size of $t$-tests testing $\hat{\alpha} = \alpha_0$ (i.e. the true value) against two sided alternatives for a size of 5%. Results are reported for $p_0 = 1$ (lin-lin) and $p_0 = 2$ (quad-quad) using only a constant as an instrument, i.e. $V_t = 1$. Size is well controlled overall, even when $\alpha_0$ is far from one half (on average). Size is less well controlled for the quad-quad loss function. The reason for this is straightforward: for the asymmetric models the forecast 'errors' are less well balanced above and below the true value so we obtain asymmetric small sample distributions and require a larger $n$ for the central limit theorem to provide a good approximation.[10] Additional in-sample or out-of-sample observations help to control the size.

Panel B employs the two time-varying instruments, $W_{1t}, W_{2t}$ in addition to the constant, i.e. $V_t = W_t$. Including extra instruments results in larger size distortions across the board. The problem is again more of an issue for the quad-quad than for the lin-lin loss function. As expected, size distortions are less of a problem when more observations are available. As before, additional out-of-sample observations play a particularly important role in controlling size. Problems are again greater, the further $\alpha_0$ is from one half.

The proposed tests for overidentification that examine whether the moment conditions are compatible with *some* $\alpha_0$ are reported in Panel C. Size is generally well controlled although the tests tend to be undersized rather than oversized, and departures from nominal size (5%) are larger when $\alpha_0$ is further away from one half. When $\alpha_0 = 1/2$, empirical size is very close to nominal size for all samples. Increasing the sample helps, adding more out of sample observations once again appearing to be more useful.

# 5 Government Deficit Forecasts

In this section we apply our estimation methods and tools for inference to the optimality of forecasts of government budget deficits for the G7 countries produced by two international organizations, namely the IMF and the OECD. This application is well suited to demonstrate our methods since, as pointed out by Artis and Marcellino (2001) "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). A similar point is made by Campbell and Ghysels (1995) in the context of an analysis of federal budget projections.

Our data comprises budget deficit forecasts, reported as a percentage of GDP, for the G7 countries and is reported as budget surpluses so that a budget deficit takes a negative value.[11] 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

[10] This is identical to the usual result in applying the central limit theorem to Bernoulli outcomes.

[11] We are grateful to Massimiliano Marcellino for providing the first part of the data. The data source is the IMF's World Economic Outlook and the OECD's Economic Outlook.

predicted deficit than the one that actually occurred. We refer to this as an overprediction of the budget deficit (underprediction of the budget surplus). In all 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 OECD data cover France, Germany, Italy and the UK, contains between 24 and 27 data points and goes from 1975 - 2001. The IMF sample has information on all G7 countries, goes from 1976 to 2000 and thus contains 25 observations. These are not large samples, so some caution should be exercised in the interpretation of the results.

In our empirical tests we first assume that the loss function is lin-lin ($p_0 = 1$). Authors such as Granger and Newbold (1986) have argued that lin-lin loss approximates other classes of asymmetric loss functions. For robustness we report results for four separate sets of instruments: (i) a constant; (ii) a constant and the lagged 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. For robustness we also conduct empirical tests under the assumption of quad-quad loss ($p_0 = 2$).

## 5.1 Evidence of Asymmetric Loss

Inspection of the forecast errors showed that overpredictions of budget deficits (positive average forecast errors) are most common - between 19 and 21 of 25 current-year IMF forecast errors are positive for Italy, Japan, UK and the US - although for Canada we found evidence of underpredictions (negative average forecast errors). Under the assumption that the loss function is piecewise linear (lin-lin), Table 2 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 are indicative of the forecaster's objectives.

First consider the current-year IMF forecasts when the model is exactly identified and a constant is the only instrument. Five of seven countries generate $\alpha$-estimates below one-half, one country (France) has an estimate (0.52) close to one-half and another country (Canada) has an $\alpha-$estimate of 0.60. The null of symmetry ($\alpha = 0.5$) is strongly rejected for Italy, Japan, UK and the US. Similar results are obtained for the 1-year-ahead IMF predictions, where the $\alpha$-estimates are significantly different from one-half for Italy, 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 for $\hat{\alpha}$ tend to decline. Hence, the null of symmetric loss is rejected with $p$-values less than 0.01 for Italy, Japan, UK and the US. In each case the point estimates for these four countries are below 0.25, thus suggesting economically strong evidence of asymmetry. At the 1-year horizon the null of symmetric loss continues to be rejected at or below the 5% level for Italy, Japan, UK and the US.

Turning to the OECD forecasts, for the current year predictions all four countries generate estimates of $\alpha$ below one half. Irrespective of the set of instruments used, the null of symmetric loss is rejected at the five percent significance level for France, Germany and Italy although the evidence of asymmetric loss is somewhat weaker at the 1-year horizon.

These results suggest that the IMF and OECD systematically overpredict government budget deficits. This is consistent with a loss function that penalizes underpredictions more heavily than overpredictions. The point estimates of $\alpha$ suggest strong asymmetries in the loss function both from an economic and a statistical point of view. For some countries they indicate that under-predictions of budget deficits are viewed as up to three times costlier than over-predictions.

## 5.2 Tests of Forecast Rationality

The shape parameters of the loss function provide important information about the forecaster's objectives. Ultimately, however, we are interested in testing whether the IMF and OECD forecasts are consistent with rationality. To test this, and to investigate what is driving our empirical results, 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. We can test this hypothesis by imposing $\alpha = 1/2$ and examining the $J$-test (7) which follows a $\chi^2_d$-distribution under this restriction.

The outcome of the joint tests of rationality and $\alpha = 1/2$ is reported in panel A of Table 3. The null hypothesis is rejected at the 5% level in exactly half of the tests (44 out of 88 cases). In the IMF data there is strong evidence against the composite null hypotheses for Italy, Japan, UK and the US, while the OECD data leads to rejections of the null in the current-year data for France, Germany and Italy and, in the 1-year forecasts, for Germany.

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 ($\alpha \neq 1/2$). The results - shown in Panel B of Table 3 - are very interesting and in complete contrast to those found in panel A. There is only very weak evidence against the composite null hypothesis of forecast rationality and a loss function belonging to the family (1). Overall there are only six cases where the null gets rejected at the 5% significance level. The only cases where two instrument sets lead to a rejection for the same country are Japan (1-year IMF forecasts) and France (current-year OECD forecasts). Comparing the results in Panels A and B it appears that the systematic rejections of the composite null hypothesis of symmetric loss and forecast optimality can be attributed to asymmetric loss in the current-year forecasts for Italy, Japan, the UK and the US and, in the case of 1-year forecasts, for Italy, UK and the US.

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, Table 4 reports empirical results for the quad-quad loss function. In the current-year IMF forecasts the joint hypothesis of MSE loss and rationality (Panel A) is strongly rejected for Italy, Japan, UK and the US. This null gets rejected for France, Germany and Italy in the current-year OECD forecasts. 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, UK and the US and, in the OECD data, also for Germany. Overall, the null continues to get rejected at the 5 percent level in nearly half of all tests (42 of 88 cases). Allowing for asymmetric quadratic loss (Panel B), the evidence against rationality is far weaker. The null gets rejected at the 5% level for the current-year data only in a single case. At the 1-year horizon, the null is strongly rejected by the IMF predictions only in three cases and in a single case in the OECD data. In total the null is only rejected at the 5% level in five cases.

Overall our conclusions thus appear to be 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 class over the other. Consistent with our findings under lin-lin loss, the tests of 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 under asymmetric quadratic loss.

# 6 Conclusion

This paper provided theory for identification and estimation of the parameters of loss functions applicable to situations where time-series data on point forecasts is available but the underlying model used by the forecaster is unknown. We also provided test statistics that can be used when testing the composite null that loss belongs to a general family of loss functions and that information is used efficiently in the computation of the forecasts.

Our estimator and test statistics are easy to compute and should find a number of practical applications. Once the limitations and restrictiveness of MSE loss are acknowledged, it clearly becomes more attractive to allow for more general classes of loss when testing forecast rationality. Most often the forecast producer's loss function is unobserved and a reasonable approach will not want to impose too much structure on this unknown loss function. Since the vast majority of work in the empirical forecasting literature has maintained MSE loss, many empirical results need to be revisited using methods such as those advocated here.

# Appendix A: Assumptions and Notation

**Notation**:

Given $p_0 \in \mathbb{N}^*$, for every $t$, $\tau \leqslant t \leqslant T+\tau-1$, and any $(\theta, \alpha_0) \in \Theta \times (0,1)$, we let $G_{t+1}(\theta) \equiv V_t 1(Y_{t+1} - \theta' W_t < 0)|Y_{t+1} - \theta' W_t|^{p_0-1}$, $H_{t+1}(\theta) \equiv V_t |Y_{t+1} - \theta' W_t|^{p_0-1}$ and $M_{t+1}(\theta, \alpha_0) \equiv G_{t+1} - \alpha_0 H_{t+1}$, and denote by $g_{t+1}$, $h_{t+1}$ and $m_{t+1}$, respectively, their realizations. Further, we let $g_T$, $h_T$ and $m_T$ denote the sample means of $G_{t+1}$, $H_{t+1}$ and $M_{t+1}$, respectively, i.e. $g_T \equiv T^{-1} \sum_{s=\tau+1}^{T+\tau} g_s$, $h_T \equiv T^{-1} \sum_{s=\tau+1}^{T+\tau} h_s$ and $m_T \equiv T^{-1} \sum_{s=\tau+1}^{T+\tau} m_s$, and we denote by $g$, $h$ and $m$ the expected values of $G_{t+1}$, $H_{t+1}$ and $M_{t+1}$, respectively. To shorten the notation: when $\theta = \hat{\theta}_t$ we add a "hat" to all the above quantities, i.e we use the notation $\hat{G}_{t+1}$ ($\hat{g}_T$), $\hat{H}_{t+1}$ ($\hat{h}_T$) and $\hat{M}_{t+1}$ ($\hat{m}_T$). Also, we let $\hat{g} \equiv E[\hat{G}_{t+1}]$, $\hat{h} \equiv E[\hat{H}_{t+1}]$ and $\hat{m} \equiv E[\hat{M}_{t+1}]$. Similarly, when $\theta = \theta^*$ we add a "star" to $G_{t+1}$, $H_{t+1}$ and $M_{t+1}$ and their sample means, i.e we use the notation $G^*_{t+1}$ ($g^*_T$), $H^*_{t+1}$ ($h^*_T$) and $M^*_{t+1}$ ($m^*_T$). In that case, the expected values are denoted $g^* \equiv E[G^*_{t+1}]$, $h^* \equiv E[H^*_{t+1}]$ and $m^* \equiv E[M^*_{t+1}]$.

**Assumptions**:

(A0) $\Theta$ is a compact subset of $\mathbb{R}^h$ and $\theta^*$ is interior to $\Theta$, 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 any $\theta^* \in \mathring{\Theta}$, $E[W_t|Y_{t+1} - \theta^{*\prime} W_t|^{p_0-1}] \neq 0$ element by element and $E[W_t W_t']$ exists and is positive definite;

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

(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 $K > 0$ such that for every $t$, $\tau \leqslant t \leqslant T+\tau-1$, $|W_t|^2 = W_t' W_t \leqslant K$, $a.s. - P$;

(A4) for every $t$, $\tau \leqslant t \leqslant T+\tau-1$, $\hat{\theta}_t$ is a consistent estimator of $\theta^*$, with $\theta^* \in G \subseteq \mathring{\Theta}$;

(A5) the stochastic processes $\{Y_t\}$ and $\{W_t\}$ are strictly stationary and $\alpha$-mixing with mixing coefficient $\alpha$ of size $-r/(r-1)$, $r > 1$, and, given $p_0 \in \mathbb{N}^*$, there exist some $\delta_Y > 0$ and $\Delta_Y > 0$ such that $\sup_{\theta \in \Theta} E[(Y_{t+1} - \theta' W_t)^{2(r+\delta_Y)(p_0-1)}] \leqslant \Delta_Y < \infty$ and some $\delta_W > 0$ and $\Delta_W > 0$ such that $E[|W_t|^{2(r+\delta_W)}] \leqslant \Delta_W < \infty$;

(A5') the stochastic processes $\{Y_t\}$ and $\{W_t\}$ are strictly stationary and $\alpha$-mixing with mixing coefficient $\alpha$ of size $-r/(r-2)$, $r > 2$, and, given $p_0 \in \mathbb{N}^*$, there exist some $\Delta_Y > 0$ such that $\sup_{\theta \in \Theta} E[(Y_{t+1} - \theta' W_t)^{2r(p_0-1)}] \leqslant \Delta_Y < \infty$ and some $\Delta_W > 0$ such that $E[|W_t|^{2r}] \leqslant$

$\Delta_W < \infty$;

(A6) for every $t$, $\tau \leqslant t \leqslant T + \tau - 1$, the density of $Y_{t+1}$ conditional on $\mathcal{F}_t$ is bounded, i.e. there exists some $C > 0$ such that $\sup_{y \in \mathbb{R}} f_t^0(y) \leqslant C < \infty$;

(A7) for some small $\varepsilon, \varepsilon \in (0,1)$: (i) $\tau^{1-2\varepsilon}/T \to \infty$ and (ii) $\sup_{\tau \leqslant t \leqslant T+\tau-1} |t^{1/2-\varepsilon}(\hat{\theta}_t - \theta^*)| \xrightarrow{p} 0$, as $\tau \to \infty$ and $T \to \infty$.

# Appendix B: Proofs

**Proof of Proposition 1.** We first show that (3) is a necessary condition for optimality of $f_{t+1}^* = \theta^{*\prime} W_t$. From (2) we know that $\theta^* \in \mathring{\Theta}$ is a solution to $\min_{\theta \in \Theta} \Sigma(\theta)$, where $\Sigma(\theta) \equiv E[\Sigma_{t+1}(\theta)]$ and $\Sigma_{t+1}(\theta) \equiv [\alpha_0 + (1 - 2\alpha_0)1(Y_{t+1} - \theta' W_t < 0)]|Y_{t+1} - \theta' W_t|^{p_0}$. 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 $\Theta \backslash A_{t+1}$. By the law of iterated expectations $\Sigma(\theta) = E\{E_{t+1}[\Sigma_{t+1}(\theta)]\}$, so that $\nabla_\theta \Sigma(\theta) = E\{\nabla_\theta \Sigma_{t+1}(\theta) E_{t+1}[1(\theta \in A_{t+1}^c)]\}$ $+E\{\nabla_\theta \Sigma_{t+1}(\theta) E_{t+1}[1(\theta \in A_{t+1})]\}$, where $E_{t+1}[1(\theta \in A_{t+1}^c)] = 1$ and $E_{t+1}[1(\theta \in A_{t+1})] = 0$. Hence, $\Sigma(\theta)$ is continuously differentiable on $\Theta$ and we have $\nabla_\theta \Sigma(\theta) = (1 - 2\alpha_0) E[\nabla_\theta 1(Y_{t+1} - \theta' W_t < 0)|Y_{t+1} - \theta' W_t|^{p_0}]$ $+p_0 E[W_t(1(Y_{t+1} - \theta' W_t < 0) - \alpha_0)|Y_{t+1} - \theta' W_t|^{p_0 - 1}]$. Note that $\nabla_\theta 1(Y_{t+1} - \theta' W_t < 0) = W_t \cdot \delta(\theta' W_t - Y_{t+1})$ where $\delta$ is the Dirac function, so that $E[W_t \cdot \delta(\theta' W_t - Y_{t+1}) \cdot |Y_{t+1} - \theta' W_t|^{p_0}] = 0$, for any non-zero $p_0$. Thus,

$$\nabla_\theta \Sigma(\theta) = p_0 E[W_t(1(Y_{t+1} - \theta' W_t < 0) - \alpha_0)|Y_{t+1} - \theta' W_t|^{p_0 - 1}]. \tag{8}$$

For given values of $p_0$, $p_0 \in \mathbb{N}^*$, and $\alpha_0$, $\alpha_0 \in (0,1)$, if $\theta^* \in \mathring{\Theta}$ is the minimum of $\Sigma(\theta)$, then $\theta^*$ is a solution to $\nabla_\theta \Sigma(\theta^*) = 0$ (c.f. Theorem 3.7.13 in Schwartz, 1997, vol 2, p 168), i.e. (3) holds for $f_{t+1}^* = \theta^{*\prime} W_t$, which completes the neccessity part of the proof. We now derive a set of sufficient conditions for $\theta^* \in \mathring{\Theta}$ to be a solution to the minimization problem (2). We know that $\theta^*$ 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). The first order condition $\nabla_\theta \Sigma(\theta^*) = 0$ is implied by the orthogonality condition (3). We now show that $\Delta_{\theta\theta} \Sigma(\theta^*)$ is positive definite. By an argument similar to that above we have

$$\begin{aligned} \Delta_{\theta\theta} \Sigma(\theta) = \; & p_0 E[W_t W_t' \cdot \delta(\theta' W_t - Y_{t+1}) \cdot |Y_{t+1} - \theta' W_t|^{p_0 - 1}] \\ & + p_0(p_0 - 1) E\{W_t W_t' [\alpha_0 + (1 - 2\alpha_0)1(Y_{t+1} - \theta' W_t < 0)]|Y_{t+1} - \theta' W_t|^{p_0 - 2}\}. \end{aligned} \tag{9}$$

Consider the following two cases separately: $p_0 = 1$ and $p_0 > 1$. When $p_0 = 1$ then $\Delta_{\theta\theta}\Sigma(\theta) = E[W_t W_t' \cdot \delta(\theta' W_t - Y_{t+1})] = E[W_t W_t' \cdot f_t^0(\theta' W_t)]$, where $f_t^0$ is the density of $Y_{t+1}$ conditional on $\mathcal{F}_t$. By (A2) $f_t^0 > 0$ and by (A1) $E[W_t W_t']$ is positive definite giving the result. For any $\theta \in \mathring{\Theta}$ the matrix $\Delta_{\theta\theta}\Sigma(\theta)$ is positive definite, therefore it is positive definite at $\theta^*$. When $p_0 > 1$, $\Delta_{\theta\theta}\Sigma(\theta) = p_0(p_0-1)E\{W_t W_t' \cdot E_t[(\alpha_0 + (1-2\alpha_0)1(Y_{t+1} - \theta' W_t < 0))|Y_{t+1} - \theta' W_t|^{p_0-2}]\}$, with $E_t[(\alpha_0 + (1-2\alpha_0)1(Y_{t+1} - \theta' W_t < 0))|Y_{t+1} - \theta' W_t|^{p_0-2}] > 0$, $a.s.-P$, for any $(\alpha_0, \theta) \in (0,1) \times \Theta$. So by this and (A1) for every $\theta \in \mathring{\Theta}$ the matrix $\Delta_{\theta\theta}\Sigma(\theta)$ is positive definite, then so must be for $\theta^*$. Thus any $f_{t+1}^* = \theta^{*\prime} W_t$ which satisfies the moment condition (3) is a solution to (2), which completes the sufficiency part of the proof. We now use the implicit function theorem to show that for any realization of $W_t$, the function $f_{t+1}^* = \theta_{p_0}(\alpha)' W_t$ defined implicitly by (3) is a one-to-one mapping from the set of asymmetry parameters $(0,1)$ to the set of forecasts $\mathbb{R}$. Given $p_0 = 1, 2$, define $\varphi_{p_0}(\alpha, \theta) \equiv E[W_t(1(Y_{t+1} - \theta' W_t < 0) - \alpha)|Y_{t+1} - \theta' W_t|^{p_0-1}]$, so (3) is $\varphi_{p_0}(\alpha_0, \theta^*) = 0$. The function $\varphi_{p_0} : (0,1) \times \Theta \to \mathbb{R}^k$ is continuously differentiable on $(0,1) \times \Theta$, and we have $\partial\varphi(\alpha, \theta)/\partial\alpha = -E[W_t|Y_{t+1} - \theta' W_t|^{p_0-1}]$, and $\partial\varphi_{p_0}(\alpha, \theta)/\partial\theta = \Delta_{\theta\theta}\Sigma(\theta)$ where $\Delta_{\theta\theta}\Sigma(\theta)$ as in (9). For every $\alpha_0 \in (0,1)$, the $\mathbb{R}^k \times \mathbb{R}^k$-matrix $\partial\varphi_{p_0}(\alpha_0, \theta^*)/\partial\theta$ is nonsingular, given that $\Delta_{\theta\theta}\Sigma(\theta^*)$ is positive definite. 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 exists an open interval $E_0$ containing $\alpha_0$ and an open set $G_0$ containing $\theta^*$, $G_0 \equiv \{\theta \in \mathring{\Theta} : |\theta - \theta^*| < \delta_0\}$ with $\delta_0 > 0$, such that for every $\alpha \in E_0$, the equation $\varphi_{p_0}(\alpha, \theta) = 0$ has a unique solution $\theta$ in $G_0$, and the function $\theta = \theta_{p_0}(\alpha)$ defined implicitly by $\varphi_{p_0}(\alpha, \theta_{p_0}(\alpha)) = 0$ is continuously differentiable from $E_0$ to $G_0$ with $d\theta_{p_0}(\alpha)/d\alpha = -[\partial\varphi_{p_0}(\alpha, \theta_{p_0}(\alpha))/\partial\theta]^{-1} \cdot \partial\varphi_{p_0}(\alpha, \theta_{p_0}(\alpha))/\partial\alpha$, i.e.

$$d\theta_{p_0}(\alpha)/d\alpha = [\Delta_{\theta\theta}\Sigma(\theta_{p_0}(\alpha))]^{-1} E[W_t|Y_{t+1} - \theta_{p_0}(\alpha)' W_t|^{p_0-1}]. \tag{10}$$

We now extend the previous implicit function argument by continuity to the entire open interval $(0,1)$. Let $G \equiv \bigcup_{\alpha_0 \in (0,1)} G_0$. $G$ being a union of open sets 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 $\varphi_{p_0}(\alpha_0, \theta) = 0$ has a unique solution $\theta^*$ in $G$ and the implicit function $\theta^* = \theta_{p_0}(\alpha_0)$ is continuously differentiable from $(0,1)$ to $G$ with $d\theta_{p_0}(\alpha)/d\alpha$ as given in (10). In particular, for any realization of $W_t$, the function $f_{t+1}^* = \theta_{p_0}(\alpha_0)' W_t$ is continuously differentiable from $(0,1)$ to $\mathbb{R}$. Finally, we show that $\theta_{p_0}(\alpha)$ is a one-to-one mapping (or bijective) from $(0,1)$ to

$G$. It is surjective by construction, so we only need to show that it is injective on $(0,1)$, i.e. $\theta_{p_0}(\alpha_1) = \theta_{p_0}(\alpha_2)$ implies $\alpha_1 = \alpha_2$. If $\theta_{p_0}(\alpha_1) = \theta_{p_0}(\alpha_2)$ then from (3) we know $0 = (\alpha_2 - \alpha_1)E[W_t|Y_{t+1} - \theta_{p_0}(\alpha_2)'W_t|^{p_0-1}]$, which, by (A1), implies $\alpha_1 = \alpha_2$. Using identifiability of a linear forecast rule, we know that for any realization of $W_t$, there exists a unique $\theta^* \in G$ such that $f^*_{t+1} = \theta^{*\prime}W_t$, hence by using the previous result there is a unique $\alpha_0 \in (0,1)$ such that $f^*_{t+1} = \theta_{p_0}(\alpha_0)'W_t$. This completes the proof of Proposition 1 ■

**Proof of Lemma 2.** If $S$ (and hence $S^{-1}$) is positive definite, then by using convexity we have that $Q_0(\alpha) \equiv m^{*\prime}S^{-1}m^*$ admits a unique minimum in $(0,1)$. It follows directly that $\alpha^* = (h^{*\prime}S^{-1}h^*)^{-1}(h^{*\prime}S^{-1}g^*)$. We need to verify that $\alpha^*$ lies 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. $V_t > 0_d$, $a.s. - P$. In that case we have $0 \leqslant G^*_{t+1} \leqslant H^*_{t+1}$, $a.s. - P$, so that $0 \leqslant g^* \leqslant h^*$. Using (A1) we know that $0 < g^*$ since $V_t$ is a sub-vector of $W_t$. Knowing that $S^{-1}$ is positive definite, we then have $0 < g^*S^{-1}g^* \leqslant g^*S^{-1}h^* \leqslant h^*S^{-1}h^*$. Hence $\alpha^* > 0$. Also, for all $\alpha \in (0,1)$, $Q_0(\alpha) > 0$ so that the reduced discriminant of $Q_0(\alpha)$ is negative. Hence, $h^*S^{-1}g^* < \sqrt{h^*S^{-1}h^*g^*S^{-1}g^*} \leqslant h^*S^{-1}h^*$ so $\alpha^* < 1$. So, if $V_t > 0_d$, then $\alpha^* \in (0,1)$. Now consider a case where the first element of $V_t$ is a constant 1 and there exists some constant $c > 0$ such that $V_t > -c \cdot 1_d$, $a.s. - P$, where $1_d$ is a $d$-vector of ones. This inequality is implied by (A3). 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 $\bar{V}_t = KV_t > 0$, $a.s. - P$. As $K$ is positive definite, $(K^{-1})'S^{-1}K^{-1}$ is positive definite if $S^{-1}$ is positive definite. Now, note that

$$\alpha^* = \frac{E[\bar{V}_t|Y_{t+1} - f^*_{t+1}|^{p_0-1}]'(K^{-1})'S^{-1}K^{-1}E[\bar{V}_t 1(Y_{t+1} - f^*_{t+1} < 0)|Y_{t+1} - f^*_{t+1}|^{p_0-1}]}{E[\bar{V}_t|Y_{t+1} - f^*_{t+1}|^{p_0-1}]'(K^{-1})'S^{-1}K^{-1}E[\bar{V}_t|Y_{t+1} - f^*_{t+1}|^{p_0-1}]},$$

so if $\alpha^*$ is the minimum of $Q_0(\alpha)$ then $\alpha^*$ is also a minimum of the quadratic form $\bar{Q}(\alpha)$, with $\bar{Q}(\alpha) \equiv E[\bar{V}_t(1(Y_{t+1} - f^*_{t+1} < 0) - \alpha)|Y_{t+1} - f^*_{t+1}|^{p_0-1}]'\ K^{-1}S^{-1}(K^{-1})'\ E[\bar{V}_t(1(Y_{t+1} - f^*_{t+1} < 0) - \alpha)|Y_{t+1} - f^*_{t+1}|^{p_0-1}]$. From the results above we then know that $\alpha^* \in (0,1)$ since $\bar{V}_t > 0$, $a.s. - P$. Hence, under (A0)-(A3), $Q_0(\alpha)$ is uniquely minimized at $\alpha^* \in (0,1)$. We now show that $\alpha_0$ is also a minimum of $Q_0(\alpha)$: given concavity of $Q_0(\alpha)$, any solution to the first order condition $0 = h^* S^{-1} h^* - \alpha h^* S^{-1} g^* = h^{*\prime} S^{-1} m^*$ is a minimum of $Q_0(\alpha)$. We know that if $V_t$ is a sub-vector of $W_t$ (A3) then $h^* \neq 0$ (A1). Moreover, $S^{-1}$ is nonsingular, so $h^{*\prime} S^{-1} m^* = 0$ implies $m^* = 0$. We know from (3) that $\alpha_0$ satisfies $m^* = 0$, so $\alpha_0$ is a minimum of $Q_0(\alpha)$. By uniqueness, we conclude that $\alpha_0 = \alpha^*$, which completes the proof. ■

**Proof of Proposition 3.** First, let us show that $S$ is positive definite. Given $p_0 \in \mathbb{N}^*$, we have $S \equiv E[M^*_{t+1} M^{*\prime}_{t+1}] = E[V_t V_t' (1(Y_{t+1} - f^*_{t+1} < 0) - \alpha)^2 |Y_{t+1} - f^*_{t+1}|^{2p_0 - 2}]$, so that for every $\xi \in \mathbb{R}^d$ we have $\xi' S \xi = E[\xi' V_t V_t' \xi (1(Y_{t+1} - f^*_{t+1} < 0) - \alpha)^2 |Y_{t+1} - f^*_{t+1}|^{2p_0-2}]$. Note that $(1(Y_{t+1} - f^*_{t+1} < 0) - \alpha)^2 |Y_{t+1} - f^*_{t+1}|^{2p_0-2} > 0$, $a.s. - P$, so $\xi' S \xi = 0 \Rightarrow \xi' V_t V_t' \xi = 0$, $a.s. - P$ $\Rightarrow \xi' E[V_t V_t'] \xi = 0$. The positive definiteness of $E[W_t W_t']$ (A1) implies $E[V_t V_t']$ is positive definite, hence $\xi' E[V_t V_t'] \xi = 0 \Rightarrow \xi = 0$, which shows that $S$ is positive definite. Recall from (6) that we have $\hat{\alpha}_T \equiv (\hat{h}_T \hat{S}^{-1} \hat{h}_T)^{-1} \hat{h}_T \hat{S}^{-1} \hat{g}_T$. To show $\hat{\alpha}_T \xrightarrow{p} \alpha_0$ it is sufficient to show that: *(i)* $\hat{h}_T - \hat{h} \xrightarrow{p} 0$, and *(ii)* $\hat{g}_T - \hat{g} \xrightarrow{p} 0$. Then, by using Lemma 2, the consistency of $\hat{S}$, $\hat{S} \xrightarrow{p} S$, the positive definiteness of $S$ (and thus of $S^{-1}$), (A1) and (A3) which ensure that $\hat{h} \neq 0$ and $\hat{g} \neq 0$, and the continuity of the inverse function (away from zero), we have that $\hat{\alpha}_T \xrightarrow{p} \alpha_0$. By the triangle inequality we have $|\hat{g}_T - g^*| \leqslant |\hat{g}_T - \hat{g}| + |\hat{g} - g^*|$ and $|\hat{h}_T - h^*| \leqslant |\hat{h}_T - \hat{h}| + |\hat{h} - h^*|$. We first show that as $T \to \infty$, $|\hat{g}_T - \hat{g}| \xrightarrow{p} 0$ and $|\hat{h}_T - \hat{h}| \xrightarrow{p} 0$ by using a law of large numbers (LLN) for $\alpha$-mixing sequences (e.g., Corollary 3.48 in White 2001). From Theorem 3.49 in White (2001) measurable functions of mixing processes are mixing of the same size. Hence, by (A5) we have $\{\hat{\theta}_t\}$, $\{\hat{H}_{t+1}\}$ and $\{\hat{G}_{t+1}\}$ are $\alpha$-mixing of size $-r/(r-1)$ with $r > 1$. Note that if $\hat{\theta}_t$ were constructed with a rolling window, i.e. as a function of a constant number of past observations, then Theorem 3.35 in White (2001) would apply and we could also say that $\{\hat{\theta}_t\}$, $\{\hat{H}_{t+1}\}$ and $\{\hat{G}_{t+1}\}$ are strictly stationary. Let $\delta_H \equiv \min(\delta_Y, \delta_W)/2 > 0$. By (A5), the Cauchy-Schwartz inequality, and using $E[|V_t|^{2(r+\delta_H)}] \leqslant E[|W_t|^{2(r+\delta_H)}]$, we know that for any $t$, $\tau \leqslant t \leqslant T + \tau - 1$, $E[|\hat{H}_{t+1}|^{r+\delta_H}] \leqslant (E[|V_t|^{2r+2\delta_H}])^{1/2} (E[(Y_{t+1} - \hat{f}_{t+1})^{2(r+\delta_H)(p_0-1)}])^{1/2} \leqslant (E[|V_t|^{2r+2\delta_H}])^{1/2} \max(1, \{\sup_{\theta \in \Theta} E[(Y_{t+1} - \theta' W_t)^{2(r+\delta_H)(p_0-1)}]\}^{1/2})$. Hence $E[|\hat{H}_{t+1}|^{r+\delta_H}] \leqslant \max(1, \Delta_W^{1/2}) \max(1, \Delta_Y^{1/2}) < \infty$, for any $t$, $\tau \leqslant t \leqslant T + \tau - 1$. Similarly, let $\delta_G \equiv \min(\delta_Y, \delta_W)/2 > 0$. We then have $E[|\hat{G}_{t+1}|^{r+\delta_G}] \leqslant (E[|V_t 1(Y_{t+1} - \hat{f}_{t+1} < 0)|^{2r+2\delta_G}])^{1/2} (E[(Y_{t+1} - \hat{f}_{t+1})^{2(r+\delta_G)(p_0-1)}])^{1/2}$, and, since $E[|V_t 1(Y_{t+1} - \hat{f}_{t+1} < 0)|^{2(r+\delta_G)}] \leqslant E[|V_t|^{2(r+\delta_G)}]$, by the same reasoning as previously, we get $E[|\hat{G}_{t+1}|^{r+\delta_G}] < \infty$, for any $t$, $\tau \leqslant t \leqslant T + \tau - 1$. Hence, $\hat{g}_T \xrightarrow{p} g^*$ and $\hat{h}_T \xrightarrow{p} h^*$ as $T \to \infty$. Next we need to show that the same holds for $|\hat{g}-g^*| \xrightarrow{p} 0$ and $|\hat{h}-h^*| \xrightarrow{p} 0$. We treat the two cases $p_0=1$ and $p_0=2$ separately. When $p_0=1$ we have $\hat{h}=h^*$. By the triangular and Cauchy-Schwartz inequalities, we have $|\hat{g}-g^*|^2 \leqslant T^{-1}\sum_{t=\tau}^{T+\tau-1} E[|V_t|^2]E[(1(Y_{t+1}-\hat{f}_{t+1}<0)-1(Y_{t+1}-f^*_{t+1}<0))^2]$. For every $t$, $\tau \leqslant t < T+\tau$, we have $E\{[1(Y_{t+1}-\hat{f}_{t+1}<0)-1(Y_{t+1}-f^*_{t+1}<0)]^2\} = E\{E_t[1(f^*_{t+1}\leqslant Y_{t+1}<\hat{f}_{t+1})+1(\hat{f}_{t+1}\leqslant Y_{t+1}<f^*_{t+1})]\}$, where $E_t[1(f^*_{t+1}\leqslant Y_{t+1}<\hat{f}_{t+1})+1(\hat{f}_{t+1}\leqslant Y_{t+1}<f^*_{t+1})] = |\int_{\theta^{*\prime}W_t}^{\hat{\theta}_t^{\prime}W_t} f_t(y)dy| \leqslant |\hat{\theta}_t-\theta^*|\cdot|W_t|\cdot\sup_{y\in\mathbb{R}} f_t(y)$. By (A3) and (A6) $|\hat{g}-g^*|^2 \leqslant T^{-1}\sum_{t=\tau}^{T+\tau-1} E[|V_t|^2]E[|\hat{\theta}_t-\theta^*|]\cdot K\cdot C \leqslant K^3\cdot C\cdot T^{-1}\sum_{t=\tau}^{T+\tau-1} E[|\hat{\theta}_t-\theta^*|]$ which shows that when, for every $t$, $\tau\leqslant t\leqslant T+\tau-1$, $\hat{\theta}_t$ is a consistent estimate of $\theta^*$ (A4), $|\hat{g}-g^*| \xrightarrow{p} 0$ as $\tau\to\infty$. Hence, when $p_0=1$, we have shown that $\hat{\alpha}_T \xrightarrow{p} \alpha_0$ as both $\tau\to\infty$ and $T\to\infty$. When $p_0=2$, by the triangular and Cauchy-Schwartz inequalities, $|\hat{h}-h^*| \leqslant T^{-1}\sum_{t=\tau}^{T+\tau-1}|E[V_t|f^*_{t+1}-\hat{f}_{t+1}|]| \leqslant T^{-1}\sum_{t=\tau}^{T+\tau-1}E[|V_t|\cdot|W_t|\cdot|\hat{\theta}_t-\theta^*|]$ $\leqslant K^2\cdot T^{-1}\sum_{t=\tau}^{T+\tau-1}E[|\hat{\theta}_t-\theta^*|]$, so that by the same argument as previously, $|\hat{h}-h^*| \xrightarrow{p} 0$ as $\tau\to\infty$. Moreover $|\hat{g}-g^*| \leqslant K\cdot T^{-1}\sum_{t=\tau}^{T+\tau-1}\{E[1(f^*_{t+1}\leqslant Y_{t+1}<\hat{f}_{t+1})|Y_{t+1}-\hat{f}_{t+1}|]$ $+E[1(\hat{f}_{t+1}\leqslant Y_{t+1}<f^*_{t+1})|Y_{t+1}-f^*_{t+1}|]\}$. By the Cauchy-Schwartz inequality, (A4) and (A5) $E[1(f^*_{t+1}\leqslant Y_{t+1}<\hat{f}_{t+1})|Y_{t+1}-\hat{f}_{t+1}|] \leqslant (E[1(f^*_{t+1}\leqslant Y_{t+1}<\hat{f}_{t+1})])^{1/2}\max(1,\Delta_Y^{1/2})$, for any $t$, $\tau\leqslant t\leqslant T+\tau-1$. As previously, by (A3) and (A6) we have $E[1(f^*_{t+1}\leqslant Y_{t+1}<\hat{f}_{t+1})] \leqslant K\cdot C\cdot E[|\hat{\theta}_t-\theta^*|]$ so that $|\hat{g}-g^*|^2 \leqslant K^3\cdot C\cdot\max(1,\Delta_Y)\cdot T^{-1}\sum_{t=\tau}^{T+\tau-1}E[|\hat{\theta}_t-\theta^*|]$, and so $|\hat{g}-g^*| \xrightarrow{p} 0$ as $\tau\to\infty$ when (A4) holds. Hence, for $p_0=2$ we have $\hat{\alpha}_T \xrightarrow{p} \alpha_0$, as both $\tau\to\infty$ and $T\to\infty$, which completes the proof. ■

**Proof of Proposition 4.** To show that $T^{1/2}(\hat{\alpha}_T-\alpha_0)$ is asymptotically normal, note

$$\sqrt{T}(\hat{\alpha}_T-\alpha_0) = (\hat{h}_T^{\prime}\hat{S}^{-1}\hat{h}_T)^{-1}\hat{h}_T^{\prime}\hat{S}^{-1}\{\sqrt{T}m_T^* + \sqrt{T}\hat{m} + \sqrt{T}(\hat{m}_T-\hat{m}-m_T^*)\}. \quad (11)$$

The idea then is to show that the second and third terms in the curly brackets are $o_p(1)$. We first show that the second term in the curly brackets is $o(1)$. A mean value expansion around $\theta^*$ yields $0=\sqrt{T}m^* = \sqrt{T}\hat{m} - E[T^{-1}\sum_{t=\tau}^{T+\tau-1}(\partial\tilde{M}_{t+1}/\partial\theta)^{\prime}\sqrt{T}(\hat{\theta}_t-\theta^*)]$, where for every $t$, $\tau\leqslant t\leqslant T+\tau-1$, we have $\tilde{\theta}_t \equiv c_t\hat{\theta}_t+(1-c_t)\theta^*$ with $c_t\in(0,1)$ and where $\tilde{M}_{t+1}$ denotes the value of $M_{t+1}$ obtained when $\theta=\tilde{\theta}_t$. We now show that $T^{-1/2}\sum_{t=\tau}^{T+\tau-1}(\partial\tilde{M}_{t+1}/\partial\theta)^{\prime}(\hat{\theta}_t-\theta^*) \xrightarrow{p} 0$ as $\tau\to\infty$ and $T\to\infty$: we have

$$\begin{aligned} |T^{-1/2}\textstyle\sum_{t=\tau}^{T+\tau-1}(\partial\tilde{M}_{t+1}/\partial\theta)^{\prime}(\hat{\theta}_t-\theta^*)| &= |T^{-1/2}\textstyle\sum_{t=\tau}^{T+\tau-1} t^{-1/2+\varepsilon}(\partial\tilde{M}_{t+1}/\partial\theta)^{\prime}t^{1/2-\varepsilon}(\hat{\theta}_t-\theta^*)| \\ &\leqslant \sup_{\tau\leqslant t\leqslant T+\tau-1}|t^{1/2-\varepsilon}(\hat{\theta}_t-\theta^*)|T^{-1/2}\textstyle\sum_{t=\tau}^{T+\tau-1}(|\partial\tilde{M}_{t+1}/\partial\theta|t^{-1/2+\varepsilon}). \end{aligned}$$

Note that (A2), (A3) and (A5') imply $E(\sup_{\theta\in\mathring{\Theta}}|\partial M_{t+1}/\partial\theta|)<\infty$, so that, for any given $\nu>0$, by (A7) and Chebyshev's inequality we have

$$\begin{aligned} P(T^{-1/2}\textstyle\sum_{t=\tau}^{T+\tau-1}|\partial\tilde{M}_{t+1}/\partial\theta|t^{-1/2+\varepsilon} &> \nu)\leqslant E(\sup_{\theta\in\mathring{\Theta}}|\partial M_{t+1}/\partial\theta|)/\nu\cdot T^{-1/2}\textstyle\sum_{t=\tau}^{T+\tau-1}t^{-1/2+\varepsilon} \\ &\leqslant E(\sup_{\theta\in\mathring{\Theta}}|\partial M_{t+1}/\partial\theta|)/\nu\cdot(T/\tau^{1-2\varepsilon})^{1/2}\to 0 \end{aligned}$$

as $\tau\to\infty$ and $T\to\infty$. Hence $\sqrt{T}\hat{m}\to 0$ as $\tau\to\infty$ and $T\to\infty$. The third term in the curly brackets in (11) is $o_p(1)$ provided that $m_t$ satisfies a certain Lipshitz condition (given below) and $\hat{\theta}_t\xrightarrow{p}\theta^*$ for all $t\geqslant\tau$, as $\tau\to\infty$. This follows because for any given $\eta>0$ and $\varepsilon>0$, there exists $\delta_\tau>0$ such that

$$\begin{aligned} &\lim_{\tau,T\to\infty}P(\sqrt{T}|\hat{m}_T-\hat{m}-m_T^*|>\eta) \\ &\leqslant \lim_{\tau,T\to\infty}P(\sqrt{T}|\hat{m}_T-\hat{m}-m_T^*|>\eta,\sup_{\tau\leqslant t\leqslant T+\tau-1}|\hat{\theta}_t-\theta^*|\leqslant\delta_\tau) \\ &\quad+\lim_{\tau,T\to\infty}P(\sup_{\tau\leqslant t\leqslant T+\tau-1}|\hat{\theta}_t-\theta^*|>\delta_\tau) \\ &\leqslant \lim_{\tau,T\to\infty}P(\sqrt{T}|\hat{m}_T-\hat{m}-m_T^*|>\eta,\sup_{\tau\leqslant t\leqslant T+\tau-1}|\hat{\theta}_t-\theta^*|\leqslant\delta_\tau), \end{aligned}$$

where the last inequality uses (A4). Now, let $r_T(\delta_\tau)\equiv\sup_{|\hat{\theta}_t-\theta^*|\leqslant\delta_\tau,\tau\leqslant t\leqslant T+\tau-1}r_{t+1}(\hat{\theta}_t)$,where for all $\theta\in\mathring{\Theta}$ we let

$$r_{t+1}(\theta)\equiv|m_{t+1}-m_{t+1}^*-\Delta_{t+1}^*\cdot(\theta-\theta^*)|/|\theta-\theta^*|, \tag{12}$$

where $\Delta_{t+1}^*$ is as defined as

$$\begin{aligned} \Delta_{t+1}^*\equiv\ & v_tw_t'\cdot\delta(\theta^{*\prime}w_t-y_{t+1})\cdot|y_{t+1}-\theta^{*\prime}w_t|^{p_0-1} \\ &+(p_0-1)\{v_tw_t'[\alpha_0+(1-2\alpha_0)1(y_{t+1}-\theta^{*\prime}w_t<0)]|y_{t+1}-\theta^{*\prime}w_t|^{p_0-2}\}. \end{aligned}$$

Then, by the definition of $r_{t+1}(\theta)$

$$\begin{aligned} \sqrt{T}|\hat{m}_T-\hat{m}-m_T^*| &\leqslant \sqrt{T}\{|\frac{1}{T}\textstyle\sum_{t=\tau}^{T+\tau-1}\Delta_{t+1}^*(\hat{\theta}_t-\theta^*)-E[\Delta_{t+1}^*(\hat{\theta}_t-\theta^*)]| \\ &\quad+\frac{1}{T}\textstyle\sum_{t=\tau}^{T+\tau-1}r_{t+1}(\hat{\theta}_t)|\hat{\theta}_t-\theta^*|+E[r_{t+1}(\hat{\theta}_t)|\hat{\theta}_t-\theta^*|]\} \\ &\leqslant \sqrt{T}\{\frac{1}{T}\textstyle\sum_{t=\tau}^{T+\tau-1}|\Delta_{t+1}^*-E[\Delta_{t+1}^*]|\sup_{\tau\leqslant t\leqslant T+\tau-1}|\hat{\theta}_t-\theta^*| \\ &\quad+[r_T(\delta_\tau)+E(r_T(\delta_\tau))]\sup_{\tau\leqslant t\leqslant T+\tau-1}|\hat{\theta}_t-\theta^*|\}. \end{aligned}$$

Using standard arguments for stochastic equicontinuity such as those given in Andrews (1994), we can show that for any $\theta\in\mathring{\Theta}$, $r_{t+1}(\theta)\to 0$ as $\theta\to\theta^*$, so that $r_T(\delta_\tau)\to 0$ with probability one, which by the dominated convergence theorem ensures $E(r_T(\delta_\tau)) \to 0$ as $\delta_\tau \to 0$. Next, we show that locally at any $\theta^* \in \mathring{\Theta}$, the sample mean of $\{\Delta^*_{t+1}\}$ converges in probability to its expected value. By (A5') we know that for every $\theta^* \in \mathring{\Theta}$, $\{\Delta^*_{t+1}\}$ is strictly stationary and $\alpha$-mixing with $\alpha$ of size $-r/(r-2)$ with $r > 2$ (see Theorems 3.35 and 3.49 in White, 2001). (A2), (A3) and (A5') moreover imply $E(\sup_{\theta^* \in \mathring{\Theta}} |\Delta^*_{t+1}|^{r+\epsilon}) < \infty$, for all $t$, $\tau \leqslant t \leqslant T+\tau-1$. Using the weak LLN for $\alpha$-mixing sequences (e.g., Corollary 3.48 in White, 2001) then gives $T^{-1}\sum_{t=\tau}^{T+\tau-1} \Delta^*_{t+1} \xrightarrow{p} \Delta^* \equiv E[\Delta^*_{t+1}]$ as $T \to \infty$, locally at $\theta^*$, for all $\theta^* \in \mathring{\Theta}$. Then, by using the Markov inequality $\lim_{T\to\infty} P(\sqrt{T}|\hat{m}_T - \hat{m} - m^*_T| > \eta, \sup_{\tau \leqslant t \leqslant T+\tau-1} |\hat{\theta}_t - \theta^*| \leqslant \delta_\tau) = 0$ and the third term in (11) is $o_p(1)$ as $\tau \to \infty$ and $T \to \infty$. Next we use the central limit theorem (CLT) for strictly stationary and $\alpha$-mixing sequences (e.g., Theorem 5.20 in White, 2001) to show that $\sqrt{T} m^*_T \xrightarrow{d} \mathcal{N}(0, S)$. Using Theorems 3.35 and 3.49 in White (2001), which together show that time-invariant measurable functions of strictly stationary and mixing sequences are strictly stationary and mixing of the same size, we know by (A5') that $\{M^*_{t+1}\}$ is strictly stationary and $\alpha$-mixing with mixing coefficient of size $-r/(r-2)$, $r > 2$. The Cauchy-Schwartz inequality and (A5') imply $E[|M^*_{t+1}|^r] \leqslant (E[|V_t|^{2r}])^{1/2} \cdot (E[(Y_{t+1} - f^*_{t+1})^{2r(p_0-1)}])^{1/2} \leqslant (E[|V_t|^{2r}])^{1/2} \max(1, \{\sup_{\theta\in\Theta} E[(Y_{t+1} - \theta' W_t)^{2r(p_0-1)}]\}^{1/2}) \leqslant \Delta$, for $\Delta \equiv \max(1, \Delta_W^{1/2}) \max(1, \Delta_Y^{1/2}) > 0$, $\Delta < \infty$. The CLT (e.g., Theorem 5.20 in White, 2001) then ensures

$$\sqrt{T} m^*_T = \sqrt{T}(g^*_T - h^*_T \cdot \alpha_0) \xrightarrow{d} \mathcal{N}(0, S). \tag{13}$$

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 $\hat{h}_T \xrightarrow{p} h^*$ as $\tau \to \infty$ and $T \to \infty$, together with (A1) and (A3), ensure that $h^{*\prime} S^{-1} h^* \neq 0$ and $\hat{h}'_T \hat{S}^{-1} \hat{h}_T \neq 0$ with probability one, so by using $\sqrt{T}(\hat{\alpha}_T - \alpha_0) = (\hat{h}'_T \hat{S}^{-1} \hat{h}_T)^{-1} \hat{h}'_T \hat{S}^{-1} \{\sqrt{T} m^*_T + o_p(1)\}$, the limit result in (13) and the Slutsky theorem we have $\sqrt{T}(\hat{\alpha}_T - \alpha_0) \xrightarrow{d} \mathcal{N}(0, (h^{*\prime} S^{-1} h^*)^{-1})$, which completes the proof. ■

# References


[1] ANDREWS, D.W.K. (1994), “Empirical Process Methods in Econometrics”, in R. F. Engle and D. L. McFadden (eds.), Handbook of Econometrics, 4, 2247-2294, Elsevier Science.

[2] ARTIS, M. and MARCELLINO, M. (2001), “Fiscal Forecasting: The Track Record of the IMF, OECD and EC”, Econometrics Journal, 4, S20-36.

[3] CAMPBELL, B. and GHYSELS, E. (1995), “Federal Budget Projections: A Nonparametric Assessment of Bias and Efficiency”, Review of Economics and Statistics, 77, 17-31.

[4] CHRISTOFFERSEN, P.F. and DIEBOLD, F.X. (1996), “Further Results on Forecasting and Model Selection Under Asymmetric Loss”, Journal of Applied Econometrics, 11, 561-572.

[5] CHRISTOFFERSEN, P.F. and DIEBOLD, F.X. (1997), “Optimal Prediction under Asymmetric Loss, Econometric Theory”, 13, 808-817.

[6] CORRADI, V. and SWANSON, N.R. (2002), “A Consistent Test for Nonlinear out of Sample Predictive Accuracy”, Journal of Econometrics, 110, 353-381.

[7] DIEBOLD, F.X. and LOPEZ, J.A. (1996), “Forecast Evaluation and Combination”, in G.S. Maddala and C.R. Rao, eds., Handbook of Statistics, 14, 241-268, Amsterdam: North-Holland.

[8] DIEBOLD, F.X., GUNTHER, T. and TAY, A. (1998), “Evaluating Density Forecasts, with Applications to Financial Risk Management”, International Economic Review, 39, 863-883.

[9] DONALD, S.G. and NEWEY, W.K. (2001), “Choosing the Number of Instruments”, Econometrica, 69, 1161-1192.

[10] ELLIOTT, G., KOMUNJER, I. and TIMMERMANN, A. (2004), "Biases in Macroeconomic Forecasts: Irrationality or Asymmetric Loss?", unpublished manuscript, UCSD and Caltech.

[11] GRANGER, C.W.J. and NEWBOLD, P. (1986), Forecasting Economic Time Series, Second Edition, Academic Press.

[12] GRANGER, C.W.J. and PESARAN, M.H. (2000), "A Decision Theoretic Approach to Forecast Evaluation", in W.S. Chan, W.K. Li and H. Tong (eds), Statistics and Finance: An Interface, 261-278, Imperial College Press, London.

[13] HANSEN, L.P. and SINGLETON, K.J. (1982), "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models", Econometrica, 50, 1269-1286.

[14] MCCRACKEN, M.W. (2000), "Robust Out of Sample Inference", Journal of Econometrics, 99, 195-223.

[15] NEWEY, W.K. and POWELL, J. (1987), "Asymmetric Least Squares Estimation and Testing", Econometrica, 55, 819-847.

[16] NEWEY, W.K. and WEST, K.D. (1987), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", Econometrica, 55, 703-708.

[17] SCHWARTZ, L. (1997), Analyse, Hermann: Paris.

[18] WEST, K.D. (1996), "Asymptotic Inference about Predictive Ability", Econometrica, 64, 1067-84

[19] WEST, K.D. and MCCRACKEN, M.W. (1998), "Regression-Based Tests of Predictive Ability", International Economic Review, 39, 817-840.

[20] WEST, K.D., EDISON, H.J. and CHO, D. (1993), "A Utility-based Comparison of Some Models of Exchange Rate Volatility", Journal of International Economics, 35, 23-46.

[21] WHITE, H. (2001), Asymptotic Theory for Econometricians, Second Edition, Academic Press, San Diego: California.

[22] ZELLNER, A. (1986), "Bayesian Estimation and Prediction Using Asymmetric Loss Functions", Journal of the American Statistical Association, 81, 446-451.

Table 1: Size of two-sided t-tests and j-tests (nominal size 5%)

A. Lin-Lin ($p_0 = 1$)

t-test, only a constant as instrument

| $N_0$ | $n_f$ | 0.2 | 0.4 | 0.5 | 0.6 | 0.8 |
|---|---|---|---|---|---|---|
| 50 | 50 | 0.053 | 0.059 | 0.061 | 0.064 | 0.057 |
| 50 | 100 | 0.066 | 0.050 | 0.057 | 0.052 | 0.060 |
| 100 | 50 | 0.056 | 0.056 | 0.066 | 0.066 | 0.049 |
| 100 | 100 | 0.063 | 0.055 | 0.057 | 0.052 | 0.058 |
| 100 | 200 | 0.061 | 0.053 | 0.053 | 0.055 | 0.054 |

t-test, two instruments

| $n_0$ | $n_f$ | 0.2 | 0.4 | 0.5 | 0.6 | 0.8 |
|---|---|---|---|---|---|---|
| 50 | 50 | 0.144 | 0.094 | 0.078 | 0.090 | 0.098 |
| 50 | 100 | 0.090 | 0.073 | 0.064 | 0.072 | 0.069 |
| 100 | 50 | 0.162 | 0.095 | 0.083 | 0.091 | 0.106 |
| 100 | 100 | 0.096 | 0.076 | 0.063 | 0.068 | 0.072 |
| 100 | 200 | 0.077 | 0.065 | 0.055 | 0.063 | 0.065 |

j-test, two instruments

| $n_0$ | $n_f$ | 0.2 | 0.4 | 0.5 | 0.6 | 0.8 |
|---|---|---|---|---|---|---|
| 50 | 50 | 0.029 | 0.047 | 0.049 | 0.048 | 0.036 |
| 50 | 100 | 0.044 | 0.048 | 0.047 | 0.047 | 0.044 |
| 100 | 50 | 0.033 | 0.047 | 0.046 | 0.049 | 0.033 |
| 100 | 100 | 0.041 | 0.052 | 0.049 | 0.047 | 0.041 |
| 100 | 200 | 0.049 | 0.047 | 0.048 | 0.052 | 0.047 |

B. Quad-Quad ($p_0 = 2$)

t-test, only a constant as instrument

| $n_0$ | $n_f$ | 0.2 | 0.4 | 0.5 | 0.6 | 0.8 |
|---|---|---|---|---|---|---|
| 50 | 50 | 0.065 | 0.069 | 0.071 | 0.071 | 0.063 |
| 50 | 100 | 0.082 | 0.062 | 0.063 | 0.061 | 0.074 |
| 100 | 50 | 0.063 | 0.068 | 0.072 | 0.072 | 0.065 |
| 100 | 100 | 0.077 | 0.057 | 0.059 | 0.057 | 0.068 |
| 100 | 200 | 0.127 | 0.055 | 0.057 | 0.052 | 0.121 |

t-test, two instruments

| $n_0$ | $n_f$ | 0.2 | 0.4 | 0.5 | 0.6 | 0.8 |
|---|---|---|---|---|---|---|
| 50 | 50 | 0.105 | 0.121 | 0.118 | 0.120 | 0.102 |
| 50 | 100 | 0.076 | 0.083 | 0.087 | 0.085 | 0.077 |
| 100 | 50 | 0.109 | 0.120 | 0.116 | 0.121 | 0.113 |
| 100 | 100 | 0.080 | 0.077 | 0.080 | 0.080 | 0.073 |
| 100 | 200 | 0.104 | 0.066 | 0.069 | 0.066 | 0.102 |

j-test, two instruments

| $n_0$ | $n_f$ | 0.2 | 0.4 | 0.5 | 0.6 | 0.8 |
|---|---|---|---|---|---|---|
| 50 | 50 | 0.020 | 0.038 | 0.043 | 0.040 | 0.023 |
| 50 | 100 | 0.032 | 0.044 | 0.045 | 0.042 | 0.026 |
| 100 | 50 | 0.026 | 0.042 | 0.041 | 0.040 | 0.022 |
| 100 | 100 | 0.033 | 0.049 | 0.050 | 0.046 | 0.030 |
| 100 | 200 | 0.030 | 0.046 | 0.051 | 0.050 | 0.035 |

Note: $n_0$ is the initial sample used to estimate the parameters of the forecasting model while $n_f$ is the size of the out-of-sample forecasting period used to test the model. 0.2, 0.4, 0.5, 0.6 and 0.8 are the values of $\alpha_{o,}$, the population asymmetry parameter.

Table 2: Parameter Estimates Under Lin-lin Loss and Tests of Symmetry

| | | IMF | | | | | | | OECD | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Canada | France | Germany | Italy | Japan | UK | US | France | Germany | Italy | UK |
| | | | | | | Current year | | | | | | |
| Inst=1 | α | 0.60 | 0.52 | 0.40 | 0.16 | 0.24 | 0.24 | 0.20 | 0.27 | 0.22 | 0.29 | 0.48 |
| | s.e. | 0.10 | 0.10 | 0.10 | 0.07 | 0.09 | 0.09 | 0.08 | 0.09 | 0.08 | 0.09 | 0.10 |
| | p-value | 0.31 | 0.84 | 0.31 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.02 | 0.85 |
| Inst=2 | α | 0.58 | 0.54 | 0.40 | 0.14 | 0.19 | 0.20 | 0.18 | 0.28 | 0.22 | 0.28 | 0.50 |
| | s.e. | 0.10 | 0.10 | 0.10 | 0.07 | 0.08 | 0.08 | 0.08 | 0.09 | 0.08 | 0.09 | 0.10 |
| | p-value | 0.39 | 0.67 | 0.33 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.02 | 0.98 |
| Inst=3 | α | 0.59 | 0.54 | 0.42 | 0.15 | 0.24 | 0.24 | 0.19 | 0.13 | 0.12 | 0.29 | 0.50 |
| | s.e. | 0.10 | 0.10 | 0.10 | 0.07 | 0.09 | 0.09 | 0.08 | 0.07 | 0.06 | 0.09 | 0.10 |
| | p-value | 0.39 | 0.68 | 0.40 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.99 |
| Inst=4 | α | 0.59 | 0.54 | 0.40 | 0.13 | 0.20 | 0.19 | 0.17 | 0.11 | 0.11 | 0.26 | 0.49 |
| | s.e. | 0.10 | 0.10 | 0.10 | 0.07 | 0.08 | 0.08 | 0.08 | 0.06 | 0.06 | 0.09 | 0.10 |
| | p-value | 0.39 | 0.67 | 0.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.92 |
| | | | | | | 1-year ahead | | | | | | |
| Inst=1 | α | 0.54 | 0.48 | 0.44 | 0.30 | 0.40 | 0.24 | 0.24 | 0.44 | 0.32 | 0.46 | 0.52 |
| | s.e. | 0.10 | 0.10 | 0.10 | 0.10 | 0.10 | 0.09 | 0.09 | 0.10 | 0.09 | 0.10 | 0.10 |
| | p-value | 0.68 | 0.84 | 0.55 | 0.04 | 0.31 | 0.00 | 0.00 | 0.55 | 0.05 | 0.68 | 0.84 |
| Inst=2 | α | 0.54 | 0.50 | 0.45 | 0.27 | 0.18 | 0.14 | 0.14 | 0.41 | 0.33 | 0.43 | 0.55 |
| | s.e. | 0.10 | 0.10 | 0.10 | 0.09 | 0.08 | 0.07 | 0.07 | 0.10 | 0.09 | 0.10 | 0.10 |
| | p-value | 0.67 | 0.96 | 0.62 | 0.01 | 0.00 | 0.00 | 0.00 | 0.38 | 0.08 | 0.52 | 0.59 |
| Inst=3 | α | 0.57 | 0.50 | 0.46 | 0.27 | 0.37 | 0.24 | 0.24 | 0.39 | 0.27 | 0.43 | 0.54 |
| | s.e. | 0.10 | 0.10 | 0.10 | 0.09 | 0.10 | 0.09 | 0.09 | 0.10 | 0.09 | 0.10 | 0.10 |
| | p-value | 0.50 | 1.00 | 0.66 | 0.01 | 0.18 | 0.00 | 0.00 | 0.24 | 0.01 | 0.52 | 0.66 |
| Inst=4 | α | 0.57 | 0.50 | 0.44 | 0.27 | 0.26 | 0.17 | 0.13 | 0.35 | 0.24 | 0.43 | 0.57 |
| | s.e. | 0.10 | 0.10 | 0.10 | 0.09 | 0.09 | 0.08 | 0.07 | 0.10 | 0.09 | 0.10 | 0.10 |
| | p-value | 0.48 | 0.98 | 0.57 | 0.01 | 0.01 | 0.00 | 0.00 | 0.13 | 0.00 | 0.51 | 0.45 |

Note: The four instrument sets labeled from inst = 1 to inst = 4 are the following: (i) a constant; (ii) a constant and the lagged forecast error; (iii) a constant and the lagged budget deficit; (iv) a constant, the lagged forecast error and the lagged budget deficit.

Table 3: Tests of the Joint Hypothesis of Lin-Lin Loss and Forecast Rationality

| | | IMF | | | | | | | OECD | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Canada | France | Germany | Italy | Japan | UK | US | France | Germany | Italy | UK |
| A. Symmetric (MAE) loss | | | | | | Current year | | | | | | |
| Inst=1 | j-stat | 1.04 | 0.04 | 1.04 | 21.50 | 9.27 | 9.27 | 14.06 | 7.04 | 12.05 | 5.04 | 0.04 |
| | p-value | 0.31 | 0.84 | 0.31 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.02 | 0.85 |
| Inst=2 | j-stat | 0.82 | 0.41 | 2.62 | 26.22 | 16.61 | 14.41 | 18.22 | 6.02 | 12.30 | 6.71 | 1.52 |
| | p-value | 0.67 | 0.81 | 0.27 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 | 0.00 | 0.03 | 0.47 |
| Inst=3 | j-stat | 0.91 | 0.17 | 0.75 | 23.93 | 8.67 | 9.28 | 14.77 | 34.37 | 40.39 | 6.03 | 0.42 |
| | p-value | 0.63 | 0.92 | 0.69 | 0.00 | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 | 0.05 | 0.81 |
| Inst=4 | j-stat | 0.93 | 0.41 | 3.49 | 30.16 | 14.81 | 16.76 | 20.14 | 43.90 | 43.85 | 8.95 | 2.51 |
| | p-value | 0.82 | 0.94 | 0.32 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.03 | 0.47 |
| | | | | | | 1-year ahead | | | | | | |
| Inst=1 | j-stat | 0.17 | 0.04 | 0.37 | 4.16 | 1.04 | 9.27 | 9.27 | 0.37 | 3.72 | 0.17 | 0.04 |
| | p-value | 0.68 | 0.84 | 0.55 | 0.04 | 0.31 | 0.00 | 0.00 | 0.55 | 0.05 | 0.68 | 0.84 |
| Inst=2 | j-stat | 3.13 | 2.58 | 2.16 | 6.35 | 23.02 | 28.06 | 28.42 | 1.31 | 3.09 | 0.44 | 4.29 |
| | p-value | 0.21 | 0.28 | 0.34 | 0.04 | 0.00 | 0.00 | 0.00 | 0.52 | 0.21 | 0.80 | 0.12 |
| Inst=3 | j-stat | 4.06 | 0.10 | 0.57 | 5.91 | 2.06 | 9.01 | 9.78 | 3.64 | 9.38 | 0.50 | 0.43 |
| | p-value | 0.13 | 0.95 | 0.75 | 0.05 | 0.36 | 0.01 | 0.01 | 0.16 | 0.01 | 0.78 | 0.81 |
| Inst=4 | j-stat | 4.24 | 3.06 | 3.31 | 6.49 | 14.68 | 21.97 | 33.47 | 7.02 | 13.12 | 0.60 | 5.25 |
| | p-value | 0.24 | 0.38 | 0.35 | 0.09 | 0.00 | 0.00 | 0.00 | 0.07 | 0.00 | 0.90 | 0.15 |
| B. Allowing for Asymmetric loss | | | | | | Current year | | | | | | |
| Inst=2 | j-stat | 0.12 | 0.24 | 1.72 | 0.79 | 2.17 | 1.66 | 1.18 | 0.01 | 0.55 | 1.59 | 1.52 |
| | p-value | 0.73 | 0.63 | 0.19 | 0.37 | 0.14 | 0.20 | 0.28 | 0.94 | 0.46 | 0.21 | 0.22 |
| Inst=3 | j-stat | 0.20 | 0.00 | 0.06 | 0.62 | 0.23 | 0.42 | 0.52 | 5.65 | 3.81 | 0.89 | 0.42 |
| | p-value | 0.66 | 0.99 | 0.81 | 0.43 | 0.63 | 0.52 | 0.47 | 0.02 | 0.05 | 0.35 | 0.52 |
| Inst=4 | j-stat | 0.21 | 0.24 | 2.46 | 1.27 | 1.97 | 2.20 | 1.49 | 6.56 | 4.12 | 2.13 | 2.50 |
| | p-value | 0.90 | 0.89 | 0.29 | 0.53 | 0.37 | 0.33 | 0.48 | 0.04 | 0.13 | 0.34 | 0.29 |
| | | | | | | 1-year ahead | | | | | | |
| Inst=2 | j-stat | 2.96 | 2.57 | 1.93 | 0.25 | 6.65 | 3.18 | 4.42 | 0.56 | 0.06 | 0.04 | 4.02 |
| | p-value | 0.09 | 0.11 | 0.16 | 0.61 | 0.01 | 0.07 | 0.04 | 0.45 | 0.81 | 0.83 | 0.05 |
| Inst=3 | j-stat | 3.62 | 0.10 | 0.38 | 0.08 | 0.34 | 0.49 | 0.58 | 2.23 | 2.90 | 0.09 | 0.25 |
| | p-value | 0.06 | 0.75 | 0.54 | 0.78 | 0.56 | 0.48 | 0.45 | 0.13 | 0.09 | 0.76 | 0.62 |
| Inst=4 | j-stat | 3.76 | 3.06 | 3.01 | 0.33 | 7.88 | 4.56 | 5.35 | 4.77 | 4.28 | 0.19 | 4.71 |
| | p-value | 0.15 | 0.22 | 0.22 | 0.85 | 0.02 | 0.10 | 0.07 | 0.09 | 0.12 | 0.91 | 0.09 |

Note: The four instrument sets labeled from inst = 1 to inst = 4 are the following: (i) a constant; (ii) a constant and the lagged forecast error; (iii) a constant and the lagged budget deficit; (iv) a constant, the lagged forecast error and the lagged budget deficit.

Table 4: Tests of the Joint Hypothesis of Quad-Quad Loss and Forecast Rationality

| | | IMF | | | | | | | OECD | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Canada | France | Germany | Italy | Japan | UK | US | France | Germany | Italy | UK |
| A. Symmetric (MSE) loss | | | | | | Current year | | | | | | |
| Inst=1 | j-stat | 2.49 | 0.95 | 0.27 | 3.66 | 2.51 | 18.81 | 4.56 | 1.64 | 9.65 | 4.20 | 0.41 |
| | p-value | 0.11 | 0.33 | 0.61 | 0.06 | 0.11 | 0.00 | 0.03 | 0.20 | 0.00 | 0.04 | 0.52 |
| Inst=2 | j-stat | 2.37 | 1.10 | 1.12 | 6.53 | 77.19 | 44.75 | 10.14 | 1.43 | 11.44 | 12.09 | 4.34 |
| | p-value | 0.31 | 0.58 | 0.57 | 0.04 | 0.00 | 0.00 | 0.01 | 0.49 | 0.00 | 0.00 | 0.11 |
| Inst=3 | j-stat | 2.73 | 2.41 | 0.25 | 24.97 | 2.63 | 14.46 | 12.04 | 9.00 | 40.33 | 6.59 | 0.04 |
| | p-value | 0.25 | 0.30 | 0.88 | 0.00 | 0.27 | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 | 0.98 |
| Inst=4 | j-stat | 3.20 | 2.41 | 1.23 | 25.72 | 18.69 | 31.81 | 24.66 | 41.67 | 42.09 | 21.39 | 4.39 |
| | p-value | 0.36 | 0.49 | 0.75 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.23 |
| | | | | | | 1-year ahead | | | | | | |
| Inst=1 | j-stat | 2.52 | 1.72 | 0.44 | 6.86 | 0.65 | 1.74 | 3.44 | 0.14 | 6.65 | 0.42 | 0.01 |
| | p-value | 0.11 | 0.19 | 0.51 | 0.01 | 0.42 | 0.19 | 0.06 | 0.71 | 0.01 | 0.52 | 0.94 |
| Inst=2 | j-stat | 6.56 | 7.38 | 1.01 | 157.83 | 116.64 | 12.34 | 12.98 | 2.26 | 5.54 | 0.98 | 3.76 |
| | p-value | 0.04 | 0.02 | 0.60 | 0.00 | 0.00 | 0.00 | 0.00 | 0.32 | 0.06 | 0.61 | 0.15 |
| Inst=3 | j-stat | 11.17 | 4.26 | 0.10 | 143.02 | 0.93 | 1.16 | 3.35 | 0.45 | 16.76 | 0.96 | 0.08 |
| | p-value | 0.00 | 0.12 | 0.95 | 0.00 | 0.63 | 0.56 | 0.19 | 0.80 | 0.00 | 0.62 | 0.96 |
| Inst=4 | j-stat | 11.42 | 8.23 | 1.21 | 154.90 | 14.26 | 12.44 | 15.80 | 7.62 | 21.48 | 1.16 | 5.26 |
| | p-value | 0.01 | 0.04 | 0.75 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 | 0.00 | 0.76 | 0.15 |
| B. Allowing for Asymmetric loss | | | | | | Current year | | | | | | |
| Inst=2 | j-stat | 0.11 | 0.00 | 1.03 | 0.58 | 2.90 | 2.66 | 1.67 | 0.19 | 0.95 | 1.55 | 4.11 |
| | p-value | 0.74 | 0.98 | 0.31 | 0.45 | 0.09 | 0.10 | 0.20 | 0.66 | 0.33 | 0.21 | 0.04 |
| Inst=3 | j-stat | 0.05 | 1.89 | 0.18 | 2.96 | 0.03 | 0.07 | 1.91 | 3.24 | 2.93 | 0.87 | 0.02 |
| | p-value | 0.83 | 0.17 | 0.67 | 0.09 | 0.85 | 0.79 | 0.17 | 0.07 | 0.09 | 0.35 | 0.89 |
| Inst=4 | j-stat | 0.32 | 1.91 | 1.13 | 2.98 | 4.07 | 3.05 | 3.30 | 6.85 | 3.14 | 1.53 | 4.19 |
| | p-value | 0.85 | 0.39 | 0.57 | 0.23 | 0.13 | 0.22 | 0.19 | 0.03 | 0.21 | 0.47 | 0.12 |
| | | | | | | 1-year ahead | | | | | | |
| Inst=2 | j-stat | 2.58 | 3.09 | 0.82 | 1.95 | 4.34 | 3.79 | 3.66 | 2.18 | 0.22 | 0.02 | 3.75 |
| | p-value | 0.11 | 0.08 | 0.36 | 0.16 | 0.04 | 0.05 | 0.06 | 0.14 | 0.64 | 0.88 | 0.05 |
| Inst=3 | j-stat | 3.52 | 1.24 | 0.02 | 1.91 | 0.39 | 0.08 | 0.17 | 0.33 | 2.33 | 0.00 | 0.03 |
| | p-value | 0.06 | 0.27 | 0.90 | 0.17 | 0.53 | 0.78 | 0.68 | 0.57 | 0.13 | 0.96 | 0.86 |
| Inst=4 | j-stat | 3.51 | 4.83 | 0.98 | 2.49 | 7.62 | 5.48 | 4.39 | 7.48 | 4.02 | 0.04 | 5.25 |
| | p-value | 0.17 | 0.09 | 0.61 | 0.29 | 0.02 | 0.06 | 0.11 | 0.02 | 0.13 | 0.98 | 0.07 |

Note: The four instrument sets labeled from inst = 1 to inst = 4 are the following: (i) a constant; (ii) a constant and the lagged forecast error; (iii) a constant and the lagged budget deficit; (iv) a constant, the lagged forecast error and the lagged budget deficit.