Social Learning, Strategic Incentives and Collective Wisdom:
An Analysis of the Blue Chip Forecasting Group

J. Peter Ferderer
Department of Economics
Macalester College
St. Paul, MN 55105
ferderer@macalester.edu

Adam Freedman
Chicago, IL 60601
freedman.adamj@gmail.com

July 22, 2015

ABSTRACT: Using GDP growth forecasts from the Blue Chip survey between 1977 and 2011, we measure absolute consensus errors and forecast dispersion for each of the 24 monthly forecast horizons and provide three main findings. First, the size of consensus errors and forecast dispersion are negatively correlated over longer-term forecast horizons (from 24 to 13 months). This finding is consistent with the hypothesis that the Lawrence Klein forecasting award, which is based on performance of 12-month-head forecasts, increased the group's collective wisdom by raising the incentive to anti-herd. Second, absolute consensus errors and forecast dispersion display significant negative temporal variation for the longest horizons (24 to 20 months). Third, after the early 1990s (i) there was a dramatic decline in forecast dispersion associated with a significant increase in the size of longer-term consensus errors, and (ii) forecasts bracket realized GDP growth much less frequently. The final two results suggest that increased herding or reduced model diversity caused collective wisdom to diminish in the second part of the sample.

JEL Classifications:
E37 Forecasting and Simulation: Models and Applications
D70 Analysis of Collective Decision-Making, General
D83 Search; Learning; Information and Knowledge; Communication; Belief
Keywords: social learning, herding, strategic incentives, reputation, consensus forecasts, collective wisdom, forecast errors, forecast dispersion, Great Moderation

The key lesson I would draw from our experience is the danger of relying on a single tool, methodology or paradigm. Policymakers need to have input from various theoretical perspectives and from a range of empirical approaches. Open debate and a diversity of views must be cultivated.

Jean-Claude Trichet (former ECB President), 2010

1. Introduction

In their seminal paper, Bates and Granger (1969) showed that mean–squared prediction errors could be reduced by combining forecasts. Collective wisdom emerges because each individual sees the truth plus some independent error and these errors cancel out when averaged.[1] This fundamental insight has inspired the formation of various aggregation services which survey forecasters and provide consensus forecasts – simple cross-sectional averages of individual forecasts – to the public. One such service, the *Blue Chip Economic Indicators* newsletter, started publishing consensus forecasts in 1976. Another, Consensus Economics, began reporting Consensus Forecasts™ in 1989. Both are widely discussed in the media and inform policy-makers, businesses and other decision makers in the economy.

While there is considerable evidence that consensus forecasts generally outperform individual forecasts over time, some have suggested that interaction between forecasters undermines their independence and resulting collective wisdom.[2] For instance, Zarnowitz (1992) speculated that large forecast errors around business cycle turning points "cannot be explained away by a general reference to random disturbances" and occur, in part, because "few forecasters take the risk of signaling a recession prematurely ahead of others." Gallo, Granger, and Jeon (2002) show that members of the Consensus Economics survey incorporate the average forecast from the previous period into their forecast revisions and conclude that "the forecasting performance of these groups may be severely affected by the detected imitation behavior and lead to convergence to a value that is not the 'right' target." Others have speculated that the failure of economists to anticipate the Great Recession resulted from a lack of model diversity (see reference to Jean-Claude Trichet above).

There are two basic features of the social environment that influence forecaster independence and their collective wisdom: social learning and market incentives. If a forecaster can observe the predictions made by others, it is rational to mimic them when the information they contain overwhelms the forecaster's own information.[3] Herding causes private information to be aggregated less efficiently and can lead a group to converge on an incorrect target. In essence, a tragedy of the commons occurs because private incentives cause forecasters to overuse the group's collective wisdom and collective wisdom declines as a consequence. Social learning can also lead individuals to adopt similar conceptual frameworks and the loss of diversity may undermine collective wisdom.[4] For example, Keynesian models may have been superior to monetarist ones in the 1960s but this does not necessarily imply that the consensus would be more accurate if all forecasters adopted a Keynesian perspective.

[1] Alternatively, members of the group use models that are misspecified in different ways and combining their predictions lowers mean-squared errors because each model captures an important element of the underlying system (Hendry and Clements (2004) and Hong and Page (2012)). Surveys of the forecast combination literature are provided by Clemen (1989), Armstrong (2001), Newbold and Harvey (2002) and Timmermann (2006). James Surowieki has popularized the idea in his 2005 book *The Wisdom of Crowds*.

[2] Bauer, et al. (2003) show that the Blue Chip consensus has been more accurate than most of the individuals who have contributed forecasts to the survey.

[3] This rational herding arises in the *information cascade* models of Banerjee (1992), Bikhchandani, Hirschleifer, and Welch (1992). See Bikhchandani and Sharma (2001) for a review.

[4] See Armstrong (2001). In *The Difference* (2007), and a number of articles with coauthors, Scott Page discusses the implications of model diversity for collective wisdom.

Forecasters, like other economic agents, also face market incentives that encourage them to engage in strategic behavior and report "untruthful" predictions that deviate from their supposed goal of error minimization.[5] For example, a long period of time may need to pass before the accuracy of forecasts can be evaluated by the market. In the interim, forecasters concerned about their reputations have an incentive to ignore private information and herd when the market believes that "bright minds think alike." In contrast, agents may elect to anti-herd and exaggerate the weight they place on their private information when they need publicity to attract new clients or participate in winner-take-all forecasting contests. In contrast to herding, anti-herding increases the dispersion of forecasts and reduces the public knowledge bias inherent in forecast averages (Lichtendahl, et al. (2013) and (Hong, et al., 2012)).[6] That is, anti-herding promotes collective wisdom.

This paper examines whether changes in social learning and strategic incentives have influenced the collective wisdom of the Blue Chip forecasting group over time and across forecasting horizons. Our empirical strategy is straightforward. If forecasters are not influenced by social learning and strategic incentives and simply attempt to minimize forecast errors, we expect to observe a positive relationship between the size of consensus errors and the dispersion of their forecasts. For instance, consensus errors and forecast dispersion should both decline as forecast horizons shorten and information accumulates. Similarly, temporal changes in the volatility of the economic environment will produce a positive relationship between the size of consensus errors and forecast dispersion. For example, consensus errors and forecast dispersion should have both decreased during the Great Moderation if forecasters were only concerned about error minimization.[7]

In contrast, we expect to observe a negative relationship between the size of consensus errors and forecast dispersion over time or across forecast horizons if social learning and strategic incentives influence behavior. If, for example, the incentive to herd increases over time we should observe a reduction in forecast dispersion accompanied by an increase in consensus errors. A reduction in model diversity will produce a similar result. A negative relationship between the size of consensus errors and forecast dispersion will materialize over the forecast horizons if strategic incentives change across this dimension as well.

We use real GDP forecasts from the Blue Chip survey to examine the covariation between the size of consensus errors and forecast dispersion between 1976 and 2011. There are several reasons why the Blue Chip group is an ideal one for this purpose. First, the forecasts are published in a monthly newsletter shortly after they are issued and this enables members of the group to learn what others have predicted. Second, they issue fixed target forecasts updated over a two year period. Thus the information contained in the forecasts remains relevant for many months. Third, members of the Blue Chip group compete in a winner-take-all forecasting contest which has the potential to promote anti-herding. Importantly, the contest's prize – the Lawrence R. Klein Award – is based on performance for only one of the 24 monthly forecasting horizons. Thus it creates an exogenous change in strategic incentives across the horizons which we use to test the hypothesis that changes in herding propensities jointly affect the collective wisdom of the group and the dispersion of their forecasts.

---

[5] Marinovic, Ottaviani and Sørensen (2013) review the literature on strategic forecasting.

[6] The public knowledge bias arises because the same public information is contained in all individual forecasts and, therefore, too much weight is placed on public information relative to private signals when individual forecasts are averaged.

[7] The Great Moderation has been documented by a number of scholars including Kim and Nelson (1990), McConnell, Perez and Quiros (2000), Stock and Watson (2002) and others. Gamber, *et al.*, (2010) show that the standard deviation of annualized real GDP growth rates, measured at quarterly frequencies, fell from nearly 5% between 1947 and 1983 to 2.2% from 1984 to 2008.

The paper provides three main findings. First, there is evidence that the incentive to herd changes over the longer-term forecast horizons and influences the collective wisdom of the Blue Chip group. In particular, during the first six months of the 24-month forecasting cycle, forecast dispersion remains relatively constant and there is little reduction in the size of consensus errors. However, forecasts become more dispersed and consensus errors decline dramatically as the 12-month horizon used to determine the winner of the Klein award approaches. This finding is consistent with the hypothesis that the award promotes anti-herding and collective wisdom.

Second, we provide evidence that temporal variation in herding or model diversity influences the collective wisdom of longer-term forecasts. Simple regression analysis shows that there is a negative and statistically significant relationship between absolute consensus errors and forecast dispersion for each forecast horizon from 24 to 20 months. This indicates that the Blue Chip group makes larger forecast errors during periods when there is greater agreement among its members.

Third, the negative relationship between the size of the consensus errors and forecast dispersion observed for long-term horizons appears to be driven by shifts in behavior over two distinct periods. In particular, forecast dispersion declined precipitously beginning in the early 1990s, which is clearly associated with an increase in consensus errors. In addition, after the early 1990s longer-term forecasts bracketed realized GDP growth much less frequently than in earlier periods. These findings suggest that increased herding or reduced model diversity caused the collective wisdom of the Blue Chip group to start declining in the 1990s.

The paper is organized as follows. Section 2 provides a model of collective wisdom to show how various forces affect consensus accuracy. Section 3 discusses previous research on economic forecasts. Section 4 presents the data and empirical strategy used in this paper. The main findings are provided in Section 5 while Section 6 offers possible explanations for the deterioration in collective wisdom over time. Section 7 concludes the paper.

2. A Model of Collective Wisdom

We begin by presenting a simple model of collective wisdom. Our goal is to show how changes in the interaction between members of a forecasting group, such as a shift from anti-herding to herding or the adoption of similar models, generate a negative relationship between the size of its consensus error and the dispersion of forecasts. In contrast, a positive relationship between these two variables manifests when forecasters operate independently but the volatility of the economic environment changes.

Assume there is a group of n individual forecasters each attempting to predict the future value of a variable. Each individual *i* receives a signal s_i that is distributed conditional on the true value of the variable, *A*. Formally, the signal distribution is given by $f_i(s_i|A)$. Forecasters use their signals to generate predictions with squared prediction errors given by

(1) $$SqE(s_i) = (s_i - A)^2$$

These squared errors reflect forecaster ability as well as the nature of the forecasting environment; if the latter becomes more volatile, squared errors will increase.

The average of the squared individual errors is given by

(2) $$SqE(\vec{s}) = \frac{1}{n}\sum_{i=1}^{n}(s_i - A)^2$$

where $\vec{s} = (s_1, s_2, \dots s_n)$ is the vector of signals received by the n forecasters.

The consensus is the equal-weighted average of the individual predictions

(3) $$c = \frac{1}{n}\sum_{i=1}^{n} s_i$$

and the squared consensus error is

(4) $$SqE(c) = (c - A)^2$$

We can view the squared consensus error as an *ex post* measure of collective wisdom, with smaller consensus errors implying greater collective wisdom.

Finally, the dispersion of predictions is the variance of individual predictions around the consensus

(5) $$Disp(\vec{s}) = \frac{1}{n}\sum_{i=1}^{n}(s_i - c)^2$$

Dispersion measures the *ex post* disagreement or predictive diversity of the forecasters. As we see below, it is determined by the signal distribution variances and covariances.

The *Prediction Diversity Theorem* (PDT) states that the squared consensus error equals the average of the squared individual errors minus the dispersion of predictions[8]

(6) $$SqE(c) \ = \ SqE(\vec{s}) \ - \ Disp(\vec{s})$$

As Page (2007) points out, a striking implication of the PDT is that collective wisdom depends equally on the average individual squared errors and predictive diversity. The first effect is obvious. Collective wisdom is high when the group includes many high-ability individuals or the economic environment is stable and easy to forecast. Less obviously, collective wisdom is high when there is greater predictive diversity. In standard models, this diversity stems from the assumption that signals are independently distributed across forecasters. By taking an average of the forecasts the noise they contain cancels out, decreasing the squared consensus error below the average of the squared individual errors.

Following Hong and Page (2012), we consider a richer framework that allows for bias and focuses on *ex ante* collective wisdom. We define individual *i's* bias as $b_i = \mu_i - A$, where μ_i is the mean of *i's* signal distribution, which is conditional on *A*. Second, the variance of individual *i's* signal is given by $v_i = \mathrm{E}[(s_i - \mu_i)^2]$ where is E[·] is the expectation over all possible signal realizations given the outcome. We now define three additional variables. First, the average bias across the group is

(7) $$\bar{b} \ = \ \frac{1}{n}\sum_{i=1}^{n}(\mu_i - A)$$

In this setting an optimist is a forecaster who has $\mu_i > A$, while a pessimist has $\mu_j < A$.

Second, the average variance of forecaster signals, or average individual uncertainty, is given by

[8] For a proof, see Hong and Page (2012). Engle (1983) and Bomberger (1996) provide earlier derivations in the context of macroeconomic forecasting.

$$(8) \qquad \bar{V} = \frac{1}{n}\sum_{i=1}^{n} E[s_i - \mu_i]^2$$

Third, the average covariance of forecaster signals is

$$(9) \qquad \overline{COV} = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} E[s_i - \mu_i][s_j - \mu_j]$$

Using these terms, the bias-variance-covariance decomposition (BVCD) can be expressed as[9]

$$(10) \qquad E[SqE(c)] = \bar{b} + \frac{1}{n}\bar{V} + \frac{n-1}{n}\overline{COV}$$

where the term on the left side represents *consensus uncertainty*, an *ex ante* measure of collective wisdom.

This identity illustrates three important results. First, consensus uncertainty falls when forecasters are biased in offsetting ways. That is, collective wisdom rises when there are both optimists and pessimists in the group.

Second, consensus uncertainty is increasing in average individual uncertainty. Both decline when forecasters have higher ability (e.g., they use better models and techniques) or the forecast environment becomes more stable. We saw this result in the Prediction Diversity Theorem as a positive relationship between $SqE(c)$ and $SqE(\vec{s})$. Notice, however, that the coefficient of $1/n$ in front of $\bar{V}$ in the BVCD is much smaller than the coefficient of one in front of $SqE(\vec{s})$ in the PDT. This is because an increase in $\bar{V}$ causes both $SqE(\vec{s})$ and $Disp(\vec{s})$ to rise. That is, an increase in the variance of the signal noise causes (i) individual errors to be larger and (ii) predictions to be more dispersed. The latter effect largely offsets the former effect on the size of the *ex post* consensus error. This is reflected in (10) with the small coefficient of *1/n* on $\bar{V}$.

Third, the BVCD shows that the signal covariance plays a fundamental role in collective wisdom. The fact that the coefficient in front of $\overline{COV}$ is much larger than the coefficient in front of $\bar{V}$ (*(n-1)/n* versus *1/n*) makes this point clear. Ceteris paribus, an increase in the covariance of signals leads to smaller forecast dispersion and larger consensus uncertainty. The ensuing reduction in collective wisdom occurs because positively correlated errors are less likely to cancel out. As we discuss below, positive covariation occurs if forecasters herd or use similar predictive models. In contrast, anti-herding produces negative signal covariances, greater forecast dispersion and smaller consensus uncertainty.

3. Previous Research

This section reviews previous research that has explored the forces driving the collective wisdom and predictive diversity of forecasting groups. The first two sub-sections examine how social learning and incentive structures can generate a negative relationship between the size of consensus errors and dispersion over time and across forecasting horizons, while the third sub-section examines how the volatility of the economic environment can produce a positive relationship between these two variables.

3.A. Herding and Anti-Herding

[9] See Hong and Page (2012) for a proof.

In information cascade models, social learning is possible because it is assumed that agents act sequentially. In this setting an agent mimics the actions of earlier movers when the information inferred from their actions overwhelms his own information (Banerjee (1992) and Bikhchandani, Hirschleifer, and Welch (1992)). While ignoring one's private information and herding can be rational for the individual, it is socially inefficient because it prevents private information from being aggregated and can cause the group to converge on the incorrect action. Eyster and Rabin (2010) consider models where agents are inferentially naïve (i.e., they underestimate the extent to which earlier movers have mimicked others) and show that this departure from full rationality increases the likelihood of herding on the wrong target.

A common feature of information cascade models is that agents are assumed to pursue the singular objective of error minimization and the predictions they report are *truthful* in the sense that they reflect this objective. However, it has long been recognized that professional forecasters operate in markets and that strategic or principal-agent considerations may induce them to report *untruthful* predictions intended to optimize a more complex payoff function (Batchelor and Dua (1990a), Laster, et al. (1999), Ottaviani and Sørensen (2006), Lichtendahl and Winkler (2007), Batchelor (2007) and Marinovic, et al., (2013)).

An important example is the model developed by Scharfstein and Stein (1990). Similar to what is observed in information cascade models, they show that managers may ignore private information and mimic the investment decisions of other managers. However, the managers are motivated to herd, not because they are exploiting the information obtained by others, but because they desire to manipulate the labor market's perception of their ability. The market makes a more favorable assessment of a manager when he herds because "smart" managers are assumed to receive informative signals that are correlated (all smart managers observe a piece of the same "truth"), while "dumb" managers receive signals that are uninformative and uncorrelated. Applying this logic to the market for forecasting, it suggests that a forecaster may ignore private information and mimic others to enhance his reputation. Although this behavior is rational from the individual's perspective, it reduces collective wisdom and is inefficient from a social standpoint.

While strategic considerations can lead to herding, they can also induce agents to anti-herd by exaggerating the weight they place on private information. For example, Ottaviani and Sørensen (2006) consider a model where forecasters participate in winner-take-all forecasting contests. Because payoffs decline sharply when forecasters have to share the spot light with others, they have an incentive to locate their forecast away from the consensus. They do so when "the first-order reduction in the expected number of winners with who the prize must be shared more than compensates for the second-order reduction in the probability of winning" (Marinovic, et al. (2013)). Importantly, anti-herding increases collective wisdom because the public knowledge bias associated with the consensus is moderated when forecasters exaggerate their private signals (see Lichtendahl, *et al.,* (2013)).

There is strong cross-sectional evidence for herding in financial markets. For example, Graham (1999) shows that investment newsletters herd on the advice of *Value Line* (the best known newsletter) and, consistent with reputation-based models, analysts with less ability herd more. Clement and Tse (2005) find that stock market analysts with more experience and greater prior accuracy make bolder earnings forecasts and those who issue bolder forecasts tend to be more accurate. They conclude that "bold forecasts incorporate analysts' private information more completely and provide more relevant information to investors than herding forecasts" (p. 307).

Welch (2000) tests time-series implications of herding behavior among stock market analysts. He finds that recommendations are influenced by the consensus from the previous period, but the impact is not significantly stronger when the consensus turns out to be correct in its prediction of subsequent stock price movements. This suggests that analysts herd to protect reputations rather to exploit the fundamental information contained in the forecasts of others. Walsh also finds that herding is stronger when stock prices are rising and concludes that "up-markets may aggregate less information, and therefore… could be more "fragile" than down-markets" (p. 371).

While there is strong evidence of herding in financial markets, the results are mixed for economic forecasting (Batchelor (2007)). For instance, Batchelor and Dua (1990a) argue that forecasters in the Blue Chip survey consistently issue predictions above or below the consensus to differentiate their product in the monopolistically competitive market for forecasting services. According to Batchelor (2007), biased predictions are not the result of biased models. Rather, they result from judgmental manipulations made by forecasters to cultivate reputations as optimists or pessimists so that they can be more attractive to particular client groups.[10] Batchelor and Dua (1992) show that the Blue Chip forecasters revise their predictions toward the lagged consensus over time, but the adjustments are inefficiently small so that initial optimism and pessimism persists as the forecast horizon shrinks. Batchelor (2007) argues that individual biases do not cause the consensus to be biased because the drive for product differentiation produces approximately equal numbers of optimists and pessimists.

Ehrbeck and Waldmann (1996) examine whether forecasters attempt to manipulate beliefs about their ability by mimicking the pattern of predictions made by better forecasters. Their model of rational cheating implies that less able forecasters will report predictions closer to the consensus to exploit its collective wisdom. Using interest rate forecasts from the Blue Chip survey, they find that the model is not supported by the data; less able forecasters (those with large errors) issue forecasts further from the consensus. They conclude that these patterns are consistent with behavioral biases where, for example, overconfident forecasters are less accurate and deviate more from the consensus.

Laster, *et al.,* (1999) provide evidence of anti-herding. They construct a model where forecasters' wages are a function of their accuracy and ability to generate publicity for their firm. These objectives conflict because, as the logic of collective wisdom dictates, predictions that minimize expected errors are likely to cluster. A cross-sectional implication of their model is that forecasters working in industries that provide the largest relative reward for publicity will make predictions that deviate more from the consensus. They argue that independent forecasters in the Blue Chip group benefit more from favorable publicity and provide evidence that they make more extreme predictions compared to those employed by banks, industrial corporations or other organizations.

Lamont (2002) also examines strategic incentives that cause forecaster to issue extreme predictions. Consistent with Laster, *et al.,* (1999), he finds that forecasters who participate in the *Business Week* survey issue bolder GDP growth predictions after they establish their own firm. In addition, predictions become bolder as forecasters age. Lamont explains this finding by arguing that the market develops "tighter priors" about a forecaster's ability as he gains experience so that herding becomes an ineffective strategy for manipulating beliefs. He also finds that the bolder forecasts are also less accurate and concludes: "Clearly, it would not be socially useful if all forecasters sought to minimize mean squared error by mimicking consensus… It may well be that as forecasters age they contribute more information to the collective process that establishes consensus forecasts" (p. 279).

Overall, the evidence suggests that economic forecasters anti-herd rather than herd. This finding is important because, as discussed earlier, anti- herding reduces the public knowledge bias inherent in the consensus and increases collective wisdom. However, it is important to recognize that these findings are largely based on short samples that end in the late 1980s or 1990s and focus on only one forecast horizon. There is no guarantee that the incentives which led to anti-herding in the past have remained intact over time or that these incentives are the same for each forecast horizon.

3.B. Model Diversity

Herding and anti-herding can occur in sequential decision making environments where agents are able to observe the actions (forecasts) of others. However, social learning can also lead to behavioral

[10] Batchelor (2007) obtains similar findings using macroeconomic forecasts for multiple countries from *Consensus Economics*.

convergence when it induces individuals to adopt similar conceptual frameworks or models. In contrast to the herding literature, little research has examined the forces which generate model diversity and their impact on collective wisdom.

One exception is Batchelor and Dua (1990b) who find that there is a great deal of diversity in the Blue Chip forecasting group. While Keynesianism was the most popular theoretical framework used by its members in the late 1980s, they also relied on monetarist, supply-side and other types of models. In terms of forecasting methods, judgment was more popular than formal modeling. Batchelor and Dua (1995) compare the accuracy of different consensus forecasts constructed by taking averages of different subsets of individual forecasts made by members of the Blue Chip group. They find that the accuracy of the consensus forecasted increased when the component forecasts were produced by a more diverse set of models and methods.

Hong, *et al.*, (2012) use simulation methods to explore how different elements of the social environment influence individual and collective accuracy. In particular, they consider two information structures: one where individuals operate in isolation and the other where they are able to adopt other's models, and two payoff incentives: individual incentives where earnings are determined exclusively by accuracy and market incentives where earnings are proportional to the inverse of the percentage of other individuals who issue correct predictions. They find that collective wisdom is maximized when predictions are made in isolated environments and payoffs are determined by market incentives. Importantly, this is the same setting which generates the least accurate individual predictions. They reconcile these seemingly contradictory results by arguing that mechanisms which induce individuals to rely on their own models increase collective wisdom by promoting diversity.[11]

This research raises additional questions. What changes in the social environment make it easier for forecasters to adopt the models of others? Does increased reliance on computer algorithms facilitate this by making it easier for individuals to transfer models to others? How do market forces influence payoff structures and do they change over time? How do broader developments in the field of macroeconomics affect model diversity? As Page (2014) points out, it is not at all clear that there is an invisible hand process at work guaranteeing that diversity equilibrates at some ideal level.

3.C. Volatility of the Economic Environment

The preceding discussion suggests that changes in herding behavior over time or across forecasting horizons produce a negative relationship between the size of consensus errors and forecast dispersion. A negative relationship is also generated by changes in model diversity. In contrast, we should observe a positive relationship between the size of consensus errors and forecast dispersion if there is no interaction between forecasters, but the volatility of the economic environment changes over time or across horizons.

Zarnowitz and Lambros (1987) provide the seminal work linking forecast dispersion to economic volatility. Using density forecasts for inflation from the NBER-ASA survey, they find a positive relationship between the standard deviation of point forecasts across survey respondents and uncertainty measured as the average standard deviation of the respondents' predictive probability distributions. They speculate that increases (decreases) in macroeconomic volatility cause both forecaster uncertainty and the dispersion of point forecasts across forecasters to rise (fall). Giordani and Sönderlin (2003) provide similar results and Bomberger (1996) uses point forecasts for inflation from the Livingston survey and finds a positive relationship between forecast dispersion and the conditional variance of the consensus forecast error. Several scholars have challenged these finding by showing that there is little evidence of a positive relationship between forecast dispersion and uncertainty when longer sample periods or different estimation techniques were used (Rich and Butler (1999), Lahiri and Liu (2006) and Rich and Tracy

[11] This result might also help us understand Lamont's finding if more experience forecaster operate more in isolation of their peers and care winning forecasting competitions.

(2010)). Boero, Smith and Wallis (2012) reconcile the conflicting results by arguing that longer sample periods include the Great Moderation and high volatility in the economic environment is necessary to generate a positive relationship between forecaster dispersion and uncertainty.

Lahiri and Sheng (2010) use a Bayesian learning model to examine the link between forecast dispersion and uncertainty. They assume that individuals form expectations by combining private signals with a common public signal. When private signals are independent of one another and uncorrelated with the public signal, forecast dispersion is simply the average of the private signal variances while the average level of individual uncertainty is equal to forecast dispersion plus the variance of the common public signal. In this case changes in the volatility of the economic environment do not affect forecast dispersion. When these assumptions are relaxed and the variances of private and public signals are correlated, forecast dispersion and the volatility of the economic environment are positively related.

Lahiri and Sheng (2008) use fixed-target GDP growth forecasts from Consensus Economics Inc. to explore the behavior of forecast dispersion across forecast horizons. They find that dispersion is relatively high for long-term forecasts and declines as the target date approaches and horizons shorten. They argue that dispersion is initially high because forecasters hold divergent prior beliefs. As the horizon shortens, uncertainty declines and this causes forecasters to disagree less. Patton and Timmermann (2010) also conclude that high dispersion at long horizons is due to the heterogeneity of models and priors and find that the dispersion decreases as horizons shorten.

Overall, theory suggests that changes in the uncertainty will produce a positive relationship between the size of consensus errors and forecast dispersion under certain conditions. Empirically, there is mixed evidence for a positive relationship over time and across forecast horizons.

4. Data and Empirical Strategy

The data used in this study are taken from the *Blue Chip Economic Indicators* newsletter. Since 1976, the editors of the newsletter have surveyed professional forecasters at the beginning of each month and reported their forecasts for a variety of macroeconomic variables a few weeks later. The newsletter's founder, Bob Eggert, was a pioneer in the use of combination methods to improve forecast accuracy (see Clemen (1989)) and reported the consensus along with the individual forecasts made by each respondent. The survey initially included about 35 forecasters with the number of contributors rising to around 50 in the 1980s. To promote a diversity of perspectives, forecasters from different sectors of the economy have always been included in the survey.

To examine whether social learning and strategic incentives have influenced the collective wisdom of the Blue Chip group, we focus on the fixed-target forecasts for annual real GDP growth made between 1976 and 2011. The survey respondents make their first forecast 24 months prior to the end of target year t. That is, they start forecasting GDP growth for year t in January of year t-1. They revise these forecasts each month and make their final one in early December of year t.

There are several reasons why the Blue Chip group is ideal for studying the impact of social learning and strategic incentives on collective wisdom. First, the participants are able to observe the forecasts of others in the group with a short delay of a few weeks and they know the identity of the other forecasters. Second, the information contained in these forecasts remains relevant for many months because they are for fixed targets.[12]

Third, the Blue Chip respondents provide both *current-year* and *year-ahead* forecasts each month. That is, they report forecasts with horizons of h and $h+12$ months where h is the number of months

[12] In contrast, the information in fixed-horizon forecasts quickly become stale because the target changes.

remaining in the year (e.g., h equals 11 in February).[13] This feature is important because uncertainty is increasing in the forecasting horizon and the incentive to herd is influenced by uncertainty.[14]

Fourth, while publication of the fixed-target forecasts makes social learning possible other features influence strategic incentives. One is the Lawrence R. Klein Award given annually to the participant with the most accurate forecasts over the past four years.[15] The award garners media attention and likely enhances the reputation and career prospects of winners. For example, Laurence H. Meyer won the award in 1993 and 1995 and was appointed to the Federal Reserve Board in 1996. As we discussed in Section 3, the Klein award should promote anti-herding to the extent that it creates winner-take-all incentives. However, the award is based exclusively on the accuracy of 12-month-ahead forecasts made in January of year t and this raises the possibility that the incentives influencing strategic behavior vary over the 24 forecasting horizons.

Finally, the Blue Chip data provide a long time-series which includes four recessions, structural changes such as the Great Moderation, major advances in computing power, the rise of the internet, paradigm shifts in macroeconomics, and so on. It is possible that these developments have influenced the opportunities for social learning, strategic incentives and changes in macroeconomic volatility.

The empirical strategy for this paper is straightforward. As discussed in Sections 2 and 3, we hypothesize that there are two basic forces driving variation in consensus accuracy and forecast dispersion and each predicts a different relationship between these two variables. If the volatility of the economic environment changes over time and is the primary force influencing the size of consensus errors and dispersion of forecasts across members of the group, we expect these two variables to be positively related. That is, periods of high (low) volatility should be associated with large (small) consensus errors and high (low) forecast dispersion. The same logic holds across forecast horizons: as the target date approaches and information accumulates, consensus errors and forecast dispersion should both decline. If, in contrast, there are significant changes in the propensity to herd or model diversity over time or across forecasting horizons, we should observe a negative relationship between the size of consensus errors and forecast dispersion. That is, increased herding over time should cause forecast dispersion to decline and consensus errors to increase. If the Lawrence R. Klein Award promotes anti-herding, forecast dispersion should increase and consensus errors decline as the forecast horizon declines to 12 months.

To examine the relationship between forecast dispersion and the magnitude of consensus errors, we construct measure of each using data from the Blue Chip survey. The consensus is measured as the cross-sectional mean of individual forecasts

$$(11) \qquad f_{th} = \frac{1}{N_t}\sum_{i=1}^{N_t} f_{ith} \qquad\qquad t = 1977, 1978, \dots, 2011; h = 1,2,\dots,24.$$

[13] The newsletter began collecting and reporting a full set of 12 current-year and 12 year-ahead forecasts each year in 1984. Prior to 1984, the panelists made current-year forecasts between January and June and year-ahead forecasts between July and December.

[14] Krishnan, Lim and Zhou (2006) show that stock market analysts herd more when the forecasting horizon increases. Marinovic, et al., (2013) argue that this is because private signals are relatively imprecise at longer horizons.

[15] The winner is determined based on accuracy for four variables: real GDP growth, CPI inflation rate, T-bill rate and unemployment rate. The absolute difference between the current-year forecasts issued in January and actual values of the variables are used to measure accuracy. Actual real GDP growth is measured using the revised figures released in July of the following year. Overall accuracy is then calculated by taking an equally-weighted average of the absolute errors for the four variables over the previous four years. The authors thank Lee McPheters for sharing information about the Lawrence R. Klein Award.

where f_{ith} is member i's forecast for real GDP growth over year t, made h months prior to the end of year t. Forecasts with *h = 1, 2, …12* are referred to as *current-year forecasts* and those with *h = 13, 14,… 24* as *year-ahead forecasts*. The number of forecasters changes over time and is given by N_t. Forecast dispersion is the cross-sectional variance of individual forecasts around the consensus

$$(12) \qquad d_{th} = \frac{1}{N_t}\sum_{i=1}^{N_t}(f_{ith} - f_{th})^2 \qquad t = 1977, 1978, \dots, 2011; h = 1,2,\dots,24.$$

Finally, the absolute value of the consensus error is

$$(13) \qquad ACE_{th} = |A_t - f_{th}| \qquad t = 1977, 1978, \dots, 2011; h = 1,2,\dots,24.$$

where A_t is actual year-over-year real GDP growth rate for year t and f_{th} is the consensus forecast issued h months prior to the end of year t.

It is well known that forecast errors for GDP growth are sensitive to the vintage of data used. To explore the robustness of our results, we use two vintages of measured real GDP growth. The first, which we refer to as the *Early Vintage*, is reported in the June issue of the *Survey of Current Business* published in year t+1. The second, the *Later Vintage*, is reported in the *Survey of Current Business* in June of year t+2.

5. Empirical Results

To evaluate whether the collective wisdom in the Blue Chip consensus is influenced by herding and model diversity, we examine the relationship between consensus errors and the dispersion of forecasts across the group's members. Our theory posits that fluctuations in consensus errors produced by changes in the volatility of the economic environment imply a positive relationship between the magnitude of the errors and the dispersion of forecasts. In contrast, fluctuations in consensus errors produced by changes herding or model diversity imply a negative relationship between the magnitude of consensus errors and forecast dispersion. We begin by examining the covariation in these two variables across forecasting horizons and then turn to examine the relationships over time.

5.A. Covariation Across Horizons

If there is no interaction between forecasters and they simply attempt to minimize prediction errors, forecast dispersion and absolute consensus errors should be positively correlated over forecasting horizons. As the horizon shortens (i.e., the forecasting target period nears), information accumulates, leading to less dispersion and smaller forecast errors.

Figure 1 illustrates averages of the absolute consensus errors and forecast dispersion for each of the 24 monthly horizons.[16] Clearly, as the current-year forecasting horizons shorten the absolute consensus errors and forecast dispersion both decline sharply. This is what we expect to observe when information accumulates and forecasters focus exclusively on error minimization. However, it is important to note that the rate of information accumulation over the current-year is very high because GDP announcements

[16] The averages are calculated over the 1986-2011 sample period—the longest span of time that includes forecast errors for all horizons in each year. Early Vintage GDP estimates were used to measure the consensus errors seen in Figure 1.

throughout the year reveal part of the forecasted variable.[17] Thus the rapid reduction in uncertainty likely dominates the impact of any changes in the incentive to herd that might occur over current-year horizons.

This type of rapid information accumulation is not possible prior to the target year because GDP announcements do not reveal direct information about GDP growth over the following year. However, Figure 1 shows that the Blue Chip consensus errors begin to fall sharply in August prior to the target year (i.e., in year t-1) after remaining stable over the first seven months. One explanation for this pattern suggested by Lahiri and Sheng (2008) is that information becomes more precise during the second half of t-1. In particular, they argue that "the first major public information seems to arrive when the forecast horizon is around 15 months…" (p. 326).

An alternative explanation for the large decline in year-ahead consensus errors in the second part of the year is that the incentives forecasters face change over the horizons. In particular, increases in the incentive to anti-herd could lead to smaller consensus errors as forecasters choose to place greater weight on their private information, thereby reducing the public information bias contained in the consensus. The fact that forecast dispersion begins to rise in August of year t-1 – the same month in which the absolute consensus error begins to plummet – is consistent with this hypothesis. If information accumulation was causing the consensus error to decline we would expect forecast dispersion to also decline. The finding that dispersion increases when the consensus error falls is consistent with the hypothesis that the incentive to anti-herd has risen.

Why would anti-herding increase during the second half of year t-1? As we discussed above, the Lawrence Klein Forecasting Award creates a winner-take-all competition that may promote anti-herding. Moreover, the prize is awarded only on the basis of 12-month-ahead forecasts issued in January of the current year; performance in the other months does not matter. Thus we would expect to observe increased forecast dispersion in the months leading up to the 12-month horizon if the Klein Award caused forecasters to anti-herd. This is what we observe in Figure 1. Therefore, the finding that forecast dispersion and the absolute consensus errors diverge as the horizon shrinks from 17 to 12 months – and converge after this pivotal point in the forecasting cycle – is consistent with the hypothesis that the award induces anti-herding, thereby increasing consensus accuracy.

Patton and Timmerman (2010) use real GDP growth forecasts from Consensus Economic Inc. and also find that consensus errors and forecast dispersion move in the opposite direction starting in August of t-1. They conclude that this result is "difficult to explain within the confines of our model, or indeed any model assuming a quadratic penalty for forecast errors and efficient use of information, and thus poses a puzzle" (p. 13). Our analysis suggests that variations in the incentives for strategic interaction over the forecast horizons can explain this puzzle.

While the absolute consensus errors decline precipitously during the last part of year t-1, they change very little during the first seven months of year t-1 (as the forecast horizon falls from 24 to 18 months). Moreover, forecast dispersion remains stable over this period and below the level observed later in the year. One explanation for these patterns is that little information is revealed over this period. Alternatively, it is possible that increased herding over the first half of t-1 reduced forecast dispersion and prevented consensus errors from declining.[18] In fact, there is evidence that signals are less precise at

[17] For example, in January the survey respondents forecast GDP growth for the entire current-year. However, by May they know first-quarter GDP growth and only need to forecast growth over the final three quarters. In August, GDP growth over the first two quarters is known and uncertainty about annual growth falls further.

[18] Krishnan, Lim and Zhou (2006) show that herding increases in the length of the forecasting horizon.

longer horizons[19] and lower signal precision increases the likelihood of incorrect information cascades where forecasters collectively converge on the wrong target (see Figure 1 in Bikhchandani, et al. (1992)).

Overall, Figure 1 provides evidence that strategic interaction impacted the collective wisdom contained in longer-term forecasts. It appears that herding prevented the absolute consensus errors from declining during the first half of year t-1, while anti-herding reduced the errors during the second half of the year. The strong positive correlation between the two series over year t is consistent with the hypothesis that increases in the private and public information precisions caused both forecast dispersion and the absolute consensus errors to decline.

5.B. Covariation Over Time

The previous section focused on the variation of absolute consensus errors and forecast dispersion over forecasting horizons. This section presents evidence about the covariation of these variables over time. To do so, we estimate the following regression

$$(14) \qquad ACE_{th} = \alpha_h + \gamma_h d_{th} + u_{th}, \qquad t = 1977, 1978, \ldots, 2011; h = 1,2, \ldots, 24.$$

where, recall, ACE_{th} is the absolute consensus error of real GDP growth for target year *t*, based on forecasts made *h* months prior to the end of target year *t*, d_{th} is the dispersion of forecasts with a horizon of *h* months, α_h is a constant term for horizon *h*, γ_h is a coefficient which measures the relationship between the absolute consensus error and forecast dispersion at horizon *h*, and u_{th} is the regression residual. Given the results of the previous section, we allow for the possibility that the coefficients differ across the horizons and estimate separate regressions for each *h*. If forecast dispersion is driven predominantly by changes in the variability of the economic environment, we should find that $\gamma_h > 0$. If the propensity to herd and anti-herd changes over time and are quantitatively more important determinants of forecast dispersion than are changes in the variability of the economic environment, we should observe that $\gamma_h < 0$.

Table 1 shows estimates of (14) for current-year forecasts.[20] Results are shown for both early and later vintage estimates of actual GDP growth. The table shows two important findings. First, the constant term declines as the forecast horizon shortens. For instance, when using early vintage data, the average absolute consensus error for GDP growth is 0.67 percent at the 12-month horizon and 0.16 percent at the 1-month horizon. This finding is consistent with what we saw in Figure 1 and provides additional evidence that consensus forecast errors shrink as the target date approaches and more information becomes available.

The second important finding is a positive relationship between forecast dispersion and the absolute consensus error. In all but one of the 24 regressions, the slope coefficient is positive. While none of the coefficients are significantly different from zero when we use early vintage data, it is important to point out that forecast errors, by their nature, are difficult to explain using observable variables. Also, we are working with relatively few observations. When later vintage data are used, the slope coefficient is positive and significantly different from zero for four of the 12 horizons (6, 7, 11 and 12). Overall, the positive relationship between current-year forecast dispersion and absolute consensus errors is consistent

---

[19] Using GDP growth forecasts from Consensus Economics Inc., Lahiri (2011) finds that forecasters do not receive dependable information to update their forecasts over 18- to 24-month horizons.

[20] Note that the number of observations used to estimate the regressions changes over the forecast horizon. As discussed earlier, current-year forecasts made in January through May (horizons 12 to 8) are available for the entire 1976-2011 period. Current-year forecasts with 7-month horizons first became available in 1980 while those with 6- to 1-month horizons were first published in 1984.

with the hypothesis that changes in the variability of the economic environment are the primary force driving both variables.

Table 2 shows estimates of (14) for year-ahead forecasts.[21] These results stand in stark contrast to those presented in Table 1. As we saw for the current-year forecasts, the constant term generally declines as the forecast horizon falls from 24 to 13 months. However, Table 2 shows negative slope coefficients for 18 of the 24 regressions shown and five of the 18 are significantly different from zero at the five percent level or higher. All of the slope coefficients that are significantly less than zero are for forecasts with horizons of 20 months or greater. Recall that the slope coefficient should be positive if changes in forecast dispersion and the absolute consensus errors are both being driven by changes in the volatility of the economic environment. Negative slope coefficients, in contrast, are consistent with the hypothesis that time-varying herding or model diversity jointly affect the dispersion of forecasts and the collective wisdom of the Blue Chip group.

The regression estimated with 23-month-ahead forecasts provides an important case. For both data vintages, the slope coefficients and associated t-statistics take on their most negative values at this horizon. In addition, the intercept increases relative to the value observed in the regression using forecasts with 24-month horizons. That is, the consensus is less accurate as the group moves closer to the target date. These two findings are consistent with the hypothesis that increased herding or reduced model diversity over time had a large detrimental impact on the collective wisdom of the group at the 23-month horizon.

Why are these results strongest at the 23-month horizon? It is important to note that this is the first horizon where members of the Blue Chip group have the opportunity to observe what others forecasted in the previous month and the incentives to anti-herd are likely weak at this horizon. Therefore if the propensity to herd changed over the sample period we would expect it to change the most at the horizon where herding is more likely. In addition, Patton and Timmermann (2010) show that forecasters disagree at the longer horizons largely because of the different models they utilize. Therefore, changes in the diversity of models over time are more likely to impact the collective wisdom of longer-term forecasts. Taken together, these observations suggest that it is not surprising that the strongest negative relationship between absolute consensus errors and forecast dispersions is observed at the 23-month horizon.

5.C. Temporal Clustering

One possible explanation for the correlations shown in the previous section is that forces driving consensus errors cluster temporally. In particular, the Great Moderation could account for the positive relationship between forecast dispersion and the absolute consensus errors for current-year forecasts shown in Table 1. If this is the case, we should observe high values of forecast dispersion and absolute consensus errors in the early part of the sample and lower values of both variables after the mid-1980s when the macroeconomy became less volatile. In contrast, increased herding over time could account for the negative relationship between longer-term forecast dispersion and absolute consensus errors shown in Table 2. A reduction in model diversity over time could also be driving this negative relationship.

To examine these issues, forecaster dispersion and the absolute consensus errors for four different horizons are presented in Figures 2 through 5. Two important findings emerge.

First, forecast dispersion decreased dramatically for all horizons beginning in the early 1990s. For instance, consider the cross-sectional variance of 18-month forecasts – the longest horizon for which data are available back to the 1970s – shown in Figure 3. It exceeds 1.0 six times between 1977 and 1990, but

[21] Once again, note that the number of observations used to estimate the regressions changes over the forecast horizon. Year-ahead forecasts made in July through December (horizons 18 to 13) are available for the entire 1976-2011 period. On the other hand, year-ahead forecasts with 19-month horizons first became available in 1982 while those with 24- to 20-month horizons were first published in 1985.

rose above 0.5 on only one occasion after 1990. Even during Great Recession, 18-month-ahead forecast dispersion was less than half the level reached in the mid-1980s. Similarly, 6-month-ahead forecasts exceeded 0.4 three times between 1977 and 1990 and remained well below 0.2 after 1990 (Figure 5).

Second, the decline in the dispersion of longer-term forecasts over the sample appears to be associated with an increase in the magnitude of consensus errors. For instance, Figure 2 shows that there were large 24-month-ahead consensus errors associated with the dot-com boom and bust of the late 1990s and early 2000s and the Great Recession of 2008-2009, while forecast dispersion remained at very low levels during both episodes. These patterns are consistent with the hypothesis that increased herding or reduced model diversity caused the collective wisdom of the Blue Chip group to decline starting in the 1990s.

To formally examine whether the level of forecast dispersion changed over the sample, we test for structural breaks in the series. To do this, Table 3 provides estimates of pooled regressions of the cross-forecaster variance of real GDP forecasts on a constant term, 11 month dummy variables (February through December), and a dummy variable that takes on values of one starting in different possible break years (*Break Dummy*). The constant term provides an estimate for the average level of forecast dispersion in January.[22] Based on the patterns seen in Figures 2-5, we consider each year from 1989 to 1996 as a potential point of structural break.

The results in Table 3 provide strong evidence that forecast dispersion declined in the early 1990s. For each possible break year, and both current-year and year-ahead forecasts, the coefficient on the break dummy is negative and statistically significant. Using the R-squared as the criterion for determining structural change, we conclude that the structural break in the level of dispersion mostly likely occurred in 1992 for current-year forecasts and 1991 for year-ahead forecasts (forecasts issued in 1990). The magnitude of the declines was large in both cases. For example, the average level of dispersion for the 12-month forecasts was 0.54 from 1977 to 1991 and 0.33 (0.54 minus 0.21) from 1992 to 2011. This is a 39 percent decline. The decline at the 24-month horizon was even more dramatic with dispersion falling from 1.02 prior to 1991 to 0.34 starting in that year, a 67 percent decline.

The next question we examine is whether the absolute consensus errors rise or fall after the break dates identified for forecast dispersion. To do this we estimate pooled regressions of the absolute consensus errors on a constant term, 11 month dummy variables (February through December), and a dummy variable that takes on values of one starting in years identified in Table 3 as break years for forecast dispersion: *Post-1991* for current-year forecasts and *Post-1990* for year-ahead forecasts. The constant term provides an estimate for the average level of the absolute consensus error in January. Tables 4 and 5 provide results for these regressions using two the different data vintages discussed above and both the full sample (1977-2011) and a shorter one which is the longest possible sample that includes data for each forecast horizon for each year (1986-2011).

The results in Table 4 provide no evidence of a change in the size of current-year consensus errors after 1991. The coefficient on the Post-1991 dummy variable is near zero and the t-statistics are low in each of the four regressions. The finding that the size of the consensus errors did decline is surprising given the Great Moderation and its presumed impact on the predictability of GDP growth after the min1980s. One possible explanation for this is that the positive impact of the Great Moderation on collective wisdom is offset by the negative impact of herding or reduced model diversity.

[22] That is, in regressions using current-year (year-ahead) forecasts this is the average level of forecast dispersion over the entire sample for 12-month (24-month) horizons. Inclusion of the month dummies in the pooled regressions allows the average level of dispersion to vary across the forecast horizons. Estimates of the coefficients on the month dummies are not shown in the table.

The results in Table 5 provide strong evidence that the size of year-head consensus errors increased after 1990. The coefficient on the Post-1990 dummy is positive in each of the four regressions and significantly different from zero in all cases. Moreover, the magnitudes are large and meaningful. For example, consider the case when early vintage data and the full sample are used. The mean January absolute consensus error is 0.88 prior to 1991 and 1.22 (0.88 plus 0.34) from 1991 onward, a 39 percent increase. The increase is even more dramatic when we consider the 1986-2011 sample period. In this case, the absolute consensus error rises 85 percent (from 0.68 to 1.26). These results are consistent with the hypothesis that increased herding or reduced model diversity caused collective wisdom of the Blue Chip group to decline starting in the early 1990s.

5.D. Forecast Bracketing

An alternative way to measure the collective wisdom of the Blue Chip group is to examine whether the range of forecast "bracket" or include the realized value of the variable being forecast.[23] When realized GDP growth does not fall between the most optimistic and most pessimistic forecast – that is, when forecasts do not bracket the truth – the consensus cannot be any more accurate than the average member of the group. In contrast, the consensus improves upon the average member when forecasts bracket the truth. This can also be seen in the bias-variance-covariance decomposition given in (10); the expected squared consensus error is increasing in average bias.

To illustrate, consider a simple example with two forecasters. Alice predicts that GDP will grow at 7 percent and Bill predicts 10 percent growth. If actual GDP growth is 9 percent, the absolute error for Alice and Bill are 2 and 1 percent, respectively, and their average error is 1.5 percent. The consensus is 8.5 percent and has an absolute error of 0.5 percent. Clearly, the accuracy of the consensus exceeds the average individual accuracy; bracketing increases collective wisdom.[24] Contrast this case to one where Alice continues to predict 7 percent, but Bill mimics Alice to some extent and predicts 8 percent. The average individual error continues to be 1.5 percent, but predictions are less dispersed and no longer bracket the truth (9 percent). The consensus error rises to 1.5 percent and collective wisdom diminishes. In general, the accuracy of the consensus cannot exceed average individual accuracy when predictions do not bracket the truth.

To examine whether the Blue Chip forecasts bracket the truth, Figures 6A and 6B show the range of GDP growth forecasts – the difference between the most optimistic (high) and pessimistic (low) forecast – and actual GDP growth. Figure 6A shows these variables for 9-month-ahead forecasts made in April of year t and Figure 6B shows them for 18-month-ahead forecasts made in July of year t-1. Two important findings emerge.

First, consistent with what we observed for forecast dispersion (Figures 2-5) the range of forecasts declined dramatically over the sample period. For the 9-month-ahead forecasts shown in Figure 6A, ranges of 3-5 percentage points were common prior to 1991 but rare after this year. Starting in 2005 ranges of around 1 percentage point were common and even during the depths of the Great Recession in 2009 the difference between the most optimistic and pessimistic forecast was only about 3 percentage points. A similar decrease in the range of forecasts is seen at the 18-month horizon in Figure 6B. While the most optimistic and pessimistic 18-month forecasts differed by over five percentage points seven times prior to 1990, we observe a range this large only once from 1990 onward (2000). The smallest forecast ranges, of around one percentage point, are observed in 2007 and 2008 at the start of the Great Recession.

---

[23] See Soll and Larrick (2009).

[24] Alternatively, if we were to randomly select one of individual forecaster the expected value of their absolute error (0.5 x 1% + 0.5 x 2% = 1.5 %) is greater than the absolute consensus error (9% - [0.5 x 7% + 0.5 x 10%] = 9% - 8.5% = 0.5 %).

Second, the frequency with which GDP growth forecasts bracket the realized value diminishes significantly starting in 1990. Prior to that year, realized GDP growth fell between the most optimistic and most pessimistic 9-month-ahead forecast each year. In contrast, forecasts failed to bracket GDP growth on four occasions starting in the mid-1990s (1995, 1997, 1998 and 2011). Results for the 18-month horizon are more extreme. The forecasts failed to bracket realized GDP growth on only one occasion prior to the 1990s (1982) and 10 times after 1990 (1991, 1994, 1997, 1998, 1999, 2000, 2001, 2008, 2009 and 2011). Note that the most optimistic forecaster underestimated growth four years in a row between 1997 and 2000. Then, in 2001, the entire group overestimated growth. It is noteworthy that the forecasts failed to bracket realized GDP growth in 2008 and 2009 by large margins, two years when forecasts were highly clustered.

When evaluating the accuracy of forecasts it must be kept in mind that forecasting economic growth is challenging and unforeseen shocks can produce large errors. However, the increase in the frequency with which the Blue Chip group failed to bracket GDP growth after the early 1990s provides additional evidence that the accuracy gain obtained by combining forecasts has fallen in recent years. This is particularly the case for longer-term forecasts where the deterioration in bracketing has been more pronounced. Lorenz, et al., (2011) argue that a decline in bracketing "corrupts the wisdom of the crowd from an observer's perspective in the sense that the group becomes less reliable in guiding decision-makers." This point raises the possibility that the decline in the collective wisdom of the Blue Chip group over the past two decades may have played a role in the Great Recession by leading to poorer decision making.

6. Discussion

The previous section showed that forecast dispersion decreased dramatically after the early 1990s and was accompanied by larger long-term consensus errors. One explanation for this finding is that the propensity to herd (anti-herd) has increased (decreased) in recent decades and reduced the collective wisdom of the Blue Chip Group. Similarly, it can be explained by a reduction in the diversity of models utilized by forecasters. However, these explanations beg the obvious questions: Why would herding increase and model diversity decline over time?

On the issue of model diversity, Batchelor and Dua (1995) have shown that combinations of Blue Chip forecasts produce smaller errors when the individual forecasts are based on a more diverse set of modeling assumptions (Keynesian, monetarist, supply-side, etc.) and different forecasting methods (judgment, econometric modeling, time-series analysis, etc.). If forecasters have come to rely on a narrower set of assumptions and techniques over time, this development could reduce the dispersion of forecasts and the collective wisdom they produce. In fact, Woodford (2009) has concluded that paradigm shifts and other developments in the field of economics have reduced diversity:

> While macroeconomics is often thought of as a deeply divided field, with less of a shared core and correspondingly less cumulative progress than other areas of economics, in fact, there are fewer fundamental disagreements among macroeconomists now than in past decades. This is due to important progress in resolving seemingly intractable debates. (p. 267)

An increasingly "shared core" has the potential to increase collective wisdom if it improves the predictive ability of individuals. However, the loss of diversity that it engenders may offset this positive effect and cause collective wisdom to decline. This is the central message of the Prediction Diversity Theorem discussed in Section 2.

Another explanation for reduced model diversity is that advances in computing technology have made it easier for forecasters to share models. This possibility was recognized long ago by Victor Zarnowitz (1992) who points out that there are multiple inputs into the forecasting process. One is the expert human judgment necessary to analyze data and events. It is innate or acquired through experience and cannot be easily transferred to others. Another input is the computer. It facilitates the measurement and codification of complex historical relationships in data, which are used to predict the future. Unlike human judgment, computer algorithms can be easily transferred between forecasters and greater reliance on them can lead to reduced diversity:

> A free market exists in economic data and ideas, and advances in forecasting technology are soon open to all practitioners. This openness tends to reduce the diversity of individual predictions created by the undoubted fact that forecasters differ greatly in theoretical orientation and training and talent and experience. (p. 133)

Although increased use of computers may allow individual forecaster to be more accurate, it could come at the cost of reduced diversity which reduces collective wisdom.

In the early years of the Blue Chip survey a distinction was made between "econometric modelers" (e.g., DRI, Wharton Econometrics, the Michigan Quarterly U.S. Model, etc.) and other forecasters who, presumably, did not use large computer models to generate forecasts. However, the emergence of low-cost computing has made it possible for all forecasters to utilize complex models which can be easily shared.

A third explanation for the reduction in predictive diversity is that the incentive to herd has changed over time. As discussed above, the Lawrence Klein forecasting award creates a winner-take-all contest and we provided evidence that it increased the incentive to anti-herd as the 12-month horizon approached. While it is not clear how the prestige associated with the award has changed over time, the real value of the cash prize has declined. When the award was first given in 1981 the winner received $5,000. This amount did not change over the next two decades so its purchasing power fell due to persistent inflation. The award was not given in 2001 and 2002 following the 9/11 terrorist attack and the cash prize has recently been discontinued. To the extent that the award's value has diminished over time, the incentive to anti-herd should have also declined.

A fourth explanation for a loss of diversity is that the types of organizations that employ forecasters have change over time and this has altered forecaster payoff functions. Laster, et al., (1999) showed that independent forecasters issue bolder forecasts than other members of the Blue Chip group and argue that this behavior is driven by the fact that independent forecasters have greater need for publicity. If this is true and the number of independent forecasters in the Blue Chip group decreased over time, we would expect the amount of anti-herding within the group to decrease.

To examine changes in the composition of the Blue Chip group, Figure 7 shows the different types of forecasters that reported predictions in odd numbered year between 1977 and 2011. Following Laster, et al., (1995, 1999), forecasters are categorized into ten types: banks, securities firms, industrial corporations, independent forecasters, econometric modelers, financial publications, government agencies, industry associations, insurance companies, ratings agencies.[25] The last five types are combined into the "other" category.

Figure 7 reveals two important findings. First, the number of forecasters increased sharply between 1977 and 1985 before stabilizing around 50. All else held equal, the increase in the number of forecasters should have reduced the size of consensus errors by enabling more idiosyncratic noise in individual forecasts to cancel out when they are averaged. If this effect was operative for the long-term forecasts, it was offset by other forces which caused longer-term consensus errors to increase in size.

Second, the number of independent forecasters has fluctuated over time. It increased rapidly during the first decade and peaked at 12 in 1989. The number then fell during the 1990s, reached a low of 7 in 1999, and stabilized around 10 in the 2000s. To the extent that independent forecasters have stronger

[25] See the Appendix for the categorization of specific forecasters.

incentives to anti-herd, the dispersion of forecasts and accuracy of the consensus should have been higher in second half of the 1980s and lower in the 1990s. This is consistent with the pattern of long-term absolute consensus errors observed in our sample.

However, it is possible that the incentive inducing independent forecasts to anti-herd might has declined over time and this would help explain the reduction in forecast dispersion and increase in long-term consensus errors. In fact, Lahiri and Sheng (2008) show that independent forecasters no longer make more extreme predictions than other members of the Blue Chip group when more recent data is considered. One possible explanation for this change in behavior is that the internet has made it possible for independent forecasters to generate publicity in other ways.

Finally, it is possible that a natural cycle in collective wisdom is created as a result of the interaction between the incentives to herd, independence, and collective wisdom (cf. Hirshleifer and Teoh (2008)). Individuals have an incentive to herd to the consensus to increase the accuracy of their predictions. However, widespread use of this strategy causes independence and the collective wisdom embodied in the consensus to diminish. When forecasters discover this, they rely more on private information and this increases independence and collective wisdom. Once again forecasters have an incentive to herd and the cycle repeats itself. From this perspective, the low levels of forecast dispersion and large consensus errors in the 1990s could simply reflect the collective wisdom of the Blue Chip group in the 1980s.

7. Conclusion

Professional forecasters compete in a market for their services and face incentives to be accurate as well as different. These competitive pressures promote diversity and collective wisdom. However, a well-known result in social psychology is that people often mimic others' actions when they are uncertain and economists have shown that it is rational for forecasters to herd to exploit the valuable information obtained by others or protect their reputations. Social learning can also cause forecasters to adopt similar models and techniques. In either case, forecasts become less dispersed and collective wisdom declines.

This paper examined whether these forces have influenced the collective wisdom of the Blue Chip forecasting group over time and across forecasting horizons. We provide three main findings. First, the size of the consensus errors falls and forecast dispersion rises as horizons decrease from 17 to 12 months. This finding is consistent with the hypothesis that the Lawrence Klein forecasting award – which is based on the performance of 12-month-head forecasts – increased the group's collective wisdom by raising the incentive to anti-herd. Second, there was a dramatic decline in forecast dispersion beginning in the early 1990s associated a significant increase in the size of longer-term consensus errors. Third, forecasts bracket realized GDP growth much less frequently after the early 1990s. For example, the most optimistic 18-month-ahead forecast underestimated GDP growth four years in a row between 1997 and 2000. These last two findings are consistent with the hypothesis that increased herding or reduced model diversity caused the collective wisdom of the Blue Chip crowd to decline starting in the 1990s.

These results are based on relatively few observations and should be viewed with some caution. Nevertheless, they suggest that market forces will not necessarily guarantee that an optimal level of diversity materializes in the market for professional forecasting and that collective wisdom may suffer as a consequence. The failure of professional forecasters to anticipate the Great Recession illustrates the potential cost of reduced collective wisdom. To the extent that decision-makers in the economy rely on these forecasts, strengthening incentives to promote independence could improve social welfare.

Appendix A

Forecasters Participating In the Blue Chip Economic Indicators
1976-2011

BANKS

Bank of America/Bank of America - Merrill Lynch
Bankers Trust Co.
Brown Brothers Harriman
Chase Manhattan Bank
Chemcial Banking
Citibank
Comercia
Connecticut National Bank
CoreStates Financial Corp.
First Fidelity Bancrop
First Interstate Bank
First National Bank of Chicago
Fleet Financial Group
Harris Trust and Saving
Irving Trust Company
J P Morgan
LaSalle National Bank
Manufacturers Hanover
Manufacturers National Bank of Detroit
Marine Midland
Mellon Bank
National City Bank of Cleveland
Northern Trust Company
Philadelphia National Bank/ PNC Bank
Provident National Bank
Security Pacific Bank
Shawmut National Corp.
United California Bank
U.S. Trust Co.
Wells Fargo Bank
Bank of Tokyo-Mitsubishi UFJ
Barnett Banks, Inc.
Huntington National Bank
Societe Generale
Wachovia

INDUSTRIAL CORPORATIONS

B. F. Goodrich
Catepillar
Chrysler Corp./**DaimlerChrysler AG**
Conrail
Eaton Corp.
DuPont
Ford Motor Company
General Electric Company
General Motors
International Paper
Machinery & Allied Products
Monsanto Company
Motorola, Inc.
Pennzoil Company
Predex Corp.
Sears Roebuck
Union Carbide

SECURITIES FIRMS

American Express/Shearson Lehman Company
Arnhold and S. Bleichroeder
A. G. Becker/Becker Associates
A.G. Edwards & Company
Chicago Capital, Inc.
CRT Government Securities
C. J. Lawrence, Inc.
Dean Witter Reynolds, Inc.
Goldman, Sachs CO.
Ladenburg, Thalmann & Co.
Loeb Rhoades, Hornblower & Co.
Morgan Stanley & Co.
NationsBank Capital Markets Inc.
Prudential Securities, Inc.
Barclays Capital
Bear Steams & Co.
BMO Nesbitt Burns/BMO Capital Markets
Chase Securities, Inc.
Credit Suisse First Boston/Credit Suisse
Daiwa Securities America/Daiwa Capital Markets America
Deutsche Bank Securities
J.W. Coons Advisors
Lehman Brothers
Loomis Sayles & Co., LP
Mesirow Financial
Nomura Securities
Pierpont Securities
RBS Greenwich Capital/RBS
Russell Investments
Scudder Kemper Investments
Schwab/Stanford Washington Research Group
Soleil Securities Group
UBS Warburg/UBS
Wells Capital Management
Wintrust Wealth Management

INDEPENDENT FORECASTERS

Albert T. Sommers
Argus Research
Arthur D. Little
Ben E. Laden Assoc.
Business Economics, Inc.
Center for Study of American Business
Computer Aided Production Planning Systems, Inc.
DePrince & Associates
DeWolf Associates
Econonclast
Econoviews International, Inc.
Evans Economics, Inc.
George Gols
Hagerbaurner Economics
Heinemann Economic Research
Helming Group
Herman I. Leibling & Assoc.

Forecasters Participating In the Blue Chip Economic Indicators cont.

Weyerhaeuser Co.
W. R. Grace
FedEx Corporation

ECONOMETRIC MODELERS
Chase Econometrics
Data Resources, Inc.
Fairmodel-Economics, Inc.
Georgia State University
Gil Heebner, Eastern College
Inforum - University of Maryland
Laurence H. Meyer & Associates/Macroeconomic Advisers, LCC
Merrill Lynch Economics
Michigan Quarterly U.S. Model
UCLA Business Forecasting
University of Illinois (B.T.)
Wharton Econometrics/WEFA Group/Global Insight

Other
FINANCIAL PUBLICATIONS
Cahners Economics
Financial Times Currency Forecaster
Eggert Economics Enterprises, Inc.
Fortune Magazine
Economist Intelligence Unit

GOVERNMENT AGENCIES
Bush Administration
Clinton Administration
Congressional Budget Office
Office of Management and Budget
Fannie Mae

INDUSTRY ASSOCIATIONS
Conference Board
Mortgage Bankers Association
National Association of Home Builders
U.S. Chamber of Commerce
National Assn. of Realtors

INSURANCE COMPANIES
Equitable Life
Metropolitan Life Insurance Co.
Prudential Insurance Co.
AIG
Swiss Re

RATINGS AGENCIES
Dun & Bradstreet
Standard and Poor's Corp.
Moody's Analytics
Moody's Capital Markets

InfoMetrica, Inc.
Joel Popkin & Co.
Juodeika Allen & Co.
Leonarnd Silk, NY Times
MAPI
Morris Cohen & Associates
Moseley, Hallgarten & Estabrook
Oxford Economics, USA
Peter L. Bernstein, Inc.
Polyconomics
Reeder Associates (Charles)
Robert Genetski and Associates, Inc.
Rutledge & Co.
Schroder, Naess, and Thomas
Sindlinger Company, Inc.
SOM Economics, Inc.
Statistical Indicators Associates
Stotler Economics
The Bostonian Group - HHG
Turning Points Micrometrics
Wayne Hummer & Company - Chicago
Action Economics
ClearView Economics
Kellner Economic Advisers
Lexington Economic Forecasting
MacroFin Analytics
MMS International
Naroff Economic Advisors
Perna Associates
RDQ Economics
Potomac Research Group
Woodley Park Research

Note: Forecaster categorization from Laster, et al. (1996). New forecasters in bold with categorization made by the authors.

References

Armstrong, J.S. 2001. "Combining Forecasts." In J. Scott Armstrong (ed.) *Principles of Forecasting: A Handbook for Researchers and Practitioners*, Norwell, MA: Kluwer Academic Publishers: 1-12.

Ashiya, M., and T. Doi. 2001. "Herd Behavior of Japanese Economists," *Journal of Economic Behavior and Organization* 46: 343-46.

Banerjee, Abhijit. 1992. "A Simple Model of Herding Behavior," *Quarterly Journal of Economics*, 107: 797-817.

Barron, Orie, Oliver Kim, Steve C. Lim and Douglas E. Stevens. 1998. "Using Analysts' Forecasts to Measure Properties of Analysts' Information Environment," *The Accounting Review* 73(4): 421-433.

Barry, Christopher B., and Robert H. Jennings. 1992. "Information and Diversity of Analyst Opinion," *Journal of Financial and Quantitative Analysis* 27: 169-83.

Batchelor, R. 2007. "Bias in Macroeconomic Forecasts." *International Journal of Forecasting*" 23(2): 189-203.

Batchelor, R., Dua, P. 1990a. "Product Differentiation in the Economic Forecasting Industry." *International Journal of Forecasting* 6: 311-16.

Batchelor, R., Dua, P. 1990b. "Forecaster ideology, forecasting technique, and the accuracy of economic forecasts." *International Journal of Forecasting* 6(1): 3-10.

Batchelor, R., Dua, P. 1992. "Conservatism and Consensus-seeking among Economic Forecasters," *Journal of Forecasting* 11: 169-81.

Batchelor, R., Dua, P. 1995. "Forecaster diversity and the benefits of combining forecasts." *Management Science* 41(1): 68-75.

Bates. J.M., Granger, C.W.J. 1969. "The Combination of Forecasts," *Operational Research Quarterly* 20(4): 451-68.

Bauer, A., Eisenbeis,, R., Waggoner, D., Zha, T. 2003. "Forecast Evaluation with Cross-Sectional Data: The Blue Chip Surveys." Federal Reserve Bank of Atlanta *Economic Review*: 17-31.

Bikhchandani, Sushil; Hirshleifer, David and Welch, Ivo. 1992. "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," *Journal of Political Economy* 100 (5): 992-1026.

Bikhchandani, S., Sharma, S. 2001. "Herd Behavior in Financial Markets," IMF Staff Papers 47(3): 279-310.

Boero, G, J. Smith and K. Wallis. 2014. "The Measurement and Characteristics of Professional Forecasters' Uncertainty." *Journal of Applied Econometrics*,

Bomberger, William. 1996. "Disagreement as a Measure of Uncertainty," *Journal of Money, Credit and Banking*, 38(3): 381-92.

Clemen. R. 1989. "Combining Forecasts: A Review and annotated bibliography." *International Journal of Forecasting* 5: 559-583.

Clement, M. and S. Tse. 2005. "Financial Analyst Characteristics and Herding Behavior in Forecasting." *The Journal of Finance* 60(1): 307-341.

Crowe, Christopher. 2010. “Consensus forecasts and inefficient information aggregation,” IMF Working Paper 178.

Dovern, Jonas, and Ulrich Fritsche, and Jirir Slacalek. 2012. “Disagreement among Forecasts in G7 Countries, Review of Economics and Statistics, 94(4): 1081-1096

Ehrbeck, T. and R. Wladmann. 1996. “Why are Professional Forecasters Biased? Agency versus Behavioral Explanations,” *Quarterly Journal of Economics* 111: 21-40.

Engle, R. 1983. “Estimates of the Variance of U.S. Inflation Based Upon the ARCH Model.” *Journal of Money, Credit, and Banking 15*: 286-301.

Eyster, Erik and Mathew Rabin. 2010. “Naïve Herding in Rich-Information Settings,” *American Economic Journal: Microeconomics*, 2(4): 221-243.

Gallo, M., Granger, C., Jeon, Y. 2002. “Copycats and Common Swings: The Impact of the Use of Forecasts in Information Sets.” IMF Staff Papers, 49(1): 4-21.

Gamber, Edward, JK Smith and MA Weiss. 2010. “Forecast errors before and during the Great Moderation,” *Journal of Economics and Business*, 63(4): 278–289.

Graham, John R. 1999. “Herding among Investment Newsletters: Theory and Evidence,” *The Journal of Finance*, 54(1): 237–268.

Gregory, A.W., Yetman, K. 2004. “The Evolution of Consensus in Macroeconomic Forecasting.” *International Journal of Forecasting* 20(3): 461-73.

Hendry, D. and M. Clements. 2004. “Pooling of Forecasts.” Econometrics Journal 7: 1-31.

Hirshleifer, David A. and Teoh, Siew Hong. 2009. “Thought and Behavior Contagion in Capital Markets.” In Thorsten Hens IV, Klaus Reiner and Schenk-Hoppe (eds.) *Handbook of Financial Markets: Dynamics and Evolution*: 1-46.

Holthausen, Robert W. and Robert E. Verrecchia. 1990). “The Effect of Informedness and Consensus on Price and Volume Behavior,” *The Accounting Review*, 65(1): 191-208.

Hong, H., J. Kubik and A. Solomon. 2000. “Security analysts’ career concerns and herding of earnings forecasts.” *RAND Journal of Economics* 31(1): 121-44.

Hong, L., Page, S. 2012. “Some Microfoundations of Collective Wisdom.” In Hélène Landemore and Jon Elster (Ed.), *Collective Wisdom: Principles and Mechanisms*. Cambridge University Press.

Hong, L., Page, S., Riolo, M. 2012. “Incentives, Information, and Emergent Collective Accuracy.” *Managerial and Decision Economics* 33: 323-334.

Kandel, E. Pearson, N. 1995. “Differential Interpretation of Information and Trade in Speculative Markets.” *Journal of Political Economy* 103(4), 831-72.

Kandel, E. Zilberfarb, B. 1999. “Differential Interpretation of Information in Inflation Forecasts.” *The Review of Economics and Statistics* 81(2): 217-226.

Kim, Oliver, Steve C. Lim and Kenneth W. Shaw. 2001. “The Inefficiency of Mean Analyst Forecast as a Summary of Forecast of Earnings,” *Journal of Accounting Research* 39: 329-335.

Kim, Chang Jin, and Charles R. Nelson. 1999. Has the U.S. Economy Become More Stable? A Bayesian Approach Base on a Markov-Switching Model of the Business Cycle.” Review of Economics and Statistics, 81(4): 1-10.

Krishnan, Murugappa, Steve C. Lim, and Ping Zhou. 2006. “Analysts’ Herding Propensity: Theory and Evidence from Earnings Forecasts, Working Paper.

Lahiri, Kajal. 2011. "Limits to Economic Forecasting." In Advances in Economic Forecasting, Matthew L. Higgins, ed. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, pp. 25-50.

Lahiri, K. and F. Liu. 2006. "Modeling Multi-period Inflation Uncertainty using a Panel of Density Forecasts." *Journal of Applied Econometrics* 21: 1199-1219.

Lahiri, Kajal and Xuguang Sheng. 2008. "Evolution of forecast disagreement in a Bayesian learning model." *Journal of Econometrics* 144: 325-340.

Lahiri, Kajal and Xuguang Sheng. 2010. "Learning and heterogeneity in GDP and inflation forecasts." *International Journal of Forecasting* 26(2): 265-292.

Lahiri, Kajal and Xuguang Sheng. 2010. "Measuring Forecast Uncertainty by Disagreement: The Missing Link," *Journal of Applied Econometrics*, 25: 514-538.

Lamont, O. 2002. "Macroeconomic Forecasts and Microeconomic Forecasters." *Journal of Economic Behavior & Organization* 48: 265-80.

Larrick, R. P., Mannes, A. E., and Soll, J. B. 2012. "The social psychology of the wisdom of crowds." In J. I. Krueger (Ed.), *Frontiers of Social Psychology: Social Psychology and Decision Making.* Philadelphia: Psychology Press: 227-42.

Laster, David, Paul Bennett and In Sun Geoum. 1997. "Rational Bias in Macroeconomic Forecasts," The Federal Reserve Bank of New York *Staff Reports* 21.

Laster, David, Paul Bennett and In Sun Geoum. 1999. "Rational Bias in Macroeconomic Forecasts," *The Quarterly Journal of Economics*, 114(1): 293-318.

Lichtendahl, Kenneth C. Jr. and Robert L. Winkler. 2007. "Probability Elicitation, Scoring Rules, and Competition Among Forecasters," *Management Science*, 53(11): 1745-1755.

Lichtendahl, Kenneth C. Jr., Yael Grushka-Cockayne, and Phillip E. Pfeifer. 2013. "The Wisdom of Competitive Crowds," *Operations Research*, 61(6): 1383-1398.

Lorenz, Jan, Keiko Rauhut, Frank Schweitzer and Dirk Helbing. 2011 "How social influence can undermine the wisdom of crowd effect." *Proceedings of the National Academy of Sciences*, 108(22): 9020-25.

Marinovic, Iván, Marco Ottaviani, and Peter Norman Sørensen. 2013. Forecasters' Objectives and Strategies, in Handbook of Economic Forecasting, Vol. 2, Part B (Graham Elliot and Allan Timmermann, eds.). Elsevier.: 690-720.

Maug, Ernst., & Naik, Narayan. 2011. "Herding and delegated portfolio management: the impact of relative performance evaluation on asset allocation." *The Quarterly Journal of Finance, 1*(02): 265-292.

McConnell, Margaret M. and Gabriel Perez and Quiros. 2000. "Output Fluctuations in the United States: What Has Changed Since the Early 1980's?" *The American Economic Review*, 90(5): 1464-1476.

Newbold, P. and D.I. Harvey. 2002. "Forecast Combination and Encompassing." In Clements, M.P. and Hendry (eds), A Comparison of Economic Forecasting: Blackwell Press: Oxford: 268-83.

Ottaviani, Marco and Peter Norman Sørensen. 2006. “The strategy of professional forecasting,” *Journal of Financial Economics*, 81(2): 441–466.

Ottaviani, Marco and Peter Norman Sørensen. 2006. “Reputational cheap talk, *The RAND Journal of Economics*, 37(1): 155–175.

Page, S. 2007. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. Princeton University Press.

Page, S. 2014. “Where diversity comes from and why it matters?” *European Journal of Social Psychology* 44: 267-279.

Patton, A., Timmermann, A. 2010. “Why do forecasters disagree? Lessons from the terms structure of cross-sectional dispersion.” *Journal of Monetary Economics* 57(7): 803-820.

Predegast, Canice, and Lars Stole. 1996. “Impetuous Youngsters and Jaded Old-Timers: Acquiring a Reputation for Learning,” *Journal of Political Economy*, 104(6): 1105-1134.

Rich, R. and J. Tracy. 2010. “The Relationships among Expected Inflation, Disagreement, and Uncertainty: Evidence from Matched Point and Density Forecasts.” *The Review of Economics and Statistics* 92(1): 200-207.

Scharfstein, David, and Jeremy Stein. 1990. “Herd Behavior and Investment,” *American Economic Review*, 80: 465-79.

Soll, J., Larrick, R. 2009. “Strategies for Revising Judgment: how (and how well) people use others’ opinions.” *Journal of Experimental Psychology: Learning, Memory and Cognition*, 35(3): 780-805.

Stock, James, and Mark Watson. 2003. “Has the Business Cycle Changed? Evidence and Explanations,” In *Monetary Policy and the Economy: Adapting to a Changing Economy*. Federal Reserve Bank of Kansas City: 9-56.

Surowiecki, J. 2004. *The Wisdom of Crowds: Why the Many are smarter than the few and how collective wisdom shapes business, economics, society, and Nations*. London: Little, Brown.

Timmermann, Allan. 2006. Forecast Combinations, in Handbook of Economic Forecasting, Vol. 1, (Graham Elliot, Clive W. J. Granger and Allan Timmermann, eds.). Elsevier.

Trichet, Jean-Claude. 2010. “Reflections on the nature of monetary policy non-standard measures and finance theory,” Speech given on the occasion of the ECB Central Banking Conference Frankfurt, 18 November.

Wallis, Kenneth F. 2011. “Combining – Forty Years Later,” *Applied Financial Economics*, 21:33-41.

Welch, I. 2000. “Herding among security analysts.” *Journal of Financial Economics* 58: 369-96.

Woodford, Michael. 2009. "Convergence in Macroeconomics: Elements of the New Synthesis." *American Economic Journal: Macroeconomics*, 1(1): 267-79.

Zarnowitz, V. 1992. “Has Macro-Forecasting Failed?” *Cato Journal*, 12(1): 129-160.

Zarnowitz, V., Braun, P. 1993. “Twenty-two Years of the NBER-ASA Quarterly Economic Outlook Surveys: Aspects and Comparisons of Forecasting Performance.” In *Business Cycles, Indicators and Forecasting*, University of Chicago Press.

Zarnowitz, V., and L. Lambros. 1987. “Consensus and Uncertainty in Economic Prediction.” *Journal of Political Economy* 95(3): 591-621.

Table 1

Regressions of Blue Chip Absolute Consensus Errors on Forecast Dispersion
Current-Year Forecasts

Horizon (months)	Early Vintage				Later Vintage			
	Constant	Dispersion	obs	R^2	Constant	Dispersion	obs	R^2
12	0.67	0.38	35	0.06	0.72	0.58	35	0.10
	(4.41)***	(1.35)			(4.08)***	(2.00)**		
11	0.59	0.41	35	0.06	0.61	0.70	35	0.11
	(4.21)***	(1.14)			(4.01)***	(1.87)*		
10	0.51	0.46	35	0.06	0.60	0.58	35	0.07
	(3.93)***	(1.42)			(4.13)***	(1.53)		
9	0.46	0.32	35	0.03	0.57	0.53	35	0.06
	(4.16)***	(1.02)			(4.94)***	(1.38)		
8	0.43	0.21	35	0.01	0.57	0.47	35	0.03
	(4.36)***	(0.55)			(5.55)***	(1.15)		
7	0.39	0.55	32	0.07	0.54	0.88	32	0.12
	(5.25)***	(1.64)			(6.47)***	(3.79)***		
6	0.33	0.49	28	0.03	0.51	1.13	28	0.06
	(4.92)***	(1.14)			(5.91)***	(2.86)***		
5	0.29	0.34	28	0.01	0.50	1.00	28	0.01
	(4.35)***	(0.60)			(4.35)***	(0.81)		
4	0.22	1.77	28	0.08	0.48	1.74	28	0.02
	(3.75)***	(1.28)			(4.37)***	(1.22)		
3	0.23	0.32	28	0.00	0.48	1.47	28	0.01
	(3.84)***	(0.26)			(4.44)***	(0.53)		
2	0.18	1.40	28	0.01	0.44	2.65	28	0.01
	(4.18)***	(0.74)			(3.61)***	(0.55)		
1	0.16	0.66	28	0.00	0.48	-1.29	28	0.00
	(3.28)***	(0.29)			(4.59)***	(-0.21)		

Notes: The table contains results for regressions of the absolute consensus error for real GDP growth on a constant and forecast dispersion. The consensus is the cross-sectional mean forecast for real GDP growth. The error is calculated using actual real GDP growth estimates reported in the June issue of the *Survey of Current Business* in the year following the target year (Early Vintage) or second year following the target year (Late Vintage). Forecast dispersion is the cross-sectional variance of real GDP forecasts. t-statistics are shown in parentheses with significance at the one, five and ten percent level denoted by ***, ** and *, respectively. Robust standard errors are used to correct for possible heteroscedasticity in the residuals.

Table 2

Regressions of Blue Chip Absolute Consensus Errors on Forecast Dispersion
Year-Ahead Forecasts

Horizon (months)	Early Vintage				Later Vintage			
	Constant	Dispersion	obs	R^2	Constant	Dispersion	obs	R^2
24	1.45	-0.62	26	0.07	1.65	-0.72	26	0.07
	(4.38)***	(-1.96)*			(4.30)***	(-2.03)*		
23	1.52	-0.80	26	0.07	1.76	-1.00	26	0.08
	(5.03)***	(-2.92)***			(4.97)***	(-3.70)***		
22	1.37	-0.42	26	0.02	1.69	-0.83	26	0.06
	(4.81)***	(-1.08)			(4.82)***	(-2.46)**		
21	1.35	-0.39	26	0.02	1.69	-0.93	26	0.07
	(4.79)***	(-1.00)			(4.84)***	(-2.59)**		
20	1.35	-0.37	26	0.02	1.66	-0.88	26	0.06
	(4.72)***	(-1.03)			(4.72)***	(-2.16)**		
19	1.24	-0.18	29	0.01	1.44	-0.53	29	0.04
	(4.72)***	(-0.73)			(4.19)***	(-1.29)		
18	1.33	-0.24	34	0.01	1.56	-0.62	34	0.04
	(4.85)***	(-0.72)			(4.61)***	(-1.59)		
17	1.11	-0.03	35	0.00	1.30	-0.28	35	0.01
	(4.41)***	(-0.08)			(4.15)***	(-0.63)		
16	1.02	0.00	35	0.00	1.17	-0.16	35	0.00
	(3.95)***	(0.00)			(3.64)***	(-0.36)		
15	0.97	0.01	35	0.00	1.06	-0.06	35	0.00
	(4.10)***	(0.01)			(3.62)***	(-0.14)		
14	0.70	0.45	35	0.07	0.79	0.43	35	0.04
	(3.70)***	(1.41)			(3.13)***	(1.16)		
13	0.66	0.42	35	0.06	0.75	0.48	35	0.06
	(3.77)***	(1.42)			(3.42)***	(1.41)		

Notes: The table contains results for regressions of the absolute consensus error for real GDP growth on a constant and forecast dispersion. The consensus is the cross-sectional mean forecast for real GDP growth. The error is calculated using actual real GDP growth estimates reported in the June issue of the Survey of Current Business in the year following the target year (Early Vintage) or second year following the target year (Late Vintage). Forecast dispersion is the cross-sectional variance of real GDP forecasts. t-statistics are shown in parentheses with significance at the one, five and ten percent level denoted by ***, ** and *, respectively. Robust standard errors are used to correct for possible heteroscedasticity in the residuals.

Table 3
Structural Break Regressions for Blue Chip Forecast Dispersion

	Year-Ahead Forecasts			Current-Year Forecasts		
Break Year	**Mean**	**Break Dummy**	**R^2**	**Mean**	**Break Dummy**	**R^2**
1989	1.06	-0.66	0.48	0.56	-0.22	0.52
	(11.51)***	(-13.93)***		(8.89)***	(-8.26)***	
1990	1.05	-0.67	0.54	0.54	-0.20	0.50
	(12.92)***	(-15.59)***		(8.66)***	(-8.29)***	
1991	1.02	-0.68	0.58	0.54	-0.20	0.51
	(13.61)***	(-17.28)***		(8.81)***	(-9.12)***	
1992	0.97	-0.64	0.55	0.54	-0.21	0.54
	(13.34)***	(-16.52)***		(9.10)***	(-10.35)***	
1993	0.91	-0.6	0.51	0.53	-0.21	0.53
	(12.48)***	(-15.76)***		(9.07)***	(-10.79)***	
1994	0.86	-0.56	0.45	0.51	-0.19	0.51
	(11.19)***	(-14.69)***		(8.62)***	(-10.29)***	
1995	0.81	-0.51	0.38	0.50	-0.18	0.49
	(9.86)***	(-13.30)***		(8.27)***	(-9.86)***	
1996	0.76	-0.46	0.33	0.49	-0.16	0.47
	(8.94)***	(-12.30)***		(8.00)***	(-9.48)***	

Notes: The table contains results for pooled regressions of forecast dispersion on a constant (Mean), 11 month dummy variables (not shown), and a dummy variable (Break Dummy) that takes on values of one starting in the break year. Forecast dispersion is the cross-sectional variance of real GDP forecasts. t-statistics are shown in parentheses with significance at the one, five and ten percent level denoted by ***, ** and *, respectively. Robust standard error estimates are used to correct for possible heteroscedasticity in the regression residuals.

Table 4
Structural Break Regressions for Blue Chip Absolute Consensus Errors: Currrent-Year Forecasts

	Early Vintage Data		Later Vintage Data	
Sample Period	1977-2011	1986-2011	1977-2011	1985-2011
Observations	375	312	375	312
Constant	0.86 (8.57)***	0.81 (7.12)***	0.95 (8.07)***	1.00 (6.87)**
Post-1992 Dummy	-0.05 (-1.28)	-0.01 (-0.20)	0.02 (0.40)	-0.04 (-0.70)
R^2	0.26	0.26	0.09	0.09

Notes: The table contains results for pooled regressions of the absolute consensus error on a constant, a dummy variable that takes on values of one starting in 1993, and 11 month dummy variables (not shown). The consensus is the cross-sectional mean forecast for real GDP growth. The error is calculated using actual real GDP growth estimates reported in the June issue of the *Survey of Current Business* in the year following the target year (Early Vintage) or second year following the target year (Late Vintage). t-statistics are shown in parentheses with significance at the one, five and ten percent level denoted by ***, ** and *, respectively. Robust standard error estimates are used to correct for possible heteroscedasticity in the regression residuals.

Table 5
Structural Break Regressions for Blue Chip Absolute Consensus Errors: Year-Ahead Forecasts

	Early Vintage Data		Late Vintage Data	
Sample Period	1977-2011	1986-2011	1977-2011	1985-2011
Observations	368	312	375	312
Constant	0.88 (3.84)***	0.68 (3.18)***	0.87 (3.33)***	0.71 (2.80)***
Post-1990 Dummy	0.34 (3.43)***	0.58 (6.83)***	0.54 (4.71)***	0.75 (6.72)***
R^2	0.04	0.07	0.05	0.07

Notes: The table contains results for pooled regressions of the absolute consensus error on a constant, a dummy variable that takes on values of one starting in 1991, and 11 month dummy variables (not shown). The consensus is the cross-sectional mean forecast for real GDP growth. The error is calculated using actual real GDP growth estimates reported in the June issue of the *Survey of Current Business* in the year following the target year (Early Vintage) or second year following the target year (Late Vintage). t-statistics are shown in parentheses with significance at the one, five and ten percent level denoted by ***, ** and *, respectively. Robust standard error estimates are used to correct for possible heteroscedasticity in the regression residuals.

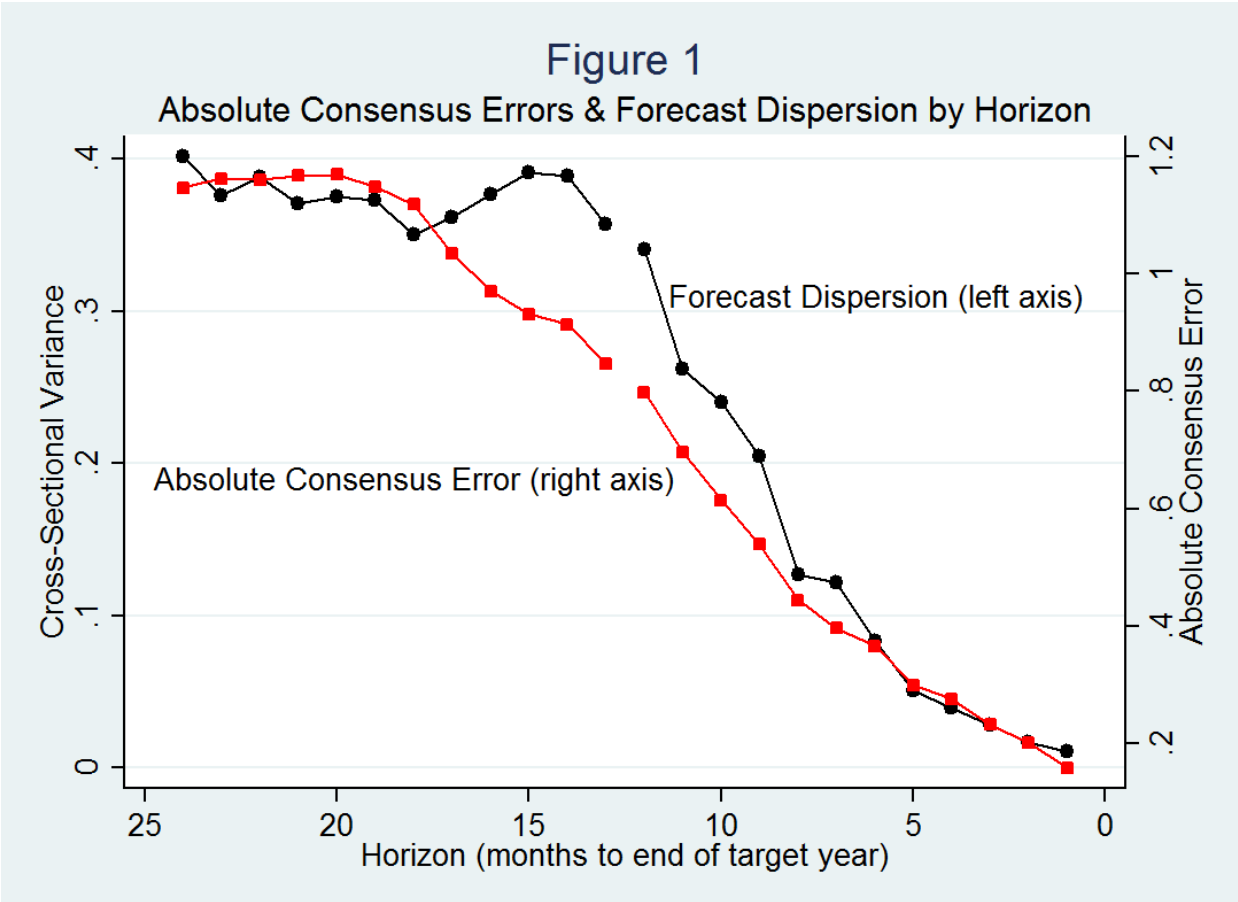
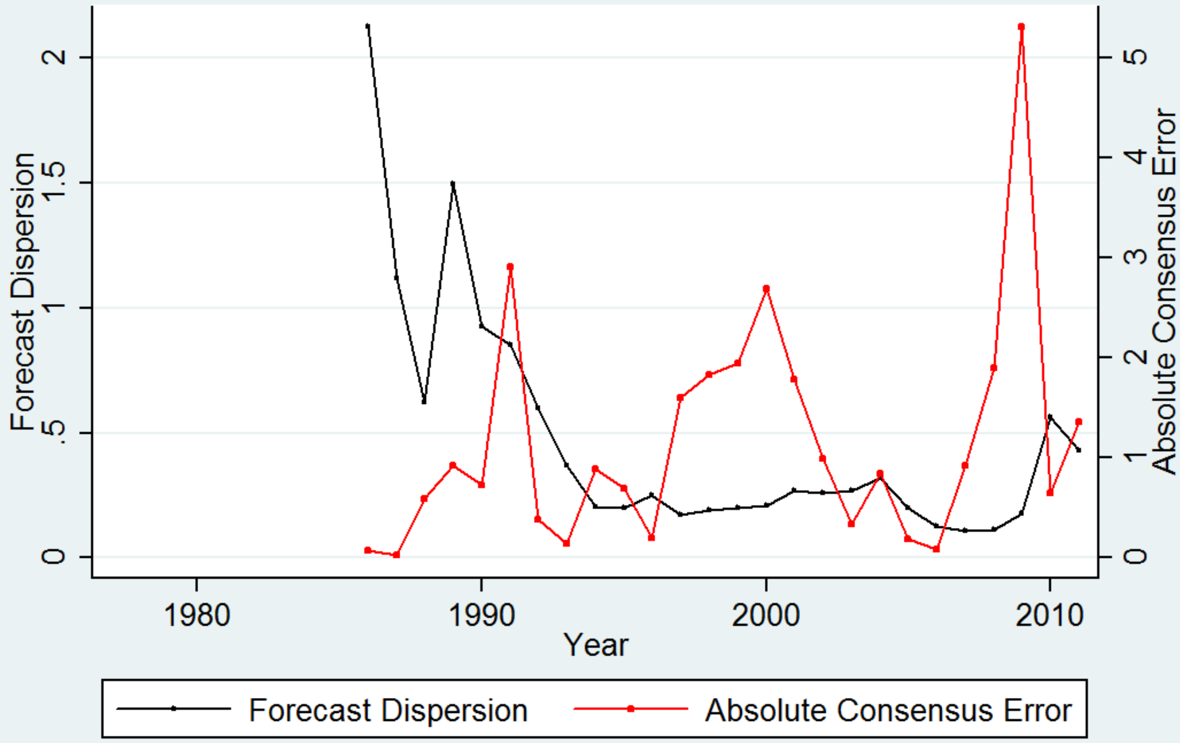


Figure 1
Absolute Consensus Errors & Forecast Dispersion by Horizon
Forecast Dispersion (left axis)
Absolute Consensus Error (right axis)
Cross-Sectional Variance
0
.1
.2
.3
.4
Absolute Consensus Error
.2
.4
.6
.8
1
1.2
25
20
15
10
5
0
Horizon (months to end of target year)

Figure 2

Consensus Errors & Forecast Dispersion: 24-Month Horizon



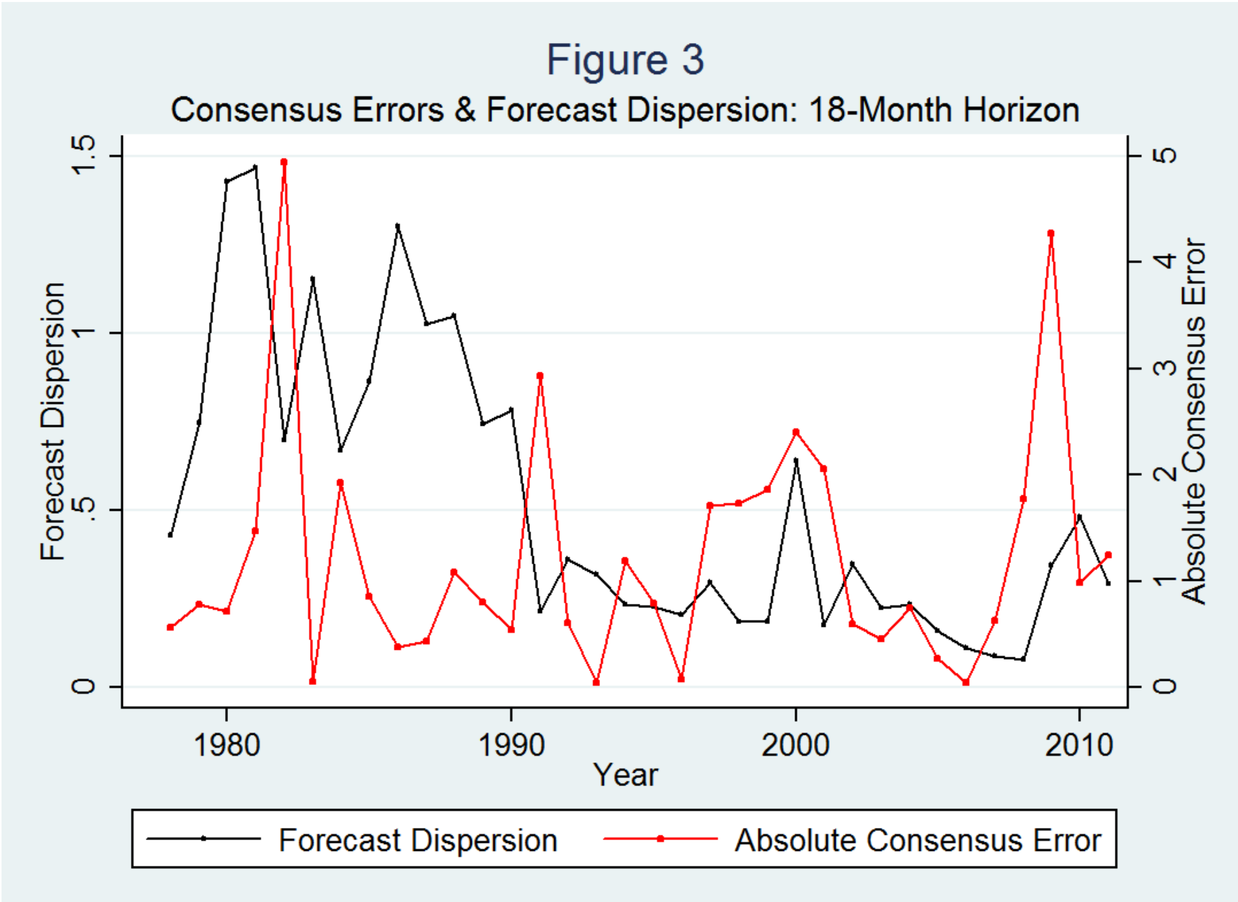


Figure 3
Consensus Errors & Forecast Dispersion: 18-Month Horizon
Forecast Dispersion
Absolute Consensus Error
Year
0
.5
1
1.5
0
1
2
3
4
5
1980
1990
2000
2010
Forecast Dispersion
Absolute Consensus Error

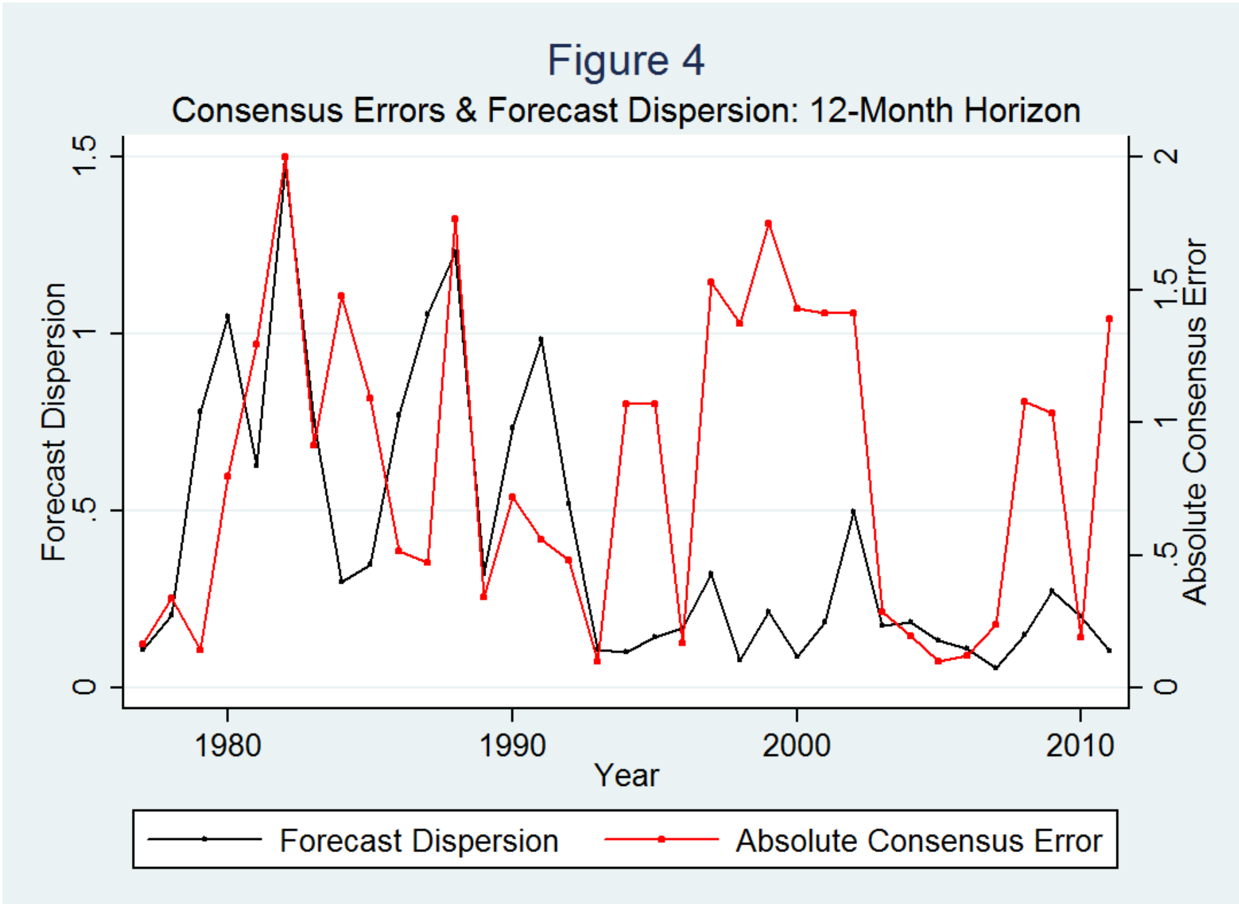


Figure 4
Consensus Errors & Forecast Dispersion: 12-Month Horizon
Forecast Dispersion
Absolute Consensus Error
0
.5
1
1.5
2
1980
1990
2000
2010
Year
Forecast Dispersion
Absolute Consensus Error

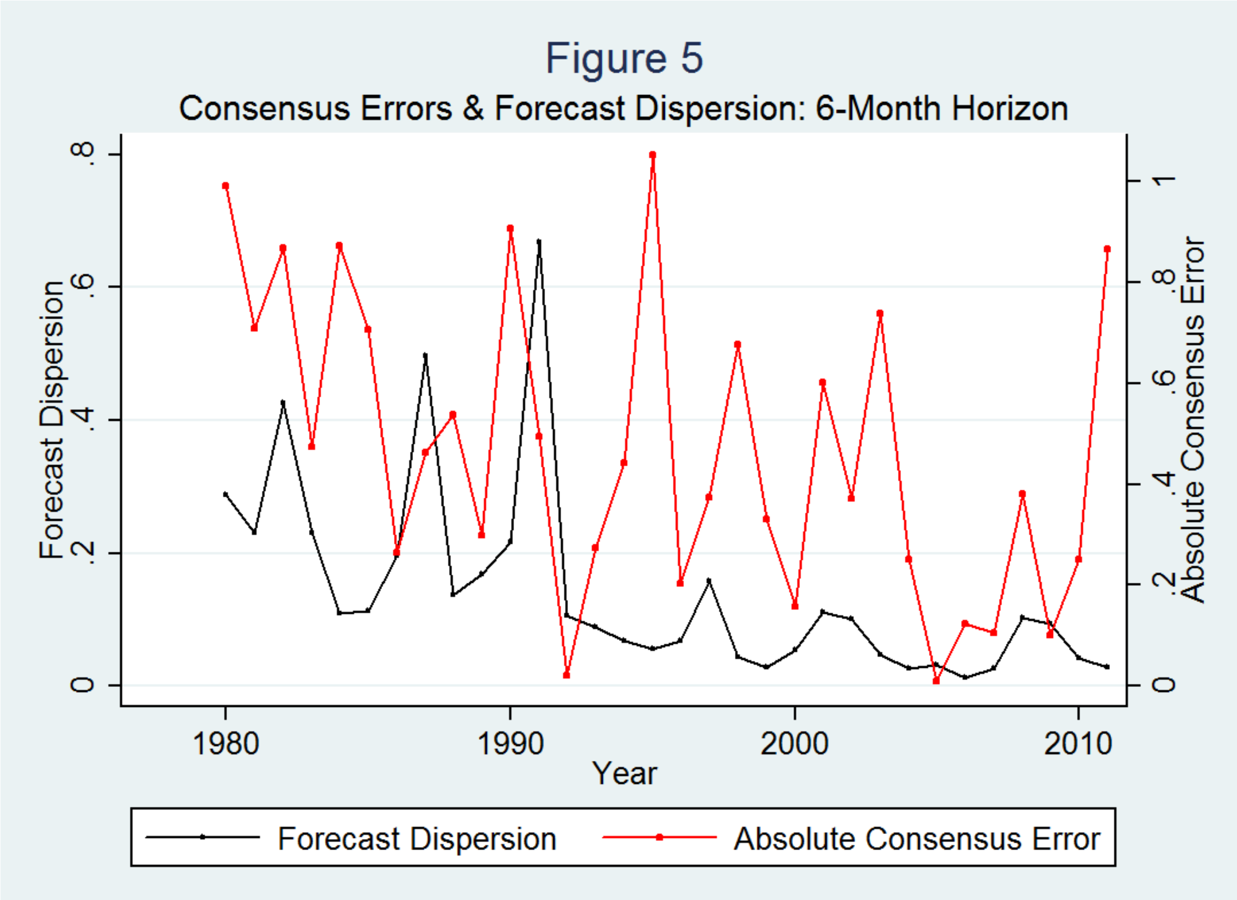


Figure 5
Consensus Errors & Forecast Dispersion: 6-Month Horizon
Forecast Dispersion
0
.2
.4
.6
.8
Absolute Consensus Error
0
.2
.4
.6
.8
1
1980
1990
2000
2010
Year
Forecast Dispersion
Absolute Consensus Error

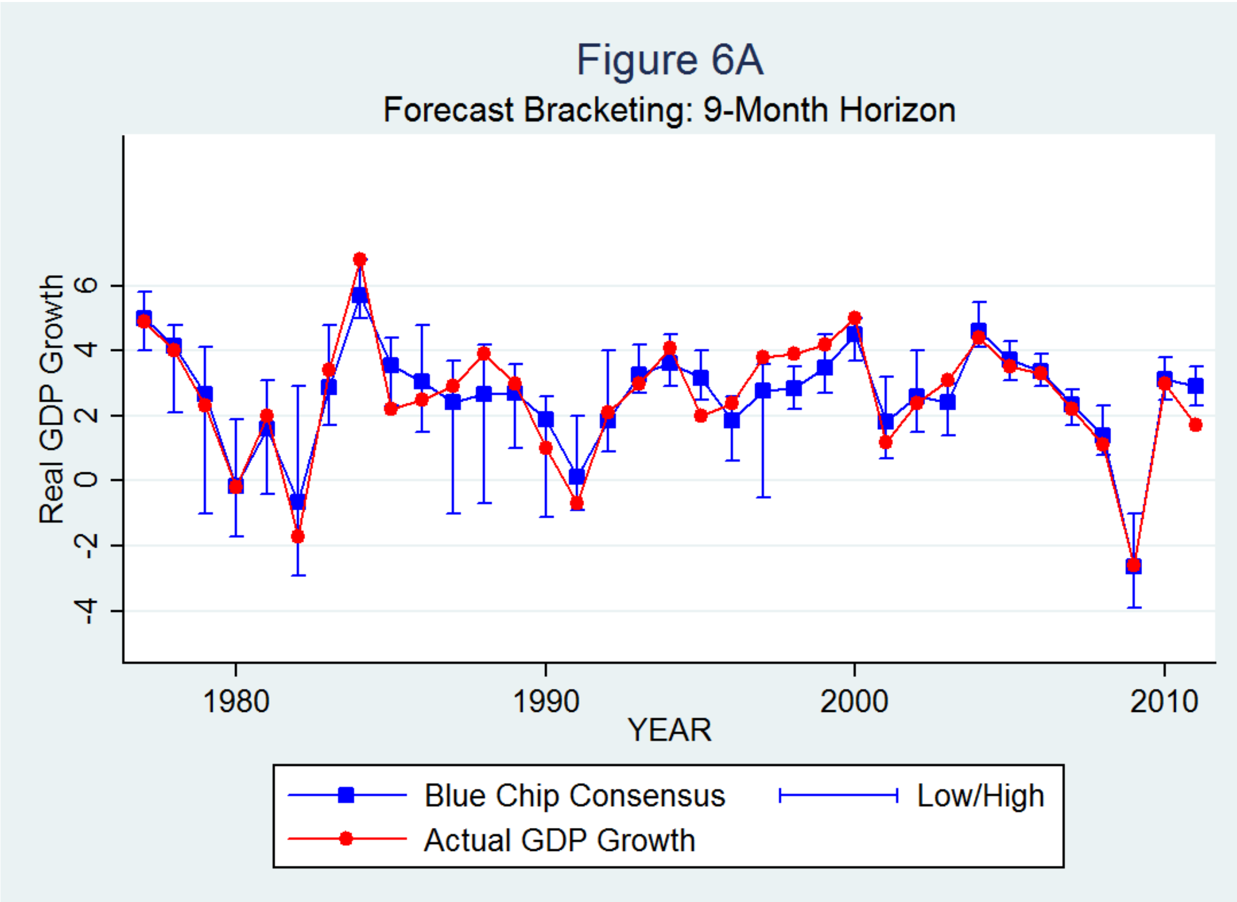


Figure 6A
Forecast Bracketing: 9-Month Horizon
Real GDP Growth
-4
-2
0
2
4
6
1980
1990
2000
2010
YEAR
Blue Chip Consensus
Low/High
Actual GDP Growth

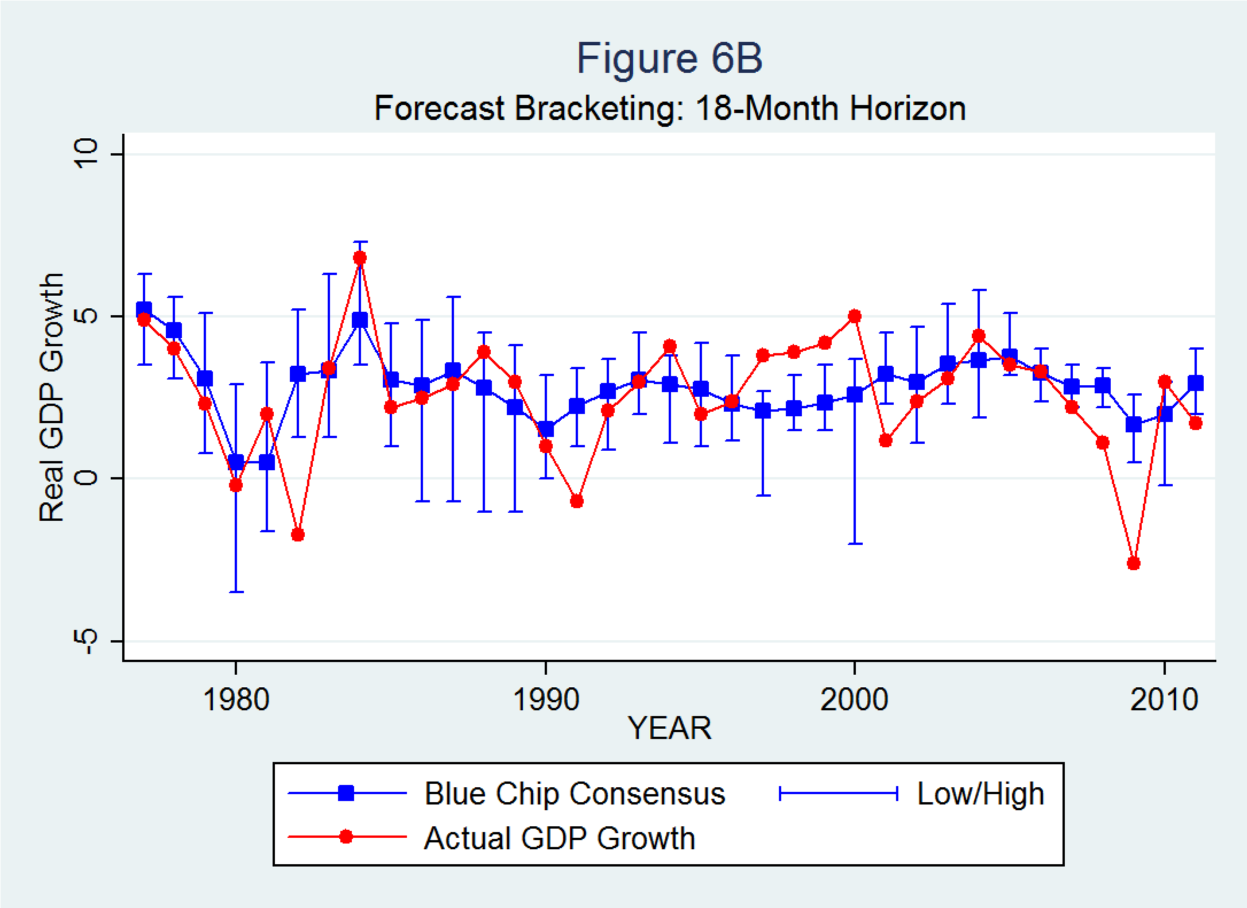


Figure 6B
Forecast Bracketing: 18-Month Horizon
Real GDP Growth
10
5
0
-5
1980
1990
2000
2010
YEAR
Blue Chip Consensus
Low/High
Actual GDP Growth

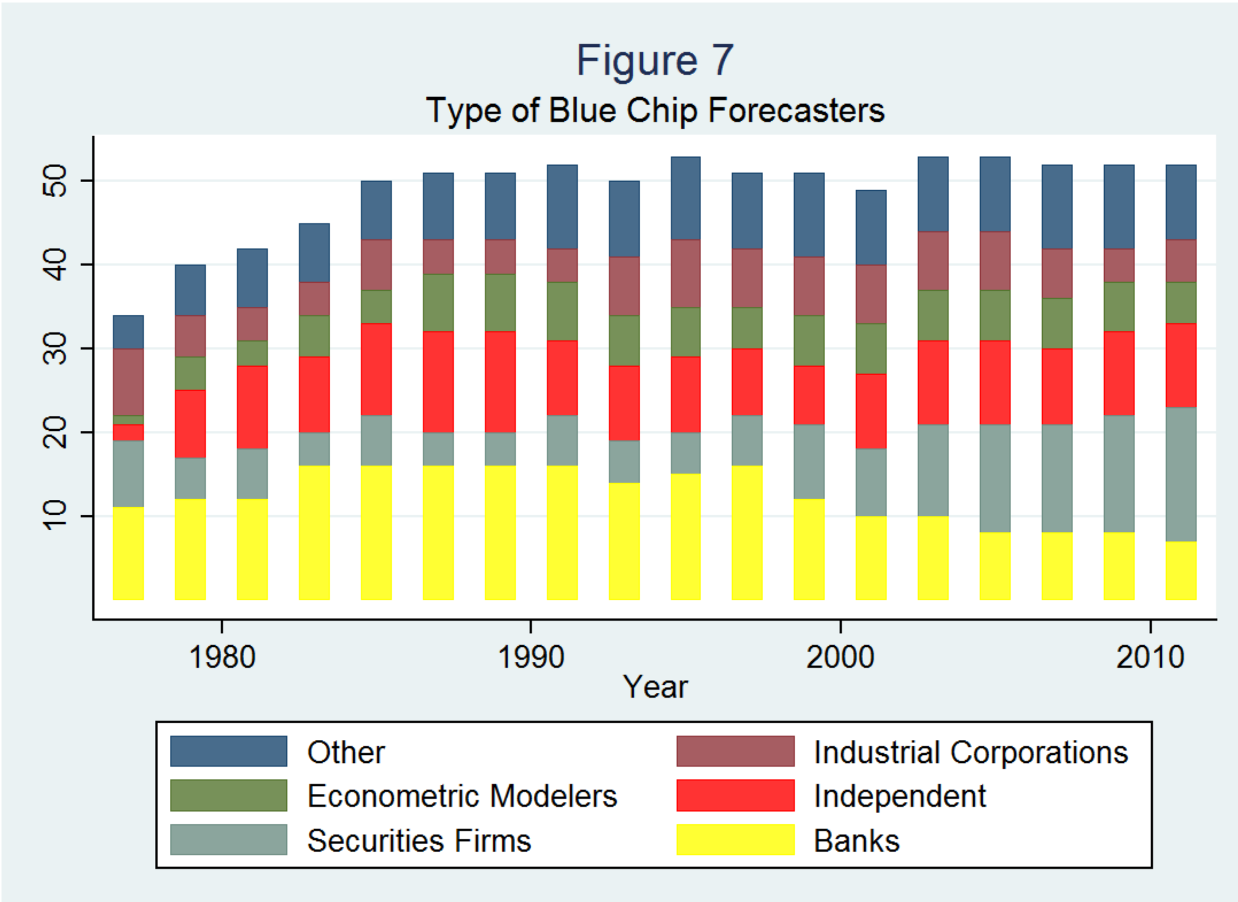


Figure 7
Type of Blue Chip Forecasters
10
20
30
40
50
1980
1990
2000
2010
Year
Other
Industrial Corporations
Econometric Modelers
Independent
Securities Firms
Banks