Herding in Probabilistic Forecasts

Ville Satopää

Technology and Operations Management
INSEAD

Joint work with
Yanwei Jia (Columbia) and Jussi Keppo (NUS)

Outline

1. Introduction
2. Model Setup
   - Morris and Shin (2002)
   - Probabilistic Forecasts
3. Main Results
   - Equilibrium
   - Information Disclosure
4. Estimation
   - Model Identification
   - Economic Forecasts
5. Conclusion

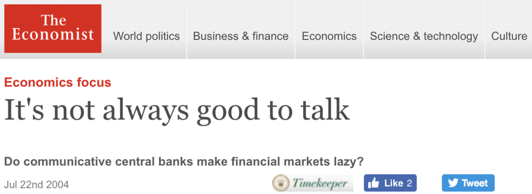
Outline

Herding Effect

- Herding effect is a well-documented social phenomenon.
  - Fund managers: Chevalier and Ellison (1999), Hong et al. (2000), Clement and Tse (2005);
  - Models of analysts/experts: Hong and Kubik (2003), Ottaviani and Sørensen (2006);
  - Scientific peer review: Park et al. (2014);
  - Social influence: Muchnik et al. (2013);
- Has been studied in different disciplines.
  - Biological sciences: Baddeley (2010);
  - Cognitive sciences: Raafat et al. (2009);
  - Economics: Keynes (1936), Scharfstein and Stein (1990), Morris and Shin (2002).
- Career concerns, peer pressure, social psychological reasons...

Expert Herding in Point Forecasting

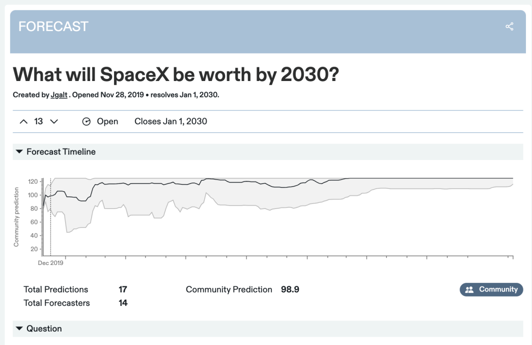
- Morris and Shin (2002) develop a model of information and herding among experts who make point forecasts simultaneously.
- Disclosure of public information reduces the typical forecasting accuracy when herding is strong and public information is sufficiently inaccurate relative to private information.
- Many follow-up papers on the externality of public information; e.g., Angeletos and Pavan (2007), Bergemann and Morris (2013) and Ui and Yoshizawa (2015).



Probabilistic Forecasting

Probabilistic forecasting is common in many applications, such as weather (Gneiting and Raftery, 2005), energy (Hong et al., 2019), macroeconomic forecasting (Garratt et al., 2003).

Example:



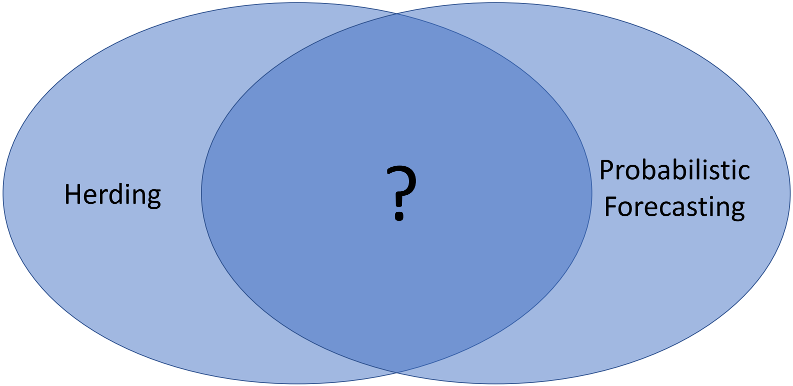
Point forecast: 100 billions.
Probabilistic forecast: normal distribution with mean 100 billions and std. dev. 10 billions.

Probabilistic Forecasting

Compared to point estimation, a probabilistic forecast is

- more informative (conveying uncertainties and confidence);
- necessary for decision makers:
  - weather, macroeconomic, financial forecast, banking regulation: managing tail risk;
  - newsvendor problem: estimating demand distribution;
  - multi-objective decision making: evaluating correlations.

Scope of the Current Work



Contributions (1/2)

- We generalize the model in Morris and Shin (2002) from point to probabilistic forecasts.
- Under our model, experts can herd in two different ways:
  1. Shrink the mean towards public information.
  2. Spread out the variance.
- Conclusions in Morris and Shin (2002) are no longer true when experts consider variance.
  - Disclosure of public information improves typical forecasting accuracy when herding is strong or public information is sufficiently inaccurate relative to private information.
  - Numerical results suggest that more accurate information (both public and private) always improves typical forecasting accuracy.

Contributions (2/2)

- Both point and probabilistic accuracy become less accurate as herding strengthens.
- We explore model identification in a one-shot forecasting setup. Our model can be identified
  - up to two parameter values based on probabilistic forecasts of a single outcome, and
  - uniquely based on probabilistic forecasts of two or more outcomes.
- We implement a Bayesian procedure and estimate herding in professional forecasts of economic indicators.

Outline

Experts' Information Structure

Morris and Shin (2002) consider the following model:

- Consider K experts, indexed by $k = 1, \ldots, K$, forecasting $\theta \in \mathbb{R}$.
- A flat prior $\pi(\theta) \propto 1$
- Public information $y|\theta \sim \mathcal{N}(\theta, \tau_y^{-1})$.
- Private information $x_k|\theta \sim \mathcal{N}(\theta, \tau_x^{-1})$.
- Expert k's Bayes-optimal belief

$$\theta | x_k, y \sim \mathcal{N}\left(\frac{\tau_x}{\tau_x + \tau_y} x_k + \frac{\tau_y}{\tau_x + \tau_y} y, \frac{1}{\tau_x + \tau_y}\right).$$

Point Forecasts: Dual Objective

Based on x_k and y, expert k chooses a value $\mu_k^{MS} \in \mathbb{R}$ that minimizes

1. forecasting error and
2. deviation from others.

Expert k chooses

$$\mu_k^{MS} = \arg\min_{\mu \in \mathbb{R}} \left\{ \mathbb{E} \left[(1-r) \underbrace{(\mu - \theta)^2}_{\text{Error}} + r \underbrace{D_k}_{\text{Deviation}} \middle| x_k, y \right] \right\},$$

where $D_k = \frac{1}{K-1} \sum_{j \neq k} \left(\mu - \mu_j^{MS} \right)^2$ and $r \in [0, 1]$ defines the strength of herding.

Point Forecasts: Equilibrium Solution

The (unique) equilibrium solution:

$$\mu_k^{MS} = \frac{(1-r)\tau_x}{(1-r)\tau_x + \tau_y} x_k + \frac{\tau_y}{(1-r)\tau_x + \tau_y} y. \qquad (1)$$

- Herding experts overemphasize public information.
- Under no herding ($r = 0$), μ_k^{MS} is the Bayes-optimal mean prediction.
- Under full herding ($r = 1$), μ_k^{MS} is the public signal y.

Point Forecasts: Value of Public Information

- Typical forecasting error

$$\mathbb{E}[(\mu_k^{MS} - \theta)^2 | \theta]$$

  is monotone decreasing in private signal precision τ_x but not in public signal precision τ_y.

- Series of discussions on this topic: Morris and Shin (2002), Hellwig (2005), Svensson (2006), Morris et al. (2006), Cornand and Heinemann (2008) and James and Lawler (2011).

Probabilistic Forecasts: Accuracy

- Expert k describes their belief with $\mathcal{N}(\mu_k, \sigma_k^2)$ by reporting $\boldsymbol{a}_k = (\mu_k, \sigma_k^2)$.
- Experts measure accuracy with a proper scoring rule (Gneiting and Raftery, 2007).
- The (negative) logarithmic score

$$E(\boldsymbol{a}_k; \theta) = \frac{(\mu_k - \theta)^2}{2\sigma_k^2} + \frac{1}{2}\log\sigma_k^2 + \frac{1}{2}\log 2\pi, \qquad (2)$$

has many desirable properties (Du, 2021), is common in the information theory, economics, and decision science literature (e.g., Sims 2003 and Prelec 2004), and has become one of the most popular proper scoring rules for continuous variables in practice (Jordan et al., 2018).

Probabilistic Forecasts: Deviation

- Other personal objectives besides accuracy.
- The score divergence from $\boldsymbol{a}_j$ to $\boldsymbol{a}_k$:

$$\mathbb{E}_{z\sim\mathcal{N}(\mu_j,\sigma_j^2)} E(\boldsymbol{a}_k; z) - \mathbb{E}_{z\sim\mathcal{N}(\mu_j,\sigma_j^2)} E(\boldsymbol{a}_j; z).$$

- "How much worse is my forecast if I am wrong and they are right?"
- The score divergence associated with the logarithmic score is the Kullback-Leibler (KL) divergence:

$$D_{KL}(\boldsymbol{a}_j, \boldsymbol{a}_k) = \frac{1}{2}\left[\frac{\sigma_j^2}{\sigma_k^2} + \frac{(\mu_k - \mu_j)^2}{\sigma_k^2} - 1 + \log\sigma_k^2 - \log\sigma_j^2\right].$$

- Expert k's average deviation from the other experts' forecasts then is

$$D(\boldsymbol{a}_k, \boldsymbol{a}_{-k}) = \frac{1}{K-1}\sum_{j\neq k} D_{KL}(\boldsymbol{a}_j, \boldsymbol{a}_k). \qquad (3)$$

Probabilistic Forecasts: Dual Objective

Following Morris and Shin (2002), we model the experts' strategic complementarity with a dual objective:

$$\boldsymbol{a}_k = \arg\min_{\boldsymbol{a}} \mathbb{E}\left\{ (1-r)\underbrace{E(\boldsymbol{a};\theta)}_{\text{Error}} + r\underbrace{D(\boldsymbol{a},\boldsymbol{a}_{-k})}_{\text{Deviation}} \middle| x_k, y \right\}, \tag{4}$$

where $r \in [0,1]$ quantifies the level of herding.

Outline

Equilibrium Solution

Theorem 1 (Equilibrium Forecasts)

The unique equilibrium for our probabilistic forecasting model is a normal distribution with mean

$$\mu_k = \mu_k^{MS} = \frac{(1-r)\tau_x}{(1-r)\tau_x + \tau_y} x_k + \frac{\tau_y}{(1-r)\tau_x + \tau_y} y, \tag{5}$$

and variance

$$\sigma_k^2 = a(x_k - y)^2 + d \tag{6}$$

for all $k \in \{1, ..., K\}$*, where*

$$a = \frac{r(1-r)\tau_x^2\tau_y^2}{[(\tau_x+\tau_y)^2 - r\tau_x^2][(1-r)\tau_x + \tau_y]^2} \quad d = \frac{(1+r)\tau_x + \tau_y}{(\tau_x+\tau_y)^2 - r\tau_x^2}.$$

Equilibrium Solution: Remarks

- The equilibrium mean aligns with Morris and Shin (2002).
- If $r = 0$, $\boldsymbol{a}_k$ is the Bayes-optimal forecast.
- If $r = 1$, $\boldsymbol{a}_k = (y, \tau_y^{-1})$.
- Herding spreads variance: $\sigma_k^2 \geq \hat{\sigma}_\theta^2$.
- The expected variance prediction is

$$\mathbb{E}[\sigma_k^2|\theta] = \frac{(1 - r^2)\tau_x + \tau_y}{[(1 - r)\tau_x + \tau_y]^2},$$

which is increasing in r and decreasing in τ_x and τ_y.

Typical Forecasting Error

Theorem 2

Typical forecasting error $\mathbb{E}\left[E(\boldsymbol{a}_k;\theta)|\theta\right]$

- *rises in the herding level* r*; and*
- *falls in* τ_y *and* τ_x *if one the following holds:*

  (a) $\frac{\tau_y}{\tau_x}$ *is sufficiently small,*
  (b) $\frac{\tau_x}{\tau_y}$ *is sufficiently small,*
  (c) r *is sufficiently close to 0,*
  (d) r *is sufficiently close to 1.*

Forecast Under Herding

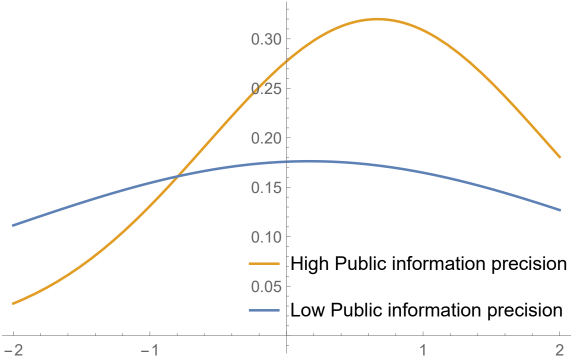


Figure 1: **The reported distribution in equilibrium**. The equilibrium solution under low and high public information precision. The public signal precision is taken as $\tau_y = 0.05,\ 0.5$ respectively. The other model primitives are $r = 0.75$ and $\tau_x = 1$. In addition, we assume $\theta = 0$ and the realized signals are $x_i = 0,\ y = 1$. The mean and variance parameter is $(0.167, 5.121)$ and $(0.667, 1.556)$ respectively.

Outline

Related Work on Model Identification

Structural estimation of one-shot incomplete information games is notoriously challenging:

- Bergemann and Morris (2013) prove that a wide class of normally distributed incomplete information games are unidentifiable.
- The model in Morris and Shin (2002) is only identifiable up to infinitely many specifications of the parameters.

Quadratic Regression Model

Our structural model can be written as:

$$\sigma_k^2 = \frac{a}{w^2}(\mu_k - y)^2 + d,$$

for a constant weight term $w = (1 - r)\tau_x / [(1 - r)\tau_x + \tau_y]$.

Assumption A1 (Structural Error)

Denote each expert k's reported mean and variance forecasts with m_k and v_k, respectively. Then, in estimation, we assume for all $k \in \{1, ..., K\}$ that

$$\log(v_k) = \log\left[\frac{a}{w^2}(m_k - y)^2 + d\right] + \epsilon_k, \tag{7}$$

where each error term $\epsilon_k \sim \mathcal{N}(0, \tau_\epsilon^{-1})$ is independent of all other model variables.

Model Identifiability: One Outcome

The parameters are $\boldsymbol{\Theta} = (\theta, y, \tau_x, \tau_y, \tau_\epsilon, r)$.

Theorem 3 (One-Shot Identification)

Under an increasingly large crowd forecasting a single outcome, we have the following:

- (i) *Given mean forecasts alone, $\boldsymbol{\Theta}$ is identifiable up to infinitely many specifications.*
- (ii) *Given mean and variance forecasts, $\boldsymbol{\Theta}$ is identifiable up to two possible specifications. Furthermore, in the specification with a larger level of herding, private information is more precise.*

Model Identifiability: Multiple Outcomes

Theorem 4 (Identifiability Under Multiple Outcomes)

Consider multiple simultaneous forecasting games with common level of herding.

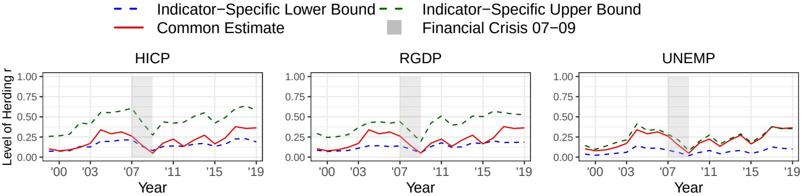
If information levels are not all identical, then the model parameters can be identified with sufficiently many probabilistic forecasts per outcome.

However, under the same conditions, the model parameters are identifiable up to infinitely many specifications based on point forecasts alone.

ECB Survey

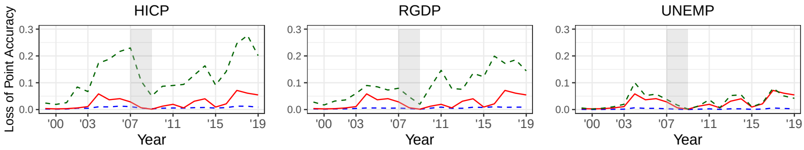
- Implement a Bayesian procedure for estimating the model.
- Apply to Surveys of Professional Forecasters by the European Central Bank (ECB):
  - HICP, RGDP, and UNEMP between 1999 - 2019.
  - Current and next 4 years.
  - On average 40 forecasts per outcome.

ECB: Level of Herding

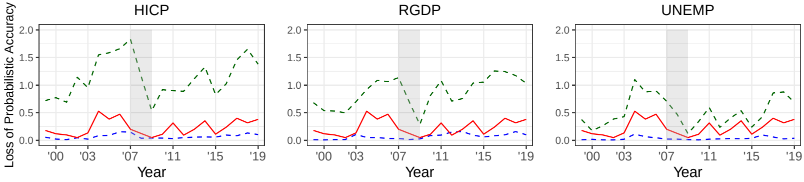


(a) Annual average estimate of the herding level r.

ECB: Impact of Herding on Typical Accuracy



(a) Annual average relative loss in the typical point accuracy due to herding.



(b) Annual relative loss in the typical probabilistic accuracy due to herding.

Remarks on Data Analysis

- Level of herding fluctuates over time.
- Herding can be more detrimental to probabilistic accuracy than to point accuracy
- Under longer horizons (four-year and five-year horizons), the level of herding fluctuates less and remains at a lower level.
- Robustness: Similar results under data from the Federal Reserve Bank of Philadelphia.

Outline

Summary of Results

- Herding is a well-known bias in forecasting. Probabilistic forecasting has become more common. We investigated the largely unexplored intersection.
- In probabilistic forecasting, an expert can herd in two ways:
  1. Shrink the mean towards the public information.
  2. Spread out the variance.
- Both point and probabilistic forecasts under herding are less accurate ...

  ... but probabilistic forecasts react more naturally to new information than point forecasts.
  - Robustness: Several extensions in the Electronic Companion.
- Probabilistic forecasts hold more information about the experts' information structure, which facilitates estimation.

Extensions and Robustness

- Normally distributed signals:
  - Log-score + Rényi's α-divergence
  - Quadratic accuracy score + L^2 distance
- Beta-binomial model
- Multinomial model
- General model of first two moments.

Future Studies

Some future directions:

1. Allow heterogeneity in experts' parameters.
2. Mechanism design for reducing herding behavior.
3. Aggregation of forecasts under herding.

Thank you!

Any questions?

Reference I

G.-M. Angeletos and A. Pavan. Efficient use of information and social value of information. *Econometrica*, 75(4):1103–1142, 2007.

M. Baddeley. Herding, social influence and economic decision-making: socio-psychological and neuroscientific analyses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1538):281–290, 2010.

D. Bergemann and S. Morris. Robust predictions in games with incomplete information. *Econometrica*, 81(4):1251–1308, 2013.

J. Chevalier and G. Ellison. Career concerns of mutual fund managers. *The Quarterly Journal of Economics*, 114(2): 389–432, 1999.

M. B. Clement and S. Y. Tse. Financial analyst characteristics and herding behavior in forecasting. *The Journal of Finance*, 60(1): 307–341, 2005.

Reference II

C. Cornand and F. Heinemann. Optimal degree of public information dissemination. *The Economic Journal*, 118(528): 718–742, 2008.

H. Du. Beyond strictly proper scoring rules: The importance of being local. *Weather and Forecasting*, 36(2):457–468, 2021.

A. Garratt, K. Lee, M. H. Pesaran, and Y. Shin. Forecast uncertainties in macroeconomic modeling: An application to the UK economy. *Journal of the American Statistical Association*, 98(464):829–838, 2003.

T. Gneiting and A. E. Raftery. Weather forecasting with ensemble methods. *Science*, 310(5746):248–249, 2005.

T. Gneiting and A. E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.

C. Hellwig. Heterogeneous information and the welfare effects of public information disclosures. *Technical Report*, 2005.

Reference III

H. Hong and J. D. Kubik. Analyzing the analysts: Career concerns and biased earnings forecasts. *The Journal of Finance*, 58(1): 313–351, 2003.

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

T. Hong, J. Xie, and J. Black. Global energy forecasting competition 2017: Hierarchical probabilistic load forecasting. *International Journal of Forecasting*, 2019.

J. G. James and P. Lawler. Optimal policy intervention and the social value of public information. *American Economic Review*, 101(4):1561–74, 2011.

A. Jordan, F. Krüger, and S. Lerch. Evaluating probabilistic forecasts with scoringRules. Available at arXiv: `https://arxiv.org/abs/1709.04743`, 2018.

Reference IV

J. M. Keynes. *The General Theory of Employment, Interest and Money*. Macmillan, 1936. 14th edition, 1973.

S. Morris and H. S. Shin. Social value of public information. *American Economic Review*, 92(5):1521–1534, 2002.

S. Morris, H. S. Shin, and H. Tong. Social value of public information: Morris and shin (2002) is actually pro-transparency, not con: Reply. *American Economic Review*, 96(1):453–455, 2006.

L. Muchnik, S. Aral, and S. J. Taylor. Social influence bias: A randomized experiment. *Science*, 341(6146):647–651, 2013.

A. O'Hagan, C. E. Buck, A. Daneshkhah, J. R. Eiser, P. H. Garthwaite, D. J. Jenkinson, J. E. Oakley, and T. Rakow. *Uncertain Judgements: Eliciting Experts' Probabilities*. John Wiley & Sons, 2006.

Reference V

M. Ottaviani and P. N. Sørensen. The strategy of professional forecasting. *Journal of Financial Economics*, 81(2):441–466, 2006.

I.-U. Park, M. W. Peacey, and M. R. Munafo. Modelling the effects of subjective and objective decision making in scientific peer review. *Nature*, 506(7486):93, 2014.

D. Prelec. A bayesian truth serum for subjective data. *Science*, 306(5695):462–466, 2004.

R. M. Raafat, N. Chater, and C. Frith. Herding in humans. *Trends in Cognitive Sciences*, 13(10):420–428, 2009.

V. Satopää. Regularized aggregation of one-off probability predictions. *Available at SSRN 3769945*, 2021.

D. S. Scharfstein and J. C. Stein. Herd behavior and investment. *American Economic Review*, 80(3):465–479, 1990.

C. A. Sims. Implications of rational inattention. *Journal of Monetary Economics*, 50(3):665–690, 2003.

Reference VI

L. E. Svensson. Social value of public information: Comment: Morris and shin (2002) is actually pro-transparency, not con. *American Economic Review*, 96(1):448–452, 2006.

T. Ui and Y. Yoshizawa. Characterizing social value of information. *Journal of Economic Theory*, 158:507–535, 2015.

Backup Slides

Simulation Results

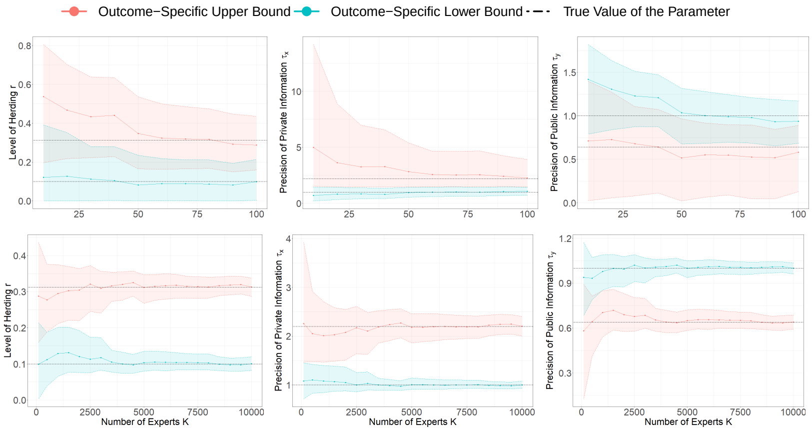
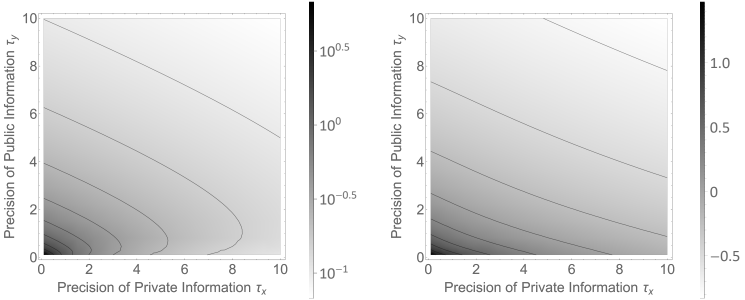


Figure 4: **Estimated model parameters under increasingly large sets of simulated forecasts.** The shaded regions are the 95% (pointwise) credible intervals under varying numbers of forecasters. The solid lines in the middle of each shaded region are the estimated posterior means. Horizontal dashed lines are the true parameter values. The parameter estimates converge to the two indistinguishable specifications of the parameters as the crowd grows larger. Model parameters: $\theta = 0.2$, $\tau_\epsilon^{-1/2} = 0.01$, and $(r, \tau_x, \tau_y) = (0.1, 1, 1)$, with an alternative parameter specification $(r', \tau_x', \tau_y') \approx (0.31, 2.20, 0.64)$.

Contours of Typical Error



(a) Point Error

(b) Probabilistic Error

Figure 5: Typical point and probabilistic errors as measured by the expected squared error and the expected negative logarithmic score, respectively. Typical point error does not reduce monotonically in τ_y, whereas typical probabilistic error does. Here the herding level is $r = 0.75$.

Model Validation

Table 1: Model comparison between our model and the no-herding baseline model in terms of in-sample log-likelihood (In-Sample ML) and leave-one-out cross validation (LOO ELPD).

Survey	Subset	# of Tasks	In-Sample ML: Tasks with Improvement	In-Sample ML: Likelihood Ratio	LOO ELPD: Tasks with Improvement	LOO ELPD: Likelihood Ratio
ECB	All	471	58.4%	2.36	45.0%	1.66
	$\hat{r} > 0.1$	350	70.6%	2.89	57.4%	1.98
	$\hat{r} > 0.2$	227	80.6%	3.80	71.4%	2.51
	$\hat{r} > 0.3$	134	85.8%	5.13	79.1%	3.03
	$\hat{r} > 0.4$	56	87.5%	7.35	82.1%	3.47
FED	All	232	67.7%	3.80	56.5%	2.47
	$\hat{r} > 0.1$	167	81.4%	4.86	71.3%	3.17
	$\hat{r} > 0.2$	82	87.8%	7.34	86.6%	4.92
	$\hat{r} > 0.3$	31	90.3%	8.10	90.3%	6.40
	$\hat{r} > 0.4$	9	88.9%	9.38	77.8%	6.60

Model Connection

A symmetric equilibrium forecast satisfies

$$\pi_k = (1-r)\hat{\pi}_k + r\frac{1}{K-1}\sum_{j\neq k}\mathbb{E}\left(\pi_j|x_k,y\right). \tag{8}$$

Morris and Shin (2002) solution is

$$\mu_k^{MS} = (1-r)\hat{\theta}_k + r\frac{1}{K-1}\sum_{j\neq k}\mathbb{E}\left(\mu_j^{MS}|x_k,y\right), \tag{5}$$

where in general $\hat{\theta}_k = \mathbb{E}\left(\theta|x_k,y\right)$.