



IMF Working Paper

Consensus Forecasts and Inefficient Information Aggregation

Christopher Crowe

IMF Working Paper

Research Department

Consensus Forecasts and Inefficient Information Aggregation

Prepared by Christopher Crowe¹

Authorized for distribution by Mr. Atish Ghosh

July 2010

Abstract

This Working Paper should not be reported as representing the views of the IMF.

The views expressed in this Working Paper are those of the author(s) and do not necessarily represent those of the IMF or IMF policy. Working Papers describe research in progress by the author(s) and are published to elicit comments and to further debate.

Consensus forecasts are inefficient, over-weighting older information already in the public domain at the expense of new private information, when individual forecasters have different information sets. Using a cross-country panel of growth forecasts and new methodological insights, this paper finds that: consensus forecasts are inefficient as predicted; this is not due to individual forecaster irrationality; forecasters appear unaware of this inefficiency; and a simple adjustment reduces forecast errors by 5 percent. Similar results are found using US nominal GDP forecasts. The paper also discusses the result's implications for users of forecaster surveys and for the literature on information aggregation.

JEL Classification Numbers: C53, D84, E17

Keywords: Consensus Forecasts, Information Aggregation; Forecast Efficiency

Author's E-Mail Address: CCrowe@imf.org

¹ Subject to the usual caveats, the author is extremely grateful to Kajal Lahiri, David Romer, Allan Timmermann, and Ken Wallis, seminar participants at the IMF Research Department and George Washington University and conference participants at the European Central Bank's 6th Workshop on Forecasting Techniques for useful comments on an earlier draft, and to Prakash Loungani, Alin Mirestean, Marco Ottaviani, Mauro Roca, Hyun Shin, Natalia Tamirisa and Ken West for helpful discussions.

Contents	Page
I. Introduction	3
II. Model and Empirical Strategy.....	8
A. Basic Model and Results.....	8
B. Assumed Information Structure of Data	9
C. Empirical Methodology.....	11
III. Application to Cross-Country Growth Forecasts.....	15
A. Data	15
B. Efficiency Tests.....	16
C. Efficiency Gains from adjusted Consensus Forecasts.....	19
D. Robustness Checks.....	23
IV. Discussion.....	28
A. Morris and Shin (2002): A Reassessment.....	28
B. Groupthink, Bank Behavior and the Credit Crunch.....	32
V. Conclusions.....	33
Tables	
Table 1. Correlation Coefficients.....	4
Table 2. Naïve vs. Rational Priors	17
Table 3. Baseline Efficiency Tests.....	18
Table 4. Efficiency Tests: Iterative Error Adjustment.....	19
Table 5. In-Sample Efficiency Gains, by Country.....	21
Table 6. In-Sample Efficiency Gains, by Forecast Horizon	22
Table 7. Out of Sample Efficiency Gains, by Country	25
Table 8. Out of Sample Efficiency Gains, by Forecast Horizon	26
Table 9. Efficiency Tests: Median Forecasts.....	27
Table 10. Efficiency Tests: Real Time Growth Data.....	28
Table 11. SPF Nominal GDP Forecasts.....	29
Appendix.....	35
Appendix Tables	
Table A1. Individual Regression Results Summary.....	38
Table A2. Individual SPF Regression Results Summary	40
References.....	41

I. INTRODUCTION

Consensus forecasts—mean or median forecasts from a panel of individual forecasters—are always inefficient, as long as forecasters make individually rational forecasts based on different information sets.² However, the empirical relevance of this fact has yet to be tested. This paper provides such a test using the well-known and widely used forecaster survey dataset produced by Consensus Economics. It demonstrates that consensus forecasts are indeed inefficient, with out of sample efficiency gains (reduction in root mean square forecast error) from a simple adjustment technique of around 5 percent. The paper goes on to discuss some potential ramifications, both for users of this kind of survey data and for information aggregation more generally.

To understand this paper’s empirical strategy, one must first appreciate the intuition behind the theoretical result. Consider a simple stylized model of the forecasting environment, in which a group of agents form independent forecasts based on a private signal and a common prior. Each individual forecaster, if producing an honest minimum variance forecast, will derive his or her forecast using Bayesian updating, that is, from a weighted average of the prior and the forecaster’s private signal, where the weights reflect their relative variance. The consensus forecast will then weight the prior and the average private signal using the same relative weights (on average) as the private forecasters. But the mean private signal has a lower variance than each of the individual signals (this is the argument for employing the consensus forecast in the first place). In other words, the consensus forecast over-weights the prior, and as a result is inefficient. Notably, it is (negatively) correlated with its own forecast error. This result is a general one and does not rely on strategic forecasting behavior or heterogeneous signal precision.

One can test the empirical relevance of the result by analyzing the relationship between the *forecast errors* associated with the consensus forecast (the difference between the consensus forecast and the actual realization) and the *revisions* to the consensus forecast (the difference between the current consensus forecast and the previous one). To provide some initial evidence, Table 1 presents simple bivariate correlations between forecast updates and forecast errors, for the full sample as well as for different forecast horizons, using the Consensus Economics cross-country dataset of real GDP growth forecasts.³ The first column provides correlation coefficients for the individual forecasts, while the second provides coefficients for the consensus forecasts. Column 2 provides strong evidence for negative

²Kim, Lim and Shaw (2001) appears to be the only published reference on the theoretical finding, although an early draft of Ottaviani and Sørensen (2006) also makes reference to it (see also Wallis, forthcoming). Other authors have discussed in more general terms the fact that pooling forecasts is inferior to pooling information sets (see, e.g., Granger, 1989 and Timmermann, 2006), which is the underlying cause of the consensus forecasts’ inefficiency. A related issue—that consensus forecasts should not be used for testing individual forecaster rationality—has received more attention (Figlewski and Wachtel, 1983; Keane and Runkle, 1990; Bonham and Cohen, 2001).

³See section 3.1 for a full description of the data and sample.

correlation between consensus forecast updates and forecast errors, particularly at forecast horizons of under 13 months. Estimated negative correlation coefficients are as high as 40 percent in absolute value. Comparison with Column 1 suggests that this negative correlation for the consensus forecasts is not driven by individual forecaster irrationality.⁴

Table 1. Correlation Coefficients

Horizon	Individual	Consensus
0	0.03 ***	-0.13 ***
1	0.10 ***	-0.19 ***
2	0.05 ***	-0.21 ***
3	-0.01	-0.28 ***
4	0.06 ***	-0.24 ***
5	0.02	-0.36 ***
6	0.01	-0.28 ***
7	0.03 ***	-0.40 ***
8	0.03 ***	-0.21 ***
9	0.06 ***	-0.32 ***
10	0.06 ***	-0.25 ***
11	0.14 ***	-0.07
12	0.02	-0.35 ***
13	0.14 ***	-0.18 ***
14	0.15 ***	-0.09 *
15	0.21 ***	0.03
16	0.20 ***	-0.08 *
17	0.21 ***	-0.02
18	0.23 ***	0.03
19	0.19 ***	-0.12 **
20	0.24 ***	0.03
21	0.22 ***	-0.06
22	0.24 ***	-0.05
Pooled	0.13 ***	-0.15 ***

Correlation between forecast errors and forecast updates

Significance Level denoted by *** (1 percent); ** (5 percent); * (10 percent).

⁴There is evidence for some positive correlation for the individual forecasters, which would suggest over weighting of new information. However, this appears due to the positive bias created by pooling across forecasters with common forecast targets noted by Zarnowitz (1985) and others. The appendix outlines a method for assessing rationality among individual forecasters that is not subject to this bias, which finds some very limited evidence of under weighting of new information.

This initial evidence of consensus forecast inefficiency is subjected to a more rigorous econometric investigation later in the paper. To this end, the paper carefully outlines the assumed data structure and several important econometric factors that have hindered testing in this area. It builds on insights in the existing literature on efficiency/rationality testing using forecast data (Zarnowitz, 1985; Keane and Runkle, 1990; Batchelor and Dua, 1991; Davies and Lahiri, 1995; Bonham and Cohen, 2001) and provides some new insights. In particular, it shows that the assumed forecast structure in Davies and Lahiri (1995), while approximately correct for analyzing forecaster behavior in the context of repeated forecasts at different time horizons and multiple forecasters per horizon, is inconsistent with rational forecaster behavior in a fully specified model of the information structure. In a fully specified model, rational forecaster behavior creates additional econometric challenges, which this paper's empirical strategy addresses.

The results of this empirical exercise provide unambiguous support for the intuition given in Table 1. There is very strong evidence that consensus forecasts are indeed inefficient, and that this inefficiency is quantitatively important, particularly at forecast horizons of around 12 months or less. These results are robust to different definitions of the consensus forecast, different vintages of the actual data and different forecast datasets. Notably, very similar results are obtained using forecasts of quarterly nominal U.S. GDP from the Survey of Professional Forecasters.

These empirical results have a number of interesting implications. Consensus forecasts—the mean or median forecast from a panel of professional forecasters—are a widely used tool for researchers and policymakers. For instance, central banks frequently present consensus private sector forecasts alongside their own in-house forecasts as a means of increasing the credibility of their own forecasts and policymaking.⁵ Consensus forecasts are also employed in the academic literature as proxies for expectations more broadly (e.g. Johnson, 2002) or as a benchmark against which to test individual public or private sector forecasters (Barakchian and Crowe, 2009; Romer and Romer, 2000). However, this paper argues that more caution should be exercised when using consensus forecasts. In particular, the consensus forecast does not typically represent the best (most efficient) estimate of the true state of the world, conditional on the aggregate information set. Its use as a benchmark may therefore flatter the performance of alternative forecasts.

This paper also has implications for the debate over the efficacy of releasing more transparent public information sparked in part by the contribution of Morris and Shin (2002) and, in the policy sphere, by moves toward greater transparency by central banks, notably in

⁵Examples include the ECB's Survey of Professional Forecasters (<http://www.ecb.int/stats/prices/indic/forecast/html/index.en.html>); and the Survey of Professional Forecasters carried out by the Federal Reserve Bank of Philadelphia (<http://www.phil.frb.org/research-and-data/real-time-center/survey-of-professional-forecasters/>), which is employed in a robustness check in this paper. The Bank of England also carries out a quarterly survey of external forecasters (see, for instance, the August 2009 *Inflation Report*, p.50).

the context of inflation targeting (IT) regimes. One can show that—even without Morris and Shin’s Beauty Contest component to the forecasting process—more transparent public information can increase the forecast error for the *consensus* forecast, even as the forecast errors of *individual* forecasts are reduced.⁶ This lesson that better public information can lead to worse collective judgments is a general one with many applications: the paper discusses an application to the U.S. financial crisis.

This paper’s results can also be related to discussions of groupthink or collective delusion (e.g., Benabou, 2008). The model outlined here differs from models of social learning that generate herd behavior (e.g. Banerjee, 1992) in that the inefficiency of the consensus forecast does not stem from forecasts being made public sequentially, but from the communication of posterior beliefs rather than raw signals. Nevertheless, it shares the core feature of these models identified by Benabou (2008): the key problem is a failure to aggregate private signals and its cure resides in more communication. A subtle insight from this model is that the *form* of communication is central to efficient aggregation. Agents must be induced to communicate not their best guess, but rather their idiosyncratic information. Moreover, communication may not be sufficient: the empirical evidence suggests that agents fail to optimally use the aggregate information that does become available.

Finally, this paper’s results point to an additional kind of aggregation problem—moving between microeconomic and macroeconomic models. For instance, it may be invalid to use macroeconomic or average data to derive insights into individual behavior: as demonstrated in this paper, the empirical finding that consensus forecasts are inefficient does not imply that *individual* forecasts are not rational. More significantly, the results suggest that insights based on individual behavior do not necessarily carry over to a macroeconomic (multi-agent) environment. Even when agents’ behavior imposes no externality, insights derived from individual rationality (e.g. that expectations are rational) may not hold in aggregate. The economy does not behave as if it were populated by a single rational representative agent.⁷ Since asset prices also reflect the aggregation of idiosyncratic private information, the results in this paper could also shed light on asset price phenomena (for instance, the success of momentum trading strategies detailed in Jegadeesh and Titman, 1993 and 2001, and Hong and Stein, 1999), although a treatment of this subject is beyond the scope of the current paper.

Section II provides a formal discussion of the paper’s key argument. It goes on to demonstrate that an efficient forecast can be recovered by reducing the overweighting on the

⁶Amato and Shin (2006) and Roca (2009) arrive at similar results in a New Keynesian DSGE setting. By contrast, this paper analyzes the impact of public information on consensus forecasts in the context of Morris and Shin’s (2002) global games model, and its special case (without the beauty contest element)—the simple signal extraction model with public and private signals.

⁷The fact that this aggregation problem shows up with respect to expectations is particularly worrisome given the central role ascribed to rational expectations in most modern macroeconomic models.

prior using a linear combination of the consensus forecast and the prior. The size of the required adjustment is increasing in the relative variance of the individual forecasters' idiosyncratic private information.⁸ This section also illustrates how the paper's predictions can be formulated as empirical hypotheses in the context of the data, and discusses some important econometric concerns that additionally dictate the empirical strategy.

Section III illustrates the argument using a large dataset of individual monthly forecasts of annual economic growth for a cross-section of countries between 1989 and 2008. It describes the data, then presents evidence of the existence and size of the inefficiency associated with employing consensus forecasts for different countries and time horizons. There is strong evidence of inefficiency for consensus forecasts, much less so for individual forecasts, as predicted by the model.⁹ The results are robust to employing a median rather than a mean forecast as the measure of the consensus forecast, to using real time (early) GDP estimates rather than revised data to measure the actual GDP out-turn, and to using an alternative dataset (nominal GDP data for the US from the *Survey of Professional Forecasters*). Using the adjustment method outlined above and employing the most recent previous consensus forecast as a proxy for the prior, one can obtain adjusted forecasts with mean squared errors on average more than 5 percent lower than the raw consensus forecasts, for forecasting horizons of a year or less, even out of sample. This effect may appear small; however, the reduction in forecast errors associated with the adjustment is typically greater than the reduction associated with using the next month's consensus forecasts (i.e., with one month's additional data available to forecasters). Moreover, since consensus forecasts are generally thought to be at, or near, the frontier in terms of forecast accuracy (see, e.g., Ang, and others, 2007, for an application to inflation forecasts), the fact that a simple adjustment technique can reduce forecast errors *at all* is significant. However, there is no evidence that forecasters make the necessary adjustment to the previous month's consensus forecast in updating their prior.

Section IV provides a fuller discussion of how the results relate to some of the existing literature discussed above. Section V then concludes with a discussion of implications for optimal forecasting methods and the use of consensus forecasts by central banks and other institutions.

⁸This section replicates the key results in Kim, Lim and Shaw (2001) for the case with a continuum rather than a finite number of individual forecasters.

⁹The difference in performance between consensus and individual forecast rationality tests has been noted before. For instance, Batchelor and Dua (1991) argue that rationality tests on the consensus forecasts provide a poor guide to the extent of rationality among individual forecasters. Similarly, Harvey, and others, (2001) note the strong autocorrelation of revisions to consensus forecasts, which, they argue, imply that the consensus forecasts cannot be optimal. However, this phenomenon was far less apparent when the records of particular individual forecasters were examined.

II. MODEL AND EMPIRICAL STRATEGY

A. Basic Model and Results

The following model captures the basic features of this discussion. A variable y is distributed iid with mean μ and variance σ_v^2 :

$$y = \mu + v. \quad (1)$$

A continuum of agents $i \in [0, 1]$ each receives a private signal of y , given by x_i :

$$x_i = y + u + \varepsilon_i, \quad (2)$$

where u is an aggregate error, iid mean-zero with variance σ_u^2 and ε_i are iid mean-zero idiosyncratic errors with variance σ_ε^2 . The information structure, including all variances, is common knowledge.

Agents form forecasts following the usual Bayesian updating:

$$f_i = F_i x_i + (1 - F_i) \mu, \quad (3)$$

where F_i minimizes the root mean squared forecast error and is therefore given by:

$$F_i = F = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_u^2 + \sigma_\varepsilon^2} \forall i \quad (4)$$

Proposition 1: The consensus forecast is inefficient, even as individual forecasts are all efficient.

Proof: The individual forecasts are uncorrelated with their forecast errors:

$$\begin{aligned} E[f_i(f_i - y)] &= E[(\mu + F(u + v + \varepsilon_i))(F(u + \varepsilon_i) - (1 - F)v)] \\ &= F(F(\sigma_u^2 + \sigma_\varepsilon^2) - (1 - F)\sigma_v^2) \\ &= 0 \end{aligned} \quad (5)$$

Define the consensus forecast as the mean forecast:

$$\bar{f} = E[f_i] = F(y + u) + (1 - F)\mu \quad (6)$$

Then the consensus forecast is correlated with its forecast errors:

$$\begin{aligned} E[\bar{f}(\bar{f} - y)] &= E[(\mu + F(u + v))(Fu - (1 - F)v)] \\ &= F(F\sigma_u^2 - (1 - F)\sigma_v^2) \\ &= -F^2\sigma_\varepsilon^2 \leq 0 \end{aligned} \quad (7)$$

Q.E.D.

This result applies equally to the median if ε_i is symmetric (or, more generally, if its median is zero), and generalizes to the case with heterogeneous σ_ε^2 or a finite number of agents (Kim, Lim and Shaw, 2001 present a proof for the latter case). The easiest way to understand this result is to compare the individually-rational relative weight F on the individual signal $(y + u + \varepsilon_i)$ with the optimal weight F^* that would arise from minimizing the root mean squared forecast error associated with the aggregate signal $(y + u)$:

$$F^* = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_u^2} = \left(1 + \frac{\sigma_\varepsilon^2}{\sigma_v^2 + \sigma_u^2}\right)F > F \quad (8)$$

Because the aggregate signal has a lower variance than the individual signals, it should receive a higher weight. Then the optimal forecast f^* is uncorrelated with its forecast errors:

$$\begin{aligned} E[f^*(f^* - y)] &= E[(\mu + F^*(u + v))(F^*u - (1 - F^*v))] \\ &= F^*(F^*\sigma_u^2 - (1 - F^*)\sigma_v^2) \\ &= 0 \end{aligned} \quad (9)$$

If the prior μ and the variances of the various shocks are known, then once the consensus forecast is observed the optimal forecast can be derived as a linear combination of the consensus forecast and the prior:

$$\begin{aligned} f^* &= \bar{f} + \frac{F^* - F}{F}(\bar{f} - \mu) \\ &= \bar{f} + \frac{\sigma_\varepsilon^2}{\sigma_v^2 + \sigma_u^2}(\bar{f} - \mu) \end{aligned} \quad (10)$$

B. Assumed Information Structure of Data

To test the main proposition of the paper, one must first modify the model to match the information structure of the data. The key difference between the simple model outlined in the previous section and the forecasting environment of the data is that repeated forecasts are made for each forecast period. Hence, for each forecast at horizon $h \in \{1, 2, \dots, 23\}$ for country c and period τ , the public information reflected in the prior $\mu_{ch\tau}$ will include previously published consensus forecasts.¹⁰

The assumed information structure is similar to that in Davies and Lahiri (1995). The key

¹⁰Forecasts are available monthly, and are for annual data up to two years ahead. Hence, there are 24 forecasts for each country and time period, with a forecast horizon of between 1 and 24 months. The structure of the data is described in more detail in section III.A.

difference is that this paper derives forecasters' optimal forecasts based wholly on the underlying information structure, whereas Davies and Lahiri have an explicit information structure for common shocks but assume idiosyncratic forecasting errors rather than deriving forecast dispersal based on idiosyncratic signals. In fact, one can show that the forecast structure assumed by Davies and Lahiri is inconsistent with optimal forecasts. Once one derives dispersed signals endogenously, then optimal forecasting behavior with respect to the common shocks changes, and this then has implications for the behavior of consensus forecasts. With respect to more cosmetic changes, this paper ignores forecaster-specific bias (since this paper is not concerned with testing for bias) and also allows for a final shock to the variable outside the forecasting window (to capture forecast revisions).

The forecast target, output y_{ct} , is assumed to be a function of 25 iid monthly shocks, one per month during the two year maximum forecast horizon and a final shock after the forecast horizon is complete due to data revision. That is:

$$y_{ct} = \sum_{h=0}^H v_h, \\ v_h \sim iid(0, \sigma_v^2) \quad (11)$$

where h indexes the forecast horizon (the horizon 0 shock is due to data revision, and $H = 24$ since the maximum forecasting horizon is two years).

At the start of each month, agents i observe two signals of y_{ct} . The first is a common prior μ_h , based on the shocks observed through the previous month:¹¹

$$\mu_h = \sum_{j=h+1}^H v_j. \quad (12)$$

Agents also receive an idiosyncratic, mean-zero signal of the current month's shock, which provides a second signal of y_{ct} :

$$x_{ih} = \mu_h + v_h + \varepsilon_{ih} \quad (13)$$

With this information structure, agents' optimal forecast of y_{ct} is given by:¹²

¹¹To match the apparent data generating process (DGP) more closely one could assume heterogeneous priors. However, the model outlined here approximates the DGP reasonably well while maintaining traditional assumptions about forecaster rationality.

¹²Davies and Lahiri (1995) implicitly set $F = 1$ (that is, forecasters fully absorb the common shock on average) but forecasts are also dispersed around the mean. This is inconsistent with optimal forecasting behavior: the common shock v_h will only be fully absorbed if it is observed without any noise, otherwise it is optimal to place some weight on the prior ($v_h = 0$). If forecasters have access to a signal without noise, then they will ignore any noisy signal and the noise term ε_{ih} should therefore not enter into their forecast. The forecast structure that Davies and Lahiri propose is consistent only with a measurement error explanation for

$$f_h^i = \mu_h + F(v_h + \varepsilon_{ih}),$$

$$F = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_\varepsilon^2} \quad (14)$$

with consensus forecasts denoted:

$$\tilde{f}_h = \mu_h + Fv_h \quad (15)$$

Here, the key characteristic of consensus forecasts—underreaction to aggregate new information—is clearly observable. By contrast, the optimal forecast based on the aggregate signal fully absorbs the new shock v_h :

$$f_h^* = \mu_h + v_h = \tilde{f}_h + \frac{(1-F)}{F}(\tilde{f}_h - \mu_h) \quad (16)$$

C. Empirical Methodology

This subsection outlines the empirical methodology for testing for consensus forecast inefficiency, within the context of the assumed information structure. The Appendix has a parallel discussion with respect to testing for individual forecast rationality, where similar issues arise. The first step is to define the consensus forecasts efficiency test, which is based on a regression of forecast errors on the difference between the forecast and the prior. Under the hypothesis that consensus forecasts under-react to new information, the estimated slope coefficient will be negative.

There are a number of challenges to implementing this test in practice. The first is arriving at a suitable empirical proxy for the prior (which is unobserved by the researcher). I outline two cases, a ‘naive priors’ case in which forecasters are assumed to use the previous consensus forecast, and a ‘rational priors’ case in which forecasters use the actual prior μ_h as in the model. Testing for forecast efficiency is simpler in the former case than in the latter, as under rational priors *all* past updates to forecasts, not only the most recent update, must be included in the regression. One can demonstrate that falsely rejecting the rational case in favor of the naive case leads one to under-estimate the slope coefficient in absolute terms, so that in this case a significant negative coefficient can still be treated as evidence for inefficiency. The naive priors case is therefore treated as the baseline for the empirical analysis. This decision is supported by the empirical analysis, which finds little evidence for the rational priors case.

forecast dispersal: i.e., all forecasters have the same forecast and differences simply reflect errors in reporting forecasts to the survey. This seems implausible. Note that optimal forecasts have properties that differ from the assumed forecast structure in Davies and Lahiri. In particular, they have a richer autocorrelation structure: the common component follows an MA(1) process rather than being white noise. This poses additional econometric challenges for rationality tests, beyond those identified by Davies and Lahiri.

The unique panel structure of forecast data poses additional econometric challenges, which are also addressed here. Finally, this section outlines a measure of the inefficiency embedded in consensus forecasts.

In the context of the information structure outlined above, the assertion in that consensus forecasts are inefficient still holds:

$$E\left[\left(\tilde{f}_{ch\tau} - \mu_{ch\tau}\right)\left(\tilde{f}_{ch\tau} - y_{c\tau}\right)\right] = -F(1-F)\sigma_v^2 \leq 0 \quad (17)$$

To test this hypotheses in a regression setting, note that:

$$\begin{aligned} \tilde{f}_{ch\tau} - y_{c\tau} &= \tilde{f}_{ch\tau} - f_{ch\tau}^* + f_{ch\tau}^* - y_{c\tau} \\ &= -\left(\frac{1-F}{F}\right)\left(\tilde{f}_{ch\tau} - \mu_{ch\tau}\right) + e_{ch\tau}, \end{aligned} \quad (18)$$

where $e_{ch\tau} = -\sum_{j=0}^{h-1} v_h$ is orthogonal to $\tilde{f}_{ch\tau} - \mu_{ch\tau}$ (and indeed to anything known through horizon h).

As discussed above, one can make two alternative assumptions about the prior $\mu_{ch\tau}$ that is adopted in each forecasting period. The first assumption is that forecasters know the true prior μ_h , either because they perfectly observe the flow of past shocks v_{h+1}, \dots, v_H or because they adjust the previous month's consensus forecast to reflect its underreaction to new information (this is the 'rational priors' case). The second assumption is that forecasters naively use the previous consensus forecast ($\mu_{ch\tau} = \tilde{f}_{c(h+1)\tau}$), perhaps due to bounded rationality (the 'naive priors' case).

Assuming that forecasters know the true prior, then substituting observables for $\mu_{ch\tau}$ yields:

$$\begin{aligned} \tilde{f}_{ch\tau} - y_{c\tau} &= -\left(\frac{1-F}{F}\right)\left(\tilde{f}_{ch\tau} - \tilde{f}_{c(h+1)\tau}\right) + \left(\frac{1-F}{F}\right)^2\left(\tilde{f}_{c(h+1)\tau} - \tilde{f}_{c(h+2)\tau}\right) \\ &\quad - \left(\frac{1-F}{F}\right)^3\left(\tilde{f}_{c(h+2)\tau} - \tilde{f}_{c(h+3)\tau}\right) + \dots + e_{ch\tau} \\ &= \sum_{j=h}^{H-1} B_{jh} \Delta \tilde{f}_{cj\tau} + e_{ch\tau}; B_{jh} = \left(-\left(\frac{1-F}{F}\right)\right)^{j+1-h} \end{aligned} \quad (19)$$

That is, forecast errors should be a function of *all* past forecast updates $\Delta \tilde{f}_{ch\tau}, \dots, \Delta \tilde{f}_{c(H-1)\tau}$. With naive priors, forecast errors are a function only of the most recent forecast update:¹³

¹³In this case, if forecasters are otherwise rational (conditional on using the wrong prior), then F need no longer be constant over h , even assuming that the underlying shocks are constant. For simplicity, this paper assumes that forecasters use the wrong prior but weight new information as in the rational case, based on the relative variance of the true prior.

$$\bar{f}_{ch\tau} - y_{c\tau} = -\left(\frac{1-F}{F}\right)\left(\bar{f}_{ch\tau} - \bar{f}_{c(h+1)\tau}\right) + e_{ch\tau} \quad (20)$$

If the rational case is falsely rejected, assuming the naive case will introduce some bias if forecasters actually use the true prior, since $\bar{f}_{c(h+1)\tau}$ is correlated with the error term:¹⁴

$$\begin{aligned} \bar{f}_{ch\tau} - y_{c\tau} &= -\left(\frac{1-F}{F}\right)\left(\bar{f}_{ch\tau} - \bar{f}_{c(h+1)\tau}\right) + \frac{(1-F)^2}{F}v_{h+1} - \sum_{j=0}^{h-1} v_j \\ &= -\left(\frac{1-F}{F}\right)\Delta\bar{f}_{ch\tau} + \hat{e}_{ch\tau} \end{aligned} \quad (21)$$

$$\text{where } Cov\left(\left(\bar{f}_{ch\tau} - \bar{f}_{c(h+1)\tau}\right), \hat{e}_{ch\tau}\right) = \frac{(1-F)^3}{F}\sigma_v^2 \geq 0$$

In this case, a regression of the form:

$$\bar{f}_{ch\tau} - y_{c\tau} = a - b\Delta\bar{f}_{ch\tau} + e_{ch\tau} \quad (22)$$

would yield the following (biased) coefficient estimate:

$$E[b] = \frac{(1-F)F}{F^2 + (1-F)^2} < \frac{1-F}{F} \quad (23)$$

Note that the direction of bias is towards zero.

The following regressions are therefore run to assess consensus forecast efficiency. The most general specification includes all lagged forecast updates:

$$\bar{f}_{ch\tau} - y_{c\tau} = a_h + \sum_{j=h}^{H-1} B_{jh}\Delta\bar{f}_{cj\tau} + e_{ch\tau}. \quad (24)$$

The naive priors case corresponds to the special case:

$$\bar{f}_{ch\tau} - y_{c\tau} = a_h - b_h\Delta\bar{f}_{cj\tau} + e_{ch\tau}, \quad (25)$$

while the rational priors case implies a distinctive pattern of alternating signs predicted for the B_{jh} coefficients:

$$B_{hh} \leq 0, B_{(h+1)h} \geq 0, B_{(h+2)h} \leq 0, B_{(h+3)h} \geq 0, \dots \quad (26)$$

While the first inequality is implied by the model with both rational and naive priors, the subsequent inequalities are necessary under the rational priors case. Hence, evidence against

¹⁴This is distinct from the case with testing individual forecaster rationality, where $\bar{f}_{c(h+1)\tau}$ serves as a valid empirical proxy even if forecasters use the actual prior, assuming rational forecasts (see the Appendix).

significant coefficients on lagged forecast updates and this pattern of alternating signs would lead one to reject the hypothesis that priors are rational, in favor of the naive case.

Under both rational and naive priors, the error terms $e_{ch\tau}$ have a complex structure over forecast horizons h and adjacent forecast periods τ . Because $e_{ch\tau} = -\sum_{j=0}^{h-1} v_j$ includes only subsequent shocks that are orthogonal to information through horizon h , the errors are contemporaneously uncorrelated with the regressors. However, neither (24) nor the restricted version (25) can be estimated over a sample that pools across forecast horizons, since $e_{ch\tau}$ includes subsequent shocks that are correlated with the regressors when subsequent forecasts are included (the Appendix provides a further discussion of this issue, which is common to consensus and individual forecasts).

Parameter estimates from separate regressions (per h) should yield consistent estimates. However, to improve the efficiency of our baseline estimates this paper adopts the following iterative procedure. Assuming that the naive priors case is the appropriate one, (25) is estimated for $h = 1$. In this case,

$$e_{c1\tau} = -v_0 \quad (27)$$

One can then subtract the estimated residuals $\hat{e}_{c1\tau}$ from the left hand side of (25) estimated for $h = 2$:

$$\begin{aligned} \tilde{f}_{c2\tau} - y_{c\tau} - \hat{e}_{c1\tau} &= a_2 - b_2 \Delta \tilde{f}_{c2\tau} + e_{c2\tau}, \\ \text{where } e_{c2\tau} &= -v_1. \end{aligned} \quad (28)$$

The residuals $\hat{e}_{c2\tau}$ can then be used as a proxy for $-v_2$ in equation estimated for $h = 3$:

$$\begin{aligned} \tilde{f}_{c3\tau} - y_{c\tau} - \hat{e}_{c1\tau} - \hat{e}_{c2\tau} &= a_3 - b_3 \Delta \tilde{f}_{c3\tau} + e_{c3\tau}, \\ \text{where } e_{c3\tau} &= -v_2, \end{aligned} \quad (29)$$

and so on. Since the iterated use of residuals could generate more noise, results from this procedure are presented alongside results from simple regressions of (25).

Finally, an estimate of the efficient forecast $f_{ch\tau}^*$ can be obtained by transforming the consensus forecast using the estimated regression coefficient from (25):¹⁵

$$\hat{f}_{ch\tau}^* = \tilde{f}_{ch\tau} + b_h \Delta \tilde{f}_{ch\tau} \quad (30)$$

To provide a metric for the efficiency gain that can be obtained via this transformation, one can compare the root mean square error (RMSE) of the raw forecasts $\tilde{f}_{ch\tau}$ with the RMSE of

¹⁵Since the theory predicts that consensus forecasts are unbiased (as long as the underlying individual forecasts are unbiased), the adjustment does not take into account the estimated bias coefficient a_h .

the adjusted forecasts:

$$\text{Efficiency Gain (\%)} = 100 \times \frac{RMSE(\tilde{f}_{ch\tau}) - RMSE(\hat{f}_{ch\tau}^*)}{RMSE(\tilde{f}_{ch\tau})} \quad (31)$$

This measure can be compared across forecast horizons and countries to ascertain in which environments the adjustment is most useful for improving forecasting efficiency.

III. APPLICATION TO CROSS-COUNTRY GROWTH FORECASTS

A. Data

This study employs forecasts of economic growth for 38 countries over the period 1989-2008 at the level of the individual forecaster, collected by Consensus Economics. Forecasters are surveyed at the start of each month and their responses are published around the 10th of the month. Country coverage is more limited at the start of the sample period; moreover, some countries are surveyed only once every two months. The forecasts are for annual economic growth over the current and next calendar year: the forecast horizon (dated to the end of the forecast year) therefore varies from 1 to 24 months. The number of forecasters per country, forecast period and forecast horizon varies between 5 and 39, with an interquartile range of 14 to 21 and a mean value of around 18. The baseline sample drops outliers from the underlying dataset of individual forecasters, to prevent coding errors or extreme values from distorting the results.¹⁶ The dataset includes 167,802 individual forecasts. Efficiency gains are calculated using a truncated sample of forecasts up to 2006, in order to allow the calculation of gains out of sample. The out of sample dataset includes all forecasts for 2008 (made in 2007 and 2008) and forecasts for 2007 made in 2007 (i.e. with a forecast horizon of less than 12 months).

Consensus forecasts are obtained by taking a simple arithmetic average across all forecasters for the forecast year (τ), time horizon (h) and country (c) in question (as a robustness check, Section III.D presents results using the median). The consensus dataset is therefore a three-dimensional panel covering up to 38 countries, 20 years and 24 forecast horizons. Forecast errors are calculated with respect to annual GDP growth in the calendar year in question, $y_{c\tau}$ taken from the IMF's *World Economic Outlook* (WEO) database. The baseline results ignore questions of data revision by using final data.¹⁷

¹⁶Outliers are defined as the bottom and top 0.5 percent of the sample of individual forecast errors. In fact the results are very similar when these outliers are not dropped.

¹⁷However, as a robustness check I assess results using GDP estimates for each year τ taken from the May, year $\tau + 1$, version of the IMF's WEO dataset as a measure of real time data, following Loungani (2001). Results are reported in section III.D. Assuming that data revisions are unpredictable, using final rather than real time data should not bias the estimates, and the results of the robustness exercise support this conclusion.

B. Efficiency Tests

This subsection presents the results of the efficiency tests outlined in section II.C. First, it estimates equation (24), including the full set of lagged forecast updates as in the rational priors case, to assess evidence of the aggregation problem. These results suggest that the rational priors case can be rejected in favor of the simpler naive priors case, and the remainder of the results assume this latter case is the relevant one. The baseline results under naive priors are then presented, both with and without the adjustment to the dependent variable using residuals from regressions for subsequent forecast horizons outlined in section II.C.

Table 2 presents a summary of results from estimating regression equations including the full set of lagged forecast updates, as in (24). This table presents the weighted mean beta coefficient from the set of relevant per-forecast horizon regressions (weighted by the inverse of the variance of the coefficient estimates).¹⁸ Significance levels are based on a z-test using the standard error of the weighted mean. The last two columns give the proportion of relevant per-forecast horizon regressions in which the beta coefficient is found to be significant at the 1 percent level and to have a positive and negative sign, respectively. The predicted sign under the rational priors case is highlighted in bold.

Looking at the mean beta coefficient estimates, only two additional lagged forecast updates (in addition to the most recent update) are statistically significant, and of these one has the wrong sign. A similar pattern emerges from the raw per-forecast horizon results. Hence, the evidence is not supportive of the rational priors case. In terms of forecaster behavior, this result suggests that forecasters do not adjust the previous month's consensus forecast in updating their forecasts, but rather use it naively. In fact, the significant negative average coefficient on the first lagged update (as opposed to the positive coefficient predicted under rational priors) suggests that even the raw consensus forecast is not fully absorbed into the prior in the first month. Given the evidence against rational priors, the remainder of the results therefore focus on the simpler naive priors case, and include only the most recent forecast update.

¹⁸The number of regressions from which the underlying beta coefficients are drawn varies across updates at different lags. E.g., the update at lag 1 is included in all 23 underlying regressions, while the update at lag 23 is included only in the regression at $h = 1$. Note that these results exclude observations for countries and years with bimonthly forecasts, since this specification requires the full set of 23 forecast horizons.

Table 2. Naïve vs. Rational Priors

Update at lag:	Beta coefficient:	Proportion of significant coefficients	
	weighted mean	Positive	Negative
1	-0.686 ***	0.00	0.30
2	-0.355 **	0.00	0.14
3	-0.051	0.10	0.10
4	0.153	0.05	0.05
5	0.146	0.11	0.05
6	0.071	0.11	0.06
7	0.105	0.24	0.00
8	0.120	0.25	0.00
9	0.098	0.07	0.00
10	0.080	0.14	0.00
11	0.071	0.15	0.00
12	0.037	0.17	0.00
13	0.003	0.18	0.00
14	0.116	0.00	0.00
15	0.194	0.00	0.00
16	0.278 *	0.00	0.00
17	0.189	0.00	0.00
18	0.084	0.00	0.00
19	-0.262	0.00	0.00
20	-0.170	0.00	0.00
21	-0.163	0.00	0.00
22	-0.137	0.00	0.00
23	0.021	0.00	0.00

This table summarizes the results of 23 regressions by forecast horizon. Each regression includes the full list of available forecast updates, plus a constant term.

The mean coefficient is weighted by the inverse of the variance.

Significance Level denoted by *** (1 percent); ** (5 percent); * (10 percent);

Based on a z-test of the null that the weighted mean is equal to zero.

"Positive" and "Negative" give the proportion of individual coefficients that are statistically significant at the 1 percent level, and >0 and <0 respectively.

Standard errors clustered by country.

The predicted sign under rational priors is highlighted in **bold**.

Tables 3 and 4 present the key results. They show regression results for univariate specifications with the consensus forecast errors as the dependent variable and the most recent update to the consensus forecast as the independent variable for each of the 23 forecast horizons, based on the methodology outlined in Section II C. Table 3 presents results for simple regressions with the raw forecast error as the dependent variable, while Table 4 presents results in which the forecast error is adjusted using residuals from prior forecast periods.¹⁹

¹⁹The results in Table 4 exclude countries and years with bimonthly forecasts, since the adjustment

Table 3. Baseline Efficiency Tests

Horizon	1	2	3	4	5	6	7	8
Update	-0.605*** (0.176)	-1.145*** (0.299)	-1.150*** (0.177)	-1.057*** (0.205)	-1.521*** (0.270)	-1.963*** (0.380)	-1.218*** (0.438)	-3.442*** (0.566)
Constant	-0.430** (0.203)	-0.514*** (0.139)	-0.512** (0.206)	-0.476*** (0.141)	-0.459** (0.194)	-0.493*** (0.153)	-0.422** (0.197)	-0.478*** (0.161)
Observations	475	495	468	489	466	489	459	470
R ²	0.016	0.035	0.046	0.08	0.06	0.129	0.076	0.158
Horizon	9	10	11	12	13	14	15	16
Update	-1.065*** (0.292)	-2.115** (0.779)	-1.572*** (0.432)	-0.253 (0.448)	-3.078*** (0.654)	-2.228** (0.976)	-0.632 (0.391)	0.226 (0.583)
Constant	-0.397* (0.199)	-0.549*** (0.166)	-0.420** (0.185)	-0.546*** (0.180)	-0.364* (0.203)	-0.582*** (0.176)	-0.305 (0.216)	-0.360* (0.192)
Observations	446	455	434	459	429	449	421	441
R ²	0.044	0.102	0.063	0.005	0.125	0.032	0.007	0.001
Horizon	17	18	19	20	21	22	23	
Update	-0.764 (0.881)	-0.259 (0.871)	0.267 (0.671)	-2.239* (1.323)	0.38 (0.507)	-0.903 (1.391)	-0.821 (1.858)	
Constant	-0.149 (0.190)	-0.317 (0.196)	-0.041 (0.199)	-0.32 (0.219)	-0.093 (0.202)	-0.386* (0.225)	-0.16 (0.204)	
Observations	417	440	410	421	398	407	353	
R ²	0.007	0	0.001	0.013	0.001	0.004	0.003	

Standard Errors clustered by country

Significance Level denoted by *** (1 percent); ** (5 percent); * (10 percent).

Dependent variable is mean (consensus) forecast error

There is very significant evidence of underweighting of new information in the consensus forecasts, as predicted. The results are strongest at shorter forecast horizons: the largest beta coefficient estimate in absolute terms is for the 8-month horizon, using both the raw and adjusted forecast errors. Coefficient estimates are somewhat larger using the raw errors as the dependent variable as in Table 3, but the resulting adjustment accounts for a much larger share of the overall forecast error, reflected in significantly higher R^2 s, for the adjusted forecast error specifications in Table 4.²⁰ Coefficient estimates are also more uniform and significant over a larger range of forecast horizons in this latter case.

methodology requires the full set of 23 forecast horizons.

²⁰This is what one would expect, since a large fraction of the uncertainty at horizon h that shows up in the residual is resolved in subsequent forecast periods, and subtracting subsequent residuals from the forecast error therefore tends to reduce the unexplained (residual) portion of the error, leading to a higher R^2 .

Table 4. Efficiency Tests: Iterative Error Adjustment

Horizon	1	2	3	4	5	6	7	8
Update	-0.816*** (0.213)	-1.208*** (0.174)	-0.697*** (0.129)	-0.634*** (0.093)	-0.984*** (0.233)	-1.072*** (0.207)	-0.817*** (0.098)	-2.195*** (0.537)
Constant	-0.330** (0.135)	-0.423*** (0.032)	-0.413*** (0.012)	-0.370*** (0.024)	-0.394*** (0.023)	-0.376*** (0.021)	-0.361*** (0.025)	-0.299*** (0.033)
Observations	423	422	416	416	416	416	409	409
R ²	0.038	0.135	0.248	0.249	0.222	0.361	0.36	0.324
Horizon	9	10	11	12	13	14	15	16
Update	-1.179*** (0.348)	-0.841*** (0.234)	-0.527*** (0.170)	-0.172*** (0.026)	-0.946*** (0.296)	-0.621*** (0.225)	-0.468*** (0.064)	-0.485 (0.295)
Constant	-0.303*** (0.022)	-0.316*** (0.022)	-0.327*** (0.017)	-0.299*** (0.017)	-0.242*** (0.052)	-0.234*** (0.031)	-0.204*** (0.012)	-0.060** (0.023)
Observations	401	394	394	391	381	380	373	372
R ²	0.187	0.154	0.092	0.103	0.085	0.052	0.204	0.093
Horizon	17	18	19	20	21	22	23	
Update	-0.249 (0.298)	-0.286** (0.133)	-0.181** (0.083)	-0.617** (0.241)	0.169* (0.086)	0.222 (0.201)	-0.083 (0.112)	
Constant	-0.021 (0.016)	0.024 (0.014)	0.038*** (0.011)	0.096*** (0.017)	0.096*** (0.010)	0.101*** (0.013)	0.095*** (0.007)	
Observations	371	371	363	358	352	345	345	
R ²	0.019	0.024	0.064	0.059	0.025	0.059	0.01	

Standard Errors clustered by country

Significance Level denoted by *** (1 percent); ** (5 percent); * (10 percent).

Dependent variable transformed by iterative subtraction of residuals, as described in text.

These results shed some light on the relative variance of idiosyncratic and shared information available to forecasters at different forecasting horizons. Using the results from Table 4 as representing the best estimates, the estimated coefficient on the consensus forecast update is generally in the range of -0.5 to -1.5, suggesting that the variance of forecasters' idiosyncratic signals σ_{ϵ}^2 , at the shorter forecast horizons where the effect is significant, is of a similar magnitude to the variance of the common signals, σ_v^2 . This implies a value of F around 0.5: the consensus forecast typically absorbs only half of the aggregate new information available each period.

C. Efficiency Gains from adjusted Consensus Forecasts

The next step is to calculate the efficiency gain associated with an appropriate adjustment of the consensus forecasts, as in (31). In all cases, the naive priors case is assumed to be the relevant one. A complicating factor is that the relative variance of private and public signals is likely to differ systematically across countries as well as across time periods, and a good correction technique should allow for this. Allowing b_j to differ across each country and time horizon would risk over-parameterizing the model, reducing the precision of parameter estimates and worsening the model's out of sample performance. However, analysis of the b_h coefficients in Table 3 suggests a systematic pattern across three broad time horizons, which

points to a relatively parsimonious parameterization, in which the slope coefficient b_j is assumed to differ across each country c and within the three broad time horizons $H = \{1 - 7, 8 - 14, 15 - 23\}$. The following specification is therefore run:

$$\tilde{f}_{ch\tau} - y_{c\tau} = a_{c,H} - b_{c,H}(\tilde{f}_{ch\tau} - \tilde{f}_{c(h+1)\tau}) + e_{ch\tau}, \quad (32)$$

and corrected forecasts are derived using (30).²¹

Tables 5 and 6 report the resulting (in-sample) efficiency gains by country and by forecast horizon, respectively. To aid interpretation, these tables also present the efficiency gains (reduction in RMSE) associated with using the next month's consensus forecast rather than the current month's (the efficiency gain in this case is calculated as in (31), but with $\tilde{f}_{c(h-1)\tau}$ substituted for $\hat{f}_{ch\tau}^*$). Comparing the two sets of estimated efficiency gains indicates whether adjusting this month's consensus leads to a similar reduction in forecast errors to having an additional month's information.

The most significant efficiency gains are felt at a relatively short horizon, between 4 and 14 months. The maximum gain, at 13 months, is 9.9 percent. Overall, efficiency gains average 3.7 percent, or 6.3 percent for horizons of 12 months or less. These efficiency gains are marginally higher than the gains associated with replacing this month's consensus forecast with the next month's: in other words, adjusting this month's consensus forecast is more valuable than having an additional month of leading and coincident indicators to inform the individual forecasts. Among countries, efficiency gains are generally largest for emerging markets and smallest for industrial countries. Efficiency gains may be largest for emerging market economies due to the relatively smaller time-series available for these countries, resulting in some over-fitting of the model in sample, as well as the greater growth volatility typically experienced.

²¹In this case, to enhance efficiency, regressions therefore pool across forecast horizons. This will tend to produce inconsistent estimates of the true $b_{c,H}$. However, this is not a major concern for this exercise, which is not designed to estimate $b_{c,H}$ consistently but rather to undertake the best adjustment to the consensus forecast, in the sense of minimizing the residual forecast error. In this case, OLS remains the optimal technique, despite the fact that the coefficient estimates are inconsistent. In this case the adjustment corrects not only for the inefficiency introduced by aggregation, but for the correlation across forecast horizons.

Table 5. In-Sample Efficiency Gains, by Country

Country	Efficiency Gain (Percent)	
	Adjusted Forecast	Next Month's Consensus
Argentina	3.2	6.5
Australia	1.8	2.1
Brazil	1.0	7.2
Canada	1.3	3.2
Chile	3.3	6.8
China P.R.: Hong Kong	4.9	3.6
China P.R.: Mainland	2.9	2.9
Colombia	0.3	-0.1
Czech Republic	6.7	-0.8
France	3.0	3.9
Germany	1.6	3.3
Hungary	2.9	-0.2
India	0.6	1.8
Indonesia	7.8	5.5
Italy	3.0	4.4
Japan	1.4	3.7
Korea Republic of	11.0	4.9
Malaysia	13.1	4.3
Mexico	3.3	8.7
Netherlands	1.1	2.5
New Zealand	0.5	1.5
Norway	0.2	0.3
Peru	5.4	5.5
Poland	5.6	8.6
Romania	1.1	3.1
Russian Federation	5.0	5.2
Singapore	4.7	2.8
Slovak Republic	3.1	6.6
Spain	-0.8	0.9
Sweden	2.8	2.7
Switzerland	2.0	3.0
Taiwan Province of China	6.1	3.0
Thailand	9.6	4.2
Turkey	2.5	8.1
Ukraine	2.6	6.2
United Kingdom	0.2	2.5
United States	0.0	2.7
Venezuela	1.7	4.1

Author's calculations, as in text

Forecasts made for years to 2006.

Forecast horizon = 1 not included, to facilitate comparison between columns 2 and 3.

Table 6. In-Sample Efficiency Gains, by Forecast Horizon

Forecast Horizon	Efficiency Gain (Percent)	
	Adjusted Forecast	Next Month's Consensus
1	0.0	
2	0.8	6.4
3	1.6	1.9
4	5.9	5.2
5	3.2	7.0
6	7.1	7.4
7	6.6	5.1
8	8.1	9.5
9	4.5	4.5
10	7.1	3.9
11	5.1	4.9
12	7.7	4.0
13	9.9	6.3
14	5.1	4.6
15	1.9	2.2
16	2.9	0.5
17	1.6	2.1
18	0.3	0.0
19	1.4	0.7
20	-0.4	0.3
21	-0.5	1.0
22	2.6	-0.8
23	1.1	1.2
≤12*	6.3	5.5
All*	3.7	3.0

Author's calculations, as in text

Forecasts made for years to 2006.

*: Drop forecast horizon = 1 to allow comparison between second and third columns.

Out of sample efficiency gains are somewhat lower overall, but not markedly so (Tables 7 and 8).²² Again, out of sample efficiency gains are somewhat better than the gains associated with having an additional month's worth of data to inform the individual forecasts. Some countries with very good in-sample performance (e.g. Indonesia), register a very poor out of sample performance, suggesting that some caution should be exercised in extrapolating in-sample patterns based on extreme events (such as the Asian crisis). Overall, however, one could have improved near-term forecast accuracy for forecasts made in 2007

and 2008 by more than 5 percent, with respect to the raw consensus forecast, based on the

²²In fact, gains are significantly higher for some countries and time periods.

adjustment technique outlined here and data available through end-2006.

D. Robustness Checks

This subsection briefly outlines the results of three robustness checks. The first simply replicates the results in Table 3 using the median forecast (rather than the mean) as the appropriate definition of the consensus forecast (Table 9). Results are almost identical to the baseline.

The second robustness exercise deals with the issue of data revisions. A number of authors have suggested that forecasters aim to predict early releases of the data rather than subsequent revised estimates: this issue is of particular relevance to forecasts of GDP growth, where the data go through several rounds of (often substantial) data revisions up to a year after the end of the period in question (Keane and Runkle, 1990; Loungani, 2001). Following Loungani (2001), one can obtain preliminary or “real time” estimates of GDP growth using GDP estimates taken from the relevant version of the IMF’s WEO data archive. Specifically, the data presented in the April version of the dataset typically represents a first estimate of GDP growth during the previous year. Table 10 replicates the results in Table 3 using these ‘real time’ measures of actual GDP growth, which are available for 2002-08. Results are again almost identical to the baseline: in fact, the coefficient on the forecast update is now negative and significant across a wider set of individual forecast horizons, despite the reduction in sample size.

As an additional robustness check, a similar analysis was undertaken using estimates of quarterly, seasonally adjusted US nominal GDP (in levels) taken from the Survey of Professional Forecasters (SPF).²³ The SPF is a survey of professional economic forecasters maintained by the Federal Reserve Bank of Philadelphia. Forecasts are made quarterly, and

²³Nominal as opposed to real GDP was used to avoid the additional complications associated with matching forecasts with appropriate vintages of actual GDP data, given changes in base years. Actuals are the first GDP estimate, available in the first month of the subsequent quarter. There is some ambiguity as to whether forecasters are aiming for the first GDP estimate or the second, revised, estimate available in the second month of the subsequent quarter. However, forecast errors are smaller using the first estimate as actual at all forecast horizons, suggesting that this is the appropriate benchmark. Because nominal magnitudes increase significantly over the sample period, forecast updates and forecast errors are all divided by the relevant prior (previous consensus forecast) to minimize heteroskedasticity. This is similar to taking a log transformation, and indeed results based on taking logs of individual forecasts prior to aggregating to form the consensus forecasts are almost identical. The scaling method used in the paper is preferred to taking logs as the latter could artificially introduce forecast bias (due to the Jensen inequality). The SPF data before 1992 refer to GNP rather than GDP: to avoid additional complications raised by changing output definitions, data prior to 1992 is dropped. Additional details of the SPF dataset and the analysis, including results for individual forecasts, where (as for the Consensus Economics data) there

forecasts are available for up to four quarters ahead, in addition to a nowcast of the current quarter and a backcast of the previous quarter. Since the furthest out horizon is dropped (because there is no base for calculating the forecast update), and the backcast is somewhat redundant (because initial GDP estimates are already available for that quarter when the backcast is made), there are four relevant forecast horizons available, from the current quarter to three quarters ahead.

is evidence for some minor under-weighting of new information, are in the Appendix.

Table 7. Out of Sample Efficiency Gains, by Country

Country	Efficiency Gain (Percent)	
	Adjusted Forecast	Next Month's Consensus
Argentina	2.9	5.0
Australia	5.7	3.6
Brazil	3.9	6.7
Canada	2.2	4.8
Chile	-1.6	2.8
China P.R.: Hong Kong	4.6	2.8
China P.R.: Mainland	4.7	4.8
Colombia	5.2	0.7
Czech Republic	-25.7	-1.1
France	3.7	2.7
Germany	-7.0	4.4
Hungary	-0.2	5.1
India	2.4	1.7
Indonesia	-59.6	8.5
Italy	3.1	3.0
Japan	1.6	3.2
Korea Republic of	5.6	1.6
Malaysia	9.0	2.5
Mexico	4.0	4.7
Netherlands	-6.7	2.3
New Zealand	2.0	5.1
Norway	2.0	-5.2
Peru	13.6	5.0
Poland	-2.5	2.8
Romania	-5.5	1.7
Russian Federation	3.2	0.9
Singapore	8.2	2.9
Slovak Republic	-3.4	1.4
Spain	2.2	3.9
Sweden	5.1	2.3
Switzerland	3.8	4.6
Taiwan Province of China	4.3	2.2
Thailand	-9.9	1.3
Turkey	4.0	4.0
Ukraine	1.4	2.6
United Kingdom	2.0	4.7
United States	0.8	6.0
Venezuela	-12.3	8.9

Author's calculations, as in text

Forecasts made from Jan 2007 onwards, for 2007/08.

Forecast horizon = 1 not included, to facilitate comparison between columns 2 and 3.

Table 8. Out of Sample Efficiency Gains, by Forecast Horizon

Forecast Horizon	Efficiency Gain (Percent)	
	Adjusted Forecast	Next Month's Consensus
1	6.2	
2	5.8	11.1
3	6.9	7.0
4	0.3	6.6
5	-0.1	4.7
6	5.3	3.4
7	-2.3	4.1
8	3.0	1.2
9	16.2	1.2
10	4.2	5.0
11	15.4	5.0
12	5.2	3.8
13	9.0	1.5
14	6.0	3.2
15	-1.2	2.0
16	-0.8	2.3
17	-0.3	1.5
18	0.3	-0.4
19	1.1	-1.3
20	-0.9	-1.5
21	3.5	0.7
22	-0.6	-0.4
23	0.7	-0.7
$\leq 12^*$	5.6	4.2
All*	3.5	2.7

Author's calculations, as in text

Forecasts made from Jan 2007 onwards, for 2007/08.

*: Drop forecast horizon = 1 to allow comparison between second and third columns.

Table 11 replicates the results of Table 3 for the SPF dataset. The point estimates for $-b$ have the correct (negative) sign at all four horizons, and are statistically significant for the 2 and 3 quarter ahead forecasts. Efficiency gains from adjusted forecasts are presented in the final row of the table. Gains are in a broadly similar range to those for the Consensus Economics dataset, albeit slightly lower. The largest gain, at the 3-quarter ahead horizon, is 5 percent. Interestingly, the relative magnitudes of gains at the different forecast horizons is similar to those using the CE data, with gains largest at forecast horizons of 2—3 quarters, and smaller gains at the shorter horizons. The estimates of b suggest that the under-weighting of new

Table 9. Efficiency Tests: Median Forecasts

Horizon	1	2	3	4	5	6	7	8
Update	-0.428*** (0.127)	-1.032*** (0.225)	-0.994*** (0.182)	-0.976*** (0.180)	-1.399*** (0.203)	-1.689*** (0.435)	-1.070** (0.422)	-2.948*** (0.555)
Constant	-0.421** (0.201)	-0.502*** (0.140)	-0.499** (0.207)	-0.437*** (0.140)	-0.455** (0.195)	-0.498*** (0.155)	-0.431** (0.197)	-0.457*** (0.160)
Observations	475	495	468	489	466	489	459	470
R ²	0.011	0.029	0.035	0.083	0.052	0.091	0.077	0.132
Horizon	9	10	11	12	13	14	15	16
Update	-0.915*** (0.253)	-1.329** (0.546)	-1.571*** (0.413)	-0.461 (0.432)	-2.287*** (0.416)	-1.668* (0.942)	-0.537 (0.460)	0.52 (0.428)
Constant	-0.393* (0.200)	-0.504*** (0.164)	-0.396** (0.191)	-0.556*** (0.181)	-0.363* (0.203)	-0.547*** (0.179)	-0.268 (0.230)	-0.318 (0.193)
Observations	446	455	434	459	429	449	421	441
R ²	0.033	0.05	0.078	0.014	0.081	0.024	0.006	0.005
Horizon	17	18	19	20	21	22	23	
Update	-0.232 (0.727)	0.11 (1.004)	-0.132 (0.550)	-0.319 (1.238)	0.021 (0.734)	-1.182 (1.026)	-0.904 (1.563)	
Constant	-0.095 (0.189)	-0.298 (0.200)	-0.034 (0.198)	-0.307 (0.222)	-0.076 (0.202)	-0.365 (0.226)	-0.145 (0.207)	
Observations	417	440	410	421	398	407	353	
R ²	0.001	0	0	0	0	0.007	0.005	

Standard Errors clustered by country

Significance Level denoted by *** (1 percent); ** (5 percent); * (10 percent).

Dependent variable is median, rather than mean, forecast error.

information is somewhat less pronounced for the SPF dataset than for the Consensus Economics data. Other things being equal, this suggests that forecasters in this case have less informative idiosyncratic signals, compared to the shared public information, which seems plausible given the significant quantity of public information available on the U.S. economy (the relatively small efficiency gains available for the U.S. using the CE data presented in Tables 5 and 7 points to a similar conclusion).

Table 10. Efficiency Tests: Real Time Growth Data

Horizon	1	2	3	4	5	6	7	8
Update	-0.895*** (0.127)	-1.765*** (0.187)	-1.911*** (0.473)	-1.147*** (0.329)	-2.244*** (0.487)	-1.594** (0.679)	-1.72 (1.024)	-2.659** (1.009)
Constant	0.009 (0.038)	-0.113** (0.046)	-0.049 (0.054)	-0.076 (0.062)	-0.081 (0.099)	-0.217** (0.106)	-0.097 (0.109)	-0.198** (0.097)
Observations	226	266	226	266	226	266	223	258
R ²	0.315	0.215	0.335	0.193	0.105	0.096	0.107	0.134
Horizon	9	10	11	12	13	14	15	16
Update	-1.553*** (0.422)	-1.699*** (0.585)	-0.208 (0.801)	-0.193 (0.197)	-3.417*** (0.382)	-1.713* (0.909)	-0.685* (0.384)	-2.337** (1.034)
Constant	-0.124 (0.122)	-0.251** (0.112)	-0.173 (0.128)	-0.358*** (0.118)	-0.108 (0.124)	-0.388*** (0.106)	-0.215 (0.140)	-0.235* (0.122)
Observations	217	250	210	257	218	258	218	258
R ²	0.069	0.066	0.002	0.006	0.219	0.032	0.011	0.061
Horizon	17	18	19	20	21	22	23	
Update	-2.516** (0.999)	-0.299 (1.252)	-1.578 (2.996)	-0.177 (1.316)	1.944 (1.541)	-3.128** (1.400)	-4.464*** (1.056)	
Constant	-0.066 (0.184)	-0.11 (0.141)	-0.016 (0.194)	-0.179 (0.182)	-0.099 (0.216)	-0.17 (0.187)	-0.091 (0.180)	
Observations	218	258	214	249	203	236	196	
R ²	0.070	0.000	0.008	0.000	0.035	0.058	0.152	

Standard Errors clustered by country

Significance Level denoted by *** (1 percent); ** (5 percent); * (10 percent).

Forecast Error calculated with respect to real time data (April WEO of following year).

IV. DISCUSSION

A. Morris and Shin (2002): A Reassessment

Morris and Shin (2002) use a global games framework to model Keynes's (1936) beauty contest. Keynes argues that agents' attempts to second guess the consensus forecast can damage the information-revelation role of market prices. Morris and Shin extend this logic to show that more transparent (lower variance) public signals could lead to higher forecast errors on the part of market participants. In their model, public information is over-weighted relative to agents' private signals because it is more helpful for second-guessing other agents and hence for aligning an agent's own forecast more closely with the consensus. More accurate public signals can exacerbate this over weighting problem, potentially increasing the volatility of agents' individual forecasts.

Table 11. SPF Nominal GDP Forecasts

Horizon (Q ahead)	0	1	2	3
Update	-0.089 (0.089)	-0.234 (0.151)	-0.434** (0.189)	-0.631*** (0.226)
Constant	-0.002*** (0.001)	-0.003** (0.001)	-0.004** (0.002)	-0.004 (0.003)
Observations	71	70	69	68
R ²	0.01	0.036	0.072	0.081
Efficiency Gain (percent)	0.8	1.8	3.5	5.1

Robust Standard Errors in Parentheses

Significance Level denoted by *** (1 percent); ** (5 percent); * (10 percent).

Dependent variable is mean (consensus) forecast error

Morris and Shin's arguments have been criticized, notably for the ad-hoc nature of the beauty contest element and because the parameter values necessary for their argument on the negative effect of public information provision to hold are unrealistic (Svensson, 2006). In particular, the public information must be at least eight times noisier than the private information, which seems unlikely (in the context of making central bank economic forecasts public) in light of evidence on the apparent superiority of central bank forecasts over their private sector counterparts (Barakchian and Crowe, 2009; Romer and Romer, 2000).

This apparent superiority may be exaggerated by comparing central bank forecasts with inefficient consensus forecasts rather than an efficient aggregator of the private sector's information set. However, the logic of our model suggests that, even without the beauty contest element, the provision of more accurate public information could harm the accuracy of the consensus forecast, even if it has an unambiguously positive effect on the accuracy of individual forecasts. Given the attention paid to the consensus forecast and its potential role for uninformed agents in the economy in forming their expectations (particularly if agents do not make the necessary adjustment to the consensus forecast in practice), this could argue against greater transparency.

To formalize this argument, I first illustrate the effect of changing the accuracy of public information in the context of the simple signal extraction model, before analyzing the effect in Morris and Shin's model that includes the beauty contest element. In both cases more accurate public information is modeled as a reduction in the variance of the public prior, σ_v^2 . In the context of the signal extraction model outlined in Section II.A, the mean squared consensus forecast error is given by:

$$E\left[(\bar{f}-y)^2\right] = F^2\sigma_u^2 + (1-F)^2\sigma_v^2 \quad (33)$$

Differentiating equation (33) with respect to σ_v^2 , assuming individually optimal weights F , yields:

$$\frac{\partial E[(\bar{f}-y)^2]}{\partial \sigma_v^2} = \frac{\sigma_u^2 + \sigma_\varepsilon^2}{(\sigma_v^2 + \sigma_u^2 + \sigma_\varepsilon^2)^3} \left((\sigma_u^2 + \sigma_\varepsilon^2)^2 - \sigma_v^2(\sigma_\varepsilon^2 - \sigma_u^2) \right) \quad (34)$$

which is positive if and only if:

$$(\sigma_\varepsilon^2 - \sigma_u^2)\sigma_v^2 < (\sigma_u^2 + \sigma_\varepsilon^2)^2 \quad (35)$$

Hence, more transparent public information can increase the forecast error for the consensus forecast if the public information is a relatively noisy signal and the common component of the private signal (u) is relatively unimportant compared to the idiosyncratic component (ε). The intuition behind this result is relatively simple. Increasing the accuracy of the public signal (reducing σ_v^2) has a direct, positive, effect on the accuracy of the consensus forecast, and a negative, indirect, effect, that arises because forecasters over weight the public signal and the degree of over weighting increases when the public signal becomes more accurate

(as σ_v^2 falls, the over weighting coefficient $\frac{\sigma_\varepsilon^2}{\sigma_v^2 + \sigma_u^2}$ increases). Hence, when σ_v^2 is high, the public signal receives a low weight so the direct effect of reducing its variance is small. Similarly, when σ_ε^2 is high, the degree of over weighting is high and the indirect effect of reducing the variance of the public signal is large.

In the special case where the private signals are purely idiosyncratic ($\sigma_u^2 = 0$), the condition (35) collapses to $\sigma_v^2 < \sigma_\varepsilon^2$: this implies that greater transparency can increase the *consensus* forecast error if the public information is initially more noisy than agents' own signals. In this case, the direct effect of having more accurate information is relatively small, because the public signal receives a relatively low weight, whereas the indirect effect is large because the degree of over-weighting is significant. This condition seems likely to hold in some situations. Moreover, the negative effect on the accuracy of the consensus forecast does not require the beauty contest element, whereas greater transparency always improves the accuracy of individual forecasts without the beauty contest.

However, one can include the beauty contest element, as in Morris and Shin (2002), so that forecasters now weight the public and private signals to minimize a loss function that is a weighted sum of the mean squared forecast error and the mean squared deviation of individual forecasts from the consensus. The weight on the beauty contest element is given by r :²⁴

²⁴This section only discusses Morris and Shin's key results with respect to individual forecast accuracy and the equivalent results for consensus forecasts. For the full details of the model, see Morris and Shin (2002).

$$L = (1-r)E[(f-y)^2] + rE[(f-\bar{f})^2] \quad (36)$$

As shown by Morris and Shin, agents now over weight the public signal:

$$F_b = \frac{(1-r)\sigma_v^2}{\sigma_\varepsilon^2 + (1-r)(\sigma_u^2 + \sigma_v^2)} \quad (37)$$

and differentiating (MSFE) with respect to σ_v^2 yields:

$$\begin{aligned} \frac{\partial E[(\bar{f}-y)^2]}{\partial \sigma_v^2} &= \frac{(1-r)\sigma_u^2 + \sigma_\varepsilon^2}{((1-r)(\sigma_v^2 + \sigma_u^2) + \sigma_\varepsilon^2)^3} \\ &\quad \times \left(((1-r)\sigma_u^2 + \sigma_\varepsilon^2)^2 - (1-r)\sigma_v^2(\sigma_\varepsilon^2 - (1-r)\sigma_u^2) \right) \end{aligned} \quad (38)$$

which is positive if and only if:

$$(1-r)\sigma_v^2(\sigma_\varepsilon^2 - (1-r)\sigma_u^2) < ((1-r)\sigma_u^2 + \sigma_\varepsilon^2)^2 \quad (39)$$

For the special case considered by Morris and Shin ($\sigma_u^2 = 0$), this condition collapses to $(1-r)\sigma_v^2 < \sigma_\varepsilon^2$. By contrast, the condition for more accurate public information making *individual* private forecasts more accurate is $(2r-1)(1-r)\sigma_v^2 < \sigma_\varepsilon^2$.

Comparing these two expressions yields a number of insights. First, while the condition for better public information to reduce the mean squared error of the *individual* forecasts always holds when the beauty contest is not present, it can fail to hold for consensus forecasts even without the beauty contest. This is because consensus forecasts over weight the public information even without the beauty contest element, so that the mechanism by which more public information can be harmful is also present without the beauty contest. In fact, while the beauty contest has to be relatively important ($r > 0.5$) for the harmful effect of transparency to (potentially) be felt for the individual forecasts, the harmful effect is most likely to hold for *consensus* forecasts when the beauty contest is relatively *unimportant* (r is low).²⁵ Finally, since $(2r-1) \leq 1$, the parameter space over which more accurate public information can be harmful is strictly larger for consensus forecasts than for individual forecasts.

²⁵The rationale for this is that the beauty contest is not the cause of the consensus forecast over weighting the prior: if one were to set σ_ε^2 to zero in (37) to eliminate the cause of overweighting the prior in the simple signal extraction model then the $(1-r)$ terms drop out and one is left with the expression for the efficient forecast weight F^* , regardless of the beauty contest. Hence, introducing the beauty contest reduces the impact of the mechanism by which the prior is overweighted, diluting the impact and making it less likely that an increase in transparency is harmful.

B. Groupthink, Bank Behavior and the Credit Crunch

The insights from this paper can also be related to recent work on groupthink or collective delusion by Benabou (2008). Benabou's model is based on a modification to a standard utility function to incorporate anticipatory feelings (i.e. deriving utility from expected as well as actual consumption). With this modification, Benabou is able to generate externalities from individual behavior that give rise to equilibria with self-fulfilling collective delusion. Agents rationally choose to ignore pertinent information, giving rise to collective judgments that differ systematically from reality.

The phenomenon of groupthink involves a number of features that distinguish it from the model in this paper, notably the introduction of anticipatory feelings that lead to rational self-denial (and ultimately mutually-reinforcing denial within groups) and the absence of a major role for idiosyncratic private information. Nevertheless, there are some important complementarities between the two approaches. The collective failure to rationally process the available aggregate information—leading to the discarding of useful information—is the most obvious common feature to both models. Another commonality is the role of common group beliefs. In the Groupthink model agents choose to ignore information at odds with the group prior, leading to a less efficient aggregation of private signals. In the simple signal extraction model in this paper the common prior is also over weighted collectively, although this behavior is individually rational.²⁶

Moreover, as discussed in relation to Morris and Shin's paper, a more accurate (lower variance) prior can lead to a less accurate collective forecast in our model. This result can provide insights into some of the groupthink phenomena discussed in Appendix A to Benabou's paper. An example of relevance to the recent U.S. subprime crisis is that of groups of bank employees who appeared to systematically underestimate the risks associated with complex structured assets. For instance, increased use of sophisticated quantitative models for pricing assets may have genuinely provided a more accurate assessment of expected risks and returns that led to a lower-variance prior among managers and a (positive) direct effect on the accuracy of their consensus forecast. However, this more accurate signal then caused the managers, receiving diffuse but ultimately correlated signals of potential downside risk (e.g. from their knowledge of the underlying assets' quality and experience of running traditional loan books), to rationally put greater weight on the group prior, collectively underweighting their noisy but (in aggregate) informative private signals, and therefore reducing consensus forecast accuracy.

The condition for this underweighting effect to dominate the direct effect—that the group prior be no less noisy than the managers' own idiosyncratic signals—seems plausible, given

²⁶Benabou's mechanism is also individually rational conditional on the inclusion of anticipatory feelings in the utility function.

that the models turned out to perform rather poorly out of sample.²⁷ In other words, the claims prior to the crisis that these new techniques were reducing risk, and the emerging consensus post-crisis that their use reduced the ability to foresee and forestall emerging risks, may both be accurate assessments. The first relates to the group prior; the second to the group posterior.

V. CONCLUSIONS

This paper has provided an empirical assessment and discussed the implications of a result first noted by Kim, Lim and Shaw (2001), that consensus forecasts tend to over weight the prior at the expense of new information. Testing this result with respect to forecasts of economic growth taken from one of the most widely-used cross-country forecast datasets, one finds robust evidence of this underreaction to new information. Moreover, this underreaction is clearly due to the forecast aggregation process, as *individual* forecasters appear to respond more or less appropriately to new information. Applying a suitable adjustment to the consensus forecasts using additional information contained in the change to the consensus forecast from the previous month's, one is able to obtain more accurate forecasts, particularly at horizons of a year or less (where the efficiency gain is more than 5 percent on average, even out of sample). The gain in efficiency is, on average, larger than that associated with having access to an additional month's data to inform the forecast. There is no evidence that forecasters make the necessary adjustment in practice, suggesting that this inefficiency in the consensus forecast is not well understood even by the sophisticated market participants included in the survey.

This paper has implications for consumers of consensus forecasts. Those using the forecasts need to be aware of the nature and source of their inefficiency as aggregators of the information available to market participants. Forecast aggregation services (such as the one whose data are employed in this paper) may need to adjust their methodology to provide more efficient consensus forecasts (although, for many countries, the efficiency gain is likely to be relatively small). One method is to employ an *ex post* adjustment similar to the one used in this paper. An alternative method would be to attempt to elicit individual forecasts that place a higher weight on each forecaster's private signal compared to their best forecast. One means of achieving this in practice would be to provide incentives for strategic forecasting such that forecasters attempt to differentiate their forecasts from those of others (Ottaviani and Sørensen, 2006). Alternatively, forecasters could be invited to submit two forecasts: their best guess (as currently) and a noisier forecast based only on their newest information.

²⁷Explanations for the crisis have tended to emphasize that the models performed worse than expected (i.e. their signal to noise ratio turned out to be lower than believed). However, while this may indeed have been the case, it is not necessary as an explanation for the crisis.

Although the paper's main insights relate to the aggregation of individual forecasts, the paper also has implications for those engaged in forecasting themselves. Forecasting technologies that decentralize information-gathering but centralize the production of a final forecast (as are typically employed in a central bank) may help to alleviate this problem of over-weighting the prior. For instance, when combining insights from a number of quantitative forecasting models, it is better to apply judgment (i.e. shade towards the prior) only at the level of the aggregate forecast, rather than at the level of each individual model, particularly if in the latter case the weights placed on each model's 'raw' results relative to the prior are endogenously determined by their relative noisiness, either explicitly or implicitly.

Appendix

Testing for Individual Forecast Rationality

Within the context of this paper, which is concerned primarily with establishing whether consensus forecasts are inefficient aggregators of individual information, individual forecast rationality is pertinent only to the extent that it sheds light on the performance of consensus forecasts. In particular, if one can establish that individual forecasts are approximately rational, then any inefficiency identified in consensus forecasts can be attributed to aggregation, and not to underlying forecaster behavior.

Individual forecasts are rational as long as the signals are weighted as in (14):

$$E[(f_{cht}^i - \mu_{cht})(f_{cht}^i - y_{ct})] = F(F\sigma_\varepsilon^2 - (1-F)\sigma_v^2) = 0. \quad (40)$$

Taking this relationship to the data is hindered by the fact that we do not observe the prior μ_{cht} . However, under the null of individual forecaster rationality the previous month's consensus forecast $\tilde{f}_{c(h+1)\tau}$ can be used instead, since rationality implies a zero coefficient in a regression of forecast errors $f_{cht}^i - y_{ct}$ on either $f_{cht}^i - \mu_h$ (i.e., assuming μ_h is available) or $f_{cht}^i - \tilde{f}_{c(h+1)\tau}$ (substituting $\tilde{f}_{c(h+1)\tau}$ for μ_h). The first case is intuitive, and follows from (40). In the second case:

$$\begin{aligned} E[\hat{\beta}] &= \frac{\text{Cov}(f_{cht}^i - \tilde{f}_{c(h+1)\tau}, f_{cht}^i - y_{ct})}{\text{Var}(f_{cht}^i - \tilde{f}_{c(h+1)\tau})} \\ &= \frac{F(F\sigma_\varepsilon^2 - (1-F)\sigma_v^2)}{(F^2 + (1-F)^2)\sigma_v^2 + F^2\sigma_\varepsilon^2} = 0 \end{aligned} \quad (41)$$

One could therefore run the equation,

$$f_{cht}^i - y_{ct} = \alpha + \beta(f_{cht}^i - \tilde{f}_{c(h+1)\tau}) + e_{cht}^i. \quad (42)$$

where rationality requires that $\alpha = \beta = 0$.

However, as noted by Zarnowitz (1985), Davies and Lahiri (1995) and Bonham and Cohen (2001), pooling individual forecasts that share common forecast targets y_{ct} will lead to biased and inconsistent estimates of β . This is because, for a given realization of y_{ct} , forecast errors are positively correlated with forecasts, and so the estimated coefficient $\hat{\beta}$ from a pooled regression of forecast errors on forecast updates will tend to be positive even when individual forecasts are rational and the true value is therefore zero. The pooled forecasts do not inherit the characteristics of the individual forecasts for the same reason that the consensus forecasts do not, namely that the properties of rational forecasts hold conditional on the forecaster's information set, and a pooled sample includes forecasters whose information sets differ. The only exception is where all forecasters have the same information, in which case the forecasts

for each forecast target will be the same (this corresponds to the case of $\sigma_\varepsilon^2 = 0$, which is also the only case in which the consensus forecast is efficient, and indeed there is nothing to be gained from employing pooled individual as opposed to consensus forecasts in this instance).

In fact the same argument applies to pooling forecasts across *time horizons*, both for individual forecasts and for the consensus forecast. Again, the problem is that information sets differ across time horizons, so that rationality arguments can no longer be invoked to justify the assumption of orthogonality between forecasts and forecast errors. Hence, for a given realization of $y_{c\tau}$, forecast errors are once again positively correlated with forecasts when observations from different forecast horizons are pooled. To illustrate this, note that, assuming rationality, the residuals from (42) are given by:

$$e^i_{ch\tau} = F\varepsilon^i_{ch\tau} - (1 - F)v_h - \sum_{j=0}^{h-1} v_h, \quad (43)$$

where $e^i_{ch\tau}$ are therefore uncorrelated with $f^i_{ch\tau} - \bar{f}_{c(h+1)\tau}$ for a given forecaster i . However, the presence of the last term in (43) means that $e^i_{ch\tau}$ are correlated with $f^i_{ch\tau} - \bar{f}_{c(h+1)\tau}$, for $k > h$. Hence, pooling across time periods introduces correlation between regressors and the error term.²⁸

Given these constraints, this paper follows Davies and Lahiri (2001) and exploits a further characteristic of rational individual forecasts—that forecast revisions are uncorrelated with information known at the time of the earlier forecast—to test for rationality. Specifically, I test for correlation between forecast revisions at adjacent forecast horizons, since the previous forecast revision is known at the time of the current revision and the two should therefore be uncorrelated. It turns out that this Martingale test of efficiency is not subject to some of the empirical concerns raised with respect to more conventional tests based on forecast errors. It also allows one to ignore issues of data revision, as it does not make reference to the actual data $y_{c\tau}$.

In the context of the assumed information structure, the forecast revision at horizon h is given by:

$$\Delta f^i_{ch\tau} = f^i_{ch\tau} - f^i_{c(h+1)\tau} = F(v_{ch\tau} + \varepsilon^i_{ch\tau} - \varepsilon^i_{c(h+1)\tau}) + (1 - F)v_{c(h+1)\tau} \quad (44)$$

Hence, current forecast revisions are uncorrelated with the most recent previous revision if forecasters are rational:

²⁸Since the correlation enters via the common shock terms v_h rather than the idiosyncratic terms $\varepsilon^i_{ch\tau}$, pooled consensus forecast regressions suffer from the same problem.

$$\text{Cov}\left(\Delta f_{ch\tau}^i, \Delta f_{c(h+1)\tau}^i\right) = F((1-F)\sigma_v^2 - F\sigma_\varepsilon^2) = 0 \quad (45)$$

To test for forecaster rationality, one can therefore regress the current forecast revision on the previous revision:

$$\Delta f_{ch\tau}^i = \alpha + \beta \Delta f_{c(h+1)\tau}^i + e_{ch\tau}^i. \quad (46)$$

Rationality requires that $\alpha = \beta = 0$. In particular:

$$\begin{aligned} E[\beta] &= \frac{\text{Cov}\left(\Delta f_{ch\tau}^i, \Delta f_{c(h+1)\tau}^i\right)}{\text{Var}\left(\Delta f_{c(h+1)\tau}^i\right)} \\ &= \frac{F((1-F)\sigma_v^2 - F\sigma_\varepsilon^2)}{(F^2 + (1-F)^2)\sigma_v^2 + 2F^2\sigma_\varepsilon^2} = 0 \end{aligned} \quad (47)$$

Note that the error terms $e_{ch\tau}^i$ now have a complex MA(1) structure across forecast horizons h , including separate MA(1) processes for $\varepsilon_{ch\tau}^i$ and for $v_{ch\tau}$. Because the former varies over both h and i , whereas the latter is common across i for a given h (and c, τ), (47) no longer holds when observations are pooled across forecasters. Hence, even the Martingale efficiency test needs to be run separately for each forecaster, although it can be pooled across forecast horizons h (as well as c, τ).²⁹

Running efficiency tests by forecaster requires us to identify each forecaster. This is made more difficult, in the context of the Consensus Economics dataset, by the fact that individual forecasters are identified only by their (institutional) name rather than a numerical identifier, and spelling/abbreviation changes make it hard to match forecasters consistently. This, combined with a requirement that a relatively long time series (at least 40 observations) is available for each forecaster, reduces the sample size. The sample includes 125,545 observations over 1,342 individual forecasters. Table A1 summarizes the results. For both $\hat{\alpha}$ and $\hat{\beta}$, Table A1 presents the weighted mean (weighted by the inverse variance of the coefficient estimate), the median, and the proportion of forecasters with statistically significant negative and positive coefficients (at the 1 percent level).³⁰

²⁹To see this, note that for any given (c, h, τ) , $\text{Cov}\left(\Delta f_{ch\tau}^i, \Delta f_{c(h+1)\tau}^i\right) = -F^2\sigma_\varepsilon^2 \leq 0$. The problem does not arise when forecasts are pooled across forecast horizons h for a given i because in this case both $\varepsilon_{ch\tau}^i$ and $v_{ch\tau}$ vary, so that the covariance is zero for optimal F . This issue does not arise in Davies and Lahiri (1995) because these authors assume that forecasters (on average) fully incorporate new information, so that the common component of the error term is iid rather than MA(1). However, as already discussed, this assumption is not consistent with rational forecaster behavior.

³⁰Estimated standard errors are clustered by year within each individual regression.

Table A1. Individual Regression Results Summary

	Alpha	Beta
Mean	-0.010	0.041
Median	-0.032	0.062
Proportion significant		
Positive	0.013	0.113
Negative	0.023	0.016
Forecasters	1342	
Total Observations	125,545	

Summary of 1342 Martingale regressions by forecaster
Significance defined at the 1 percent level
Individual regressions have standard errors clustered by year
Means are weighted by the inverse of the variance.

These results suggest some modest under weighting of new information on the part of individual forecasters. The (weighted) mean estimated β coefficient is positive, as is the median, while 11 percent of individual coefficients are positive and statistically significant (while less than 2 percent are negative and significant). A positive estimate of β implies that $\frac{F}{1-F} < \frac{\sigma_v^2}{\sigma_\varepsilon^2}$, so that the relative weight put on new information is below the optimal.

However, this under weighting is limited compared to that identified for the consensus forecast. It turns out that, although the Martingale efficiency test undertaken for the individual forecasters is somewhat different to the simpler test undertaken for the consensus forecasts, the estimated β and b coefficients are directly comparable:

$$E[\beta] = \frac{F((1-F)\sigma_v^2 - F\sigma_\varepsilon^2)}{(F^2 + (1-F)^2)\sigma_v^2 + 2F^2\sigma_\varepsilon^2}; E[b] = \frac{F(1-F)}{(F^2 + (1-F)^2)} \quad (48)$$

In fact, when there is no idiosyncratic uncertainty ($\sigma_\varepsilon^2 = 0$) the two expressions coincide exactly.

Given the lack of evidence in favor of rational priors, the naive priors case seems likely to hold, and the estimate of b is therefore likely to be unbiased: $E[b] = \frac{1-F}{F}$. Using the relevant expressions for β and b and the estimated coefficients, one can undertake back of the envelope calculations of the degree of under weighting of new information by both individual forecasters and the consensus. From Table 4, the estimated value of b , at least for forecast horizons of a year or less, is around 1, implying that $F = 0.5$. Given the mean estimate of β of around .04, this implies that the optimal individual weight ($\frac{\sigma_v^2}{\sigma_v^2 + \sigma_\varepsilon^2}$) is around 0.54. Thus, the average weight placed on new information (0.5) is approximately correct for the individual forecasters, but clearly inadequate for the consensus forecasts, where the optimal weight is 1.

Analysis of SPF data

The Survey of Professional Forecasters (SPF) is carried out by the Federal Reserve Bank of Philadelphia, who took over responsibility for the survey from the ASA/NBER in 1990. A panel of professional forecasters is surveyed every quarter, just after the release of initial GDP estimates for the previous quarter by the Bureau of Economic Analysis in the first month of the quarter. There is significant persistence in the make-up of the forecaster panel, at least at the institutional level, although forecasters also drop in and out of the sample. For the post-1992 sample used in the analysis, the number of forecasters each quarter varies between 30 and 52.

The survey contains forecasts of 31 economic variables. This paper uses data on one of these, nominal GDP (seasonally adjusted, annual rate, in levels). This series goes back to 1968. However, prior to 1992 the forecasts were of nominal GNP not GDP. In addition, the documentation on the data prior to the Philadelphia Fed assuming responsibility for the survey in 1990 is limited (e.g. it can be harder to consistently identify forecasters across time, and the relative timing of forecasts and data releases is not so clear). Hence, this study uses only the GDP data, covering forecasts made in 1992Q1 through 2009Q4. Since this data is not expected to contain any outliers, unlike the Consensus Economics (CE) data, it was not trimmed.

Actual GDP data for comparison is taken from the Federal Reserve Bank of St Louis's *ALFRED* database, which contains GDP estimates for different vintages. An important question is which GDP release forecasters are forecasting: the first estimate, available in the first month after the quarter in question has ended; the first revised estimate, available the next month; or subsequent revised estimates. A comparison of forecast errors suggested that forecasters aimed to match the first estimate, so this estimate was used as the actual for comparison.

Since nominal GDP has grown significantly over time, using raw data in levels would create a significant problem of heteroskedasticity. Effectively, recent data would receive a much higher relative weight. Hence, the data must be scaled appropriately to ensure that the magnitude of the observations remain stationary. For the consensus forecast regressions, both the forecast error and forecast update are divided by the previous month's consensus forecast (the proxy for the prior). For the individual forecast regressions, the most recent and the lagged forecast update are divided by the two period lagged forecast. Otherwise, the methodology is identical to that underlying the baseline results presented in Table 3 and the results for individual forecasts presented in Table A1.

Results for individual SPF forecasters are presented in Table A2. This Table summarizes results for regressions for the 57 individual forecasters in the sample for whom there were at least 40 forecast observations across the relevant forecast horizons. As with the CE data,

there is some modest evidence that forecasters under-weight new information, leading to positive correlation across forecast revisions, although the extent of under-weighting is significantly less than for consensus forecasts. Undertaking the same back of the envelope calculations as before, \hat{b} is around 0.5 for the forecast horizons at which it is significantly different from zero (Table 11). This implies that $F = 0.67$. Using the expression for $E[\hat{\beta}]$ and the estimate $\hat{\beta} = 0.15$, then the individually optimal weight on new private information is around 0.8. The estimated actual weight, 0.67, is thus too low, both from the perspective of optimal individual forecasts (where the weight should be 0.8) and from that of optimal consensus forecasts (where the weight should be 1). As with the CE dataset, the extent of under-weighting is considerably greater for consensus than for individual forecasts.

Table A2. Individual SPF Regression Results Summary

	Alpha	Beta
Mean	0.000	0.148
Median	0.000	0.085
Proportion significant		
Positive	0.000	0.035
Negative	0.000	0.000
Forecasters		57
Total Observations		5,937

Summary of 57 Martingale regressions by forecaster

Significance defined at the 1 percent level

Individual regressions have standard errors clustered by year

Means are weighted by the inverse of the variance.

Comparing results in Table A1 and Table A2, the median and mean $\hat{\beta}$ point to greater evidence of inefficiency in individual forecasts for the SPF dataset than for the CE dataset, although there is a higher proportion of significant individual $\hat{\beta}$ coefficients in the case of the latter. No forecaster in the SPF dataset significantly over-weights new information, and there is no significant evidence of bias, positive or negative, for any forecaster in the SPF dataset.

REFERENCES

- Amato, Jeffery D. and Hyun Song Shin. 2006, Imperfect common knowledge and information value of prices. *Economic Theory* Vol. 27 No. 1, pp. 213–41.
- Ang, A., G. Bekaert and M. Wei, 2007, Do Macro Variables, Asset Markets or Surveys Forecast Inflation Better? *Journal of Monetary Economics* Vol. 54, pp. 1163–1212.
- Barakchian, S. Mahdi and Christopher Crowe, 2009, Monetary Policy Matters: New Evidence from a New Shock Measure. International Monetary Fund: Mimeo.
- Batchelor, Roy and Pami Dua, 1991, Blue Chip Rationality Tests, *Journal of Money, Credit and Banking*, Vol. 23, pp. 692–705.
- Benabou, Roland, 2008, Groupthink: Collective Delusions in Organizations and Markets. Princetown University: Mimeo.
- Bonham, Carl S. and Richard H. Cohen, 2001, To Aggregate, Pool, or Neither: Testing the Rational-Expectations Hypothesis Using Survey Data. *Journal of Business and Economic Statistics*, Vol. 19, pp. 278–91.
- Davies, Anthony and Kajal Lahiri, 1995, A new framework for analyzing survey forecasts using three-dimensional panel data. *Journal of Econometrics* Vol. 68, pp. 205-27.
- Figlewski, Stephen and Paul Wachtel, Rational Expectations, Informational Efficiency, and Tests Using Survey Data: A Reply. *Review of Economics and Statistics*, Vol. 65, pp.529–31.
- Gallo, Giampiero M., Clive W. J. Granger, and Yongil Jeon (2002). Copycats and Common Swings: The Impact of the Use of Forecasts in Information Sets. *IMF Staff Papers*, Vol. 49, pp. 4–21.
- Granger, Clive W. J., 1989, Combining Forecasts—Twenty Years Later. *Journal of Forecasting* Vol. 8, pp. 167–73.
- Harvey, David I., Stephen J. Leybourne, and Paul Newbold, 2001, Analysis of a panel of UK macroeconomic forecasts. *Econometrics Journal* 4, S37-S55.
- Hong, Harrison and Jeremy C. Stein, 1999, A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets, *Journal of Finance*, Vol. 54, pp. 2143–84.
- Isiklar, Gultekin, 2005, On aggregation bias in fixed-event forecast efficiency tests, *Economics Letters* Vol. 89, pp. 312–6.

- Jegadeesh, Narasimhan and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, Vol. 48, pp. 65–91.
- Jegadeesh, Narasimhan and Sheridan Titman, 2001, Profitability of Momentum Strategies: An Evaluation of Alternative Explanations. *Journal of Finance* Vol. 56, 699–720.
- Keane, Michael P. and David E. Runkle, 1990, Testing the Rationality of Price Forecasts: New Evidence from Panel Data. *American Economic Review*, Vol. 80, pp.714–35.
- Kim, Oliver, Steve C. Lim and Kenneth W. Shaw, 2001, The Inefficiency of the Mean Analyst Forecast as a Summary Forecast of Earnings. *Journal of Accounting Research*, Vol. 39, pp. 329–36.
- Loungani, Prakash (2001). How accurate are private sector forecasts? Cross-country evidence from *consensus forecasts* of output growth. *International Journal of Forecasting* 17, 419-32.
- Morris, Stephen and Hyun Song Shin, 2002, Social Value of Public Information. *American Economic Review* Vol. 92, pp. 1521–1534.
- Ottaviani, Marco and Peter Norman Sørensen, 2006, The strategy of professional forecasting. *Journal of Financial Economics* Vol. 81, pp. 441–66.
- Ottaviani, Marco and Peter Norman Sørensen, 2009, Surprised by the Parimutuel Odds? *American Economic Review* 99 Vol. 5, pp. 2129–34.
- Roca, Mauro, 2009, Transparency and monetary policy with imperfect common knowledge. Washington, DC: International Monetary Fund (mimeo).
- Romer, Christina D. and David H. Romer, 2000, Federal Reserve Information and the Behavior of Interest Rates. *American Economic Review* Vol. 90, pp. 429–57.
- Svensson, Lars, 2006, Social Value of Public Information: Comment: Morris and Shin, 2002 is Actually Pro-Transparency, not Con. *American Economic Review* Vol. 96, pp. 448–52.
- Timmermann, Allan, 2006, Forecast Combinations. in *Handbook of Economic Forecasting*, Vol. 1 (Graham Elliot, Clive W. J. Granger and Allan Timmermann, eds.). Elsevier.
- Wallis, Kenneth F. (forthcoming) Combining Forecasts—Forty Years Later. *Applied Economics*.

Zarnowitz, Victor, 1985, Rational Expectations and Macroeconomic Forecasts. *Journal of Business and Economic Statistics* vol. 3, pp. 293–311.