Estimating Loss Function Parameters*

Graham Elliott                Ivana Komunjer
University of California San Diego    Caltech

Allan Timmermann
University of California San Diego

February 4, 2003

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. 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 seminar participants at UCLA, Atlanta Fed 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. 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 (Frankel and Froot (1987), Hansen and Hodrick (1980)), inflation forecasting (Bonham and Cohen (1995), Figlewski and Wachtel (1981), Keane and Runkle (1990), Mishkin (1981), Pesando (1975), Schroeter and Smith (1986)) and tests of the Fisher equation (Gultekin (1983)).

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.\footnote{In addition to the studies cited in the first paragraph, Hafer and Hein (1985) and Zarnovitz (1979) use mean squared error loss and mean absolute error loss as a metric for measuring forecast accuracy. This is just a small subset of papers studying forecast rationality under symmetric loss. For additional references, see www.Phil.frb.org/econ/spf/spfbin.html.} 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 the optimality of a particular forecast series an easy exercise. The analysis can be based directly on the observable forecast errors and does 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), 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 (loss of potential sales, customers and reputation) should be exactly the same as the costs from overpredicting it (added costs of
production and storage). As a second example, central banks are likely to have asymmetric preferences, as pointed out by Peel and Nobay (1998), perhaps tending to err in the direction of caution in reaching inflation targets.\(^2\)

Consequently, in economics and finance forecasting performance is increasingly evaluated under more general loss functions that account for asymmetries as witnessed by studies such as Batchelor and Peel (1998), Christoffersen and Diebold (1997), Granger and Newbold (1986), Granger and Pesaran (2000), Varian (1974), West, Edison and Choi (1996) and Zellner (1986). 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) and Patton and Timmermann (2002). 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.

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 the case under MSE loss where the properties of the observed forecast errors 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.

---

\(^2\)Peel and Nobay (1998) cite the following quote from the Sunday Times, 9 November 1997: “There is a bias towards over caution in policy built into the new arrangements, at least for a while. If George [Governor of the Bank of England] has to write to Brown [Chancellor of Exchequer] in two years time and apologize for the fact that inflation is 1%, and therefore outside his effective target range [2-5%], he would do so a happy man. If he had to do so with inflation at 5%, he would probably slip his resignation letter into the same envelope.”
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 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. These parameters are 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 decision maker has MSE loss and is irrational or that he has an asymmetric loss function and rationally overpredicts because he dislikes positive forecast errors more than he dislikes 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. 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.

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 and nest both MAD and MSE loss as special cases. 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 applied work in economics and finance to test for ‘rationality’ of expectations 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 (MSE). This choice is useful in practice for a number of reasons - it provides simple optimal properties of the forecasts and relates directly to least squares regression on forecast errors. Tests of rationality based on MSE loss therefore fit directly within the standard econometric toolbox.

Suppose, however, that we are not sure that the loss function is of the MSE type. What inference can we then draw from empirical inspection of a sequence of point forecasts? In this section we review the optimal properties of forecasts for more general loss functions than MSE. We then ‘turn the problem around’ and motivate the idea of estimating loss functions from observed forecasts.

Consider a stochastic process $X \equiv \{X_t : \Omega \rightarrow \mathbb{R}^{m+1}, m \in \mathbb{N}, t = 1, \ldots, n+1\}$ defined on a complete probability space $(\Omega, \mathcal{F}, \mathcal{P})$ where $\mathcal{F} = \{\mathcal{F}_t, t = 1, \ldots, n + 1\}$ and $\mathcal{F}_t$ is the $\sigma$-field $\mathcal{F}_t \equiv \sigma\{X_s, s \leq t\}$. In what follows, we 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. The random variable $Y_t$ is further assumed to be continuous. In standard notation the subscript $t$ on the distribution function $F(\cdot)$ of $Y_{t+1}$, its density $f(\cdot)$, and the expectation $E[\cdot]$ denotes conditioning on the information set $\mathcal{F}_t$.\textsuperscript{3} The forecasting problem considered here involves forecasting the variable $Y_{t+s}$, where $s$ is the prediction horizon of interest, $s \geq 1$. In what follows, we set $s = 1$ and examine the one-step-ahead predictions of the realization $y_{t+1}$, knowing that all results developed in this case can readily be generalized to any $s \geq 1$.

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 $\theta$ is an unknown $k$-vector of parameters, $\theta \in \Theta$, and $\Theta$ is 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., the IMF or the OECD) and they need not be known by the forecast user. As a general rule, $W_t$ should include variables that are observed by the forecaster at time $t$ and which are thought to help forecast $Y_{t+1}$ (e.g., a subset of the $m$-vector of exogenous variables in $X_t$, 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 - that is likely to be relevant in practice - we do not assume that $M = \{f_{t+1}\}$ is correctly specified, i.e. we allow for certain types of 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 $L$ defined by

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

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

\textsuperscript{3}As a general rule, we hereafter use upper case letters for random variables, i.e. $Y_t$ and $X_t$, and lower case letters for their realizations, i.e. $y_t$ and $x_t$.\null
to the forecast error $\varepsilon_{t+1}$.\footnote{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 time $t+1$ and use the notation $\mathcal{L}(p, \alpha, \theta)$ instead of $\mathcal{L}_{t+1}(p, \alpha, \theta).$} We let $\alpha_0$ and $p_0$ be the unknown true values of $\alpha$ 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 $f_{t+1}$, but also of the shape parameters $\alpha$ and $p$ of $\mathcal{L}$. Special cases of $\mathcal{L}$ include: (i) squared loss function $\mathcal{L}(2, 1/2, \theta) = (Y_{t+1} - f_{t+1})^2$, (ii) absolute deviation loss function $\mathcal{L}(1, 1/2, \theta) = |Y_{t+1} - f_{t+1}|$, as well as their asymmetrical counterparts obtained when $\alpha \neq 1/2$, i.e. (iii) quad-quad loss, $\mathcal{L}(2, \alpha, \theta)$, and (iv) lin-lin loss, $\mathcal{L}(1, \alpha, \theta)$.

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

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

(2)

We let $\varepsilon_{t+1}^*$ be the optimal forecast error, $\varepsilon_{t+1}^* \equiv y_{t+1} - f_{t+1}^* = y_{t+1} - \theta^* w_t$, 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 (Necessary Optimality Condition)** Under Assumption (A0), given $(p_0, \alpha_0) \in \mathbb{N}^* \times (0, 1)$, if $\theta^*$ is the minimum of $E[\mathcal{L}(p_0, \alpha_0, \theta)]$, then $\theta^*$ satisfies the first order condition

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

(3)

In other words, if the optimal forecast $f_{t+1}^*$ is such that $\theta^*$ is an interior point of $\Theta$ (Assumption (A0)), the sequence of optimal forecast errors $\varepsilon_{t+1}^*$ will satisfy the moment conditions $E[W_t \cdot (1(\varepsilon_{t+1}^* < 0) - \alpha_0) \cdot |\varepsilon_{t+1}^*|^{p_0-1}] = 0$.

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 to the forecaster 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 economic researchers 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 for forecasting should be included in this model, variables that could have been used to construct the forecasts can be used to test the orthogonality condition.\(^5\)

We assume that the forecast user observes a $d$-vector of variables $V_t$ that were available to the forecast producer. Under rational forecasts $V_t$ is a subvector of $W_t$. Given values for $(\alpha_0, p_0)$, the hypothesis of rational forecasts can be tested through the moment conditions

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

Under mean squared error loss, the parameters of the loss function are $(\alpha_0, p_0) = (0.5, 2)$. This choice simplifies the expression (4) since $-(1(Y_{t+1} - f_{t+1}^* < 0) - \alpha_0) \cdot |Y_{t+1} - f_{t+1}^*|^{p_0-1}) = \varepsilon_{t+1}$ so the observable forecast errors themselves should be 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 a major reason why this loss function is so popular. It is this result that is typically tested in practice with data (see, e.g., Campbell and Ghysels, 1995, Keane and Runkle, 1990, Zarnowitz, 1985). Such tests on the forecast errors are usefully divided into tests of ‘unbiasedness’ and ‘orthogonality’. Unbiasedness tests set $V_t = 1$, and hence test that $E[\varepsilon_{t+1}^*] = 0$. This can be undertaken by either directly testing that the mean from an observed sequence of forecast errors is zero, i.e. having observed a $T \times 1$ time series of forecast errors $\{\varepsilon_{t+1}^*\}_{t=1}^T$ 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 by noting that $\varepsilon_{t+1}^* \equiv y_{t+1} - f_{t+1}^*$. This suggests estimating the regression

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

\(^5\)This means we are concerned with partial rationality, i.e. the forecaster’s efficient use of a particular subset of information as opposed to full rationality which requires efficient utilization of all relevant information at the time the forecast is produced, c.f. Brown and Maital (1981).
and considering the joint test $H_0 : \beta_0 = 0, \beta_1 = 1$ against the alternative that one or both coefficients differ from their null values.

The 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.

It is important to note that these tests require that the MSE loss specification is valid. Under more general loss functions, however, it is not the forecast errors themselves that are orthogonal to time-$t$ information but a transformation of these forecast errors (i.e. the first difference of the loss function evaluated at the forecast error). Hence, as noted in passing by many 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. The economic interpretation is unclear when there is a rejection of the joint hypothesis. It may be that the power of the rationality tests 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 important to extend the class of loss functions for which the tests are valid.

It is this point that motivates the approach of our paper. Rather than assume an explicit loss function, we generalize the idea of rationality to a class of loss functions indexed by the parameter set $(\alpha, p)$. We then show that, given observed forecasts and outcomes, we can estimate the asymmetry parameter of the loss function $\alpha_0$ within the families we examine (families in which $p_0$ is given). Further, by using the additional time-$t$ information $V_t$ we are able to jointly test rationality and the class of loss functions rather than imposing a particular loss function, such as MSE loss. As we show in the following section, the resulting test is simply a test of overidentification ($J$-test) in a GMM estimation procedure.
3 Estimating Loss Function Parameters

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) from Proposition 1. The main idea behind our approach is fairly simple: if for given shape parameters $p_0$ and $\alpha_0$ the forecaster uses (3) to determine $\theta^*$, then for a given $\theta^*$ we can reverse the problem and use the same moment condition (3) to recover $p_0$ and $\alpha_0$. 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 $\alpha_0$. This creates an important difference between our problem and that considered by Hansen and Singleton (1982). Hansen and Singleton work with moment conditions of the form

$$E_t[h(Z_{t+1}, \beta_0)] = 0,$$  \hspace{1cm} (5)

where $Z_{t+1}$ is a vector of state variables observed by both agents and the econometrician at time $t+1$ and $\beta_0$ is a vector of unknown parameters. Hence, the key difference between moment conditions (4) and (5) is that Hansen and Singleton condition on the observable state variables ($Z_{t+1}$) which in our setup is the agent’s forecast error, $Z_{t+1} = Y_{t+1} - f_{t+1}$, whereas in the setup here we consider directly how this state variable was constructed and how it depends on a vector of unobservables, $W_{t+1}$. This gives a different interpretation of the estimated parameter. Employing the usual GMM framework and conditioning on the forecast error, the estimated parameter ($\beta_0 = \alpha_0$ in our notation when $p_0$ is given) estimates the parameter that is consistent with the forecasts being rational for the loss function, however stops short of linking this parameter with the actual loss function used to generate the forecasts. To provide this link we develop conditions on the form of the forecasting model and on the agent’s estimation problem ensuring that the estimated GMM parameters do identify the agent’s loss function and show a set of primitive conditions for the forecasting model where the parameters of the loss function can be consistently estimated.

The identification requirement is not easy to meet in general so we 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 $\theta^*$ to be optimal, i.e. not every value $\theta^*$ solving (3) is going to be the minimum of $E[\mathcal{L}(p_0, \alpha_0, \theta)]$ on $\hat{\Theta}$ (the interior
of $\Theta$). 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 $\theta^* \in \acute{\Theta}$ is a solution to the first order condition (3) then $\theta^*$ is a strict local minimum of $\mathcal{L}(p_0, \alpha_0, \theta)$ on $\acute{\Theta}$, i.e. there exists a neighborhood $\mathcal{V}$ of $\theta^*$ 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 $\Theta$ to be a local minimum of $\mathcal{L}$. Note that the first order condition does not necessarily hold if $\theta^*$ is on the boundary of $\Theta$, i.e. if $\theta \in \Theta \setminus \acute{\Theta}$. Also, note that the condition in Proposition 2 is slightly stronger than a necessary condition for $\theta^* \in \acute{\Theta}$ to be a local minimum of $\mathcal{L}$. Indeed $\theta^* \in \acute{\Theta}$ being a 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 $\theta$, evaluated at $\theta^*$, is positive semidefinite. In Proposition 2, however, Assumptions (A1)-(A2) imply that the Hessian is positive definite so that $\theta^*$ is a strict local minimum of $\mathcal{L}$ on $\acute{\Theta}$.

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

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 \( F_t \), s o t h a t t h e first order condition (3) is equivalent to

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

(6)

Note that for \( p_0 = 1, 2 \), and conditional on \( F_t \), the term \( |Y_{t+1} - \theta^*W_t|^{p_0-1} \) is strictly positive a.s. - \( \mathcal{P} \) so that the condition (6) can only be satisfied if \( 1(Y_{t+1} - \theta^*W_t < 0) - \alpha_0 = 0 \), a.s. - \( \mathcal{P} \). In other words, if the forecasting model \( M = \{ f_{t+1} \} \) is correctly specified, then \( E_t[1(Y_{t+1} - \theta^*W_t < 0)] = \alpha_0 \), so that the conditional \( \alpha_0 \)-quantile of the optimal forecast error \( \varepsilon^*_t \equiv Y_{t+1} - \theta^*W_t \) is exactly equal to zero. Hence, the optimal value \( \theta^* \) is unique: \( \theta^*W_t = F_t^{-1}(\alpha_0) \), where \( F_t^{-1} \) is the inverse of the conditional distribution function of \( Y_{t+1} \).

This uniqueness property allows us to compute the value of \( \alpha_0 \) used in the construction of the cost function \( L \), by inverting the preceding equation. Thus, the uniqueness of \( \alpha_0 \) follows directly from the result

\[
\alpha_0 = F_t(\theta^*W_t).
\]

(7)

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 \( F_t \)-measurable variables that are not contained in \( W_t \). In this case, the first order condition (3) is weaker than (6) and the aforementioned property of the optimal forecast errors \( \varepsilon^*_t \) is no longer true. Hence, in the presence of misspecification in the forecaster’s model we cannot deduce from (3) that the conditional \( \alpha_0 \)-quantile of the optimal forecast error \( \varepsilon^*_t \) is zero. In particular, this implies that the unicity of \( \theta^* \) is not trivially verified, which makes the true value of the probability level \( \alpha_0 \) 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 \( G \)
of $\hat{\Theta}$ such that, for any $\alpha_0 \in (0, 1)$, equation (3) has a unique solution $\theta^*$ in $G$ and that this solution is implicitly defined as a function $\theta_{p_0}(\alpha_0)$ of $\alpha_0$. This result is established in the following Proposition.

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

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

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

We now turn to the problem of estimating the true value $\alpha_0$ used in the loss function $\mathcal{L}$ minimization problem (2). As previously, we are interested in recovering $\alpha_0$ 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. Instead we assume 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) imply that

$$E[V_t \cdot (1(Y_{t+1} - f_{t+1}^* < 0) - \alpha_0) \cdot |Y_{t+1} - f_{t+1}^*|^{p_0-1}] = 0. \quad (8)$$

The following lemma will be useful in the construction of an estimator for $\alpha_0$.

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

$$Q_0(\alpha) \equiv E[V_t \cdot 1(Y_{t+1} - f_{t+1}^* < 0) \cdot |Y_{t+1} - f_{t+1}^*|^{p_0-1}] \cdot S^{-1} \cdot E[V_t \cdot (1(Y_{t+1} - f_{t+1}^* < 0) - \alpha) \cdot |Y_{t+1} - f_{t+1}^*|^{p_0-1}],$$

i.e.

$$\alpha_0 = \frac{E[V_t \cdot |Y_{t+1} - f_{t+1}^*|^{p_0-1}] \cdot S^{-1} \cdot E[V_t \cdot (1(Y_{t+1} - f_{t+1}^* < 0) - \alpha) \cdot |Y_{t+1} - f_{t+1}^*|^{p_0-1}]}{E[V_t \cdot |Y_{t+1} - f_{t+1}^*|^{p_0-1}] \cdot S^{-1} \cdot E[V_t \cdot |Y_{t+1} - f_{t+1}^*|^{p_0-1}]} \quad (9)$$
where $V_t$ is a sub-vector of $W_t$ and $S \equiv E[V_t V'_t \cdot (1(Y_{t+1} - f^*_t < 0) - \alpha)^2 \cdot |Y_{t+1} - f^*_t|^{2p_0-2}]$ is a positive definite weighting matrix.

If we observed the sequence of optimal one-step-ahead point forecasts $f^*_t \equiv \theta^* W_t$ provided by the forecaster, we could estimate $\alpha_0$ directly from equation (9). In practice, however, we only observe the sequence $\{f_{t+1}\}$ where $f_{t+1} \equiv \theta'_t w_t$ and $\theta'_t$ 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 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 $\theta^*$. 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 $\theta^*$ is obtained based on 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 provides a sequence of $T = n - \tau + 1$ forecasts, $\{\hat{f}_{t+1}\}_{\tau \leq t < T + \tau}$, where for each $t,$ $\tau \leq t < T + \tau$, the forecasts are constructed as $\hat{f}_{t+1} = \hat{\theta}_t w_t$ and where $\hat{\theta}_t$ is an estimate of $\theta^*$ that relies on the data from period $t$ and earlier, i.e. $(y_2, w'_1, \ldots, y_t, w'_{t-1})'$. This procedure is summarized in Figure 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 $\{\hat{f}_{t+1}\}_{\tau \leq t < T + \tau}$ provided by the forecaster, we now construct an estimator for $\alpha_0$ based on equation (9). 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 $\alpha_0$, $\hat{\alpha}_T$, defined as

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

where $\hat{S}$ is some consistent estimate of $S = E[V_t V_t'] \cdot (1(Y_{t+1} - \hat{f}_{t+1}^* < 0) - \alpha)^2 \cdot |Y_{t+1} - \hat{f}_{t+1}^*|^2 p_0^{-2}$ in Lemma 4. 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 (10). 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 exists with probability approaching one and is moreover consistent for the true value $\alpha_0$. 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 economic researcher can still consistently estimate the true value $\alpha_0$ used in the loss function minimization problem (2). These concerns are
important in practice where the full information set observed by the forecaster is unlikely to be available to economic researchers.

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

**Proposition 6 (Asymptotic Normality)** Given $p_0 = 1, 2$, let $\hat{\alpha}_T$ be the linear IV estimator defined in (10). 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, as in Lemma 4, $S = E[V_t V_t' \cdot (1(Y_{t+1} - f_{t+1}^* < 0) - \alpha)^2 \cdot |Y_{t+1} - f_{t+1}^*|^{2p_0-2}]$ and $h_0 = E[V_t \cdot |Y_{t+1} - f_{t+1}^*|^{p_0-1}]$.

In other words, the linear IV estimator $\hat{\alpha}_T$ is asymptotically normal, with asymptotic variance which does not depend on neither $W_t$ or $\theta^*$, 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 $\tau \leq t < T$, and $\hat{\alpha}_T$ are obtained as solutions to the same first order condition (8), which moreover is linear in $\alpha$. Were $\hat{\theta}_t$, $\tau \leq t < T$, 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 $\hat{\theta}_t$, $\tau \leq t < T$, would in that case affect the asymptotic variance of $\hat{\alpha}_T$ and make it substantially more complicated.\(^6\)

Propositions 5 and 6 focus on the cases where $p_0 = 1, 2$, which are likely to be the ones most useful in empirical analysis as they nest MAD and MSE loss. It is not clear that much is learned from using higher values of $p_0$ since the quadratic asymmetric loss function is already capable of providing large weights on large forecast errors. However, it is clear from the proofs in the Appendix that results can readily be derived for larger values of $p_0$ under

\(^6\)For general results on asymptotic inference in the presence of parameter uncertainty see, e.g., West, 1996, West and McCracken, 1998, Corradi and Swanson, 2002.
modifications of the assumptions ensuring that the relevant mixing and moment conditions hold.

In practice, the computation of the linear IV estimator \( \hat{\alpha}_T \) is done iteratively\(^7\). According to equation (10), \( \hat{\alpha}_T \) 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, for example, \( \hat{S}(\hat{\alpha}_T) \equiv T^{-1} \sum_{t=\tau}^{T+\tau-1} v_t v_t' (1(y_{t+1} - \hat{f}_{t+1} < 0) - \hat{\alpha}_T)^2 |y_{t+1} - \hat{f}_{t+1}|^{2p_0-2} \), where \( \hat{\alpha}_T \) is some consistent initial estimate of \( \alpha_0 \), or by using some heteroskedasticity and autocorrelation robust estimator, such as Newey and West’s (1987) estimator.\(^8\) 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 (10) 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 (10) 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 \( h_0 \) in Proposition 6 above, 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.\(^9\)

Finally, we can use the moment conditions (3) to test the hypothesis that the forecasts are rational with respect to available information within the class of loss functions (1) (i.e. without specifying a value for \( \alpha_0 \)). 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 \( \alpha_0 \) above). A different question that can be asked is whether, for a given \( V_t \), there exists some value of \( \alpha_0 \) for which the forecasts are rational? The usual test for overidentification of the IV estimates tests this proposition (so long as \( d > 1 \)). One degree of freedom is used in the estimation of the loss parameter, \( \hat{\alpha}_T \), so, from the results of Proposition 6, the resultant

---

\(^7\)In the case where \( d = 1 \) the estimator is independent of \( S \) and a closed form solution exists. In the case where \( p_0 = 1 \) and \( V_t = 1 \) for example the estimator is simply the proportion of negative forecast errors.

\(^8\)The consistency of \( \hat{S}(\hat{\alpha}_T) \) can be shown by an argument analogous to the one in the Proof of Proposition 5. We therefore omit it from the paper.

\(^9\)The consistency of \( \hat{h}_T \) is shown in the proof of Proposition 5.
test statistic,

\[ J = T^{-1} \left( \sum_{t=\tau}^{T+\tau-1} v_t (1(y_{t+1} - \hat{f}_{t+1} < 0 - \hat{\alpha}_T)|y_{t+1} - \hat{f}_{t+1}|^{p_0-1}) \hat{S}^{-1} \right) \]

is distributed \( \chi^2_{d-1} \) and thus rejects for large values. Tests based on an assumption of mean squared error 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 \( \alpha_0 \) were different from 0.5 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 assumption that \( \alpha_0 = 0.5 \), however there is clearly a gain in terms of directing power in the desired direction.

### 4 Simulation Results

We next examine the behavior of the proposed estimator (10) and tests (11) in a Monte Carlo experiment. Pseudo data 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.

Table 1 examines, for various samples 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 for
\( p_0 = 1 \) (lin-lin) are provided in the first panel while the second panel reports results for \( p_0 = 2 \) (quad-quad). In all cases we have not used any instruments other than the constant term, 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, hence we obtain asymmetric small sample distributions and require a larger \( n \) for the central limit theorem to provide a good approximation.\(^{10}\) More observations (either in or out of sample) tend to help control the size.

Table 2 repeats the exercise in Table 1 but now employs the two time-varying instruments, \( W_{1t}, W_{2t} \) in addition to the constant, i.e. \( V_t = W_t \). The use of the extra instruments results in larger size distortions across the board. The problem is again more of an issue for the quad-quad case than for the lin-lin loss function. As we expect, this is less of a problem when more in-sample and out-of-sample observations are available. As before, more out-of-sample observations help more than more in-sample observations in controlling size. Problems are again greater, the further \( \alpha_0 \) is from one half.

We examine the proposed tests for overidentification in Table 3. These tests examine whether or not the moment conditions are compatible with some \( \alpha_0 \), left unspecified in the testing procedure. We report the size of the tests under the null for the same cases studied as those in the previous two tables. For these tests, size is well controlled. The tests tend to be undersized rather than oversized, and the size departures from nominal size (5%) are larger when \( \alpha_0 \) is further away from one half. When \( \alpha_0 \) is at one half, 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 on the optimality of forecasts of government budget deficits for the G7 countries produced by two international

\(^{10}\)This is identical to the usual result in applying the central limit theorem to the Bernoulli distributed outcomes.
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) 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). A similar point is made by Campbell and Ghysels (1995) in the context of an analysis of federal budget projections.

The data that we use extends the data considered in Artis and Marcellino (2001) by five years and comprises budget deficit forecasts, reported as a percentage of GDP, for the G7 countries. 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 the one which 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 the G7 countries, namely France, Germany, Italy and the UK. 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 contains between 24 and 27 data points and goes from 1975 - 2001. The IMF sample has information on all G7 countries. It goes from 1976 to 2000 and thus contains 25 observations. Clearly these are not large samples, so some caution should be exercised in the interpretation of the statistical results.

Table 4 presents summary statistics for the forecast errors. The mean forecast error suggests that overpredictions of budget deficits is the typical situation (positive average forecast errors), although there are also some countries, most notably Canada, with underpredictions (negative average forecast errors). For some countries there are strong imbalances between the number of positive \((n+)\) and negative \((n−)\) forecast errors, e.g. in the case of current-year IMF forecasts for Italy, Japan, UK and the US, where between 19 and 21 of 25 forecast

\[\text{We are grateful to Massimiliano Marcellino for providing the first part of the data. The data source is the IMF’s World Economic Outlook and the OECD’s Economic Outlook. Artis and Marcellino (2001) also consider forecasts from the European Commission, but this data set is very short (14 observations) so we decided not to include it here.}\]
errors are positive. While most countries produce a majority of positive forecast errors, Canada is again the exception with a majority of negative forecast errors. The current year RMSE values vary from 0.49 (France) to 2.08 (Japan). This range extends from 0.74 to 2.32 for the 1-year-ahead forecasts. For all countries the RMSE values are higher at the 1-year horizon than at the current-year horizon. This is as we would expect since the current-year forecasts are based on more up-to-date information than the 1-year-ahead forecasts.

The previous sections show how to test forecast optimality in conjunction with the assumption that the loss function belongs to a particular class that only depends on a single unknown shape or asymmetry parameter, $\alpha$. This approach is well suited to the current data set where only a small sample of forecasts is available. In our empirical tests we adopt the strategy of first assuming that the loss function is lin-lin ($p_0 = 1$). 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 non-symmetric 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 ($p_0 = 2$).

### 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 are indicative of 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 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 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 very 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 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 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, the 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. The evidence of asymmetric loss is somewhat weaker at the 1-year 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 budget deficits. 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 $\alpha$ suggest very 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. In our context we can test this hypothesis by imposing $\alpha = 1/2$ and examining the $J$-test (11) 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 Table 6. Overall, 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 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 in the current-year data for France, Germany and Italy and, in the 1-year forecasts, also for Germany. Notice that this is the same list of countries for which we found overwhelming evidence of asymmetric loss in Table 5.

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 (11) to our forecast data when $\alpha$ is not constrained to reflect symmetric loss.

The results are very interesting and in complete contrast to those found in Table 6. 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 in Table 6 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 Tables 6 and 7 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.

5.3 Robustness to the Shape of the Loss Function

So far we have maintained $p_0 = 1$ in our tests, corresponding to a piecewise linear 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 empirical results for the quad-quad loss function.

First consider the evidence of asymmetries in the quad-quad loss function (Table 8). For 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 $\alpha$ well below one-half. Interestingly, there is now also some evidence of asymmetric loss for Canada, albeit with an $\alpha$-estimate above one-half. At the 1-year forecast horizon Canada, France, Italy, Japan, UK and the US all produce evidence of asymmetric loss. In the OECD data there is strong evidence of asymmetric loss for France, Germany and Italy in the case of current-year predictions and, at the 1-year horizon, also for Germany.

Returning again 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 Italy, Japan, UK and the US. The 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, as we do in Table 10, 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 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 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.

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 transformed 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. In a sample as small as that considered here, this could be a major concern. In fact, our results suggest that the evidence against forecast rationality and quad-quad loss is quite weak.

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. In an empirical application 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.

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.
Forecasts produced by central banks and international organizations such as the IMF and OECD are routinely used as inputs in economic agents’ decision process. Our results suggest that unless it is known that such forecast producers have symmetric loss, it is important to account for the possible effects of asymmetric loss. Most obviously, 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.

Appendix A: Assumptions

(A0) Θ is a compact subset of ℝ^h and θ^* is interior to Θ, i.e. θ^* ∈ ˚Θ;

(A1) the h-vector W_t (with the first component 1) is such that, given p_0 ∈ ℕ^*, for all θ ∈ ˚Θ, E[W_t · Y_{t+1} − θ'W_t|^{p_0−1}] ≠ 0 element by element, E[W_t · 1(Y_{t+1} − θ'W_t < 0) · Y_{t+1} − θ'W_t|^{p_0−1}] ≠ 0 element by element, and E[W_tW_t'] exists and is positive definite;

(A2) the density of Y_{t+1} conditional on ℱ_t is strictly positive, i.e. f_t (y) > 0, for every y ∈ ℝ;

(A3) the d-vector V_t is a sub-vector of the h-vector W_t (d ≤ h) with the first component 1 and there exists a constant K > 0 such that for every t, ||W_t||^2 = W_t'W_t ≤ m, a.s. − ℙ;

(A4) for every t, τ ≤ t < T + τ, ˚θ_t is a consistent estimate of θ^* and θ^* ∈ G ⊆ ˚Θ;

(A5) the stochastic processes {Y_t} and {W_t} are α-mixing with mixing coefficient α of size −r/(r − 1), r > 1, and, given p_0 ∈ ℕ^*, there exist some δ_Y > 0 and Δ_Y > 0 such that sup_{θ ∈ Θ} E[(Y_{t+1} − θ'W_t)^2(r+δ_Y)(p_0−1)] ≤ Δ_Y < ∞ and some δ_W > 0 and Δ_W > 0 such that E[||W_t||^2(r+δ_W)] ≤ Δ_W < ∞;

(A5') the stochastic processes {Y_t} and {W_t} are α-mixing with mixing coefficient α of size −r/(r − 2), r > 2, and, given p_0 ∈ ℕ^*, there exist some Δ_Y > 0 such that sup_{θ ∈ Θ} E[(Y_{t+1} − θ'W_t)^2(r+δ_Y)(p_0−1)] ≤ Δ_Y < ∞ and some Δ_W > 0 such that E[||W_t||^2r] ≤ Δ_W < ∞;

(A6) the density of Y_{t+1} conditional on ℱ_t is bounded, i.e. there exists some M > 0 such that sup_{y ∈ ℝ} f_t (y) ≤ M < ∞;
Appendix B: Proofs

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

\[
\min_{\theta \in \Theta} E\{[\alpha_0 + (1 - 2\alpha_0) \cdot 1(Y_{t+1} - \theta'W_t < 0)] \cdot |Y_{t+1} - \theta'W_t|^p \} \equiv \min_{\theta \in \Theta} \Sigma(\theta), \tag{12}
\]

with \( \Sigma(\theta) \) continuously differentiable on \( \Theta \), and \( \theta^* \in \hat{\Theta} \) (Assumption (A0)), then \( \theta^* \) 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 \( \Sigma_{t+1}(\theta) \equiv [\alpha_0 + (1 - 2\alpha_0) \cdot 1(Y_{t+1} - \theta'W_t < 0)] \cdot |Y_{t+1} - \theta'W_t|^p \). The function \( \Sigma_{t+1}(\theta) \) is continuously differentiable on \( \Theta \setminus A_{t+1} \) 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 \( \Theta \setminus A_{t+1} \). We have, by the law of iterated expectations,

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

so that

\[
\nabla_\theta \Sigma(\theta) = E\{\nabla_\theta E_{t+1}[\Sigma_{t+1}(\theta)]\} = E\{\nabla_\theta E_{t+1}[\Sigma_{t+1}(\theta) \cdot 1(\theta \in A_{t+1}^c)]\} + E\{\nabla_\theta E_{t+1}[\Sigma_{t+1}(\theta) \cdot 1(\theta \in A_{t+1})]\} = E\{\nabla_\theta \Sigma_{t+1}(\theta) \cdot E_{t+1}[1(\theta \in A_{t+1}^c)]\} + E\{\nabla_\theta \Sigma_{t+1}(\theta) \cdot E_{t+1}[1(\theta \in A_{t+1})]\},
\]

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

\[
\nabla_\theta \Sigma(\theta) = (1 - 2\alpha_0) E[\nabla_\theta 1(Y_{t+1} - \theta'W_t < 0) \cdot |Y_{t+1} - \theta'W_t|^p] - p_0 \cdot E\{[\alpha_0 + (1 - 2\alpha_0) \cdot 1(Y_{t+1} - \theta'W_t < 0)] \cdot W_t \cdot [1 - 2 \cdot 1(Y_{t+1} - \theta'W_t < 0)] \cdot |Y_{t+1} - \theta'W_t|^{p_0 - 1}\}
\]

so that

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

Note that

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

27
where \( \delta \) 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

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

for any non-zero \( p_0 \). Thus,

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

For given values of \( p_0 \) and \( \alpha_0 \), if \( \theta^* \) is the minimum of \( \Sigma(\theta) \) then \( \theta^* \) 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 \( \theta^* \in \hat{\Theta} \) to be a solution to the minimization problem

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

We know that \( \theta^* \) is a strict local minimum of \( \Sigma(\theta) \) on \( \hat{\Theta} \) 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 160). Recall that,

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

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

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

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

28
Consider the following two cases separately: \( p_0 = 1 \) and \( p_0 > 1 \).

CASE \( p_0 = 1 \): the above expression becomes

\[
\Delta_{\theta\theta} \Sigma(\theta) = E[W_t W_t' \cdot \delta(\theta' W_t - Y_{t+1})] \\
= E\{W_t W_t' \cdot E_t[\delta(\theta' W_t - Y_{t+1})]\} \\
= E[W_t W_t' \cdot f_t(\theta' W_t)],
\]

where \( f_t \) is the density of \( Y_{t+1} \) conditional on \( F_t \). Hence, for any \( \varphi \in \mathbb{R}^k \) we have

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

so that by imposing the strict positivity of the conditional density \( f_t \) (Assumption (A2)), we have

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

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

CASE \( p_0 > 1 \): the matrix of second derivatives \( \Delta_{\theta\theta} \Sigma(\theta) \) reduces to

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

since \( E[W_t W_t' \cdot \delta(\theta' W_t - Y_{t+1}) \cdot Y_{t+1} - \theta' W_t | p_0 - 1] = 0 \) if \( p_0 > 1 \). In that case we have

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

and, conditional on \( F_t \), 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

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

so that by Assumption (A1) \( \varphi = 0 \). Hence, for every \( \theta \in \hat{\Theta} \) the matrix \( \Delta_{\theta\theta} \Sigma(\theta) \) is positive definite, and \( \theta^* \) is a strict local minimum of \( \Sigma(\theta) \) on \( \hat{\Theta} \). 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

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

(13)

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). 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 $\partial h_{p_0}(\alpha_0, \theta^*)/\partial \theta$ is nonsingular, i.e. $[\partial h_{p_0}(\alpha_0, \theta^*)/\partial \theta]^{-1}$ exists.

According to equation (13) the function $h_{p_0}$ is linear in $\alpha$ 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 $\alpha$. Therefore, the function $\partial h_{p_0}(\cdot, \theta)/\partial \alpha : (0, 1) \to \mathbb{R}^k$ is continuous on $(0, 1)$. We now turn to the study of $h_{p_0}(\alpha, \cdot) : \Theta \to \mathbb{R}^k$. Note that $\frac{\partial h_{p_0}}{\partial \theta}(\alpha, \theta) = \Delta_{\theta} \Sigma(\theta)$ where $\Sigma(\theta)$ is defined as in (12), so that

$$\frac{\partial h_{p_0}}{\partial \theta}(\alpha, \theta) = 1(p_0 = 1) \cdot E[W_t W_t' \cdot f_t(\theta'' W_t)]$$

$$+ 1(p_0 = 2) \cdot E[W_t W_t' \cdot (\alpha + (1 - 2 \alpha) \cdot F_t(\theta' W_t))]
+ 1(p_0 > 2) \cdot (p_0 - 1)
\cdot E[W_t W_t' \cdot (\alpha + (1 - 2 \alpha) \cdot 1(Y_{t+1} - \theta' W_t < 0)) \cdot |Y_{t+1} - \theta' W_t|^{p_0-2}],$$

where $F_t$ and $f_t$ are the distribution function and the density of $Y_{t+1}$ conditional on $\mathcal{F}_t$. Being an integral, the function $\partial h_{p_0}(\alpha, \cdot)/\partial \theta : \Theta \to \mathbb{R}^k \times \mathbb{R}^k$ is clearly continuous on $\Theta$. 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)) and 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 $[\partial h_{p_0}(\alpha_0, \theta^*)/\partial \theta]^{-1}$ exists for every $\alpha_0 \in (0, 1)$, which verifies 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 exists an open interval $E_0$ containing $\alpha_0$ and an open set $G_0$ containing $\theta^*$, $G_0 \equiv \{ \theta \in \tilde{\Theta} : ||\theta - \theta^*|| < \delta_0 \}$ with $\delta_0 > 0$, such that for every $\alpha \in E_0$, the equation $h_{p_0}(\alpha, \theta) = 0$ has a unique solution $\theta$ in $G_0$, and 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 $G_0$ with

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

$$\theta'_{p_0}(\alpha) = \begin{cases} E[W_t W'_t]^{-1} \cdot E[W_t], & \text{if } p_0 = 1, \\
E[W_t W'_t] \cdot (\alpha + (1 - 2\alpha) \cdot F_0(\theta_{p_0}(\alpha)/W_t))^{-1} \cdot E[W_t \cdot |Y_{t+1} - \theta_{p_0}(\alpha)/W_t|], & \text{if } p_0 = 2, \\
(p_0 - 1)E[W_t W'_t] \cdot (\alpha + (1 - 2\alpha) \cdot 1(Y_{t+1} - \theta_{p_0}(\alpha)/W_t < 0)) \cdot |Y_{t+1} - \theta_{p_0}(\alpha)/W_t|^{p_0 - 2}]^{-1} \cdot E[W_t \cdot |Y_{t+1} - \theta_{p_0}(\alpha)/W_t|^{p_0 - 1}], & \text{if } p_0 > 2. 
\end{cases} \quad (14)$$

It is important to note that we can 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, $G$ is an open subset of $\tilde{\Theta}$. 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 $\theta^*$ in $G$ and the implicit function $\theta^* = \theta_{p_0}(\alpha_0)$ is continuously differentiable from $(0, 1)$ to $G$ with $\theta'_{p_0}(\alpha)$ as given in (14). We now show that $\theta_{p_0}(\alpha)$ is 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. $\alpha_1 \neq \alpha_2$ 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

$$0 = E[W_t \cdot (1(Y_{t+1} - \theta_{p_0}(\alpha_1)/W_t < 0) - \alpha_1) \cdot |Y_{t+1} - \theta_{p_0}(\alpha_1)/W_t|^{p_0 - 1}]$$

$$-E[W_t \cdot (1(Y_{t+1} - \theta_{p_0}(\alpha_2)/W_t < 0) - \alpha_2) \cdot |Y_{t+1} - \theta_{p_0}(\alpha_2)/W_t|^{p_0 - 1}]$$

$$= (\alpha_2 - \alpha_1)E[W_t \cdot |Y_{t+1} - \theta_{p_0}(\alpha_2)/W_t|^{p_0 - 1}],$$

which, by Assumption (A1), implies $\alpha_1 = \alpha_2$. Hence for a given $\theta^* \in G$ 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. First, let us show that $S$ is positive definite. Recall that, given $p_0 \in \mathbb{N}^*$, we have

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

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

$$\varphi'^t S \varphi = 0 \Rightarrow \varphi'^t V_t \varphi = 0, \text{a.s.} - \mathcal{P} \Rightarrow \varphi'^t E[V_t V_t'] \varphi = 0.$$

Now, note that the positive definiteness of $E[W_t W_t']$ (Assumption (A1)) 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 $\varphi'^t E[V_t V_t'] \varphi = E[\varphi'^t V_t \varphi \cdot (1(Y_{t+1} - f_{t+1}^*) < 0) - \alpha)^2 \cdot |Y_{t+1} - f_{t+1}^*|^{2p_0-2}] > 0$), this implies that $E[V_t V_t']$ is positive definite. Therefore $\varphi'^t E[V_t V_t'] \varphi = 0$ implies $\varphi = 0$, which shows that $S$ (and hence $S^{-1}$) is positive definite. We next show that there exists a unique minimum in $(0, 1)$ of the quadratic form

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

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

$$a \equiv E[V_t \cdot |Y_{t+1} - f_{t+1}^*|^{p_0-1}]' \cdot S^{-1} \cdot E[V_t \cdot |Y_{t+1} - f_{t+1}^*|^{p_0-1}],$$
$$b \equiv E[V_t \cdot |Y_{t+1} - f_{t+1}^*|^{p_0-1}]' \cdot S^{-1} \cdot E[V_t \cdot 1(Y_{t+1} - f_{t+1}^* < 0) \cdot |Y_{t+1} - f_{t+1}^*|^{p_0-1}],$$
$$c \equiv E[V_t \cdot 1(Y_{t+1} - f_{t+1}^* < 0) \cdot |Y_{t+1} - f_{t+1}^*|^{p_0-1}]' \cdot S^{-1} \cdot E[V_t \cdot 1(Y_{t+1} - f_{t+1}^* < 0) \cdot |Y_{t+1} - f_{t+1}^*|^{p_0-1}].$$

Since the weighting matrix $S^{-1}$ is positive definite, we know that $a > 0$ so that $Q_0(\alpha)$ is a concave function of $\alpha$. It therefore has a unique minimum at $\alpha^* = b/a$,

$$\alpha^* = \frac{E[V_t \cdot |Y_{t+1} - f_{t+1}^*|^{p_0-1}]' \cdot S^{-1} \cdot E[V_t \cdot 1(Y_{t+1} - f_{t+1}^* < 0) \cdot |Y_{t+1} - f_{t+1}^*|^{p_0-1}]}{E[V_t \cdot |Y_{t+1} - f_{t+1}^*|^{p_0-1}]' \cdot S^{-1} \cdot E[V_t \cdot |Y_{t+1} - f_{t+1}^*|^{p_0-1}]}.$$ (15)

To demonstrate that this solution is valid, we need to verify that $\alpha^*$ defined in (15) 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. \( V_t > 0_d, a.s. - \mathcal{P} \), where \( 0_d \) is a \( d \)-vector of zeros. In that case we have

\[
0 \leq V_t \cdot 1(Y_{t+1} - f^*_{t+1} \leq 0) \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1} \leq V_t \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}, a.s. - \mathcal{P},
\]

so that

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

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 \( S^{-1} \) is positive definite, we have

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

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

\[
\leq E[V_t \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}] \cdot S^{-1} \cdot E[V_t \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}],
\]

i.e. \( 0 < c < b \leq a \). Hence \( \alpha^* > 0 \). We also know that for all \( \alpha \in (0, 1) \), \( Q_0(\alpha) > 0 \) so that the reduced discriminant \( b^2 - ac < 0 \). Hence, \( b < \sqrt{ac} \leq a \) 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 Assumption (A3), which ensures that \( ||V_t|| \leq ||W_t|| \leq K \) so that the components of \( V_t \) are necessarily bounded by some constant \( c \). Now, consider the rotation of the \( d \)-vector \( V_t \),

\[
\tilde{V}_t = KV_t = \begin{pmatrix} 1 & 0 \\ c & I_{d-1} \end{pmatrix} V_t,
\]

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

\[
\alpha^* = \frac{E[V_t \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}] \cdot S^{-1} \cdot E[V_t \cdot 1(Y_{t+1} - f^*_{t+1} < 0) \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}]}{E[V_t \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}] \cdot S^{-1} \cdot E[V_t \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}]}
\]

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

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

\[
\tilde{Q}(\alpha) \equiv E[\tilde{V}_t \cdot (1(Y_{t+1} - f_{t+1}^*) < 0) - \alpha) \cdot [Y_{t+1} - f_{t+1}^*|_{p_0-1}]^T \cdot K^{-1}S^{-1}(K^{-1})' \cdot E[\tilde{V}_t \cdot (1(Y_{t+1} - f_{t+1}^*) < 0) - \alpha) \cdot [Y_{t+1} - f_{t+1}^*|_{p_0-1}].
\]

From the results above we then know that \(\alpha^* \in (0, 1)\) since \(\tilde{V}_t > 0\), a.s. - \(\mathcal{P}\). Hence, under Assumptions (A0)-(A3), we know that \(Q_0(\alpha)\) is uniquely minimized at \(\alpha^*\) defined in (15) and \(\alpha^* \in (0, 1)\).

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

\[
0 = b - \alpha a
\]

(16)

\[
= E[\tilde{V}_t \cdot |Y_{t+1} - f_{t+1}^*|_{p_0-1}]^T \cdot S^{-1}. E[\tilde{V}_t \cdot (1(Y_{t+1} - f_{t+1}^*) < 0) - \alpha) \cdot [Y_{t+1} - f_{t+1}^*|_{p_0-1}]
\]

is a minimum of \(Q_0(\alpha)\). We know that if \(V_t\) is a sub-vector of \(W_t\) (Assumption (A3)) then \(E[\tilde{V}_t \cdot |Y_{t+1} - f_{t+1}^*|_{p_0-1}] \neq 0\) (Assumption (A1)). Moreover, \(S^{-1}\) is nonsingular, so that the equality (15) implies \(E[\tilde{V}_t \cdot (1(Y_{t+1} - f_{t+1}^*) < 0) - \alpha) \cdot [Y_{t+1} - f_{t+1}^*|_{p_0-1}] = 0\). Hence, any solution to the moment condition \(E[\tilde{V}_t \cdot (1(Y_{t+1} - f_{t+1}^*) < 0) - \alpha) \cdot [Y_{t+1} - f_{t+1}^*|_{p_0-1}] = 0\) is a minimum of \(Q_0(\alpha)\). We know from (3) that \(\alpha_0\) satisfies this condition, so \(\alpha_0\) is a minimum of \(Q_0(\alpha)\). Therefore, we conclude that \(\alpha_0 = \alpha^*\), which completes the proof of Lemma 4. ■

Proof of Proposition 5. In this proof we use Assumptions (A0) to (A6).

Recall that from (10) we have

\[
\hat{\alpha}_T \equiv \frac{[T^{-1} \sum_{t=\tau}^{T+\tau-1} v_t|y_{t+1} - \hat{f}_{t+1}|_{p_0-1}]^T \cdot \hat{S}^{-1} \cdot [T^{-1} \sum_{t=\tau}^{T+\tau-1} v_t1(y_{t+1} - \hat{f}_{t+1} < 0)|y_{t+1} - \hat{f}_{t+1}|_{p_0-1}]}{[T^{-1} \sum_{t=\tau}^{T+\tau-1} v_t|y_{t+1} - \hat{f}_{t+1}|_{p_0-1}]^T \cdot \hat{S}^{-1} \cdot [T^{-1} \sum_{t=\tau}^{T+\tau-1} v_t|y_{t+1} - \hat{f}_{t+1}|_{p_0-1}]}.
\]

In order to show that \(\hat{\alpha}_T \overset{p}{\to} \alpha_0\) it is sufficient to show that: (i) \(T^{-1} \sum_{t=\tau}^{T+\tau-1} v_t|y_{t+1} - \hat{f}_{t+1}|_{p_0-1} \overset{p}{\to} E[\tilde{V}_t|Y_{t+1} - f_{t+1}^*|_{p_0-1}]\), and (ii) \(T^{-1} \sum_{t=\tau}^{T+\tau-1} v_t1(y_{t+1} - \hat{f}_{t+1} < 0)|y_{t+1} - \hat{f}_{t+1}|_{p_0-1} \overset{p}{\to} E[\tilde{V}_t \cdot 1(Y_{t+1} - f_{t+1}^* < 0) \cdot |Y_{t+1} - f_{t+1}^*|_{p_0-1}].\) Then, by using Lemma 4, the consistency of \(\hat{S},\)
that an alternative way to prove the same result would be to work with the quadratic form and consider it by using a law of large numbers (LLN) for $r/\alpha$ and for every $t, \tau \leq t < T + \tau$, let

$$g_t \equiv v_t(y_{t+1} - \hat{f}_{t+1} < 0)|y_{t+1} - \hat{f}_{t+1}|^{p_0-1},$$

$$\hat{g}_T \equiv T^{-1} \sum_{t=\tau}^{T+\tau-1} g_t,$$

$$g_0 \equiv E[V_t \cdot 1(Y_{t+1} - f^*_{t+1} < 0) \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}],$$

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

and

$$h_t \equiv v_t|y_{t+1} - \hat{f}_{t+1}|^{p_0-1},$$

$$\hat{h}_T \equiv T^{-1} \sum_{t=\tau}^{T+\tau-1} h_t,$$

$$h_0 \equiv E[V_t \cdot |Y_{t+1} - f^*_{t+1}|^{p_0-1}],$$

$$h^* \equiv E[V_t \cdot |Y_{t+1} - \hat{f}_{t+1}|^{p_0-1}].$$

We now show that conditions (i)-(ii) hold: by the triangle inequality we have $||\hat{g}_T - g_0|| \leq ||\hat{g}_T - g^*|| + ||g^* - g_0||$ and $||\hat{h}_T - h_0|| \leq ||\hat{h}_T - h^*|| + ||h^* - h_0||$. We first show that $||\hat{g}_T - g^*|| \overset{p}{\to} 0$ and $||\hat{h}_T - h^*|| \overset{p}{\to} 0$ by using a law of large numbers (LLN) for $\alpha$-mixing sequences (e.g., Corollary 3.48 in White 2001).

We first need to show that both stochastic processes $\{\hat{H}_t\}$, where $\hat{H}_t \equiv V_t \cdot |Y_{t+1} - \hat{f}_{t+1}|^{p_0-1},$ and $\{\hat{G}_t\}$, where $\hat{G}_t \equiv V_t \cdot 1(Y_{t+1} - \hat{f}_{t+1} < 0) \cdot |Y_{t+1} - \hat{f}_{t+1}|^{p_0-1},$ are $\alpha$-mixing: by Theorem 3.49 in White (2001) we know that measurable functions of mixing processes are mixing of the same size. Hence, by assumption (A5) we know that $\{\hat{H}_t\}$ and $\{\hat{G}_t\}$ are $\alpha$-mixing of size $-r/(r - 1)$ with $r > 1$. Before applying the LLN for $\alpha$-mixing sequences we need to ensure...
that the following moment conditions hold:

\[ E[|\hat{H}_t|^{r+\delta_H}] < \Delta_H < \infty \]
\[ E[|\hat{G}_t|^{r+\delta_G}] < \Delta_G < \infty, \]

for some \( \delta_H > 0 \) and \( \delta_G > 0 \). Let \( \delta_H \equiv \min(\delta_Y, \delta_W)/2 > 0 \). By Assumption (A5), the Cauchy-Schwarz inequality, and using \( E[||V_t||^{2(r+\delta_H)}] \leq E[||W_t||^{2(r+\delta_H)}] \), we know that

\[ E[|\hat{H}_t|^{r+\delta_H}] \leq (E[||V_t||^{2r+2\delta_H}])^{1/2} \cdot (E[(Y_{t+1} - \hat{t}_{t+1})^{2(r+\delta_H)(\rho_0-1)}])^{1/2} \]
\[ \leq (E[||V_t||^{2r+2\delta_H}])^{1/2} \cdot \max(1, \{ \sup_{\theta \in \Theta} E[(Y_{t+1} - \theta W_t)^2(r+\delta_H)(\rho_0-1)] \})^{1/2}, \]

by compactness of the parameter set \( \Theta \) (assumption (A0)) and consistency of \( \theta_t \) for every \( t \), \( \tau \leq t < T + \tau \) (assumption (A4)). Hence

\[ E[|\hat{H}_t|^{r+\delta_H}] \leq \max(1, \Delta_1^{1/2}) \cdot \max(1, \Delta_Y^{1/2}) < \infty. \]

Similarly, let \( \delta_G \equiv \min(\delta_Y, \delta_W)/2 > 0 \). We then have

\[ E[|\hat{G}_t|^{r+\delta_G}] \leq (E[||V_t||^{2r+2\delta_G}])^{1/2} \cdot (E[(Y_{t+1} - \hat{t}_{t+1})^{2(r+\delta_G)(\rho_0-1)}])^{1/2}, \]

and, since \( V_t^t V_t \cdot 1(Y_{t+1} - \hat{t}_{t+1} < 0) \leq V_t^{\prime} V_t, a.s. - \mathcal{P} \), so that \( E[||V_t||^{2(r+\delta_G)}] \leq E[||V_t||^{2(r+\delta_G)}] \), by the same reasoning as previously, we get \( E[|\hat{G}_t|^{r+\delta_G}] < \infty \). 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 for this case \( h^* = h_0 \) and that by the triangular and Cauchy-Schwarz inequalities, we have

\[ ||g^* - g_0||^2 = ||E[V_t \cdot (1(Y_{t+1} - \hat{t}_{t+1} < 0) - 1(Y_{t+1} - f_{t+1}^* < 0))||^2 \]
\[ \leq E[||V_t||^2] \cdot E[(1(Y_{t+1} - \hat{t}_{t+1} < 0) - 1(Y_{t+1} - f_{t+1}^* < 0))^2]. \]

For every \( t, \tau \leq t < T + \tau \), we have

\[ E\{[1(Y_{t+1} - \hat{t}_{t+1} < 0) - 1(Y_{t+1} - f_{t+1}^* < 0)]^2 \}
\[ = E\{[1(f_{t+1}^* \leq Y_{t+1} < \hat{t}_{t+1}) - 1(\hat{t}_{t+1} \leq Y_{t+1} < f_{t+1}^*)]^2 \}
\[ = E[1(f_{t+1}^* \leq Y_{t+1} < \hat{t}_{t+1}) + 1(\hat{t}_{t+1} \leq Y_{t+1} < f_{t+1}^*)]
\[ = E\{E_t[1(f_{t+1}^* \leq Y_{t+1} < \hat{t}_{t+1}) + 1(\hat{t}_{t+1} \leq Y_{t+1} < f_{t+1}^*)] \} , \]
where
\[
E_t[1(f_{t+1}^* \leq Y_{t+1} < \hat{f}_{t+1}) + 1(\hat{f}_{t+1} \leq Y_{t+1} < f_{t+1}^*)] = \int_{\theta^*W_t}^W f_t(y) \, dy
\]
\[
\leq |\hat{\theta}_t^*W_t - \theta^*W_t| \cdot \sup_{y \in \mathbb{R}} f_t(y)
\]
\[
\leq ||\hat{\theta}_t - \theta^*|| \cdot ||W_t|| \cdot \sup_{y \in \mathbb{R}} f_t(y).
\]

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

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

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

CASE \( p_0 = 2 \): we now have, by the triangular and Cauchy-Schwartz inequalities,

\[
||h^* - h_0|| = ||E[V_t \cdot (|Y_{t+1} - \hat{f}_{t+1}| - |Y_{t+1} - f_{t+1}^*|)]||
\]
\[
\leq ||E[V_t \cdot |f_{t+1}^* - \hat{f}_{t+1}|]||
\]
\[
\leq E[||V_t|| \cdot ||W_t|| \cdot ||\hat{\theta}_t - \theta^*||]
\]
\[
\leq m^2 \cdot E[||\hat{\theta}_t - \theta^*||],
\]

so that by the earlier argument, \( ||h^* - h_0|| \xrightarrow{p} 0 \). For this case,

\[
||g^* - g_0|| = ||E[V_t \cdot (1(Y_{t+1} - \hat{f}_{t+1} < 0) \cdot |Y_{t+1} - \hat{f}_{t+1}| - 1(Y_{t+1} - f_{t+1}^* < 0) \cdot |Y_{t+1} - f_{t+1}^*|)]||
\]
\[
= ||E[V_t \cdot (1(f_{t+1}^* \leq Y_{t+1} < \hat{f}_{t+1}) \cdot |Y_{t+1} - \hat{f}_{t+1}| - 1(\hat{f}_{t+1} \leq Y_{t+1} < f_{t+1}^* \cdot |Y_{t+1} - f_{t+1}^*|)]||
\]
\[
\leq E[||V_t|| \cdot (1(f_{t+1}^* \leq Y_{t+1} < \hat{f}_{t+1}) \cdot |Y_{t+1} - \hat{f}_{t+1}| - 1(\hat{f}_{t+1} \leq Y_{t+1} < f_{t+1}^* \cdot |Y_{t+1} - f_{t+1}^*|)]
\]
\[
\leq m \cdot (E[1(f_{t+1}^* \leq Y_{t+1} < \hat{f}_{t+1}) \cdot |Y_{t+1} - \hat{f}_{t+1}|] + E[1(\hat{f}_{t+1} \leq Y_{t+1} < f_{t+1}^* \cdot |Y_{t+1} - f_{t+1}^*|)]).
\]

By the Cauchy-Schwartz inequality and Assumptions (A4) and (A5)

\[
E[1(\hat{f}_{t+1}^* \leq Y_{t+1} < \hat{f}_{t+1}) \cdot |Y_{t+1} - \hat{f}_{t+1}|] \leq (E[1(f_{t+1}^* \leq Y_{t+1} < \hat{f}_{t+1})])^{1/2} \cdot (E[(Y_{t+1} - \hat{f}_{t+1})^2])^{1/2}
\]
\[
\leq (E[1(f_{t+1}^* \leq Y_{t+1} < \hat{f}_{t+1})])^{1/2} \cdot \max(1, \Delta_Y^{1/2}).
\]

As previously, by Assumptions (A3) and (A6) we have

\[
E[1(f_{t+1}^* \leq Y_{t+1} < \hat{f}_{t+1})] \leq m \cdot M \cdot E[||\hat{\theta}_t - \theta^*||]
\]
so that
\[ ||g^* - g_0||^2 \leq m^3 \cdot M \cdot \max(1, \Delta_Y) \cdot E[||\hat{\theta}_t - \theta^*||], \]
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 this proof we use Assumptions (A0)-(A4), (A5′) and (A6). Note that Assumption (A5′) is very similar to Assumption (A5), except for the mixing and moment conditions which are stronger than previously. We now show that \( T^{1/2}(\hat{\alpha}_T - \alpha_0) \) is asymptotically normal by expanding the first order condition for \( \hat{\alpha}_T \) around \( \alpha_0 \):

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

The idea then is to show that
\[
\hat{h}'_T \cdot \hat{S}^{-1} \cdot (\hat{g}_T - \hat{h}_T \cdot \alpha_0) = \bar{h}'_T \cdot \bar{S}^{-1} \cdot (\bar{g}_T - \bar{h}_T \cdot \alpha_0) + o_p(1),
\]

where we used the following notation:
\[
\bar{g}_t \equiv v_t 1(y_{t+1} - f_{t+1}^* < 0)|y_{t+1} - f_{t+1}^*|^{p_0-1},
\]
\[
\bar{g}_T \equiv T^{-1} \sum_{t=\tau}^{T+\tau-1} \bar{g}_t,
\]
and
\[
\bar{h}_t \equiv v_t |y_{t+1} - f_{t+1}^*|^{p_0-1},
\]
\[
\bar{h}_T \equiv T^{-1} \sum_{t=\tau}^{T+\tau-1} \bar{h}_t.
\]

We now show that (18) holds: by the triangle inequality we have
\[
||\hat{h}'_T \cdot \hat{S}^{-1} \cdot (\hat{g}_T - \hat{h}_T \cdot \alpha_0) - \bar{h}'_T \cdot \bar{S}^{-1} \cdot (\bar{g}_T - \bar{h}_T \cdot \alpha_0)|| \leq ||\hat{h}'_T \cdot \hat{S}^{-1} \cdot (\hat{g}_T - \bar{g}_T)||
+ ||(\hat{h}_T - \bar{h}_T)' \cdot \hat{S}^{-1} \cdot \hat{h}_T||
+ ||(\hat{h}_T - \bar{h}_T)' \cdot \bar{S}^{-1} \cdot \bar{h}_T \cdot \alpha_0||
+ ||\bar{h}'_T \cdot \bar{S}^{-1} \cdot (\bar{h}_T - \bar{h}_T) \cdot \alpha_0||
\]

38
so that
\[ ||\hat{h}_t \cdot \hat{S}^{-1} \cdot (\hat{g}_T - \hat{h}_T \cdot \alpha_0) - \bar{h}_t \cdot \bar{S}^{-1} \cdot (\bar{g}_T - \bar{h}_T \cdot \alpha_0)|| \leq ||\hat{h}_t \cdot \hat{S}^{-1}|| \cdot ||\hat{g}_T - \bar{g}_T|| + (||\hat{S}^{-1} \cdot \bar{g}_T|| + \alpha_0||\hat{S}^{-1} \cdot \hat{h}_T|| + \alpha_0||\bar{h}_t \cdot \hat{S}^{-1}||) \cdot ||\hat{h}_T - \bar{h}_T||.\]

Now note that \( ||\hat{g}_T - \bar{g}_T|| \leq ||\hat{g}_T - g_0|| + ||\bar{g}_T - g_0||. \) From the previous proof we know 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 line of argument, we can show that \( ||\hat{h}_T - \bar{h}_T|| \xrightarrow{p} 0 \). Therefore, by using 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 (18) holds.

Next we use the central limit theorem (CLT) for \( \alpha \)-mixing sequences (e.g., Theorem 5.20 in White, 2001) to show that \( T^{1/2}(\hat{g}_T - \bar{h}_T \cdot \alpha_0) \xrightarrow{d} \mathcal{N}(0, S) \). We first need to show that \( \{H_t\} \), where \( H_t \equiv V_t|Y_{t+1} - f^*_t|^{p_0-1} \), and \( \{G_t\} \), where \( G_t \equiv V_t1(Y_{t+1} - f^*_t < 0)|Y_{t+1} - f^*_t|^{p_0-1} \), are \( \alpha \)-mixing with mixing coefficient of size \(-r/(r-2), r > 2\). This follows directly from assumption (A5') and Theorem 3.49 in White (2001), which shows that measurable functions of mixing sequences are mixing of the same size. Hence \( \{G_t - \alpha_0H_t\} \) is \( \alpha \)-mixing with mixing coefficient of size \(-r/(r-2), r > 2.\ In order to apply the CLT we need to ensure that for some \( \Delta > 0 \)

\[ E[||G_t - \alpha_0H_t||^r] < \Delta < \infty. \]

Note that the Cauchy-Schwartz inequality and Assumption (A5') imply
\[ E[||G_t - \alpha_0H_t||^r] = E[||V_t \cdot (1(Y_{t+1} - f^*_t < 0) - \alpha_0) \cdot |Y_{t+1} - f^*_t|^{p_0-1}||^r] \leq \Delta < \infty. \]

for \( \Delta \equiv \max(1, \Delta_{W}^{1/2}) \cdot \max(1, \Delta_{Y}^{1/2}) > 0, \Delta < \infty. \) The CLT (e.g., Theorem 5.20 in White, 2001) then ensures \( T^{1/2}(\hat{g}_T - \bar{h}_T \cdot \alpha_0) \xrightarrow{d} \mathcal{N}(0, S) \) so that \( T^{1/2}[\hat{h}_t \cdot \hat{S}^{-1} \cdot (\hat{g}_T - \bar{h}_T \cdot \alpha_0)] \xrightarrow{d} \mathcal{N}(0, h_0' \cdot S^{-1} \cdot h_0). \) Together with (18) this implies (by Slutsky’s theorem)
\[ T^{1/2}[\hat{h}_t \cdot \hat{S}^{-1} \cdot (\hat{g}_T - \bar{h}_T \cdot \alpha_0)] \xrightarrow{d} \mathcal{N}(0, h_0' \cdot S^{-1} \cdot h_0). \]
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 (17) is equivalent to $T^{1/2}(\hat{\alpha}_T - \alpha_0) = [\hat{h}'_T \cdot \hat{S}^{-1} \cdot \hat{h}_T]^{-1} T^{1/2} [\hat{g}'_T \cdot \hat{S}^{-1} \cdot (\hat{g}_T - \hat{h}_T \cdot \alpha_0)]$. We then use the limit result in (19) and the Slutsky 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


### Table 1: Size of t-tests using only a constant as instrument (nominal size 5%, two-sided test)

<table>
<thead>
<tr>
<th>$N_0$</th>
<th>$n_0$</th>
<th>$n_f$</th>
<th>$\alpha$</th>
<th>$r_0$</th>
<th>$p_0$</th>
<th>$p_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>50</td>
<td>0.053</td>
<td>0.059</td>
<td>0.061</td>
<td>0.064</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>50</td>
<td>0.066</td>
<td>0.050</td>
<td>0.057</td>
<td>0.052</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>50</td>
<td>0.056</td>
<td>0.056</td>
<td>0.066</td>
<td>0.066</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>50</td>
<td>0.063</td>
<td>0.055</td>
<td>0.057</td>
<td>0.052</td>
</tr>
<tr>
<td>100</td>
<td>200</td>
<td>50</td>
<td>0.061</td>
<td>0.053</td>
<td>0.053</td>
<td>0.055</td>
</tr>
</tbody>
</table>

Quad-Quad ($p_0 = 2$)

<table>
<thead>
<tr>
<th>$N_0$</th>
<th>$n_0$</th>
<th>$n_f$</th>
<th>$\alpha$</th>
<th>$r_0$</th>
<th>$p_0$</th>
<th>$p_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>50</td>
<td>0.065</td>
<td>0.069</td>
<td>0.071</td>
<td>0.071</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>50</td>
<td>0.082</td>
<td>0.062</td>
<td>0.063</td>
<td>0.061</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>50</td>
<td>0.063</td>
<td>0.068</td>
<td>0.072</td>
<td>0.072</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>50</td>
<td>0.077</td>
<td>0.057</td>
<td>0.059</td>
<td>0.057</td>
</tr>
<tr>
<td>100</td>
<td>200</td>
<td>50</td>
<td>0.127</td>
<td>0.055</td>
<td>0.057</td>
<td>0.052</td>
</tr>
</tbody>
</table>

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.

### Table 2: Size of t-tests using two instruments (nominal size 5%, two-sided test)

<table>
<thead>
<tr>
<th>$n_0$</th>
<th>$n_f$</th>
<th>$\alpha$</th>
<th>$r_0$</th>
<th>$p_0$</th>
<th>$p_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>0.144</td>
<td>0.094</td>
<td>0.078</td>
<td>0.090</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>0.090</td>
<td>0.073</td>
<td>0.064</td>
<td>0.072</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>0.162</td>
<td>0.095</td>
<td>0.083</td>
<td>0.091</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>0.096</td>
<td>0.076</td>
<td>0.063</td>
<td>0.068</td>
</tr>
<tr>
<td>100</td>
<td>200</td>
<td>0.077</td>
<td>0.065</td>
<td>0.055</td>
<td>0.063</td>
</tr>
</tbody>
</table>

Quad-Quad ($p_0 = 2$)

<table>
<thead>
<tr>
<th>$n_0$</th>
<th>$n_f$</th>
<th>$\alpha$</th>
<th>$r_0$</th>
<th>$p_0$</th>
<th>$p_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>0.105</td>
<td>0.121</td>
<td>0.118</td>
<td>0.120</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>0.076</td>
<td>0.083</td>
<td>0.087</td>
<td>0.085</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>0.109</td>
<td>0.120</td>
<td>0.116</td>
<td>0.121</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>0.080</td>
<td>0.077</td>
<td>0.080</td>
<td>0.080</td>
</tr>
<tr>
<td>100</td>
<td>200</td>
<td>0.104</td>
<td>0.066</td>
<td>0.069</td>
<td>0.066</td>
</tr>
</tbody>
</table>
Table 3: Size of j-tests for overidentification using two instruments (nominal size 5%)

<table>
<thead>
<tr>
<th>Lin-Lin ($p_0 = 1$)</th>
<th>$\alpha_o$</th>
<th>0.2</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_0$</td>
<td>$n_f$</td>
<td>50</td>
<td>50</td>
<td>0.029</td>
<td>0.047</td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>100</td>
<td>0.044</td>
<td>0.048</td>
<td>0.047</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>50</td>
<td>0.033</td>
<td>0.047</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>100</td>
<td>0.041</td>
<td>0.052</td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>200</td>
<td>0.049</td>
<td>0.047</td>
<td>0.048</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Quad-Quad ($p_0 = 2$)</th>
<th>$\alpha_o$</th>
<th>0.2</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_0$</td>
<td>$n_f$</td>
<td>50</td>
<td>50</td>
<td>0.020</td>
<td>0.038</td>
<td>0.043</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>100</td>
<td>0.032</td>
<td>0.044</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>50</td>
<td>0.026</td>
<td>0.042</td>
<td>0.041</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>100</td>
<td>0.033</td>
<td>0.049</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>200</td>
<td>0.030</td>
<td>0.046</td>
<td>0.051</td>
</tr>
</tbody>
</table>

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.
<table>
<thead>
<tr>
<th></th>
<th>IMF</th>
<th>OECD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Canada</td>
<td>France</td>
</tr>
<tr>
<td><strong>Current year</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSE</td>
<td>0.87</td>
<td>0.49</td>
</tr>
<tr>
<td>MAE</td>
<td>0.70</td>
<td>0.38</td>
</tr>
<tr>
<td>Mean</td>
<td>-0.24</td>
<td>-0.09</td>
</tr>
<tr>
<td>(n^+)</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>(n^-)</td>
<td>15</td>
<td>13</td>
</tr>
<tr>
<td>(N)</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td><strong>1-year ahead</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSE</td>
<td>1.14</td>
<td>0.74</td>
</tr>
<tr>
<td>MAE</td>
<td>0.88</td>
<td>0.53</td>
</tr>
<tr>
<td>Mean</td>
<td>-0.32</td>
<td>-0.17</td>
</tr>
<tr>
<td>(n^+)</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td>(n^-)</td>
<td>13</td>
<td>12</td>
</tr>
</tbody>
</table>

Note: For each country this table shows the bias of the forecast error, the number of positive \((n^+)\) and negative \((n^-)\) forecast errors, their sum \((N)\), the mean absolute forecast error \((MAE)\) and the root mean squared error \((RMSE)\).
Table 5: Parameter Estimates Under Lin-lin Loss and Tests of Symmetry

<table>
<thead>
<tr>
<th>Inst=1</th>
<th>Lin-Lin</th>
<th>IMF</th>
<th>OECD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Current year</td>
<td></td>
<td></td>
</tr>
<tr>
<td>α</td>
<td>0.60</td>
<td>0.52</td>
<td>0.40</td>
</tr>
<tr>
<td>s.e.</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>p-value</td>
<td>0.31</td>
<td>0.84</td>
<td>0.31</td>
</tr>
<tr>
<td>Inst=2</td>
<td>α</td>
<td>0.58</td>
<td>0.54</td>
</tr>
<tr>
<td>s.e.</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>p-value</td>
<td>0.39</td>
<td>0.67</td>
<td>0.33</td>
</tr>
<tr>
<td>Inst=3</td>
<td>α</td>
<td>0.59</td>
<td>0.54</td>
</tr>
<tr>
<td>s.e.</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>p-value</td>
<td>0.39</td>
<td>0.68</td>
<td>0.40</td>
</tr>
<tr>
<td>Inst=4</td>
<td>α</td>
<td>0.59</td>
<td>0.54</td>
</tr>
<tr>
<td>s.e.</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>p-value</td>
<td>0.39</td>
<td>0.67</td>
<td>0.30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Inst=1</th>
<th>1-year ahead</th>
<th>IMF</th>
<th>OECD</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td>0.54</td>
<td>0.48</td>
<td>0.44</td>
</tr>
<tr>
<td>s.e.</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>p-value</td>
<td>0.68</td>
<td>0.84</td>
<td>0.55</td>
</tr>
<tr>
<td>Inst=2</td>
<td>α</td>
<td>0.54</td>
<td>0.50</td>
</tr>
<tr>
<td>s.e.</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>p-value</td>
<td>0.67</td>
<td>0.96</td>
<td>0.62</td>
</tr>
<tr>
<td>Inst=3</td>
<td>α</td>
<td>0.57</td>
<td>0.50</td>
</tr>
<tr>
<td>s.e.</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>p-value</td>
<td>0.50</td>
<td>1.00</td>
<td>0.66</td>
</tr>
<tr>
<td>Inst=4</td>
<td>α</td>
<td>0.57</td>
<td>0.50</td>
</tr>
<tr>
<td>s.e.</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>p-value</td>
<td>0.48</td>
<td>0.98</td>
<td>0.57</td>
</tr>
</tbody>
</table>
Table 6: Tests of the Joint Hypothesis of Symmetric Lin-lin Loss and Forecast Rationality

<table>
<thead>
<tr>
<th>Inst</th>
<th>Lin-Lin</th>
<th>IMF</th>
<th>OECD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Current year</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Canada</td>
<td>France</td>
</tr>
<tr>
<td>Inst=1</td>
<td>j-stat</td>
<td>1.04</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>0.31</td>
<td>0.84</td>
</tr>
<tr>
<td>Inst=2</td>
<td>j-stat</td>
<td>0.82</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>0.67</td>
<td>0.81</td>
</tr>
<tr>
<td>Inst=3</td>
<td>j-stat</td>
<td>0.91</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>0.63</td>
<td>0.92</td>
</tr>
<tr>
<td>Inst=4</td>
<td>j-stat</td>
<td>0.93</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>0.82</td>
<td>0.94</td>
</tr>
</tbody>
</table>

1-year ahead

<table>
<thead>
<tr>
<th>Inst</th>
<th>Lin-Lin</th>
<th>IMF</th>
<th>OECD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Current year</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Canada</td>
<td>France</td>
</tr>
<tr>
<td>Inst=1</td>
<td>j-stat</td>
<td>0.17</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>0.68</td>
<td>0.84</td>
</tr>
<tr>
<td>Inst=2</td>
<td>j-stat</td>
<td>3.13</td>
<td>2.58</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>0.21</td>
<td>0.28</td>
</tr>
<tr>
<td>Inst=3</td>
<td>j-stat</td>
<td>4.06</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>0.13</td>
<td>0.95</td>
</tr>
<tr>
<td>Inst=4</td>
<td>j-stat</td>
<td>4.24</td>
<td>3.06</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>0.24</td>
<td>0.38</td>
</tr>
</tbody>
</table>

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 7: Test of Forecast Rationality, allowing for Asymmetric Lin-lin Loss

<table>
<thead>
<tr>
<th>Inst</th>
<th>Lin-Lin</th>
<th>Canada</th>
<th>France</th>
<th>Germany</th>
<th>Italy</th>
<th>Japan</th>
<th>UK</th>
<th>US</th>
<th>France</th>
<th>Germany</th>
<th>Italy</th>
<th>UK</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>j-stat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.12</td>
<td>0.24</td>
<td>1.72</td>
<td>0.79</td>
<td>2.17</td>
<td>1.66</td>
<td>1.18</td>
<td>0.01</td>
<td>0.55</td>
<td>1.59</td>
<td>1.52</td>
<td></td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>0.73</td>
<td>0.63</td>
<td>0.19</td>
<td>0.37</td>
<td>0.14</td>
<td>0.20</td>
<td>0.28</td>
<td>0.94</td>
<td>0.46</td>
<td>0.21</td>
<td>0.22</td>
</tr>
<tr>
<td>3</td>
<td>0.20</td>
<td>0.00</td>
<td>0.06</td>
<td>0.62</td>
<td>0.23</td>
<td>0.42</td>
<td>0.52</td>
<td>5.65</td>
<td>3.81</td>
<td>0.89</td>
<td>0.42</td>
<td></td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>0.66</td>
<td>0.99</td>
<td>0.81</td>
<td>0.43</td>
<td>0.63</td>
<td>0.52</td>
<td>0.47</td>
<td>0.02</td>
<td>0.05</td>
<td>0.35</td>
<td>0.52</td>
</tr>
<tr>
<td>4</td>
<td>0.21</td>
<td>0.24</td>
<td>2.46</td>
<td>1.27</td>
<td>1.97</td>
<td>2.20</td>
<td>1.49</td>
<td>6.56</td>
<td>4.12</td>
<td>2.13</td>
<td>2.50</td>
<td></td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>0.90</td>
<td>0.89</td>
<td>0.29</td>
<td>0.53</td>
<td>0.37</td>
<td>0.33</td>
<td>0.48</td>
<td>0.04</td>
<td>0.13</td>
<td>0.34</td>
<td>0.29</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Inst</th>
<th>1-year ahead</th>
<th>j-stat</th>
<th>p-value</th>
<th>j-stat</th>
<th>p-value</th>
<th>j-stat</th>
<th>p-value</th>
<th>j-stat</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2.96</td>
<td>2.57</td>
<td>0.16</td>
<td>0.25</td>
<td>0.56</td>
<td>0.56</td>
<td>0.04</td>
<td>0.45</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>0.09</td>
<td>0.11</td>
<td>0.61</td>
<td>0.01</td>
<td>0.07</td>
<td>0.45</td>
<td>0.81</td>
<td>0.83</td>
<td>0.09</td>
</tr>
<tr>
<td>3</td>
<td>3.62</td>
<td>0.10</td>
<td>0.38</td>
<td>0.08</td>
<td>0.58</td>
<td>2.23</td>
<td>2.90</td>
<td>0.09</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>0.06</td>
<td>0.75</td>
<td>0.54</td>
<td>0.78</td>
<td>0.45</td>
<td>0.13</td>
<td>0.09</td>
<td>0.76</td>
<td>0.09</td>
</tr>
<tr>
<td>4</td>
<td>3.76</td>
<td>3.06</td>
<td>3.01</td>
<td>0.33</td>
<td>4.77</td>
<td>4.28</td>
<td>0.19</td>
<td>4.71</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>0.15</td>
<td>0.22</td>
<td>0.22</td>
<td>0.85</td>
<td>0.09</td>
<td>0.12</td>
<td>0.91</td>
<td>0.09</td>
<td>0.22</td>
</tr>
</tbody>
</table>

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 8: Parameter Estimates Under Quad-Quad Loss and Tests of Symmetry

<table>
<thead>
<tr>
<th></th>
<th>IMF</th>
<th>OECD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Quad-Quad Current year</td>
<td></td>
</tr>
<tr>
<td>Inst=1 α</td>
<td>0.67 0.62 0.43 0.24 0.27 0.18</td>
<td>0.27 0.34 0.34 0.22 0.24 0.43</td>
</tr>
<tr>
<td>s.e.</td>
<td>0.11 0.12 0.13 0.13 0.14 0.07</td>
<td>0.11 0.13 0.14 0.13 0.11 0.11</td>
</tr>
<tr>
<td>p-value</td>
<td>0.11 0.33 0.61 0.06 0.11 0.00</td>
<td>0.20 0.00 0.00 0.04 0.52 0.52</td>
</tr>
<tr>
<td>Inst=2 α</td>
<td>0.67 0.62 0.46 0.19 0.02 0.11</td>
<td>0.21 0.35 0.20 0.16 0.45 0.45</td>
</tr>
<tr>
<td>s.e.</td>
<td>0.11 0.12 0.13 0.12 0.05 0.06</td>
<td>0.10 0.13 0.09 0.10 0.11 0.11</td>
</tr>
<tr>
<td>p-value</td>
<td>0.12 0.28 0.76 0.01 0.00 0.00</td>
<td>0.26 0.00 0.00 0.00 0.63 0.63</td>
</tr>
<tr>
<td>Inst=3 α</td>
<td>0.68 0.59 0.46 0.07 0.29 0.19</td>
<td>0.19 0.22 0.11 0.20 0.48 0.48</td>
</tr>
<tr>
<td>s.e.</td>
<td>0.11 0.12 0.13 0.09 0.13 0.08</td>
<td>0.10 0.11 0.06 0.12 0.11 0.11</td>
</tr>
<tr>
<td>p-value</td>
<td>0.09 0.47 0.79 0.00 0.10 0.00</td>
<td>0.00 0.01 0.00 0.01 0.87 0.87</td>
</tr>
<tr>
<td>Inst=4 α</td>
<td>0.68 0.58 0.46 0.07 0.12 0.12</td>
<td>0.13 0.05 0.10 0.13 0.46 0.46</td>
</tr>
<tr>
<td>s.e.</td>
<td>0.11 0.12 0.13 0.09 0.10 0.07</td>
<td>0.08 0.07 0.06 0.08 0.11 0.11</td>
</tr>
<tr>
<td>p-value</td>
<td>0.08 0.47 0.74 0.00 0.00 0.00</td>
<td>0.00 0.00 0.00 0.00 0.69 0.69</td>
</tr>
<tr>
<td></td>
<td>1-year ahead</td>
<td></td>
</tr>
<tr>
<td>Inst=1 α</td>
<td>0.68 0.66 0.42 0.20 0.39 0.34</td>
<td>0.30 0.55 0.26 0.42 0.51 0.51</td>
</tr>
<tr>
<td>s.e.</td>
<td>0.11 0.12 0.12 0.11 0.14 0.12</td>
<td>0.11 0.12 0.09 0.13 0.12 0.12</td>
</tr>
<tr>
<td>p-value</td>
<td>0.11 0.19 0.51 0.01 0.42 0.19</td>
<td>0.06 0.71 0.01 0.52 0.94 0.94</td>
</tr>
<tr>
<td>Inst=2 α</td>
<td>0.72 0.74 0.45 0.06 0.02 0.20</td>
<td>0.20 0.53 0.27 0.37 0.49 0.49</td>
</tr>
<tr>
<td>s.e.</td>
<td>0.11 0.11 0.12 0.03 0.04 0.10</td>
<td>0.10 0.12 0.10 0.13 0.12 0.12</td>
</tr>
<tr>
<td>p-value</td>
<td>0.04 0.03 0.66 0.00 0.00 0.00</td>
<td>0.00 0.77 0.02 0.32 0.92 0.92</td>
</tr>
<tr>
<td>Inst=3 α</td>
<td>0.77 0.71 0.46 0.06 0.40 0.37</td>
<td>0.30 0.54 0.19 0.37 0.53 0.53</td>
</tr>
<tr>
<td>s.e.</td>
<td>0.10 0.12 0.13 0.04 0.13 0.12</td>
<td>0.11 0.12 0.08 0.13 0.11 0.11</td>
</tr>
<tr>
<td>p-value</td>
<td>0.00 0.08 0.77 0.00 0.45 0.29</td>
<td>0.07 0.72 0.00 0.32 0.82 0.82</td>
</tr>
<tr>
<td>Inst=4 α</td>
<td>0.78 0.72 0.44 0.05 0.21 0.22</td>
<td>0.18 0.46 0.17 0.37 0.51 0.51</td>
</tr>
<tr>
<td>s.e.</td>
<td>0.10 0.12 0.12 0.04 0.11 0.10</td>
<td>0.09 0.12 0.08 0.12 0.11 0.11</td>
</tr>
<tr>
<td>p-value</td>
<td>0.00 0.06 0.63 0.00 0.01 0.01</td>
<td>0.00 0.71 0.00 0.28 0.92 0.92</td>
</tr>
</tbody>
</table>
Table 9: Tests of the Joint Hypothesis of MSE Loss and Forecast Rationality

<table>
<thead>
<tr>
<th>IMF</th>
<th>OECD</th>
<th>Quad-Quad</th>
<th>Canada</th>
<th>France</th>
<th>Germany</th>
<th>Italy</th>
<th>Japan</th>
<th>UK</th>
<th>US</th>
<th>France</th>
<th>Germany</th>
<th>Italy</th>
<th>UK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current year</td>
<td></td>
<td>Inst=1</td>
<td>j-stat</td>
<td>2.49</td>
<td>0.95</td>
<td>0.27</td>
<td>3.66</td>
<td>2.51</td>
<td>18.81</td>
<td>4.56</td>
<td>1.64</td>
<td>9.65</td>
<td>4.20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p-value</td>
<td>0.11</td>
<td>0.33</td>
<td>0.61</td>
<td>0.06</td>
<td>0.11</td>
<td>0.00</td>
<td>0.03</td>
<td>0.20</td>
<td>0.00</td>
<td>0.04</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inst=2</td>
<td>j-stat</td>
<td>2.37</td>
<td>1.10</td>
<td>1.12</td>
<td>6.53</td>
<td>77.19</td>
<td>44.75</td>
<td>10.14</td>
<td>1.43</td>
<td>11.44</td>
<td>12.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p-value</td>
<td>0.31</td>
<td>0.58</td>
<td>0.57</td>
<td>0.04</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.49</td>
<td>0.00</td>
<td>0.00</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inst=3</td>
<td>j-stat</td>
<td>2.73</td>
<td>2.41</td>
<td>0.25</td>
<td>24.97</td>
<td>2.63</td>
<td>14.46</td>
<td>12.04</td>
<td>9.00</td>
<td>40.33</td>
<td>6.59</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p-value</td>
<td>0.25</td>
<td>0.30</td>
<td>0.88</td>
<td>0.00</td>
<td>0.27</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.04</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inst=4</td>
<td>j-stat</td>
<td>3.20</td>
<td>2.41</td>
<td>1.23</td>
<td>25.72</td>
<td>18.69</td>
<td>31.81</td>
<td>24.66</td>
<td>41.67</td>
<td>42.09</td>
<td>21.39</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p-value</td>
<td>0.36</td>
<td>0.49</td>
<td>0.75</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.23</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IMF</th>
<th>OECD</th>
<th>1-year ahead</th>
<th>j-stat</th>
<th>2.52</th>
<th>1.72</th>
<th>0.44</th>
<th>6.86</th>
<th>0.65</th>
<th>1.74</th>
<th>3.44</th>
<th>0.14</th>
<th>6.65</th>
<th>0.42</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>p-value</td>
<td>0.11</td>
<td>0.19</td>
<td>0.51</td>
<td>0.01</td>
<td>0.42</td>
<td>0.19</td>
<td>0.06</td>
<td>0.71</td>
<td>0.01</td>
<td>0.52</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inst=2</td>
<td>j-stat</td>
<td>6.56</td>
<td>7.38</td>
<td>1.01</td>
<td>157.83</td>
<td>116.64</td>
<td>12.34</td>
<td>12.98</td>
<td>2.26</td>
<td>5.54</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p-value</td>
<td>0.04</td>
<td>0.02</td>
<td>0.60</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.32</td>
<td>0.06</td>
<td>0.61</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inst=3</td>
<td>j-stat</td>
<td>11.17</td>
<td>4.26</td>
<td>0.10</td>
<td>143.02</td>
<td>0.93</td>
<td>1.16</td>
<td>3.35</td>
<td>0.45</td>
<td>16.76</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p-value</td>
<td>0.00</td>
<td>0.12</td>
<td>0.95</td>
<td>0.00</td>
<td>0.63</td>
<td>0.56</td>
<td>0.19</td>
<td>0.80</td>
<td>0.00</td>
<td>0.62</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p-value</td>
<td>0.01</td>
<td>0.04</td>
<td>0.75</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.05</td>
<td>0.00</td>
<td>0.76</td>
<td>0.15</td>
</tr>
</tbody>
</table>

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>
<thead>
<tr>
<th>Quad-Quad</th>
<th>IMF</th>
<th>OECD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Current year</td>
<td></td>
</tr>
<tr>
<td>Inst=2</td>
<td>j-stat 0.11  0.00 1.03 0.58 2.90 2.66 1.67 0.19 0.95 1.55 4.11</td>
<td>p-value 0.74 0.98 0.31 0.45 0.09 0.10 0.20 0.66 0.33 0.21 0.04</td>
</tr>
<tr>
<td></td>
<td>j-stat 0.05 1.89 0.18 2.96 0.03 0.07 1.91 3.24 2.93 0.87 0.02</td>
<td>p-value 0.83 0.17 0.67 0.09 0.85 0.79 0.17 0.07 0.09 0.35 0.89</td>
</tr>
<tr>
<td>Inst=4</td>
<td>j-stat 0.32 1.91 1.13 2.98 4.07 3.05 3.30 6.85 3.14 1.53 4.19</td>
<td>p-value 0.85 0.39 0.57 0.23 0.13 0.22 0.19 0.03 0.21 0.47 0.12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1-year ahead</th>
<th>IMF</th>
<th>OECD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inst=2</td>
<td>j-stat 2.58 3.09 0.82 1.95 4.34 3.79 3.66 2.18 0.22 0.02 3.75</td>
<td>p-value 0.11 0.08 0.36 0.16 0.04 0.05 0.06 0.14 0.64 0.88 0.05</td>
</tr>
<tr>
<td>Inst=3</td>
<td>j-stat 3.52 1.24 0.02 1.91 0.39 0.08 0.17 0.33 2.33 0.00 0.03</td>
<td>p-value 0.06 0.27 0.90 0.17 0.53 0.78 0.68 0.57 0.13 0.96 0.86</td>
</tr>
<tr>
<td>Inst=4</td>
<td>j-stat 3.51 4.83 0.98 2.49 7.62 5.48 4.39 7.48 4.02 0.04 5.25</td>
<td>p-value 0.17 0.09 0.61 0.29 0.02 0.06 0.11 0.02 0.13 0.98 0.07</td>
</tr>
</tbody>
</table>

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.