Biases in Macroeconomic Forecasts: Irrationality or Asymmetric Loss?*

Graham Elliott
University of California San Diego

Ivana Komunjer
Caltech

Allan Timmermann
University of California San Diego

June 24, 2004

Abstract

Empirical studies using survey data on expectations have frequently observed that forecasts are biased and have concluded that agents are not rational. We establish that existing rationality tests are not robust to even small deviations from symmetric loss and hence have little ability to tell whether the forecaster is irrational or the loss function is asymmetric. We quantify the exact trade-off between forecast inefficiency and asymmetric loss leading to identical outcomes of standard rationality tests and explore new and more general methods for testing forecast rationality jointly with flexible families of loss functions that embed quadratic loss as a special case. An empirical application to survey data on forecasts of nominal output growth demonstrates the empirical significance of our results.

*Graham Elliott and Allan Timmermann are grateful to the NSF for financial assistance under grant SES 0111238. Carlos Capistran provided excellent research assistance. We thank Ken Arrow, Dean Croushore, Ed Schlee, Clive Granger, Mordecai Kurz, Preston McAfee, Adrian Pagan, Hal White and seminar participants at the NSF/Penn conference 2003, UTS, UNSW, Rice, Texas A&M, U. Texas - Austin, Caltech, USC, UCLA, Duke and Stanford for insightful comments.

1 Introduction

How agents form expectations and, in particular, whether they are rational and efficiently incorporate all available information into their forecasts, is a question of fundamental importance in economic analysis. Ultimately this question can best be resolved through empirical analysis of expectations data. It is therefore not surprising that a large literature has been devoted to empirically testing forecast rationality based on survey data such as the Livingston data or the Survey of Professional Forecasters.[1] Summarizing this literature, Conlisk (1996, page 672) concluded that "*Survey data on expectations of inflation and other variables commonly reject the unbiasedness and efficiency prediction of rational expectations.*"

The vast majority of studies has tested forecast rationality in conjunction with an assumption of mean squared error (MSE) loss. This symmetric loss function has largely been maintained out of convenience: under MSE loss rationality implies that the observed forecast errors should have zero mean and be uncorrelated with all variables in the current information set. Yet, a reading of the literature reveals little discussion of why loss should be symmetric in the forecast error. One would, if anything, typically expect asymmetric loss as a reflection of the primitive economic conditions of the problem such as, e.g., asymmetric stockout and inventory holding costs.[2] This is potentially important since relaxing the symmetry assumption is known to profoundly change the properties of optimal forecasts, c.f. Christoffersen and Diebold (1997) and Patton and Timmermann (2002).

Many studies on forecast rationality testing are aware of the limitations of symmetric loss and indicate that rejections of rationality may be driven by asymmetries. For example, Keane and Runkle (1990, page 719) write "*If forecasters have differential costs of over- and underprediction, it could be rational for them to produce biased forecasts. If we were to find that forecasts are biased, it could still be claimed that forecasters were rational if it could be shown that they had such differential costs.*" Unfortunately, little is known about the magnitude of the problem - i.e. how much this really matters in practice. In this paper we

[1] See, e.g., Bonham and Cohen (1995), Fama (1975), Keane and Runkle (1990), Mankiw, Reis and Wolfers (2003), Mishkin (1981) and Zarnowitz (1985).

[2] Indeed, as far back as in the fifties economists distinguished between inventory and backordering costs, c.f. Arrow, Karlin and Scarf (1958).

therefore examine the theoretical and practical importance of the joint nature of tests for forecast rationality. We show that the coefficients in standard forecast efficiency tests are biased if the loss function is not symmetric and characterize this bias. Under asymmetric loss, standard rationality tests thus do not control size and may lead to false rejections of rationality. Conversely, even large inefficiencies in forecasters' use of information may not be detectable by standard tests when the true loss is asymmetric.

To demonstrate these points, we revisit the Survey of Professional Forecasters (SPF) data on US output growth and examine whether the apparently high rejection rate for rationality found in this data set can be explained by asymmetric loss. We find strong evidence of bias in the forecast errors of many individual survey participants. In fact, close to 30% of the individual predictions lead to rejections of the joint hypothesis of rationality and symmetric loss at the 5% critical level. Allowing for asymmetric loss, the rejection rate is very close to 5% which is consistent with rationality. Output forecasts thus tend to be consistent with rationality under asymmetric loss though not under symmetric loss. Furthermore, our estimates of the direction of asymmetries in loss overwhelmingly suggest that the cost of overpredicting exceeds the cost of underpredicting output growth.

The plan of the paper is as follows. Section 2 reviews the evidence against symmetric loss and rationality in forecasts of output growth from the Survey of Professional Forecasters. Section 3 adopts a flexible family of loss functions to examine standard tests for forecast rationality based on quadratic loss and shows how they can lead to biased estimates and wrong inference when loss is genuinely asymmetric. Construction of rationality tests under asymmetric loss is undertaken in Section 4, while Section 5 presents empirical results and Section 6 concludes. Technical proofs and details of the data set are provided in appendices at the end of the paper.

2 Bias in Forecasts of Output Growth

Our paper studies forecasts of US nominal output growth - a series in which virtually all macroeconomic forecasters should have some interest. Forecasts of output growth have been the subject of many previous studies. Brown and Maital (1981) studied average GNP forecasts and rejected unbiasedness and efficiency in six-month predictions of growth in GNP measured in current prices. Zarnowitz (1985) found only weak evidence against efficiency for the average forecast, but stronger evidence against efficiency for individual forecasters. Batchelor and Dua (1991) found little evidence that forecast errors were correlated with their own past values. In contrast, Davies and Lahiri (1995) conducted a panel analysis and found evidence that informational efficiency was rejected for up to half of the survey participants.

2.1 Data

The main data used in this paper is from the Survey of Professional Forecasters (SPF) which has become a primary source for studying macroeconomic forecasts.[3] Survey participants provide point forecasts of these variables in quarterly surveys. Surveys such as the SPF do not specify the objective of the forecasting exercise. This leaves open the question what the objective of the forecaster is. It is by no means clear that the forecaster simply minimizes a quadratic loss function and reports the conditional mean. For example, in a study of predictions of interest rates, Leitch and Tanner (1991) found that commercial forecasts performed very poorly according to an MSE criterion but did very well according to a sign prediction criterion linked more closely to profits from simple trading strategies based on these forecasts. Clearly, these forecasters did not use a quadratic loss function.

Survey participants are anonymous; their identity is only known to the data collectors and not made publicly available. It is plausible to expect that participants report the same forecasts that they use either for themselves or with their clients. Forecasts should therefore closely reflect the underlying loss function. Strategic behavior may also play a role and could induce bias as we briefly discuss below.

The SPF data set is an unbalanced panel. Although the sample begins in 1968, no forecaster participated throughout the entire sample. Each quarter some forecasters leave the sample and new ones are included. We therefore have very few observations on most individual forecasters. We deal with this problem by requiring each forecaster to have partic-

[3] For an academic bibliography, see the extensive list of references to papers that have used this data source maintained by the Federal Reserve Bank of Philadelphia at http://www.phil.frb.org/econ/spf/spfbib.html.

ipated for a minimum of 20 quarters. Imposing this requirement leaves us with 98 individual forecast series. The data appendix provides more details of the construction of the data.

Figure 1 shows a histogram of the average forecast errors across the 98 forecasters in the data set. The average forecast error, defined as the difference between the realized and predicted value, has a positive mean (0.16% per quarter). Out of 98 sets of forecast errors, 80 had a positive mean, suggesting systematic underpredictions of output growth.

2.2 Forecast Unbiasedness Tests

Under quadratic loss - often referred to as mean squared error (MSE) loss - forecast rationality has traditionally been studied by testing one of two conditions: (1) that the forecast under consideration is unbiased and (2) that it is efficient with respect to the information set available to the forecaster at the time the forecast was made.

Tests of forecast unbiasedness typically use the Mincer-Zarnowitz (1969) regression:

$$y_{t+1} = \beta_c + \beta f_{t+1} + u_{t+1}, \tag{1}$$

where y_{t+1} is the time $t+1$ realization of the target variable - US nominal output growth in our data - which we denote by Y_{t+1}, f_{t+1} is its one-step-ahead forecast and u_{t+1} is a realization of a scalar error random variable, U_{t+1}, satisfying $E[U_{t+1}] = 0$. Under the null hypothesis of zero bias we should have $\beta_c = 0$ and $\beta = 1$.

Table 1 shows the outcome of tests for bias in the forecast errors. Under quadratic loss, the null of no bias is rejected at the 1% critical level for 16 participants and gets rejected in 29 cases at the 5% level.[4] If MSE loss is accepted, this strongly questions rationality for a large proportion of the survey participants. On the other hand, if the forecasters incur different losses from over- and underprediction, it would be rational for them to produce biased forecasts.

[4]These numbers are a little higher than those reported by Zarnowitz (1985). This is likely to reflect our longer sample and our requirement of at least 20 observations which gives more power to the test.

3 Caveats in Rationality Tests under Quadratic Loss

In this section we study the behavior of standard tests of forecast rationality when the forecaster's loss function allows for asymmetries. To do so we require a setup that nests MSE loss as a special case and - in view of the small survey samples typically available - allows for asymmetries in a highly parsimonious way. To this end we follow Elliott, Komunjer and Timmermann (2003) and assume that the loss function only depends on the forecast error, $e_{t+1} = Y_{t+1} - f_{t+1}$, and belongs to the following two parameter family

$$L(e; \alpha, p) \equiv [\alpha + (1 - 2\alpha)1(e < 0)]e^p, \tag{2}$$

with a positive exponent p and an asymmetry parameter α, $0 < \alpha < 1$.

An attractive feature of the function in (2) is that it generalizes losses commonly used in the rationality testing literature. When $(\alpha, p) = (1/2, 2)$, loss is quadratic and (2) reduces to MSE loss. More generally, when $p = 2$ and $0 < \alpha < 1$, the family of losses L is piecewise quadratic and we call it 'Quad-Quad'. Similarly, when $p = 1$ and $0 < \alpha < 1$ we get the piecewise linear family of losses L, known as 'Lin-Lin', a special case of which is the absolute deviation or mean absolute error (MAE) loss, obtained when $(\alpha, p) = (1/2, 1)$. As α moves away from 1/2 in either direction the loss function becomes increasingly asymmetric.

Intentionally we do not take a stand on what generates asymmetries in agents' loss functions. One possibility is linked to production costs as when stockout and inventory costs differ. Another possibility is related to strategic behavior arising in situations where the forecaster's remuneration depends on factors other than the mean squared forecast error, c.f. Scharfstein and Stein (1990), Truman (1994) and Ehrbeck and Waldmann (1996). Common features of the models used by these authors is that forecasters differ by their ability to forecast, reflected by differences in the precision of their private signals, and that their main goal is to influence clients' assessment of their ability. Such objectives are common to business analysts or analysts employed by financial services firms such as investment banks, whose fees are directly related to their clients' assessment of analysts' forecasting ability. The main finding of these models is that, consistent with asymmetric loss, the forecasts need not reflect analysts' private information in an unbiased manner.

3.1 Misspecification Bias

Suppose that the exponent $p = 2$ in equation (2), so that the forecaster's loss function is piecewise quadratic or 'Quad-Quad',

$$L_2(e;\alpha) \equiv [\alpha + (1-2\alpha)1(e<0)]e^2. \tag{3}$$

L_2 is parametrized by a single shape or asymmetry parameter, α, whose true value, α_0, may be known or unknown to the forecast evaluator. This loss function offers an ideal framework to discuss how standard tests of rationality - derived under MSE loss ($\alpha_0 = 1/2$) - are affected if the true loss function is 'Quad-Quad' with $\alpha_0 \neq 1/2$.

As previously discussed, forecast errors should be unpredictable under MSE loss so it is common to test forecast rationality by means of the efficiency regression

$$e_{t+1} = \beta' v_t + u_{t+1}, \tag{4}$$

where e_{t+1} is the forecast error and v_t are the observations of a $d \times 1$ vector of variables (including a constant), denoted V_t, that are known to the forecaster at time t. Assuming that a sample of forecasts running from $t = \tau$ to $t = T+\tau-1$ is available, the regression (4) tests the orthogonality condition $E[\sum_{t=\tau}^{T+\tau-1} V_t U_{t+1}] = 0$, obtained under the assumption that the forecaster's loss is quadratic. If, in reality, the true loss function is 'Quad-Quad', the correct moment condition is $E[\sum_{t=\tau}^{T+\tau-1} V_t U_{t+1}] = (1-2\alpha_0)E[\sum_{t=\tau}^{T+\tau-1} V_t|e_{t+1}|]$. In other words, by misspecifying the forecaster's loss, we omit the variable $(1-2\alpha_0)|e_{t+1}|$ from the linear regression (4) and introduce correlation between the error term and the vector of explanatory variables. Hence, the standard OLS estimator $\hat{\beta} \equiv [\sum_{t=\tau}^{T+\tau-1} v_t v_t']^{-1}[\sum_{t=\tau}^{T+\tau-1} v_t e_{t+1}]$ will be biased away from β by a quantity which we derive in Proposition 1:

Proposition 1 *Under assumptions (A1)-(A4) given in Appendix A and under 'Quad-Quad' loss L_2, the standard OLS estimator, $\hat{\beta}$, in the efficiency regression* (4) *has a bias that equals*

$$\operatorname{plim} \hat{\beta} - \beta = (1-2\alpha_0)\Sigma_V^{-1} h_V, \tag{5}$$

where $\Sigma_V \equiv T^{-1} \sum_{t=\tau}^{T+\tau-1} E[V_t V_t']$ *and* $h_V \equiv T^{-1} \sum_{t=\tau}^{T+\tau-1} E[V_t|e_{t+1}|]$.

In other words, the misspecification bias depends on: (i) the extent of the departure from symmetry in the loss function L_2, quantified by $(1 - 2\alpha_0)$; (ii) the covariance vector h_V of the instruments used in the test, V_t, with the absolute value of the forecast error, $|e_{t+1}|$; and (iii) the covariance matrix of the instruments, Σ_V.

Some positive implications can be drawn from Proposition 1, improving our understanding of standard efficiency tests when the forecaster's loss is asymmetric $(\alpha_0 \neq 1/2)$:

- In the usual implementation of efficiency regressions such as the one in (4), a constant is included in V_t and we can write $V_t = (1, \tilde{V}_t')'$. The first element of the covariance vector h_V then equals $T^{-1} \sum_{t=\tau}^{T+\tau-1} E[|e_{t+1}|]$. Thus, when $\alpha_0 \neq 1/2$, there will *always* be bias in at least the constant term, unless the absolute forecast error is zero in expectation. This can only occur in the highly unlikely situation where the forecasts are perfect and the forecast errors are zero with probability one, so standard tests of forecast rationality will in general be biased under asymmetric loss.

- The bias of $\hat{\beta}$ decreases with the variability of the regressors. In other words, if the covariance matrix Σ_V is sufficiently large, it can 'drown out' the bias. Moreover, whenever the matrix Σ_V is nonsingular, the bias that arises through the constant term will extend directly to biases in the other coefficients for each regressor whose mean is nonzero. This follows from the interaction of Σ_V^{-1} and the first term of h_V. Hence, even when regressors have no additional information for improving the forecasts, they may still have nonzero coefficients when the loss function is misspecified, giving rise to false rejections.

- In practice, we can easily evaluate the relative biases for each coefficient in the efficiency regression (4) by simply computing the term $\Sigma_V^{-1} h_V$. For any degree of asymmetry, the latter can be consistently estimated by regressing the absolute forecast errors on V_t. Such regressions should accompany results that assume quadratic loss, especially when there are rejections. They allow us to understand how sensitive the results are to misspecifications of the loss function, at least of the form examined here.

3.2 Power of Efficiency Tests

Misspecification of the loss function not only affects the bias of the standard OLS estimator $\hat{\beta}$ in (4) but also its asymptotic distribution. Hence, rationality tests implemented by traditional MSE regression based on the null hypothesis $\beta = 0$ might well lead to incorrect inference. Proposition 2 allows us to study the magnitude of this problem:

Proposition 2 *Under Assumptions (A1)-(A5) listed in Appendix A and under 'Quad-Quad' loss L_2, the asymptotic distribution of $\hat{\beta}$ in the efficiency regression* (4) *is*

$$\sqrt{T}(\hat{\beta} - \beta^*) \xrightarrow{d} N(0, \Omega_V^*), \tag{6}$$

where $\beta^* \equiv \beta + (1 - 2\alpha_0)\Sigma_V^{-1} h_V$ and the expression for Ω_V^* and its consistent estimator $\hat{\Omega}_V$ are provided in Appendix A. For local deviations from symmetric loss, $\alpha_0 = 1/2$, given by $\alpha_0 = \frac{1}{2}(1 - aT^{-1/2})$, and local deviations from rationality, $\beta = 0$, given by $\beta = bT^{-1/2}$, with a and b fixed, the Wald test statistic based on the efficiency regression (4) *is asymptotically distributed as $T\hat{\beta}'\hat{\Omega}_V^{-1}\hat{\beta} \xrightarrow{d} \chi_d^2(m)$, a non-central chi-square with d degrees of freedom and non-centrality parameter m given by*

$$m = a^2 \varphi_V + b' {\Omega_V^*}^{-1} b + 2ab' k_V, \tag{7}$$

where the d-vector k_V and the scalar φ_V are defined in Appendix A.

What Proposition 2 shows is: (i) that for a wide range of combinations of the asymmetry parameter, $\alpha_0 \neq 1/2$, and the regression coefficient, β, efficiency tests may fail to reject even for large degrees of inefficiency ($\beta \neq 0$); (ii) when the forecaster genuinely uses information efficiently ($\beta = 0$) the efficiency test will tend to reject the null provided loss is asymmetric ($\alpha_0 \neq 1/2$). More specifically,

- Spurious rejections of the rationality hypothesis follow whenever the absolute value of the forecast error is correlated with V_t and the standard error of $\hat{\beta}$ is not too large. Since the power of the rationality test for $\beta = 0$ is driven entirely by the noncentrality parameter m, it suffices to consider this parameter to study the power of the standard rationality test in the directions of nonzero a and b. Non-zero values of a and b have very different economic interpretations: $b \neq 0$ implies that the forecasting model is misspecified, while $a \neq 0$ reflects asymmetric loss. Only the former can be interpreted as forecast inefficiency or irrationality. Yet, for given values of φ_V, Ω_V^* and k_V we can construct pairs of values (a, b) that lead to identical power (same m). Standard efficiency tests based on (4) can therefore *not* tell whether a rejection is due to irrationality or asymmetric loss - i.e., they lack robustness with respect to the shape of the loss function. A large value of m can arise even when the forecaster is *fully* rational ($b = 0$) provided that $|a|$ is large.[5]

- Conversely, suppose that the test does not reject, which would happen at the right size provided $m = 0$. This does *not* imply that the forecast is rational ($b = 0$) because we can construct pairs of non-zero values (a, b) such that $m = 0$. This will happen when the misspecification in the forecasts cancels out against the asymmetry in the loss function. The test will not have any power to identify this problem.

To demonstrate the importance of these points, Figure 2A plots iso-m - or, equivalently, iso-power - curves for different values of a and b, assuming a test size of 5% and $V_t = 1$.[6] Under MSE loss and informational efficiency $a = b = 0$. Positive values of a correspond to $\alpha_0 < 1/2$, while negative values of a represent $\alpha_0 > 1/2$. For any value of m we can solve the quadratic relationship (7) to obtain a trade-off between a and b. When $m = 0$ (the thick line in the center), the test rejects with power equal to size and the trade-off between a and b is simply $b = -a\hat{h}_V$. The two $m = 0.65$ lines represent power of 10%, the $m = 1.96$ lines give 50% power, while the $m = 3.24$ lines furthest towards the corners of the figure

[5]When $a \neq 0$, the constant term in (4) is particularly likely to lead to a rejection even when the forecasts are truly rational. This bias will be larger the less of the variation in the outcome variable is explained (since $E[|e_{t+1}|]$ is increasing in the variation of the forecast error).

[6]For this case $\Sigma_V = 1$ and h_V is a scalar that can be estimated by $\hat{h}_V = T^{-1} \sum_{t=\tau}^{t=T+\tau} |e_{t+1}|$. Hence, $\hat{k}_V = \hat{h}_V / \hat{\sigma}_u^2$ and $\hat{\varphi}_V = \hat{h}_V^2 / \hat{\sigma}_u^2$ – with $\hat{\sigma}_u^2$ being the variance of the residuals of the regression – are consistent estimators of k_V and φ_V, respectively. Values for $\hat{\sigma}_u$ and $\hat{h}_V$ were chosen to match the survey data in the empirical section.

represent power of 90%.[7] The lines slope downward since a larger value of a corresponds to a smaller value of α_0 and a stronger tendency to underpredict which cancels out against a larger negative bias in b.

Pairs of values a, b on the $m = 0$ line are such that biases in the forecasts (non-zero b) *exactly* cancel out against asymmetry in loss (non-zero a) in such a way that the standard test cannot detect the bias (in the sense that power = size) even though forecasters are irrational. For nonzero values for m, we see the converse. The point where these contours cross the $b = 0$ boundary (in the centre of the graphs) gives the asymmetry parameter that – if true for the forecaster – would result in rejections with greater frequency than size even though the forecaster is rational for that asymmetric loss function.[8]

Economic interpretation of these results is facilitated by plotting the power contours in $\alpha, \beta-$space, where the latter is reported in standard error units of the efficiency regression (4). For this plot - shown in Figure 2B - the iso-power lines become upward-sloping as larger values of α lower the loss-induced bias and hence cancel out against larger inefficiency biases, β. The figure shows that biases as large as 3.5 standard error units away from zero will be virtually undetectable provided the loss function is sufficiently asymmetric. Conversely, moving vertically along the $\beta = 0$ 'rationality line', we find that strongly asymmetric loss can lead to a 90% chance of falsely rejecting the null of rationality.

4 Rationality Tests under Asymmetric Loss

The lack of robustness of standard rationality tests to asymmetries suggests that a new set of tests is required. In this section we describe two such approaches. The first approach is applicable when the shape and parameters of the loss function are known. This set-up does not pose any new problems and least-squares estimation still applies, albeit on a

[7] The range of values for a in the figure (-10, 10) ensures that $\alpha_0 \in (0, 1)$ when $T = 100$. This range becomes more narrow (wider) for smaller (larger) sample sizes.

[8] Only if $\hat{h}_V = 0$ would asymmetric loss not cause problems to the standard test. In this case the absolute value of the forecast error is not correlated with the instrument, V_t, there is no omitted variable bias and the iso-power curves would be vertical lines, so size would only be controlled when $b = 0$.

transformation of the original forecast error. The second case arises when the parameters of the loss function are unknown and have to be estimated as part of the test. This framework requires different estimation methods which we describe below.

4.1 Known Loss

Under a loss function $L(e;\eta)$, characterized by some shape parameter η, the sequence of forecasts $\{f_{t+1}\}$ is said to be *optimal* under loss L if at any point in time, t, the forecast f_{t+1} minimizes $E[L(e_{t+1};\eta)|\mathcal{I}_t]$ - the expected value of L conditional on the information set available to the forecaster, $\mathcal{I}_t$. This implies that, at any point in time, the optimal forecast errors $\{e_{t+1}\}$ satisfy the first order condition $E[L'_e(e_{t+1};\eta)|\mathcal{I}_t] = 0$, where L'_e denotes the derivative of L with respect to the error e_{t+1}. When both L and η are known we can simply transform the observed forecast error e_{t+1} and test the orthogonality conditions by means of the 'generalized' efficiency regression

$$L'_e(e_{t+1};\eta) = \beta' v_t + u_{t+1}, \tag{8}$$

where the error term u_{t+1} satisfies $E[\sum_{t=\tau}^{T+\tau-1} V_t U_{t+1}] = 0$. Under standard regularity conditions, the linear regression parameter β can be consistently estimated by using the ordinary least squares (OLS) estimator $\tilde{\beta} \equiv [\sum_{t=\tau}^{T+\tau-1} v_t v_t']^{-1}[\sum_{t=\tau}^{T+\tau-1} v_t L'_e(e_{t+1};\eta)]$. As for the standard quadratic case in (4), forecast rationality is equivalent to having $\beta = 0$. Hence, under general loss a test for rationality can be performed by: (i) first transforming the observed forecast error e_{t+1} into $L'_e(e_{t+1};\eta)$, then (ii) regressing the latter on V_t by means of the regression (8) and finally (iii) testing the null hypothesis that all regression coefficients are zero, i.e. $\beta = 0$.

To demonstrate this type of test, suppose that it is known that the forecaster has a 'Quad-Quad' loss function with known asymmetry parameter $\alpha_0 \neq 1/2$. For this case the generalized efficiency regression takes the simple form

$$e_{t+1} - (1 - 2\alpha_0)|e_{t+1}| = \beta' v_t + u_{t+1}. \tag{9}$$

When α_0 equals one half, the previous regression collapses to the one traditionally used in tests for strong rationality (4). Assuming that the forecaster's loss is quadratic, $(\alpha_0 = 1/2)$, amounts to omitting the term $(1-2\alpha_0)|e_{t+1}|$ from the regression (9). Whenever $\alpha_0 \neq 1/2$, the estimates of the slope coefficient β in the resulting efficiency regression (4) are biased. This finding is as we would expect from the standard omitted variable bias result with the difference that we now have constructed the omitted regressor.

4.2 Unknown Loss Parameters

For many applications both L and η are unknown to the forecast evaluator. One way to proceed in this case is to relax the assumption that the true loss is known by assuming that L belongs to some flexible and known family of loss functions but with unknown shape parameter, η. Forecast rationality tests merely verify whether, under the loss L, the forecasts are optimal with respect to a set of variables V_t, known to the forecaster. They can therefore be viewed as tests of moment conditions, which arise from first order conditions of the forecaster's optimization problem. Traditional rationality tests, such as the one proposed by Mincer and Zarnowitz (1969), adopt a regression based approach to testing these orthogonality conditions. A natural alternative is to use a Generalized Method of Moments (GMM) framework as in Hansen (1982). The benefits of the latter are easily illustrated in the 'Quad-Quad' case. If the asymmetry parameter, α_0, is unknown it is impossible to compute the term $(1-2\alpha_0)|e_{t+1}|$ and hence not feasible to estimate the regression coefficient, β, in (9). However, it is still possible to test whether the moment conditions $E\{\sum_{t=\tau}^{T+\tau-1} V_t[(\alpha_0 - 1(e_{t+1}<0))|e_{t+1}| - \beta' V_t]\} = 0$ associated with the first order condition of forecast optimality under 'Quad-Quad' loss in (3) hold, with $\beta = 0$ and α_0 left unspecified.

The statistic suggested by Elliott, Komunjer and Timmermann (2003) for testing the null hypothesis that the forecasts are rational takes the form of a test for overidentification:

$$J_T \equiv T^{-1}[\sum_{t=\tau}^{T+\tau-1} v_t(\hat{\alpha}_T - 1(e_{t+1}<0))|e_{t+1}|]'\hat{S}_T^{-1}[\sum_{t=\tau}^{T+\tau-1} v_t(\hat{\alpha}_T - 1(e_{t+1}<0))|e_{t+1}|]. \quad (10)$$

Here $\hat{S}_T$ is a consistent estimator of $S \equiv T^{-1}\sum_{t=\tau}^{T+\tau-1} E[V_t V_t'(1(e_{t+1}<0)-\alpha_0)^2|e_{t+1}|^2]$, and $\hat{\alpha}_T$ is a linear Instrumental Variable (IV) estimator of α_0,

$$\hat{\alpha}_T \equiv \frac{[\sum_{t=\tau}^{T+\tau-1} v_t|e_{t+1}|]'\hat{S}_T^{-1}[\sum_{t=\tau}^{T+\tau-1} v_t(1(e_{t+1}<0)|e_{t+1}|]}{[\sum_{t=\tau}^{T+\tau-1} v_t|e_{t+1}|]'\hat{S}_T^{-1}[\sum_{t=\tau}^{T+\tau-1} v_t|e_{t+1}|]}. \quad (11)$$

In other words, the GMM overidentification test (J-test) is a consistent test of the null hypothesis that the forecasts are rational, i.e. $\beta = 0$, even if the true value of the asymmetry parameter α_0 is unknown, and the forecast errors depend on previously estimated parameters. The test is asymptotically distributed as a χ^2_{d-1} random variable and rejects for large values. Effectively, we exploit the first order conditions under forecast rationality, $E[\sum_{t=\tau}^{T+\tau-1} V_t(\alpha_0 - 1(e_{t+1} < 0))|e_{t+1}|] = 0$, with α_0 left unspecified. As a by-product, an estimate of the asymmetry parameter, $\hat{\alpha}_T$, is generated from equation (11).

Intuitively, the power of our test arises from the existence of overidentifying restrictions. In practice, for each element of V_t we could obtain an estimate for the asymmetry parameter, α_0, that would rationalize the observed sequence of forecasts. However, when the number of instruments, d, is greater than one, our method tests that the implied asymmetry parameter is the same for each moment condition. If no common value for α_0 satisfies all of the moment conditions, the test statistic J_T in (10) becomes large. This explains why the test still has power against the alternative that the forecasts were not constructed rationally and why it is *not* possible to justify arbitrary degrees of inefficiency in the forecasts by means of asymmetric loss: if the forecasts did not efficiently use the information in V_t, then $\hat{\alpha}_T$ would be very different for each of the moment conditions and the test would reject.

Although this approach does not impose a fixed value of α_0, it maintains that the loss function belongs to the family (2) with exponent $p = 2$ and the test (10) provides a joint test of rationality and this assumption. The advantage of this approach is that it loses little power since only one parameter has to be estimated. It is possible to take an even less restrictive approach and estimate the moment conditions non-parametrically. However, this is unlikely to be a useful strategy in view of the short survey data samples typically available.

5 Empirical Results

To see how asymmetric loss affects the empirical results from Section 2, derived under MSE loss, we proceed to test rationality of the output forecasts under 'Quad-Quad' loss. Results under four different sets of instruments, V_t, are considered, namely: (1) a constant and the lagged forecast error, (2) a constant and lagged actual GDP growth, (3) a constant, the lagged forecast error and the lagged value of GDP growth, and (4) a constant and the lagged absolute forecast error. These instruments are similar to those adopted in the literature and have power to detect predictability in forecast errors such as serial correlation.

5.1 Rationality Tests and Estimates of Loss Parameters

Table 2 shows the outcomes of two separate tests for rationality. The first test is for the joint hypothesis of rationality and symmetric loss ($\alpha_0 = 1/2$). The second test is for rationality but allows for asymmetry within the context of the more general family of 'Quad-Quad' loss functions. When three instruments are used, the null is rejected at the 1% level for 20 forecasters and it gets rejected for 34 forecasters at the 5% level and 42 forecasters at the 10% level. The results are very different when we no longer impose symmetry on the loss function. For this case no rejection is found at the 1% level, while four forecasts produce a rejection at the 5% level and 11 do so at the 10% level.

Standard tests of forecast rationality thus have reasonable power in the direction of detecting asymmetry in the loss function. In fact, rejections of the joint hypothesis of rationality and symmetry appear mostly to be driven by the symmetry assumption. Our rejection frequencies under asymmetric loss are almost exactly equal to the size of the test and hence suggest little evidence against the joint null of asymmetric loss and efficient forecasts.

So far we have not discussed the α-estimates although clearly there is considerable economic information in these values which should reflect the shape of the forecasters' loss function. Figure 3 shows a histogram of the 98 α-estimates computed using (11) for $V_t = 1$. The evidence is clearly indicative of asymmetric loss. Irrespective of which set of instruments is used, the proportion of α-estimates above one-half never exceeds 20%.

Importantly, the α-estimates suggested by our data do not appear to be 'extreme' and are clustered with a mode around 0.38. This corresponds to putting around one and a half times as large a weight on positive forecast errors as on negative ones. We might have found α-values much closer to zero in which case the degree of asymmetry required to explain biases in the forecasts would have to be implausibly large. Hence only a modest degree of asymmetry in the loss function is required to overturn rejections of the null hypothesis.

5.2 Bias and Type of Forecaster

The SPF data does not identify the affiliation of the forecaster. It is natural, however, to expect the extent of loss asymmetry to be different for forecasters associated with academia, banking and industry. It seems more plausible that academics have less of a reason to produce biased forecasts than, say, industry economists whose forecasts are produced for a specific firm or industry and thus - at least in theory - should put more weight on positive forecast errors if, e.g., inventory costs exceed stockout costs. It is more difficult to conjecture the size and direction of the bias for the banking forecasters. If these were produced for clients that were fully hedged with regard to unanticipated shocks to economic growth, one would expect α-estimates closer to one-half. However, if bank losses arising from over-predictions of economic growth exceed those from underpredictions, again we would expect more α-estimates below one-half than above it.

To consider this issue, we used data from the Livingston survey which lists the forecaster's affiliation. Unfortunately this data set tends to be much shorter as forecasts are only generated every six months. We therefore only required a minimum of 10 observations. This leaves us with 12 industry, five academic and 12 forecasters from the banking sector - admittedly a very small sample.

The α-estimates for these forecasters are shown in Figure 4. Academic forecasters tend to produce α-estimates closer to one-half than the forecasters from industry and banking. The joint null of rationality and $\alpha_0 = 1/2$ is not rejected for any of the academic forecasters, while this hypothesis is rejected at the 5% level for two of the 12 industry forecasters and for five of the 12 banking forecasters. While this evidence is by no means conclusive given the very small sample available here, it is indicative that differential costs associated with positive and negative forecast errors play a role in explaining forecast biases.

6 Conclusion

Empirical studies frequently find that forecasts from survey data are biased. Does this mean that forecasters genuinely use information inefficiently and hence are irrational or simply that they have asymmetric loss? We have shown in this paper that standard forecast efficiency tests often cannot distinguish between these two possible explanations. The importance of this point - which many previous papers have expressed concern about - was validated empirically as we found that rejections of rationality may largely have been driven by the assumption of symmetric loss. Conversely, we should not necessarily conclude from the failure to reject the null when we allow for asymmetric loss that forecasters are rational in view of the limited ability of rationality tests to identify inefficient use of information in survey samples on individual forecasters as small as those used here.

Our empirical findings raise the question whether the degree of asymmetry in the loss function required for forecasts to be efficient is excessive given what is known about the forecasting situation. There is, of course, a precursor for this type of question. In finance, the equity premium puzzle consists of the finding that the value of the risk aversion parameter required for historical stock returns to be consistent with a representative investor model appears to be implausibly high, c.f. Mehra and Prescott (1985). In our context, the empirical findings suggest that often only a modest degree of asymmetry is required to overturn rejections of rationality and symmetric loss. The results therefore point towards the need for collecting data on the costs associated with forecast errors of different signs and magnitude in order to better understand the forecasters' objectives.

References

[1] Arrow, K.J., S. Karlin, and H. Scarf, 1958, Studies in the Mathematical Theory of Inventory and Production. Stanford: Stanford University Press.

[2] Batchelor R. and P. Dua, 1991, Blue Chip rationality tests. Journal of Money, Credit and Banking 23, 692-705.

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

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

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

[6] Conlisk, J., 1996, Why Bounded Rationality? Journal of Economic Literature 34, 669-700.

[7] Davies, A. and K. Lahiri, 1995, A New Framework for Analyzing three-dimensional Panel Data. Journal of Econometrics 68, 205-227.

[8] Ehrbeck, T. and Waldmann, R., 1996, Why are professional forecasts biased? Agency versus behavioral explanations, Quarterly Journal of Economics 111, 21-40.

[9] Elliott, G., I. Komunjer and A. Timmermann, 2003, Estimating Loss Function Parameters. Mimeo University of California San Diego and Caltech.

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

[11] Hansen, L.P., 1982, Large Sample Properties of Generalized Method of Moments Estimators. Econometrica 50, 1029-1054.

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

[13] Leitch, G. and J.E. Tanner, 1991, Economic forecast evaluation: profits versus the conventional error measures, American Economic Review 81, 580-90.

[14] Mankiw, N.G., R. Reis and J. Wolfers, 2003, Disagreement about Inflation Expectations. NBER Macroeconomics Annual, 2003.

[15] McCracken, M.W., 2000, Robust Out-of-Sample Inference, Journal of Econometrics 99, 195-223.

[16] Mehra, R. and E. Prescott, 1985, The Equity Premium: A Puzzle, Journal of Monetary Economics 15, 145-161.

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

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

[19] Patton, A. and A. Timmermann, 2002, Properties of Optimal Forecasts. Mimeo, LSE and UCSD.

[20] Scharfstein, D. and J. Stein, 1990, Herd Behavior and Investment. American Economic Review 80, 464-479.

[21] Truman, B., 1994, Analyst Forecasts and Herding Behavior, Review of Financial Studies 7, 97-124.

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

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

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

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

Data Appendix

Our empirical application uses the growth in quarterly seasonally adjusted nominal US GDP in billions of dollars before 1992 and nominal GNP after 1992. The growth rate is calculated as the difference in natural logs. Data for the actual values come from two sources. The official (revised) data is from the BEA. To avoid using revised data that was not historically available to the forecasters we also use real-time data from the Philadelphia Fed. This provides the vintages of data available in real time and takes the following form:

	$68.IV$	$69.I$	$69.II$	$69.III$	$69.IV$
$68.IV$	NA	887.8	887.4	892.5	892.5
$69.I$	NA	NA	903.4	908.7	908.7
$69.II$	NA	NA	NA	925.1	924.8
$69.III$	NA	NA	NA	NA	942.3
$69.IV$	NA	NA	NA	NA	NA

Rows represent the dates corresponding to the index while columns track the vintage. So, in 1969.IV, a forecaster looked at a value of 942.3 for 69.III, 924.8 for 69.II and so on. This real-time data is used to construct real time instruments used in the rationality tests. Both the lagged forecast error and the lagged value of output growth are based on the historical vintages available in real time.

Data on the forecasts come from the Survey of Professional Forecasters, also maintained by the Philadelphia Fed. This data runs from the fourth quarter of 1968 to the first quarter of 2002. It provides the quarter, the number of the forecaster, the most recent value known to the forecaster (preceding), the value (most of the times forecasted) for the current quarter (current) and then forecasts for the next four quarters. We use the values corresponding to the current and the first forecast to calculate the one-step-ahead growth rate.[9]

Some forecasters report missing values while others decide to leave for a while, but then return and continue to produce forecasts. To deal with these problems we followed three steps. We eliminate individuals with less than 20 forecasts (so, from a total of 512 individuals

[9] A few forecasts were omitted from the data base. There were clear typos for forecaster number 12 (1989.II), forecasters 20 and 62 (1992.IV) and forecaster 471 (1997.II).

we keep 107 forecasters).[10] We then eliminate forecasters with missing values. This reduces the number of individual forecasters to 98.

The Livingston data considered in Section 6.3 uses the difference in the logs of the six-month forecast (two quarters ahead) over the forecast of the current quarter and starts in the second semester of 1992 which is the time when current quarter figures start to get included. Data runs through the second semester of 2002. This data contains information on the affiliation of the forecasters. Most affiliations have very few observations, so only those corresponding to Industry, Academic and Banking were considered. Individuals with implausibly large forecast errors (greater than 5 percentage points over a six-month period) and too few observations (less than ten) were excluded from the analysis. This leaves us with five Academic forecasters, 12 Banking forecasters and 12 Industry forecasters.

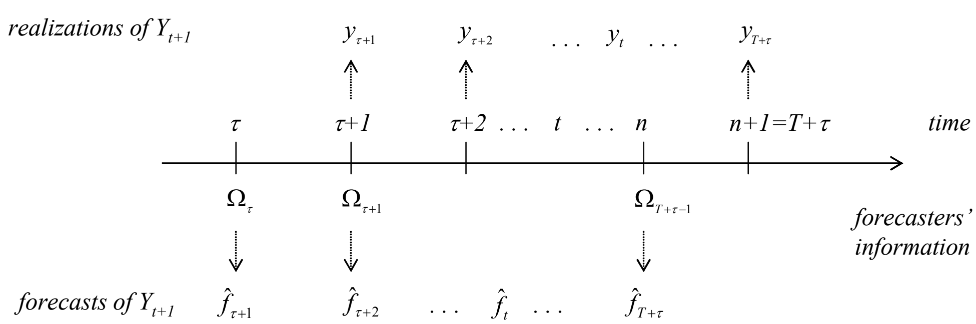
[10] 103 out of these 107 individuals have a gap of at least one quarter in their reported forecasts. Most forecasters skip one or more quarters.

Appendix A

This appendix describes the forecasting setup, lists the assumptions used to establish Propositions 1 and 2 and defines the terms referred to in the latter. We use the following notations: if v is a real d-vector, $v = (v_1, \ldots, v_d)'$, then $|v|$ denotes the standard L_2-norm of v, i.e. $|v|^2 = v'v = \sum_{i=1}^{d} v_i^2$. If M is a real $d \times d$-matrix, $M = (m_{ij})_{1 \leqslant i,j \leqslant d}$, then $|M|$ denotes the L_∞-norm of M, i.e. $|M| = \max_{1 \leqslant i,j \leqslant d} |m_{ij}|$.

Forecasting scheme:

As a preamble to our proofs, it is worth pointing out that the estimation uncertainty of the observed forecasts, which we hereafter denote $\hat{f}_{t+1}$, gives rise to complications when testing rationality. The models used by the forecasters to produce $\hat{f}_{t+1}$ are typically unknown to the econometrician or forecast user. Indeed, there are a number of different forecasting methods which can be used by the forecasters at the time they make their forecasts, most of which involve estimating (or calibrating) some forecasting model.



In addition to different models employed, forecasts may also differ according to the forecasting scheme used to produce them. For example, a fixed forecasting scheme constructs the in-sample forecasting model only once, then uses it to produce all the forecasts for the out-of-sample period. A rolling window forecasting scheme re-estimates the parameters of the forecasting model at each out-of-sample point. In order to fix ideas, we assume that all the observed forecasts are made recursively from some date τ to $\tau + T$ as depicted in the figure above, so that the sequence $\{\hat{f}_{t+1}\}$ depends on recursive estimates of the forecasting model. The sampling error caused by this must be taken into account (see, e.g., West (1996), West and McCracken (1998), McCracken (2000)). Throughout the proofs we assume that the forecasters' objective is to solve the problem

$$\min_{\{f_{t+1}\}} E[\sum_{t=\tau}^{\tau+T-1} \alpha|e_{t+1}|^p 1(e_{t+1} \geq 0) + (1-\alpha)|e_{t+1}|^p 1(e_{t+1} < 0)], \tag{12}$$

and thus define a sequence of forecasts $\{f_{t+1}^*\}$ and corresponding forecast errors $\{e_{t+1}^*\}$. It is important to note that $\{f_{t+1}^*\}$ which minimizes the above expectation is unobservable in practice. Instead we assume the econometrician observes $\{\hat{f}_{t+1}\}$ thus taking into account that the observed sequence of forecasts embodies a certain number of recursively estimated parameters of the forecasting model.

Assumptions:

(A1) $\beta \in B$ where the parameter space $B \subseteq \mathbb{R}^d$ and B is compact. Moreover $\beta^* \in \mathring{B}$;

(A2) for every t, $\tau \leqslant t < T+\tau$, the forecast of Y_{t+1} is a measurable function of an Ω_t-measurable h-vector W_t, i.e. $f_{t+1} = f_{t+1}(W_t)$, where the function f_{t+1} is unknown but bounded, i.e. $\sup_{\theta\in\Theta}|f_{t+1}(W_t)| \leqslant C < \infty$ with probability one, and $\hat{f}_{t+1} = f_{t+1}^* + o_p(1)$;

(A3) the d-vector V_t is a sub-vector of the h-vector W_t ($d \leqslant h$) with the first component 1 and for every t, $\tau \leqslant t < T+\tau$, the matrix $E[V_t V_t']$ is positive definite;

(A4) $\{(Y_t, W_t')\}$ is an α-mixing sequence with mixing coefficient α of size $-r/(r-2)$, $r > 2$, and there exist some $\Delta_Y > 0$, $\Delta_V > 0$ and $\delta > 0$ such that for every t, $\tau \leqslant t < T+\tau$, $E[|Y_{t+1}|^{2r+\delta}] \leqslant \Delta_Y < \infty$ and $E[|V_t|^{2r+\delta}] \leqslant \Delta_V < \infty$;

(A5) for some small $\varepsilon, \varepsilon \in (0,1)$: (i) $\tau^{1-2\varepsilon}/T \to \infty$ and (ii) $\sup_{\tau\leqslant t<T+\tau}|t^{1/2-\varepsilon}(\hat{f}_{t+1} - f_{t+1}^*)| \xrightarrow{p} 0$, as $\tau \to \infty$ and $T \to \infty$.

Definitions:

$\beta^* \equiv \beta + (1-2\alpha_0)\Sigma_V^{-1}h_V$,

$\Omega_V^* \equiv \Sigma_V^{-1}\Delta_V(\beta^*)\Sigma_V^{-1}$,

$\Delta_V(\beta^*) \equiv T^{-1}\sum_{t=\tau}^{T+\tau-1} E[u_{t+1}^2 V_t V_t'] + 2(1-2\alpha_0)E[u_{t+1}|e_{t+1}^*|V_t V_t'] + (1-2\alpha_0)^2 E[e_{t+1}^{*2} V_t V_t']$,

$\hat{\Delta}_V(\hat{\beta}) \equiv T^{-1}\sum_{t=\tau}^{T+\tau-1}(\hat{e}_{t+1} - \hat{\beta}' v_t)^2 v_t v_t'$ with $\hat{\Delta}_V(\hat{\beta}) \xrightarrow{p} \Delta_V(\beta^*)$,

$\hat{\Omega}_V \equiv \hat{\Sigma}_V^{-1}\hat{\Delta}_V(\hat{\beta})\hat{\Sigma}_V^{-1}$ with $\hat{\Omega}_V \xrightarrow{p} \Omega_V^*$,

$k_V \equiv \Sigma_V \Delta_V(\beta^*)^{-1}h_V$ and $\varphi_V \equiv h_V'\Delta_V(\beta^*)^{-1}h_V$.

Appendix B

Proof of Proposition 1. In the first part of this proof we show that $\hat{\beta} \xrightarrow{p} \beta^*$, where $\beta^* \equiv (\sum_{t=\tau}^{T+\tau-1} E[V_t V_t'])^{-1} \cdot (\sum_{t=\tau}^{T+\tau-1} E[V_t e_{t+1}^*])$. We then use this convergence result in the second part of the proof to derive the expression for the misspecification bias of $\hat{\beta}$. Recall from Section 4 that the standard OLS estimator is $\hat{\beta} \equiv [\sum_{t=\tau}^{T+\tau-1} v_t v_t']^{-1} [\sum_{t=\tau}^{T+\tau-1} v_t \hat{e}_{t+1}]$. In order to show that $\hat{\beta} \xrightarrow{p} \beta^*$, it suffices to show that the following conditions hold:
(i) β^* is the unique minimum on B (compact in $\mathbb{R}^d$) of the quadratic form $S_0(\beta)$ with $S_0(\beta) \equiv T^{-1} \sum_{t=\tau}^{T+\tau-1} E[(e_{t+1}^* - \beta' V_t)^2]$; (ii) $T^{-1} \sum_{t=\tau}^{T+\tau-1} v_t v_t' \xrightarrow{p} T^{-1} \sum_{t=\tau}^{T+\tau-1} E[V_t V_t']$; (iii) $T^{-1} \sum_{t=\tau}^{T+\tau-1} v_t \hat{e}_{t+1} \xrightarrow{p} T^{-1} \sum_{t=\tau}^{T+\tau-1} E[V_t e_{t+1}^*]$. From the positive definiteness of $E[V_t V_t']$, for all t (assumption A3) and the continuity of the inverse function (away from zero), it then follows that $\hat{\beta} \xrightarrow{p} \beta^*$.
We start by showing (i): note that $S_0(\beta) = T^{-1} \sum_{t=\tau}^{T+\tau-1} E[(e_{t+1}^*)^2] - 2\beta' T^{-1} \sum_{t=\tau}^{T+\tau-1} E[V_t e_{t+1}^*] + \beta' T^{-1} \sum_{t=\tau}^{T+\tau-1} E[V_t V_t'] \beta$, so $S_0(\beta)$ admits a unique minimum at β^* if for every t, $\tau \leqslant t < T+\tau$, the matrix $E[V_t V_t']$ is positive definite, which is satisfied by assumption (A3). This verifies the uniqueness condition (i).
In order to show (ii) and (iii), we use a law of large numbers (LLN) for α-mixing sequences (e.g., Corollary 3.48 in White, 2001). By assumptions (A2) and (A3) we know that for every t, $\tau \leqslant t < T + \tau$, $\hat{f}_{t+1}$ and V_t are measurable functions of W_t which by (A4) is an α-mixing sequence with mixing coefficient α of size $-r/(r-2)$, $r > 2$. Hence, by Theorem 3.49 in White (2001, p 50) we know that $\{(\hat{e}_{t+1} V_t', \mathrm{vec}(V_t V_t')')'\}$, where $\hat{e}_{t+1} = Y_{t+1} - \hat{f}_{t+1}$, is an α-mixing sequence with mixing coefficient α of the same size $-r/(r-2)$, $r > 2$. For $\delta > 0$, we have $r + \delta/2 > 2$ and $r/2 + \delta/4 > 1$ so by assumption (A4)

$$\begin{aligned} E[|V_t V_t'|^{r/2+\delta/4}] &\leqslant E[|V_t|^{r+\delta/2}] \\ &\leqslant \max\{1, \Delta_V^{1/2}\} < \infty, \end{aligned}$$

for all t, $\tau \leqslant t < T + \tau$. Hence, by applying the results from Corollary 3.48 in White (2001) to the sequence $\{\mathrm{vec}(V_t V_t')'\}$, we conclude that $T^{-1} \sum_{t=\tau}^{T+\tau-1} v_t v_t'$ converges in probability to its expected value $T^{-1} \sum_{t=\tau}^{T+\tau-1} E[V_t V_t']$, which shows that (ii) holds.

Similarly, we know by the Cauchy-Schwartz inequality that, for all t, $\tau \leqslant t < T+\tau$,

$$E[|V_t \hat{e}_{t+1}|^{r/2+\delta/4}] \leqslant (E[|V_t|^{r+\delta/2}])^{1/2} \cdot (E[|\hat{e}_{t+1}|^{r+\delta/2}])^{1/2}.$$

Hence there exists some $n_{r,\delta} \in \mathbb{R}_*^+$, $n_{r,\delta} < \infty$, such that

$$\begin{aligned} E[|V_t \hat{e}_{t+1}|^{r/2+\delta/4}] &\leqslant \max\{1, \Delta_V^{1/2}\} \cdot (n_{r,\delta} \cdot (E[|Y_{t+1}|^{r+\delta/2}] + E[|\hat{f}_{t+1}|^{r+\delta/2}]))^{1/2} \\ &\leqslant \max\{1, \Delta_V^{1/2}\} \cdot n_{r,\delta}^{1/2} \cdot (\max\{1, \Delta_Y^{1/2}\} + \max\{1, C^{r+\delta/2}\})^{1/2} \\ &< \infty, \end{aligned}$$

for all t, $\tau \leqslant t < T+\tau$, where we have used assumptions (A2) and (A4). Hence, our previous argument applies to the sequence $\{\hat{e}_{t+1} V_t'\}$ as well, and we conclude that $T^{-1}\sum_{t=\tau}^{T+\tau-1} v_t \hat{e}_{t+1}$ converges in probability to its expected value $T^{-1}\sum_{t=\tau}^{T+\tau-1} E[V_t \hat{e}_{t+1}]$. Note, however, that this does not ensure that (iii) holds, as we moreover need to show that substituting e_{t+1}^* for $\hat{e}_{t+1}$ does not affect the result, i.e. that $T^{-1}\sum_{t=\tau}^{T+\tau-1} E[V_t \hat{e}_{t+1}] - E[V_t e_{t+1}^*] \xrightarrow{p} 0$. For every t, $\tau \leqslant t < T+\tau$, we have

$$\begin{aligned} |E[V_t \cdot (\hat{e}_{t+1} - e_{t+1}^*)]| &= |E[V_t \cdot (f_{t+1}^* - \hat{f}_{t+1})]| \\ &\leqslant (E[|V_t|^2])^{1/2} \cdot (E[(f_{t+1}^* - \hat{f}_{t+1})^2])^{1/2} \\ &\leqslant \max\{1, \Delta_V^{1.2}\} \cdot (E[(f_{t+1}^* - \hat{f}_{t+1})^2])^{1/2}. \end{aligned}$$

Since by (A2) we know that $f_{t+1}^* - \hat{f}_{t+1} = o_p(1)$, for all t, we get $T^{-1}\sum_{t=\tau}^{T+\tau-1} E[V_t \hat{e}_{t+1}] - E[V_t e_{t+1}^*] \xrightarrow{p} 0$. Combined with our previous result this shows that (iii) holds. Hence, we conclude that $\hat{\beta} \xrightarrow{p} \beta^*$.

We now use this convergence result to derive the bias in $\hat{\beta}$. We know from Section 4 that the parameter β in the generalized regression (9) satisfies the set of identifying constraints

$$T^{-1}\sum_{t=\tau}^{T+\tau-1} E\{V_t \cdot [2(\alpha_0 - 1(e_{t+1}^* < 0))|e_{t+1}^*| - \beta' V_t]\} = 0,$$

so that $T^{-1}\sum_{t=\tau}^{T+\tau-1} 2E[(\alpha_0 - 1(e_{t+1}^* < 0))V_t|e_{t+1}^*|] = T^{-1}\sum_{t=\tau}^{T+\tau-1} E[V_t V_t']\beta$. Using that $2 \cdot 1(e_{t+1}^* < 0)|e_{t+1}^*| = |e_{t+1}^*| - e_{t+1}^*$, this last equality can be written $T^{-1}\sum_{t=\tau}^{T+\tau-1} E[V_t e_{t+1}^*] - T^{-1}\sum_{t=\tau}^{T+\tau-1} E[(1-2\alpha_0)V_t|e_{t+1}^*|] = T^{-1}\sum_{t=\tau}^{T+\tau-1} E[V_t V_t']\beta$, so by positive definiteness of $E[V_t V_t']$ (A2) we have $\beta = (T^{-1}\sum_{t=\tau}^{T+\tau-1} E[V_t V_t'])^{-1} \cdot \{T^{-1}\sum_{t=\tau}^{T+\tau-1} E[V_t e_{t+1}^*] - E[(1-2\alpha_0)V_t|e_{t+1}^*|]\}$.

In other words, $\beta = \beta^* - (1 - 2\alpha_0) \cdot (T^{-1} \sum_{t=\tau}^{T+\tau-1} E[V_t V_t'])^{-1} \cdot (T^{-1} \sum_{t=\tau}^{T+\tau-1} E[V_t | e_{t+1}^* |])$. This shows that $\hat{\beta} \xrightarrow{p} \beta + (1 - 2\alpha_0) \Sigma_V^{-1} h_V$ with $\Sigma_V \equiv T^{-1} \sum_{t=\tau}^{T+\tau-1} E[V_t V_t']$ and $h_V \equiv T^{-1} \sum_{t=\tau}^{T+\tau-1} E[V_t | e_{t+1}^* |]$, which completes the proof of Proposition 1.[11] ■

Proof of Proposition 2. We now show that $T^{1/2}(\hat{\beta} - \beta^*)$ is asymptotically normal by expanding the first order condition for $\hat{\beta}$ around β^*:

$$[\sum_{t=\tau}^{T+\tau-1} v_t(\hat{e}_{t+1} - \hat{\beta}' v_t)] = 0 = [\sum_{t=\tau}^{T+\tau-1} v_t(\hat{e}_{t+1} - \beta^{*\prime} v_t)] - (\sum_{t=\tau}^{T+\tau-1} v_t v_t')(\hat{\beta} - \beta^*). \tag{13}$$

The idea then is to use (i) $T^{1/2}[\sum_{t=\tau}^{T+\tau-1} v_t(\hat{e}_{t+1} - \beta^{*\prime} v_t)] = T^{1/2}[\sum_{t=\tau}^{T+\tau-1} v_t(e_{t+1}^* - \beta^{*\prime} v_t)] + o_p(1)$, together with (ii) $T^{-1/2} \sum_{t=1}^{T} v_t(e_{t+1}^* - \beta^{*\prime} v_t) \xrightarrow{d} N(0, \Delta_V(\beta^*))$, where $\Delta_V(\beta^*) \equiv T^{-1} \sum_{t=1}^{T} E[(e_{t+1}^* - \beta^{*\prime} V_t)^2 V_t V_t']$, to show by Slutsky's theorem that

$$T^{-1/2} \sum_{t=\tau}^{T+\tau-1} v_t(\hat{e}_{t+1} - \beta^{*\prime} v_t)] \xrightarrow{d} \mathcal{N}(0, \Delta_V(\beta^*)). \tag{14}$$

The remainder of the asymptotic normality proof is then similar to the standard case: the positive definiteness of Σ_V^{-1}, and the consistency of $\hat{\Sigma}_V = T^{-1} \sum_{t=\tau}^{T+\tau-1} v_t v_t'$, $\hat{\Sigma}_V \xrightarrow{p} \Sigma_V$, ensure that the expansion (13) is equivalent to $T^{1/2}(\hat{\beta} - \beta^*) = \hat{\Sigma}_V^{-1} T^{-1/2} \sum_{t=\tau}^{T+\tau-1} v_t(\hat{e}_{t+1} - \beta^{*\prime} v_t)$. We then use the limit result in (14) and Slutsky's theorem to show that

$$T^{1/2}(\hat{\beta} - \beta^*) \xrightarrow{d} \mathcal{N}(0, \Sigma_V^{-1} \Delta_V(\beta^*) \Sigma_V^{-1}),$$

which is what Proposition 2 states.

Hence, we need to show that conditions (i) and (ii) hold: For (i) it is sufficient to show that $T^{1/2} \sum_{t=\tau}^{T+\tau-1} v_t \hat{e}_{t+1} - T^{1/2} \sum_{t=\tau}^{T+\tau-1} v_t e_{t+1}^* \xrightarrow{p} 0$. We have

$$\begin{aligned} T^{-1/2} \Big| \sum_{t=\tau}^{T+\tau-1} V_t(\hat{e}_{t+1} - e_{t+1}^*) \Big| &= T^{-1/2} \Big| \sum_{t=\tau}^{T+\tau-1} t^{-1/2+\varepsilon} V_t t^{1/2-\varepsilon} (\hat{f}_{t+1} - f_{t+1}^*) \Big| \\ &\leqslant \sup_{\tau \leqslant t \leqslant T+\tau-1} |t^{1/2-\varepsilon}(\hat{f}_{t+1} - f_{t+1}^*)| \cdot T^{-1/2} \sum_{t=\tau}^{T+\tau-1} |V_t| t^{-1/2+\varepsilon}. \end{aligned}$$

[11] Moreover, these results ensure that $\hat{h}_V - h_V \xrightarrow{p} 0$ and $\hat{\Sigma}_V - \Sigma_V \xrightarrow{p} 0$ where $\hat{h}_V \equiv T^{-1} \sum_{t=\tau}^{T+\tau-1} v_t |\hat{e}_{t+1}|$ and $\hat{\Sigma}_V \equiv T^{-1} \sum_{t=\tau}^{T+\tau-1} v_t v_t'$, which makes the estimation of the bias components straightforward.

Note that (A4) implies $\sup_{\tau \leqslant t < T+\tau} E(|V_t|) < \infty$, so that, for any given $\nu > 0$, by (A5) and Chebyshev's inequality we have

$$\begin{aligned} P(T^{-1/2} \sum_{t=\tau}^{T+\tau-1} |V_t| t^{-1/2+\varepsilon} &> \nu) \leqslant \sup_{\tau \leqslant t < T+\tau} E(|V_t|)/\nu \cdot T^{-1/2} \sum_{t=\tau}^{T+\tau-1} t^{-1/2+\varepsilon} \\ &\leqslant \sup_{\tau \leqslant t < T+\tau} E(|V_t|)/\nu \cdot (T/\tau^{1-2\varepsilon})^{1/2} \to 0 \end{aligned}$$

as $\tau \to \infty$ and $T \to \infty$. Therefore, we conclude that $T^{-1/2} | \sum_{t=\tau}^{T+\tau-1} v_t \hat{e}_{t+1} - \sum_{t=\tau}^{T+\tau-1} v_t e^*_{t+1} | \xrightarrow{p} 0$, which implies that (i) holds.

We now show that (ii) holds as well, i.e. that $T^{-1/2} \sum_{t=1}^{T} v_t (e^*_{t+1} - \beta^{*\prime} v_t) \xrightarrow{d} N(0, \Delta_V(\beta^*))$ with $\Delta_V(\beta^*) \equiv T^{-1} \sum_{t=1}^{T} E[(e^*_{t+1} - \beta^{*\prime} V_t)^2 V_t V_t']$. For that, we use a central limit theorem (CLT) for α-mixing sequences (e.g. Theorem 5.20 in White, 2001): first, note that, by construction, $T^{-1} \sum_{t=\tau}^{T+\tau-1} E[V_t(e^*_{t+1} - \beta^{*\prime} V_t)] = 0$. For $r > 2$, the Cauchy-Schwartz inequality and assumption (A4) imply that for every t, $\tau \leqslant t < T+\tau$,

$$\begin{aligned} E[|V_t(e^*_{t+1} - \beta^{*\prime} V_t)|^r] &\leqslant (E[|V_t|^{2r}])^{1/2} \cdot (E[(e^*_{t+1} - \beta^{*\prime} V_t)^{2r}])^{1/2} \\ &\leqslant \max\{1, \Delta_V^{1/2}\} \cdot (E[n_r(|e^*_{t+1}|^{2r} + |\beta^*|^{2r} |V_t|^{2r})])^{1/2} \\ &\leqslant \max\{1, \Delta_V^{1/2}\} \cdot \max\{1, n_r^{1/2} (E[|e^*_{t+1}|^{2r}] + |\beta^*|^{2r} E[|V_t|^{2r}])^{1/2}\}, \end{aligned}$$

where $n_r \in \mathbb{R}^+_*$ is a constant, $n_r < \infty$, such that $E[(\hat{e}_{t+1} - \beta^{*\prime} V_t)^{2r}] \leqslant E[n_r(|e^*_{t+1}|^{2r} + |\beta^{*\prime} V_t|^{2r})]$. Knowing that for every t, $\tau \leqslant t < T+\tau$, $E[|e^*_{t+1}|^{2r}] = E[(Y_{t+1} - f^*_{t+1})^{2r}] \leqslant n_r(E[|Y_{t+1}|^{2r}] + E[|f^*_{t+1}|^{2r}]) \leqslant n_r(\Delta_Y + C^{2r})$, we get

$$E[|V_t(e^*_{t+1} - \beta^{*\prime} V_t)|^r] \leqslant \max\{1, \Delta_V^{1/2}\} \cdot \max\{1, n_r^{1/2} (n_r(\Delta_Y + C^{2r}) + |\beta^*|^{2r} \Delta_V)^{1/2}\} < \infty,$$

by assumptions (A1) (β^* is an element of a compact set) and (A2) (boundedness of $|f_{t+1}|$). Assumption (A3) moreover ensures that the matrix $\Delta_V(\beta^*)$ is positive definite, so that the CLT implies $T^{-1/2} \sum_{t=1}^{T} v_t(e^*_{t+1} - \beta^{*\prime} v_t) \xrightarrow{d} N(0, \Delta_V(\beta^*))$. This shows that (ii) holds.

The reasoning we described at the beginning of the proof then gives $\sqrt{T}(\hat{\beta} - \beta^*) \xrightarrow{d} N(0, \Omega^*_V)$ with $\Omega^*_V \equiv \Sigma_V^{-1} \Delta_V(\beta^*) \Sigma_V^{-1}$. Now note that

$$\begin{aligned} \Delta_V(\beta^*) &= T^{-1} \sum_{t=\tau}^{T+\tau-1} E[(e^*_{t+1} - \beta^{*\prime} V_t)^2 V_t V_t'] \\ &= T^{-1} \sum_{t=\tau}^{T+\tau-1} E[\varepsilon^2_{t+1} V_t V_t'], \end{aligned}$$

where $\varepsilon_{t+1} \equiv u_{t+1} + (1-2\alpha_0)|e^*_{t+1}|$ and u_{t+1} is the realization of the error term U_{t+1} in the generalized efficiency regression (9). Hence,

$$\Delta_V(\beta^*) = T^{-1}\sum_{t=\tau}^{T+\tau-1} E[u^2_{t+1}V_tV_t'] + 2(1-2\alpha_0)E[u_{t+1}|e^*_{t+1}|V_tV_t'] + (1-2\alpha_0)^2 E[|e^*_{t+1}|^2 V_tV_t'].$$

Moreover, the results above ensure that $\Delta_V(\beta^*)$ can be consistently estimated by $\hat{\Delta}_V(\hat{\beta}) \equiv T^{-1}\sum_{t=\tau}^{T+\tau-1}(\hat{e}_{t+1} - \hat{\beta}'v_t)^2 v_tv_t'$. Using that Σ_V^{-1} is positive definite, we can then show that $\hat{\Omega}_V \equiv \hat{\Sigma}_V^{-1}\hat{\Delta}_V(\hat{\beta})\hat{\Sigma}_V^{-1}$ is a consistent estimator of the asymptotic covariance matrix of the standard OLS estimator $\hat{\beta}$, i.e. $\hat{\Sigma}_V^{-1}\hat{\Delta}_V(\hat{\beta})\hat{\Sigma}_V^{-1} = \hat{\Omega}_V \xrightarrow{p} \Omega_V^* = \Sigma_V^{-1}\Delta_V(\beta^*)\Sigma_V^{-1}$.

To prove the last part of the proposition, let $\beta = bT^{-1/2}$ and $1-2\alpha_0 = aT^{-1/2}$ where a and b are fixed. We can write

$$\begin{aligned} T\hat{\beta}' \cdot \hat{\Omega}^{-1} \cdot \hat{\beta} &= T^{1/2}(\hat{\beta}-\beta^*)' \cdot \hat{\Omega}^{-1} \cdot T^{1/2}(\hat{\beta}-\beta^*) \\ &\quad + T^{1/2}\beta^{*\prime} \cdot \hat{\Omega}^{-1} \cdot T^{1/2}\beta^* \\ &\quad + 2T^{1/2}\beta^{*\prime} \cdot \hat{\Omega}^{-1} \cdot T^{1/2}(\hat{\beta}-\beta^*). \end{aligned}$$

It follows from the first part of the proposition that the first term is asymptotically χ^2_d distributed. For the second term, recall from Proposition 1 that $\beta^* = \beta + (1-2\alpha_0)\Sigma_V^{-1}h_V$ so $T^{1/2}\beta^* = (b + a\Sigma_V^{-1}h_V)$ and, moreover, $\hat{\Omega}_V \xrightarrow{p} \Omega_V^* = \Sigma_V^{-1}\Delta_V(\beta^*)\Sigma_V^{-1}$ with Ω_V^* nonsingular. We then have $T\beta^{*\prime}\hat{\Omega}_V^{-1}\beta^* \xrightarrow{p} m$ with

$$\begin{aligned} m &= (b + a\Sigma_V^{-1}h_V)'(\Sigma_V^{-1}\Delta_V(\beta^*)\Sigma_V^{-1})^{-1}(b + a\Sigma_V^{-1}h_V) \\ &= b'\Sigma_V\Delta_V(\beta^*)^{-1}\Sigma_V b + 2ab'\Sigma_V\Delta_V(\beta^*)^{-1}h_V + a^2h_V'\Delta_V(\beta^*)^{-1}h_V \\ &= b'\Omega_V^{*-1}b + 2ab'k_V + a^2\varphi_V, \end{aligned}$$

where we have let k_V be a vector $k_V \equiv \Sigma_V\Delta_V(\beta^*)^{-1}h_V$ and the scalar φ_V equals $\varphi_V \equiv h_V'\Delta_V(\beta^*)^{-1}h_V$. For the third term, application of the first part of Proposition 2 and Slutsky's theorem gives, $\hat{\Omega}_V^{-1/2} \cdot T^{1/2}(\hat{\beta}-\beta^*) \xrightarrow{d} N(0,I)$. Hence, $T\beta^{*\prime}\hat{\Omega}_V^{-1}(\hat{\beta}-\beta^*) \xrightarrow{d} N(0,s)$ where

$$\begin{aligned} s &= (b + a\Sigma_V^{-1}h_V)'\Omega_V^{*-1}(b + a\Sigma_V^{-1}h_V) \\ &= m. \end{aligned}$$

Therefore, $T\hat{\beta}'\hat{\Omega}_V^{-1}\hat{\beta} \xrightarrow{d} \chi^2_d + m + N(0,m)$, which completes the proof of Proposition 2. ■

Figure 1

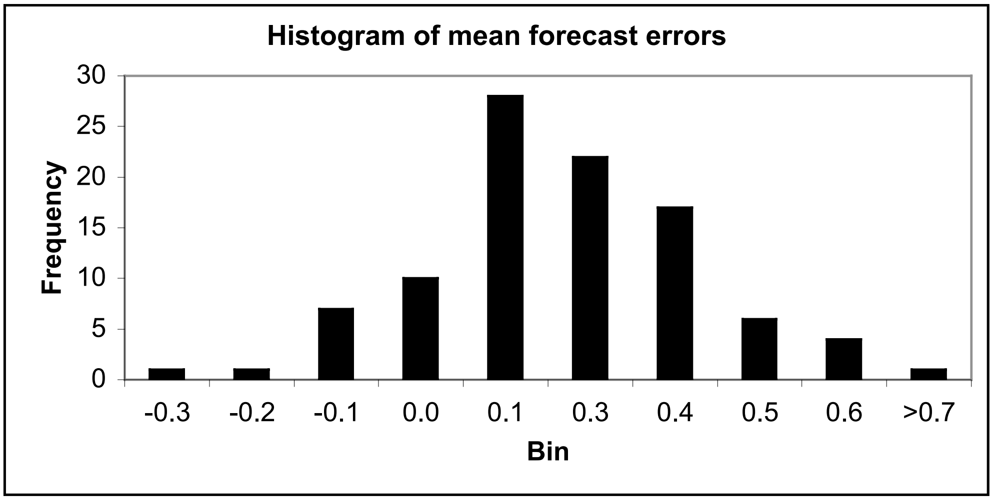


Figure 2A: Combination of asymmetry (a) and bias parameter (b) leading to identical power of rationality test

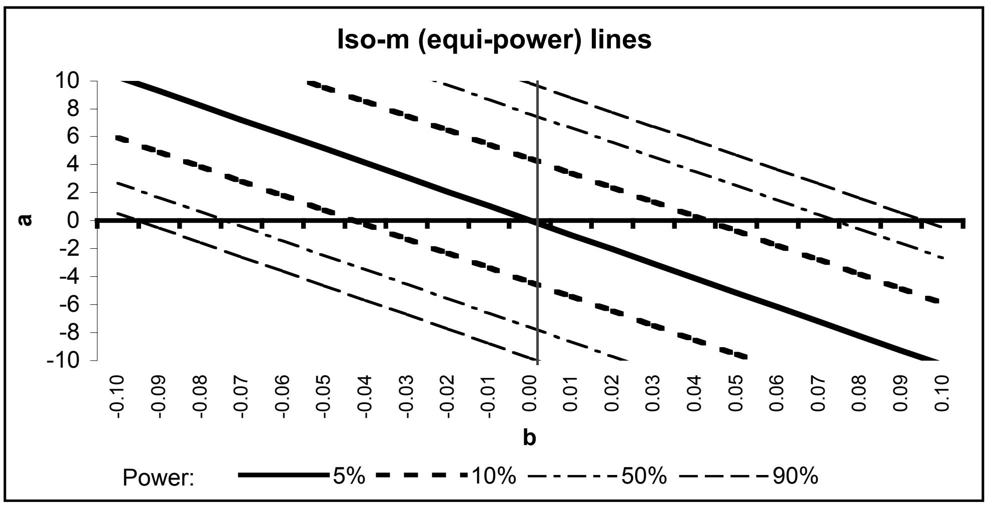


Figure 2B: Combination of asymmetry (alpha) and bias parameter (beta in s.e. units) leading to identical power of rationality test

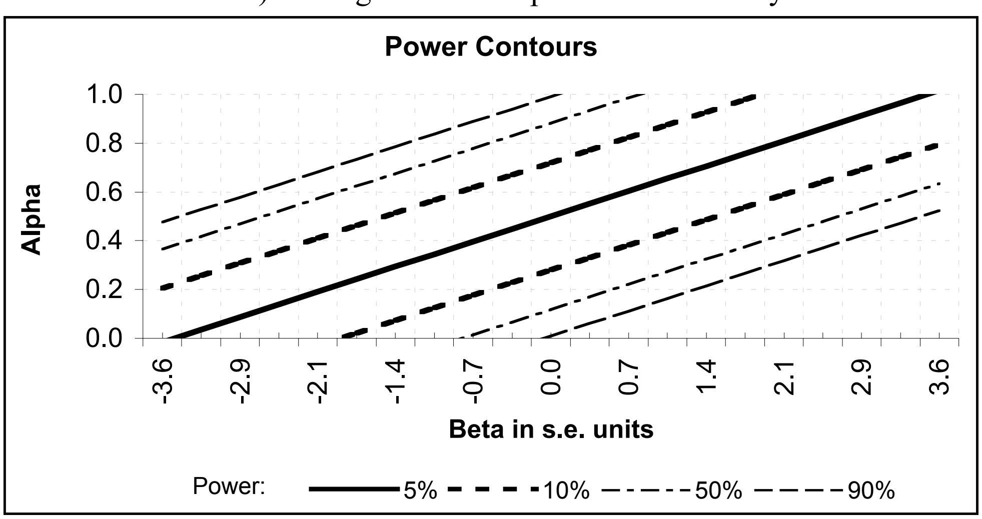


Figure 3

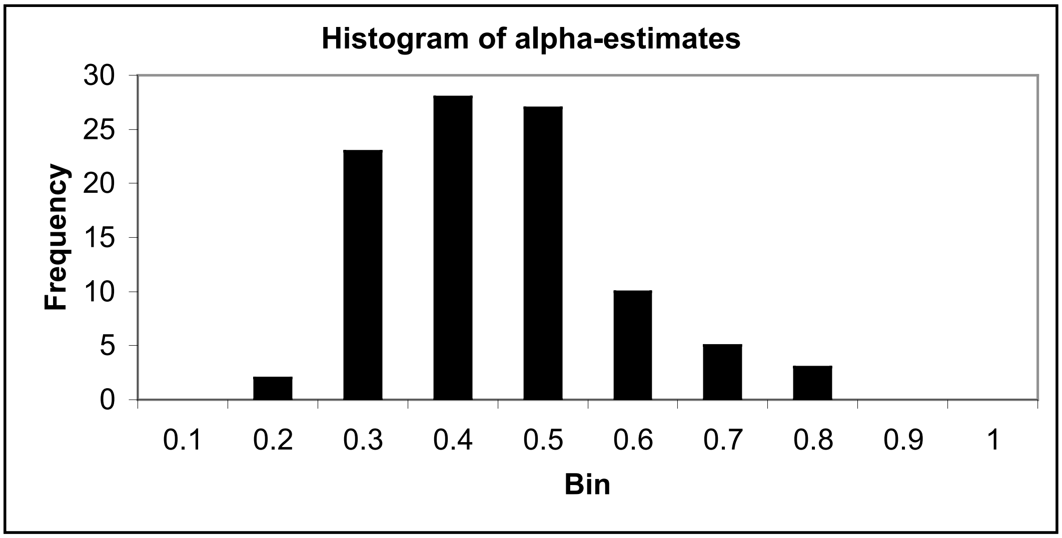


Figure 4

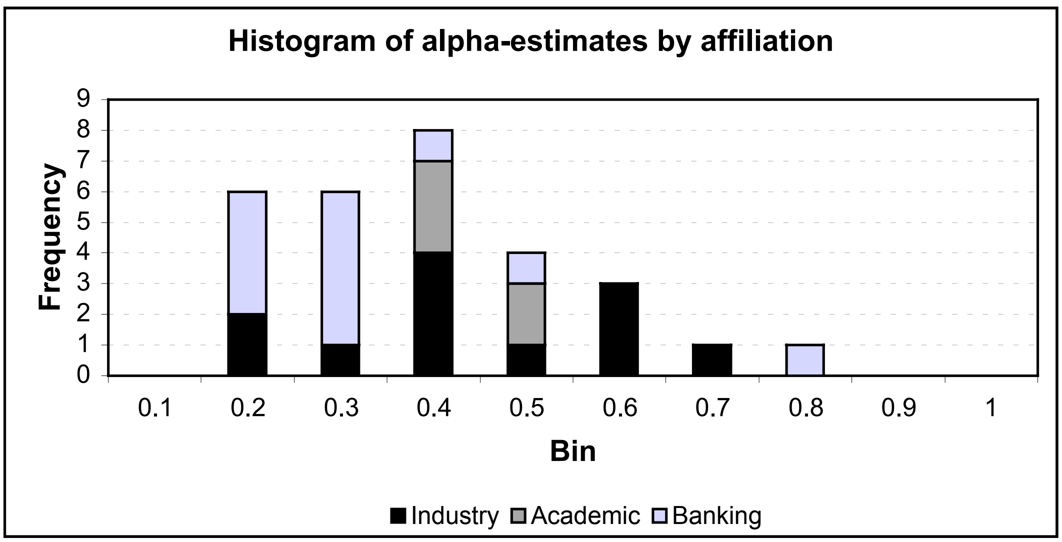


Table 1: Tests for bias under MSE loss

p-value	Unbiasedness and $\alpha = ½$
<1%	16
<5%	29
<10%	33

Note: this table reports the number of forecasters (out of a total of 98) for whom the null of symmetric loss ($\alpha = ½$) could be rejected at the specified critical values.

Table 2: J-tests of rationality and symmetry of the loss function (Quad-Quad)

	Rationality and $\alpha = ½$			Rationality and α unconstrained		
Range	<1%	<5%	<10%	<1%	<5%	<10%
Inst = 1	13	34	39	1	8	19
Inst = 2	15	30	33	0	4	12
Inst = 3	20	34	42	0	4	11
Inst = 4	12	30	35	0	5	10

Note: this table reports the number of forecasters (out of a total of 98) for whom the null of rationality and symmetry ($\alpha = ½$) and rationality alone (α unconstrained) could be rejected at the specified critical values. The instruments are as follows:
Inst = 1: constant plus lagged errors
Inst = 2: constant plus lagged actual values
Inst = 3: constant plus lagged errors and actual values
Inst = 4: constant plus lagged absolute errors