Journal of Economic Behavior & Organization
Vol. 48 (2002) 265–280

JOURNAL OF
Economic Behavior
& Organization

www.elsevier.com/locate/econbase

Macroeconomic forecasts and microeconomic forecasters

Owen A. Lamont

Graduate School of Business, University of Chicago, NBER, 1101 E. 58th Street, Chicago IL 60637, USA

Received 8 February 2000; accepted 8 January 2001

Abstract

In the presence of principal-agent problems, published macroeconomic forecasts by professional economists may not measure expectations. Forecasters may use their forecasts in order to manipulate beliefs about their ability. I test a cross-sectional implication of models of reputation and information-revelation. I find that as forecasters become older and more established, they produce more radical forecasts. Since these more radical forecasts are in general less accurate, ex post forecast accuracy grows significantly worse as forecasters become older and more established. These findings are consistent with reputational factors at work in professional macroeconomic forecasts.

JEL classification: D82; D84; E17

Keywords: Herding; Forecasters; Reputation; Strategic behavior; Incentives; Expectations

1. Introduction

Macroeconomic forecasts come in two varieties: statistical objects produced by mechanical models, and economic objects produced by human beings. The latter are "economic" in the sense that they are not necessarily designed to minimized squared forecast errors; rather, forecasts may be set to optimize profits or wages, credibility, shock value, marketability, political power (in the case of government forecasts), or more generally to minimize some loss function. This paper tests the influence of reputation in the making of economic forecasts by testing a cross-sectional implication of theories of strategic forecasting.

An extensive body of literature has examined macroeconomic and financial market forecasts, typically treating the forecasts as if they were the expectations of the forecasters and testing rationality properties. Thus, the literature tests the joint hypothesis that forecasters

E-mail address: owen.lamont@gsb.uchicago.edu (O.A. Lamont).

PII: S0167-2681(01)00219-0

have rational expectations, and that they report these expectations truthfully. For example, Keane and Runkle (1990) use a survey of professional forecasters and state:

> because these professionals report to the survey the same forecasts that they sell on the market, their survey responses provide a reasonably accurate measure of their expectations. Thus, these data are less subject to the criticism made by opponents of survey forecast rationality tests that the respondents had nothing to lose if they made bad forecasts.

But seen from a principal-agent perspective, using professional forecasters may actually be worse than using disinterested observers, depending on the rewards forecasters receive. For example, Bryan and Gavin (1986) find that forecasts produced by households have better statistical properties than those produced by professionals. Agency problems may help explain why rational expectations are often rejected in empirical work, even when using survey data from professional forecasters. For example, using the ASA–NBER dataset, Zarnowitz (1985) rejects rationality using professional's forecasts of prices.

I discuss below examples of payoff structures that provide incentives to produce forecasts that do not minimize forecast errors. By rewarding the acquisition of a reputation, these structures provide an incentive for forecasters to try to manipulate their own forecasts relative to those of rival forecasters, behavior that is sub-optimal from the standpoint of providing accurate forecasts. Since a reputation is acquired over time, the manipulation of forecasts will vary over the professional life of the forecaster.

Using panel data on published macroeconomic forecasts made by professional economists, I test this novel cross-sectional implication: that the dispersion of forecasts is related to the age and reputation of the forecaster. I find that as forecasters become older and more established, they make more radical forecasts. Strikingly, this behavior apparently causes forecast accuracy to decline over time, so that their forecasts grow worse as they become more experienced.

I first describe the theoretical motivation for the hypothesis to be tested, coming mainly from work by Scharfstein and Stein (1990) and Zwiebel (1995). I then provide anecdotal and institutional evidence that suggests some of the theoretical set-ups may be relevant to the real world. Next, I describe the data and present the results, looking both at ex ante forecast dispersal and (more briefly) at ex post accuracy. Last, I summarize and present conclusions.

2. Theory and literature

Since my focus is empirical, I discuss the underlying theory only briefly and informally (see Scharfstein and Stein and Zwiebel for a full presentation). Unlike Banerjee's (1992) model of herding, in which the information structure drives herding in agents wishing to make optimal forecasts, reputation models are driven by principal-agent concerns. Even though the principal (the consumer of the forecast) wants to receive an optimal forecast, the agent (the forecaster) has a different agenda. This agenda may lead to either excessively conservative or excessively radical forecasts.

Consider an economist, j, who wants to forecast a variable, y, based on an information set, and who competes with other economists in doing so. Suppose that the economist, after examining all relevant data, forms an expectation

$$e_j = E[y|I_j] \tag{1}$$

I_j represents not necessarily inside information, but also the idiosyncratic knowledge about the economy possessed by the individual forecaster. I assume that the clients or employer of the economist want to receive a forecast, f_j, which is error-minimizing. If the economist were acting in the best interest of the forecast purchaser, he would honestly report $f_j = e_{\mathrm{j}}$.

I assume that, for each forecaster, I_j contains the lagged forecasts of all the other forecasters. This assumption is quite realistic, since professional forecasters disseminate their forecasts through the media, newsletters, fax, electronic services such as Telerate and Market News Service, and through published surveys in periodicals including *Business Week*, *Wall Street Journal*, *Barron's*, *Institutional Investor*, and *Blue Chip Economic Indicators*. For example, subscribers (and contributors) to Blue Chip receive monthly forecasts from a panel of forecasters, with each forecast identified by forecaster name and firm. A summary measure of others' forecasts is the mean of their forecasts. This statistic, which I here denote as f_{c}, is usually called the "consensus" forecast and is widely circulated.

Even if forecaster *j* does not know precisely what each other forecaster is saying, it is reasonable to assume that he has a good idea of what consensus is at any time. In the real world of professional forecasts, forecasts are produced in an almost continuous process, where forecasters are constantly updating their forecasts to reflect new information. For example, while this study samples forecasts at annual frequencies, many forecasters produce forecasts at weekly intervals (and some produce commentary daily which is distributed electronically or via fax). Forecasters often comment on consensus and why they are deviating from it.

Empirical studies often use consensus to represent expectations (for example, Froot, 1989). The ideas presented here suggest that in some situations, f_{c} is a good measure of expectations even when the individual f_j are not.

Suppose forecaster *j* is paid proportionately to his reputation, which is formed by comparing his forecast to other forecasts and to the realized outcome:

$$w_j = R(|f_j - y|, |f_j - f_{\mathrm{c}}|) \tag{2}$$

where w_j is the wage, reputation *R* is a function of the absolute value of its arguments, and the partial derivative $R_1 \leq 0$. I assume f_{c} is known at the time that forecaster *j* makes his forecast, or can be (arbitrarily well) approximated by its lagged value. Clearly, the forecaster would be willing to set $f_j = e_j$ (report his true expectation) in the case where $R_2 = 0$, as happens if *R* is determined by forecast accuracy alone.

If $R_2 < 0$, forecasters would want to set f_j closer to the consensus, and relative to the benchmark efficiency case of error-minimizing forecasts, the f_j would be concentrated around consensus. I call this phenomenon herding. If $R_2 > 0$, then forecasters would want to move f_j away from consensus. I call this scattering.

Scharfstein and Stein present a setting in which the payoff structure is similar to Eq. (2) with $R_2 < 0$. They model the behavior of a forecaster (called a manager) who is making a forecast (an investment decision) in an environment where good forecasters observe a signal that is correlated with other good forecasters' signals, while bad forecasters observe uncorrelated noise. Employers use both forecast error and deviation from consensus to infer the type of the agent. One forecaster moves first; under some parameter values (for example, when R_1 is small or zero), the subsequent forecasters always mimic the first one, thus, forming a herd. Their model reflects a comment by Keynes (1936): "Worldly wisdom teaches that it is better for reputation to fail conventionally than to succeed unconventionally". Scharfstein and Stein also discuss the inefficiencies that result when $R_2 > 0$. If forecasters are paid according to relative ability, they might scatter, since it is hard to win when making a forecast similar to others'.

The theory so far cannot be tested. For example, McNees (1989) reports that for forecasters who use econometric models and then make judgmental adjustments before selling the forecasts, for some variables adjustments work to make the forecasts more similar to each other. This cannot be interpreted as evidence in favor of herding, however, because the judgmental adjustments contain two components: a non-optimal herding/scattering component, and a Bayesian component which properly incorporates extra model information such as commonly observed news and others' published forecasts.

I test here a hypothesis discussed in Scharfstein and Stein, who suggest two competing effects that cause the severity of herding to vary over the life cycle of the forecaster in the $R_2 < 0$ case. Eq. (2) can represent a dynamic model, if the reputation function includes as arguments all past observations and the forecaster is optimizing his discounted sum of expected future earnings (for notational simplicity I will continue to refer to R_2, although this now represents R_2's dynamic analog). Scharfstein and Stein suggest:

> . . . herding may become more or less of a problem as a manager's career progresses. One the one hand, there is apt to be less uncertainty about the manager's ability, which should reduce the incentives for herd behavior. On the other hand, later in a successful career, wages are probably higher above the outside alternative. . . This latter effect can increase the propensity to herd.

The first effect is the "tighter priors" effect. If herding is induced by uncertainty about a forecaster's talent, then as his career progresses, uncertainty diminishes and so the herding incentive diminishes. This effect is similar to the seminal Holmstrom (1999) dynamic model of reputation, which in the simplest case implies managerial effort declining over time. In the present context, this means that over time R_2 goes to zero.

The second is the "far-to-fall" effect. If R is bounded from below (as when forecasters can always earn some alternate wage by switching professions), then a forecaster at the lower bound faces $R_2 = 0$. As a forecasters' wages and reputation rise over time, his reputation becomes a more valuable asset to be protected, the magnitude of $R_2 < 0$ increases, and herding increases.

In the case of $R_2 > 0$, scattering might also vary intertemporally. A similar "far-to-fall" effect is also present in the Diamond (1991) model of reputation acquisition in debt markets. A story in the spirit of the Diamond model: young forecasters initially scatter and choose outlandish forecasts, since the expected NPV of future payoffs is low (they might drop out of the forecasting business altogether). If they are lucky enough to make good guesses, thereby acquiring a reputation for accuracy, they then survive to become mature, conservative forecasters who cease scattering.

I test, therefore, whether the pattern of forecast herding/scattering varies significantly over the professional lifetime of the forecaster. I make no attempt to test whether forecasters are herding or are scattering (that is, placing too much/little weight on the forecasts of others); I simply examine movements in a measure of forecast dispersal. The relevant null hypothesis is that the dispersal of forecasts is unrelated to the forecasters' age or other measures of reputation.

Early studies of reputation and forecasting include Lakonishok et al. (1992), who test for herding in the case of institutional money managers' stock purchases and are unable to reject the no-herding null. Ehrbeck and Waldmann (1996) study whether forecasters revise their own forecasts optimally. Other evidence is only suggestive. Ito (1990) studies exchange rate forecasts, and finds evidence for "wishful expectations"—forecasters predict events that will benefit their firms (for example, exporters forecast depreciation). Although there are many possible explanations, one can imagine agency problems that might explain this pattern. For example, suppose the forecaster gets a bonus for accuracy, but only if the firm has sufficient profit to pay bonuses. [1]

Since this paper was first circulated as Lamont (1994), the literature on reputation and herding has burgeoned. For example, Avery and Chevalier (1999) present a model that also predicts that older agents will herd less. For other work, see Prendergast (1999) who reviews recent theory and empirical papers.

Three papers are of particular interest, since they also test whether herding changes as agents age and reputations develop. First, Chevalier and Ellison (1999) study mutual fund managers. They study the relation between performance and subsequent firing, and find it may give young managers an incentive to herd and decrease risk. They then examine the portfolio holdings of young managers, and find that young managers hold portfolios that are more conventional and less risky. Second, Graham (1999) studies stock recommendations made by editors of investment newsletters. He does not find a statistically significant correlation between age and dispersion. Third, Hong et al. (2000) examine earnings forecasts made by stock analysts. They find that young analysts herd, having forecasts that are closer to consensus and also tend to produce their forecasts after observing forecasts made by older analysts. To summarize, two out of the three papers confirm the main result of this paper, that younger forecasters herd more. Thus, the finding has fared well out of sample.

I do not attempt to test related reputational models of herding-type behavior. Prendergast and Stole (1996) present a model where strategic behavior causes forecasters to update their forecasts in order to manipulate beliefs about their competence, so that young forecasters exaggerate their information and old forecasters are overly conservative (where the word "conservative" has a different meaning than that used elsewhere in this paper). Zwiebel presents a model where yardstick competition causes high-quality and low-quality forecasters to take innovative actions while middle-quality forecasters are conservative. Like, Scharfstein and Stein, Zwiebel also finds that as forecasters age, their tendency to innovate is subject to two competing effects.

I next present evidence that suggests that the *R* function above might be familiar to real-world market participants.

[1] I did not test for this pattern in our data. However, some journalists believe it is present. For example, Barron's (11/25/85) presented a forecast for higher inflation made by Drexel Burnham Lambert's chief economist and commented, perhaps, but consider whom a return to inflation would help most: highly leveraged entities, such as Drexel's junk bond underwriting clients.

3. Anecdotal evidence

There is significant anecdotal evidence that indicates forecasters are not paid according to their mean squared error. Forecasters seek to enhance their reputation, manipulate perceptions of their quality, and use their forecasts in various ways unrelated to the minimization of mean squared error. Many of the strategies discussed above appear to be used in practice.

First, I note the stochastic environment assumed in reputational models is quite realistic in the context of macroeconomic forecasting, since it appears to be difficult to infer ability from forecast track records. For example, Felix Rohatyn commented in the Wall Street Journal that "the record of very intelligent people is so bad that you have to come to the conclusion that it is not the fault of the people but that it is essentially an unpredictable situation" (WSJ 1/4/82).

Scattering appears to be a popular practice, both to generate attention and to gain credibility in the unlikely event that the forecast turns out to be accurate. For example, in discussing Wall Street economists, Henry (1989) reports:

> Another technique for seeking attention is to produce a forecast that departs sharply from the consensus... it probably will get you some press, and there seems to be room in the market for a group of "intelligent extremists"... one or two strikingly unorthodox predictions that prove accurate can make a career... If you're hot, you'll get favorable publicity and so will your firm. And, during those periods when you're consistently wrong, so what. You'll surely have plenty of company, and being right or wrong does not seem to matter... after you appear in the press a few times, you become an authority figure in customer's minds.

In terms of the notation, this passage suggests that $R_2 > 0$, and that R_1 is not particularly large in magnitude. Intertemporally, it also implies that once a forecaster becomes "an authority figure", R is permanently racheted up to a fixed higher level.[2] Scattering is also rewarded by the press, as extreme forecasts garner the most attention both ex ante and ex post, regardless of accuracy.

Herding is also frequently mentioned, by journalists and by forecasters. Celebrated forecaster Henry and Kaufman believe that $R_2 < 0$: "There is comfort in being with the crowd... you cannot be singled out for being wrong or be a target of envy or resentment for being right" (Barron's 10/17/94).

One practice is the "broken clock" strategy, which consists of always forecasting the same event. An example in the sample is A. Gary Shilling, a well-known recession-caller. Throughout the 1980s, Shilling continually predicted recession. In 15 out of 18 Wall Street Journal surveys in which he participated 1981–1992 (data which are not used elsewhere in this paper), his year-ahead long-bond yield projection was the lowest among all forecasters. Fig. 1 shows GNP growth forecasts published in Business Week (various issues). As can be seen in the figure, 8 out of 10 times his forecast is below consensus, and he is often

[2] Another example is Henry Kaufman's reputation. After achieving celebrity status for predictive accuracy in the late 1970s, Kaufman suffered a series of inaccurate forecasts, but his reputation remained intact: "Once you build up that kind of momentum, it is equally hard to dissipate,' says Robert Sinche, chief economist of Bear, Stearns, and one of his rivals" Barron's (12/19/83).

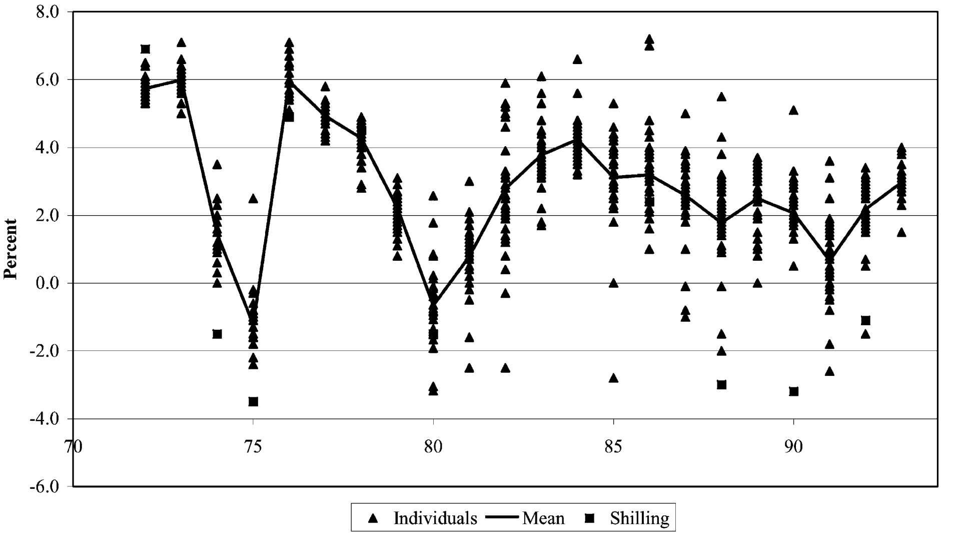

Fig. 1. Forecasts of real GNP growth, 1972–1993.

the extreme pessimistic outlier (when he is optimistic in his forecast for 1972, he is also the extreme optimistic outlier). Keane and Runkle cite Zarnowitz's (1969) finding that in a survey including non-professional forecasters, "a number of the occasional forecasters submitted extreme and rather unreasonable predictions" as an example of "inaccuracies due to lack of proper economic incentives" of the forecaster. Yet it is hard to describe Shilling's forecasts as anything, but "extreme and rather unreasonable".

Shilling also provides direct anecdotal evidence on the forecasters' manipulation of perceived ability by feeding his clients selected observations into the *R* function. The following abstract is taken from the 1987 Wall Street Journal index:

> A. Gary Shilling & Co., an economic consultant and investment strategist, recently mailed clients material that included a copy of a Wall Street Journal article with a paragraph showing that Mr. Shilling had made the best forecast of 30 years treasury bonds in a survey published about a year ago; but he covered up a paragraph noting that Mr. Shilling was tied for last place with his bond forecast of 6 months ago (1/26-23; 3).

Note that Shilling is not an outcast among forecasters: he is frequently quoted in the financial press, and has run his own firm for more than 10 years, a firm which in 1992 employed 18 people and had US$ 2 million in annual sales.

Before discussing the data and results, it is important to note the limits of the analysis. First, the hypothesis I test is not a sharp one. Since reputational models suggest (at least) two competing effects, they cannot make a strong prediction about the dispersion/age curve—it may slope either upward, downward, or be non-linear. This ambiguity makes the reputation concerns hypothesis harder to reject.

Second, no matter what I find, there will always be a non-rational psychological story to explain the findings. If I find young forecasters have high dispersal, it can be explained by the impetuousness of youth. Similarly, middle-aged dispersion could be explained by mid-life crisis, old-age dispersion by cantankerousness. On a more cognitive note, another possibility is that as forecasters become more experienced, the precision of their estimates changes.

4. Data on forecasts and forecasters

The data come from Business Week's annual year-end outlook issue from 1971–1992 (generally the last issue of the year, published in December). The surveys featured forecasts made for the subsequent year; thus, forecasts are available for the years 1972–1993. Each issue surveyed thirty or more economic forecasters and listed each forecaster's name, firm name, and forecasts for several macroeconomic variables. For every year 1971–1992, annual real GNP growth forecasts were available, and for every year but 1979, forecasts for CPI inflation and the unemployment rate were also available.

Business Week categorized the forecasters in a way that is potentially useful for testing the reputational hypothesis. Prior to 1989, forecasters were classified as being either "Economists" or "Econometric Models". For the case of econometric models, Business Week did not list a human's name but rather listed the name of model (for example, DRI or Wharton Econometrics). [3]

This categorization is useful since the reputational hypothesis has implications for humans but not for models. Thus, the contemporaneous forecasts made by non-human sources provide a benchmark for evaluating human behavior. Business Week's classification is surely imperfect, since most human forecasters use econometric models to assist them, and most commercial econometric forecasts contain significant judgmental components. But as long as the classification scheme contains some information, one would expect that any reputation-driven time profile of forecast dispersal to be less pronounced in the forecasts made by "Econometric Models".

The participants included economists from investment banks, commercial banks, money management firms, and financial consultant firms. A smaller number of participants came from regional banks, academia, and non-profit industry groups. Some economists changed firms, sometimes more than once. I did not collect other information about the forecasters that was constant over time and that might affect their forecasts (e.g. education, temperament, or ideological orientation). However, Kaufman (1984) shows that more than 50 percent of bank economists have Ph.D. and more than 75 percent have graduate degrees of some type.

Using the Business Week data I created an unbalanced panel of forecasts. Since I wanted mainly to examine changes in forecast properties over the life of the forecaster, I discarded all forecasters who did not participate in at least three surveys.

To estimate the hypothesis about the forecast deviations of each forecaster, $|f_j - f_c|$, I calculated f_c as the simple mean of the forecasts in my sample. Since I wanted to analyze

[3] The magazine stopped categorizing forecasters in this way after 1988 and instead listed the name of a human being associated with the model. For example, after 1988 the forecaster "Ray Fair" replaced "FAIRMODEL" in Business Week's survey.

Table 1
Summary statistics

	Humans	Models	Total
Number of forecasters	118	15	133
Average number of GNP observation per forecaster	4.8	10.9	5.5
Forecasters owning own firm (ever)	16	–	16
$\|f_j - f_{c(-j)}\|$ in basis points			
GNP (mean, S.D.)	76, 85	62, 62	73, 81
CPI (mean, S.D.)	53, 50	46, 44	51, 48
UNEMP (mean, S.D.)	34, 39	27, 26	32, 37

how an individual forecaster behaved taking the other forecaster's moves as given, for each forecaster j I calculated a corresponding consensus forecast, $f_{c(-j)}$, as the average forecast for that variable and time period excluding the forecast of forecaster j.[4] Fig. 1 shows the individual forecasts (and the consensus forecast) for GNP growth for each year. Table 1 shows the summary statistics for each of the three variables.

Zarnowitz and Lambros (1987) have documented that when forecast uncertainty is higher (as subjectively reported by forecasters), forecast dispersal increases (comments from the Business Week survey support this claim: one forecaster commented in 1972 that "It is only at the beginning and end of a business cycle when there is uncertainty and forecasts are spread around").[5] To control for time-varying aggregate shocks affecting the dispersal of forecasts (perhaps reflecting increased uncertainty), I calculated the variable AVGDEV($-j$) as the average of the forecast deviations $|f_j - f_{c(-j)}|$ of the forecasts produced for a given variable and a given time period by all the forecasters other than forecaster j. Thus, for each forecast, there is the mean, $f_{c(-j)}$, and average deviation, AVGDEV($-j$), of all the other competing forecasts. Fig. 2 shows the time path of AVGDEV for real GNP growth forecasts (and for comparison includes the standard deviation of forecasts as well).

5. Estimation results: forecast deviations

Since the hypothesis is about the time-varying component of strategies, I wanted to allow for forecaster-specific components of forecast deviations to cope with some of the issues suggested in Section 3. To the extent that the idiosyncratic strategies pursued by different forecasters are constant over time, they can be controlled for using fixed effects. The fixed effects also avoid "vintage effects" that might explain why old forecasters and young forecasters differ.

[4] So that the consensus forecast is the mean of the all forecasts for a given variable for a given year, after discarding the forecasts made by forecasters who had participated in less than three surveys and the forecast made by forecaster j.

[5] Of course, uncertainty is not always known ex ante. The 23 December 1972, Business Week declared flatly that one should "plan with confidence—1973 will be an excellent year for the economy". Paul Samuelson commented that "1973 is a pretty easy year to forecast"—not anticipating the Arab embargo and resulting recession in 1973/1974.

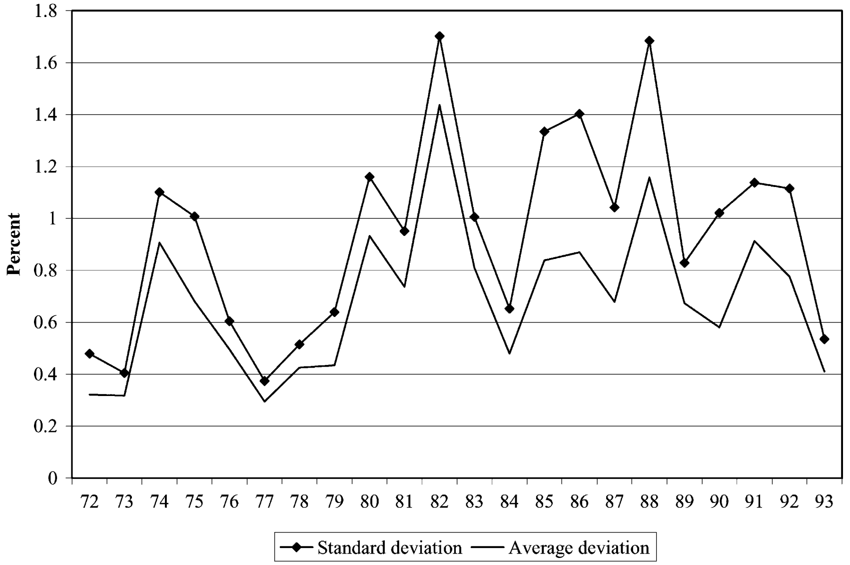

Fig. 2. GNP forecast dispersal, 1972–1993.

The dependent variable is $|f_j - f_{c(-j)}|$, the absolute deviation from forecast consensus. To capture the effects of reputation as measured by age, I regressed absolute deviations on an individual-specific constant, and on the variable AGE, the number of years since the forecaster first appeared in the survey. Thus, AGE is zero the first year that a forecaster appears, one in the subsequent year, and so on. The variable AGE is of course exactly collinear with the chronological age of the forecaster and the forecaster's experience in field.

The sample includes forecasts made both by humans and by models. Since models are not driven by reputational concerns, I allowed models to have their own time profile of dispersal by estimating both a coefficient on AGE and on AGE$*$MODEL, where MODEL is a dummy variable equal to one for forecasters Business Week classified as "Econometric Models". If humans have a reputation-driven time profile of forecast deviations but models do not, one would expect AGE and AGE$*$MODEL to have coefficients of opposite sign and equal magnitude.

Table 2A shows the basic fixed effects estimates. The positive coefficient on AGE shows that as individuals grow older, their forecasts become more radical. The negative coefficient on AGE$*$MODEL means that for models, the effect of age is approximately zero. The positive coefficient on AVGDEV shows that the more dispersed other's forecasts are in a given year, the more likely an individual's forecast will be far away from the mean forecast.

For all three of the macro variables, AGE and AGE$*$MODEL have opposite signs of roughly equal magnitude. For two of the three, AGE is significantly different from zero, so that one can reject the hypothesis that a human forecaster's deviation is unrelated to his age. For one of the three one can reject the hypothesis that humans and models have the same coefficient on age. For all three, one cannot reject the hypothesis that for models, forecast dispersal is unrelated to age.

Table 2
AGE and forecast dispersal fixed effects and GLS estimates[a]

Dataset	AGE	AGE∗MODEL	AVGDEV($-j$)	
(A) Fixed effects (dummy variable) estimates				
(i) GNP	1.80	−0.91	0.77	$R^2 = 0.43$, $N = 728$
OLS S.E.	0.74	1.47	0.10	
t-Statistics	2.44	0.62	7.54	
(ii) UNEMP	1.48	−2.28	0.67	$R^2 = 0.49$, $N = 700$
OLS S.E.	0.35	0.64	0.15	
t-Statistics	4.28	3.54	4.53	
(iii) CPI	0.53	−0.97	0.65	$R^2 = 0.42$, $N = 700$
OLS S.E.	0.44	0.90	0.14	
t-Statistics	1.21	1.08	4.51	
(B) GLS (error components) estimates				
(i) GNP	2.20	−1.56	0.77	$N = 728$
GLS S.E.	0.67	1.18	0.10	
t-Statistics	3.30	1.32	7.87	
(ii) UNEMP	1.45	−1.90	0.63	$N = 700$
GLS S.E.	0.31	0.54	0.13	
t-Statistics	4.66	3.35	4.83	
(iii) CPI	0.59	−1.17	0.65	$N = 700$
GLS S.E.	0.40	0.74	0.14	
t-Statistics	1.46	1.58	4.74	

[a] Dependent variable: $|f_j - f_{c(-j)}|$ in basis points. Independent variable: includes individual intercept for each individual forecaster (fixed effects), not displayed.

After estimating the fixed effects equations, I then computed a random effects GLS estimate which is more efficient but potentially inconsistent. For all the GLS estimates reported in this paper, a standard Hausman (1978) test does not reject the hypothesis that error components is the correct specification. The GLS estimates are shown in Table 2B and (not surprisingly, given the results of the Hausman test) show roughly the same results as the fixed effects estimates. As human forecasters become more experienced, they make less conservative forecasts.

The coefficients from Table 2 imply the following. Holding constant the dispersion of forecasts made by his competitors, as an individual human forecaster ages 10 years, he (on average) increases the distance of his GNP growth forecast from the consensus forecast by 22 basis points or 0.22 percent (from Table 2B (i)).

I also explored possible alternative specifications, including redefining the dependent variable. Since I was not sure about whether forecast deviations should be normalized, I tried dividing $|f_j - f_{c(-j)}|$ by either the consensus forecasts (so that deviations are measured as a fraction of the consensus) or by AVGDEV($-j$), Table 3A and B show the result. Again, human forecasters (but not models) exhibited increasing forecast dispersal as they grew older; the results do not depend on normalization.[6] I also considered the possibility

[6] For GNP growth, I did not normalize by the consensus forecast since this variable was not strictly positive during this period.

Table 3
AGE and forecast dispersal[a]

Dataset	AGE	AGE∗MODEL	
(A) Dependent variable: $(\lvert f_j - f_{c(-j)}\rvert)/f_{c(-j)}$ (in percent)			
(i) UNEMP	0.18	−0.17	$N = 700$
GLS S.E.	0.04	0.07	
t-Statistics	4.31	2.39	
(ii) CPI	0.36	−0.19	$N = 700$
GLS S.E.	0.09	0.16	
t-Statistics	3.82	1.12	
(B) Dependent variable: $(\lvert f_j - f_{c(-j)}\rvert)$/AVGDEV(−*j*) (in percent)			
(i) GNP	2.71	−2.28	$N = 728$
GLS S.E.	0.96	1.69	
t-Statistics	2.82	1.35	
(ii) UNEMP	3.52	−7.18	$N = 700$
GLS S.E.	0.99	1.78	
t-Statistics	3.56	4.04	
(iii) CPI	1.22	−2.89	$N = 700$
GLS S.E.	0.82	1.47	
t-Statistics	1.50	1.96	
(C) Dependent variable: $\lvert f_j - f_{c(-j)}\rvert$ in basis points (year dummies on R.H.S.)			
(i) GNP	3.49	−2.86	$N = 728$
GLS S.E.	0.69	0.92	
t-Statistics	4.99	3.12	
(ii) UNEMP	1.65	−1.72	$N = 700$
GLS S.E.	0.32	0.43	
t-Statistics	5.20	4.03	
(iii) CPI	1.05	−1.99	$N = 700$
GLS S.E.	0.43	0.57	
t-Statistics	2.44	3.46	

[a] Alternative dependent variables and alternative specifications; GLS estimates

that aggregate time-varying shocks, not captured by AVGDEV(−*j*), might be driving the results. I, therefore, tried, in Table 3C, estimating the model using year-dummies but not individual-dummies. These results rejected the no reputation null even more strongly.[7]

I checked to see if the results in Table 2B were caused by extreme observations. I reran the regressions excluding the outlier forecaster mentioned before (Shilling). I also checked to make sure Business Week's definition change in 1989 did not affect the results. Last, I excluded all models so that the results were run with only human forecasters (and excluding AGE∗MODEL). None of these changes altered the conclusions from Table 2B.

Is there a better measure of reputation than AGE? Five participants entered the sample as principals in firms bearing their own names, and eleven others founded their own firms during the sample period. It is likely that economic forecasters running their own firm have

[7] Due to the definition of AGE as an individual specific linear time-trend, I could not include both AGE, year dummies, and individual dummies; only two of these three can be included in any regression.

Table 4
OWNFIRM variable[a]

Dataset	AGE	AGE∗MODEL	AVGDEV	OWNFIRM	
(i) GNP	1.36	−0.64	0.75	51.7	$N = 728$
GLS S.E.	0.69	1.18	0.10	12.6	
t-Statistics	1.96	0.55	7.76	4.11	
(ii) UNEMP	0.90	−1.25	0.57	34.1	$N = 700$
GLS S.E.	0.32	0.53	0.13	5.9	
t-Statistics	2.84	2.33	4.51	5.79	
(iii) CPI	0.09	−0.65	0.65	29.5	$N = 700$
GLS S.E.	0.42	0.74	0.14	8.1	
t-Statistics	0.22	0.87	4.73	3.65	

[a] GLS estimates; dependent variable: $|f_j - f_{c(-j)}|$.

a well-established reputation, since Business Week must have selected them based on their personal reputation rather than their employer's reputation. I, therefore, created the dummy variable OWNFIRM, set equal to one if the last name of the (human) forecaster is currently (or in the past has been) included in his affiliation listed in Business Week. For example, forecaster Robert H. Parks was listed as being employed by Robert H. Parks & Associates. About nine percent of the sample comes from OWNFIRM forecasters. [8]

Table 4 shows the results with OWNFIRM. For each of the three macro variables, OWNFIRM is strongly positive and significant (as before, the fixed effects estimates, not shown, are very similar). [9] Controlling for age and the individual-specific effect, a forecaster who starts his own firm raises his deviation from the consensus GNP forecast by a whopping 52 basis points (compared to an average of 73 basis points for this deviation). AGE and AGE∗MODEL still have the same qualitative pattern, although the effect is somewhat weakened (AGE and OWNFIRM are by definition positively correlated).

Again, I checked to make sure the coefficient on OWNFIRM was robust to changes in the sample. Excluding outlier Shilling, excluding post-1988 data, and excluding models did not affect the main conclusion from Table 4: OWNFIRM is still large and statistically significant for all three macro variables.

6. Estimation results: accuracy

It is a well-documented fact in forecast survey data that consensus forecasts are much more accurate than the majority of individual forecast (as shown, e.g. in Zarnowitz and Braun (1993)). The results presented above on ex ante forecast deviation show that as forecasters become older and more established, they make forecasts that are farther away

[8] OWNFIRM is equal to one the first year that the forecaster's company name contained his last name, and for every year thereafter. Of the 16 OWNFIRM forecasters, three reverted from being self-employed to being employed according to Business Week.

[9] Note again that the fixed effects control for constant individual idiosyncrasies, so that the following story is not consistent with the data: erratic forecasters get fired and start their own firms. Any explanation for OWNFIRM's coefficient has to explain why the same forecaster would increase his dispersal after starting his own firm.

Table 5
Ex post accuracy[a]

Dataset	AGE	AGE∗MODEL	AVGACC	OWNFIRM	
(A) Without OWNFIRM					
(i) GNP	2.22	−2.78	0.96		$N = 728$
GLS S.E.	0.81	1.14	0.03		
t-Statistics	2.74	2.44	31.3		
(ii) UNEMP	0.96	−1.00	0.96		$N = 700$
GLS S.E.	0.36	0.58	0.04		
t-Statistics	2.70	1.72	21.5		
(iii) CPI	0.96	−0.72	1.00		$N = 700$
GLS S.E.	0.52	0.82	0.02		
t-Statistics	1.83	0.88	49.5		
(B) With OWNFIRM					
(i) GNP	1.37	−1.82	0.95	49.8	$N = 728$
GLS S.E.	0.83	1.14	0.03	13.0	
t-Statistics	1.64	1.59	31.3	3.84	
(ii) UNEMP	0.37	−0.36	0.96	32.8	$N = 700$
GLS S.E.	0.36	0.58	0.04	6.3	
t-Statistics	1.03	0.63	21.8	5.19	
(iii) CPI	0.59	−0.33	1.00	20.2	$N = 700$
GLS S.E.	0.54	0.83	0.02	9.2	
t-Statistics	1.09	0.40	49.6	2.21	

[a] GLS estimates; dependent variable: $|f_j - y|$.

from consensus. Does this imply that forecasters become less accurate over time? Using the data and framework presented here, the answer is a clear yes; forecasters become less accurate as they grow older and gain reputations.

A full and traditional test of the rationality properties of forecasts is beyond the scope of this paper. Instead, I simply looked at ex post forecast accuracy, $|f_j - y|$, for each forecaster and year.[10] For each forecaster *j*, I calculated AVGACC($-j$), the average ex post forecast accuracy of *j*'s competitors (by averaging $|f_i - y|$ for all *i* not equal to *j*). Using AVGACC-($-j$) as a control variable that captures the common component of forecast errors, I regressed ex post accuracy against AGE and OWNFIRM.

Table 5 shows the results for forecast accuracy. Consistent with the literature's findings that consensus is the most accurate forecast, the results mirror those of Tables 2–4. Panel A shows that for human beings, but not for models, accuracy worsens as forecasters age. Panel B shows that forecasters who start their own firms experience a large decrease in ex post forecast accuracy. In contrast to Table 4, given the OWNFIRM variable none of the AGE variables are statistically significant, although they retain the same pattern. Thus, there is only weak evidence that age affects accuracy independently of employment status.

OWNFIRM continues to be large and highly significant for all variables. Table 4 shows that ex ante, forecasters who start their own firm increase their absolute deviation for GNP forecasts by 52 basis points. Table 5 shows that ex post, this change in behavior decreases

[10] Using revised data on GNP, prices, and unemployment.

accuracy almost one-for-one, since forecasters who start their own firm increase their errors by 50 basis points. As a reference, average ex post accuracy for real GNP forecasts was 160 basis points during the sample period.

Again, I checked for robustness, excluding outlier Shilling, excluding data from 1989 and after, and excluding models. The conclusions from Table 5B are robust to these variations. [11]

7. Discussion and conclusion

In summary, the empirical findings are quite consistent with the reputational hypothesis. By a variety of measures, forecast dispersal exhibits a systematic pattern of the professional life of human forecasters. This pattern is not matched by contemporaneous forecasts made by econometric models. Older human forecasters make bolder forecasts compared to their own behavior when younger. Further, when human forecasters establish their own firm, their behavior changes dramatically and they produce even bolder forecasts. These bolder forecasts turn out to be less accurate, particularly for forecasters who start their own firm.

These results do not mean that reputation or yardstick competition hurt economic efficiency. Clearly, it would not be socially useful if all forecasters sought to minimize mean squared error by mimicking consensus; depending on the R function, this herding may also not be an equilibrium. It may well be that as forecasters age, they contribute more information to the collective process that establishes consensus forecasts.

Can cognitive factors explain the findings presented here? Any cognitive story (in which forecasters report their true expectations) would have to explain why owning one's own firm is correlated with having a different process of expectations formation. [12] It would also have to explain why forecast accuracy deteriorates as forecasters become more experienced.

Combined with other recent evidence, the results presented here suggest that forecasters are not uniformly acting to minimize mean squared error. What they are optimizing remains unclear. Recent developments in information economics suggest they are optimizing the market value of their reputations.

Acknowledgements

I thank Olivier Blanchard, Ray Fair (who is also one of my data points), David Scharfstein, Jeremy Stein, and seminar participants at Columbia University, the Federal Reserve Bank of Cleveland, George Washington University, Lafayette College, and the NBER Behavioral Macro Program for helpful comments. The Alfred P. Sloan Foundation, the Center for Research in Securities Prices at the University of Chicago Graduate School of Business,

[11] Although Table 5B was robust, excluding forecaster Shilling caused the AGE coefficient in Table 5A (ii) to fall and become insignificant.

[12] Since the fixed effect estimates control for idiosyncratic effects, the following story does not explain the results. Some forecasters are better than others and therefore rationally and optimally give more weight to their own signals and less weight on the consensus forecast. They produce more dispersed forecasts, are more accurate, and are then rewarded by the marketplace by getting their own firm.

and the National Science Foundation generously supplied financial support. Kevin Grundy provided excellent research assistance. I am especially grateful to John Conlisk and John Graham, who are responsible for this paper's publication.

References

Avery, C.N., Chevalier, J.A., 1999. Herding over the career. Economics Letters 63, 327–333.

Banerjee, A., 1992. A simple model of herd behavior. Quarterly Journal of Economics 107, 787–817.

Business Week. Various issues.

Bryan, M.F., Gavin, W.T., 1986. Models of inflation expectations formation: a comparison of household and economist forecasts. Journal of Money Credit and Banking 18, 539–544.

Chevalier, J., Ellison, G., 1999. Career concerns of mutual fund managers. Quarterly Journal of Economics 114, 389–432.

Diamond, D.W., 1991. Reputation acquisition in debt markets. Journal of Political Economy 97, 828–862.

Ehrbeck, T., Waldmann, R., 1996. Why are professional forecasters biased? Agency versus behavioral explanations. Quarterly Journal of Economics 111, 21–40.

Froot, K.A., 1989. New hope for the expectations hypothesis of the term structure of interest rates. Journal of Finance 44, 283–305.

Hausman, J., 1978. Specification tests in econometrics. Econometrica 46, 69–85.

Henry, G.B., 1989. Wall street economists: are they worth their salt? Business Economics 10, 44–48.

Holmstrom, B., 1999. Managerial incentive problems: a dynamic perspective. Review of Economic Studies 66, 169–182.

Hong, H.G., Kubik, J.D., Solomon, A., 2000. Security analysts' career concerns and herding of earnings forecasts. Rand Journal of Economics 31, 121–144.

Graham, J.R., 1999. Herding among investment newsletters: theory and evidence. Journal of Finance 54, 237–268.

Ito, T., 1990. Foreign exchange rate expectations: micro survey data. American Economic Review 80, 434–449.

Kaufman, G.G., 1984. The academic preparation of economists employed by commercial and investment banks. Journal of Money Credit and Banking 16, 352–359.

Keane, M.P., Runkle, D.E., 1990. Testing the rationality of price forecasts: new evidence from panel data. American Economic Review 80, 714–735.

Keynes, John Maynard, 1936. The General Theory of Employment, Interest and Money. Macmillan, London.

Lakonishok, J., Shleifer, A., Vishny, R.W., 1992. The impact of institutional trading on stock prices. Journal of Financial Economics 32, 23–43.

Lamont, Owen A., 1994. Macroeconomic Forecasts and Microeconomic Forecasters. NBER Working Paper 5284.

McNees, S.K., 1989. Why do forecasts differ? New England Economic Review 1/2, 42–54.

Prendergast, C., 1999. The provision of incentives in firms. Journal of Economic Literature 37, 7–63.

Prendergast, C., Stole, L., 1996. Impetuous youngsters and jaded oldtimers: an analysis of behavioral decision-making rules. Journal of Political Economy 104, 1105–1134.

Scharfstein, D., Stein, J., 1990. Herd behavior and investment. American Economic Review 80, 464–479.

Zarnowitz, V., 1969. The new ASA–NBER survey of forecasts by economic statisticians. American Statistician 23, 12–16.

Zarnowitz, V., 1985. Rational expectations and macroeconomic forecasts. Journal of Business and Economic Statistics 3, 293–311.

Zarnowitz, V., Braun, P., 1993. Twenty-two years of the ASA–NBER quarterly economic outlook surveys: aspects and comparisons of forecasting performance. In: Stock, J.H., Watson, M.W. (Eds.), Business Cycles, Indicators, and Forecasting. The University of Chicago Press, Chicago, pp. 11–84.

Zarnowitz, V., Lambros, L.A., 1987. Consensus and uncertainty in economic prediction. Journal of Political Economy 96, 591–621.

Zwiebel, J., 1995. Corporate conservatism and relative compensation. Journal of Political Economy 103, 1–25.