Simulating the Survey of Professional Forecasters

ECB: The Transformative Power of AI

Anne L. Hansen[1], **John J. Horton**[2], **Sophia Kazinnik**[3], **Daniela Puzzello**[4], **Ali Zarifhonarvard**[4]

[1]Federal Reserve Bank of Richmond

[2]MIT Sloan School of Management

[3]Stanford HAI

[4]Indiana University Bloomington

April 2nd, 2025

LLMs as Approximations of Humans

Growing body of literature shows that LLMs produce responses consistent with both economic theory and documented patterns of human behavior:

- behavioral econ experiments (Horton; 2023)
- consumer choice surveys (Brand, Israeli, and Ngwe; 2023)
- surveys on political biases (Argyle et al.; 2023)

Additionally, LLMs:

- can align with their Big Five assigned personality profiles (Jiang, Zhang, Cao, and Kabbara; 2023)
- and exhibit personality consistency (Frisch and Giulianelli; 2024)

LLMs are *human enough*.

Motivation

- Survey-based forecasts (e.g., SPF) are critical for policymakers and researchers (e.g., SPF is used by central banks, academia, practitioners)
- Survey data collection is costly (& infrequent); can't easily adapt questions
- LLMs can augment suvey data collection by simulating agent behavior (quickly and cheaply)

The Paper

Goal: Construct LLM-based synthetic forecasters mimicking the *Survey of Professional Forecasters* (SPF) participants

1. Build synthetic forecasters *at the individual level*, based on information of actual SPF participants
2. Use these synthetic personas, past median SPF forecasts, and real-time data as LLM inputs
3. Ask for point forecasts similar to the SPF instrument
4. Compare accuracy of LLM-based forecasts to SPF forecasts

Agenda

(1) A framework of Human and AI Forecasting

(2) Survey of Professional Forecasters

(3) Simulating the SPF with LLMs

(4) Results

Framework

- Forecasting process:

$$y_{t+H} = f(x_t, z_t) + \varepsilon_{t+H}$$

with t as current time period, H as forecast horizon, x_t as observable predictors, z_t as unobservable, and ε_t unpredictable with zero mean

- Unobservables z_t represent any additional information that can help predict y_{t+H} but is (very) hard to quantify, e.g.:
  - Private insights
  - Tacit domain knowledge
  - Internalized heuristics
  - Intuition

Humans, Algorithms, and AI

- **Humans** can access both x_t and z_t, but do so imperfectly:

$$h_{i,t} = f(x_t, z_t) + \Delta_{i,t}$$

  - $\Delta_{i,t}$ is human bias that may not have zero mean

- **Traditional algorithms** cannot access z_t but they process x_t efficiently (*direct mapping*):

$$m_t = \mathbb{E}\left[f(x_t, z_t) \mid x_t\right]$$

- **LLMs** are similar to traditional algorithms in that they only access x_t, but expectations are formed differently (*massive text-based probability distribution*):

$$m_t^{\text{AI}} = \mathbb{E}^{\text{AI}}\left[f(x_t, z_t) \mid x_t\right]$$

Humans vs. AI

- The distance between human and AI forecasts ultimately depends on the size of human bias ($\Delta_{i,t}$) relative to LLM's ($\Delta_t^{AI} = m_t^{AI} - f(x_t, z_t)$)

- We can minimize this distance by giving an LLM:

  (1) **Forecaster characteristics** to capture systematic patterns in biases:

  $$\Delta_{i,t} = \gamma(w_{i,t}) + e_{i,t},$$

  (2) **Past median SPF forecasts** to proxy unobservable z_t:

  $$\bar{h}_{t-1} = f(x_{t-1}, z_{t-1}) + \bar{\Delta}_{t-1}$$

- This helps LLMs mimic humans in their forecasting process:

$$m_{i,t}^{AI} = \mathbb{E}^{AI}\left[f(x_t, z_t) \mid x_t,\ f(x_{t-1}, z_{t-1}) + \bar{\Delta}_{t-1},\ w_{i,t}\right]$$

The Survey of Professional Forecasters

About the SPF

- Oldest quarterly survey of macroeconomic expectations in the U.S.
  - Launched in 1968
  - Conducted by the Federal Reserve Bank of Philadelphia since 1990
- Widely used by policy-makers and economic researchers
- Survey questions:
  - 23 point forecasts at nine horizons: the current quarter (nowcast), one to four quarters ahead, the current year, and one to three years ahead
- Survey responses are releases at the individual level, but without forecaster identifiers. However, published surveys include the names and affiliations of recent contributors ▸ Example of Acknowledgments

Data

- We focus on all point forecast variables:
  - U.S. business indicators (e.g., Nominal GDP; Unemployment Rate; T-Bill Rate, 3-month)
  - Real GDP and its components (e.g., Real GDP, Real Personal Consumption Expenditures)
  - Inflation measures (CPI, Core CPI, PCE, Core PCE)
- We forecast over five horizons: nowcast + one to four quarters ahead
- Sample: 1999-2023 + an out-of-sample validation for 2024

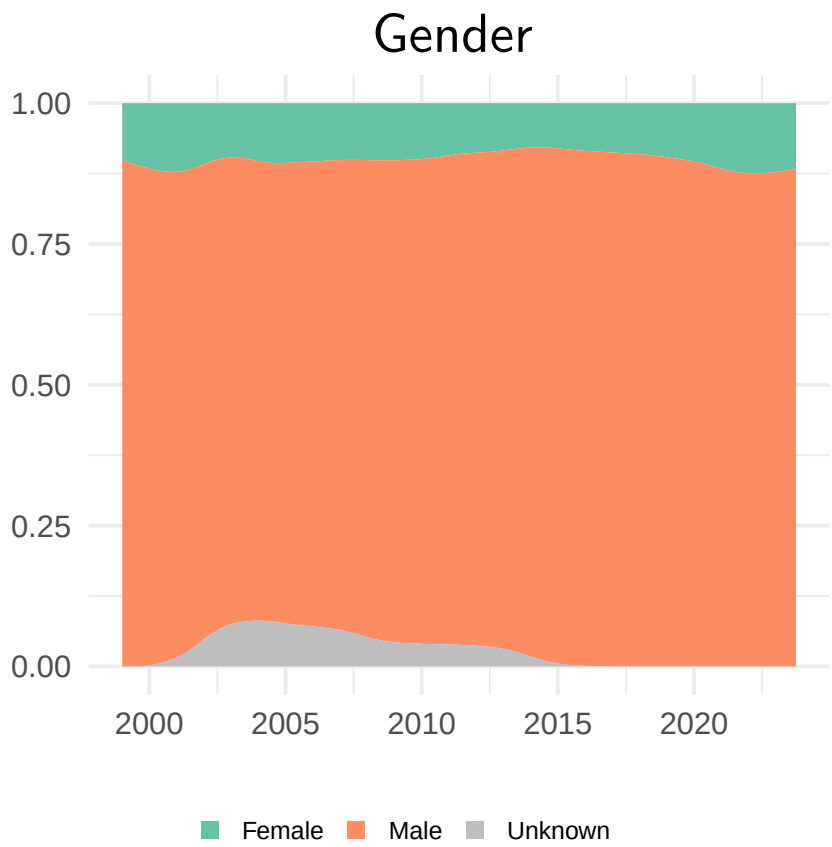
Simulating the SPF with LLMs

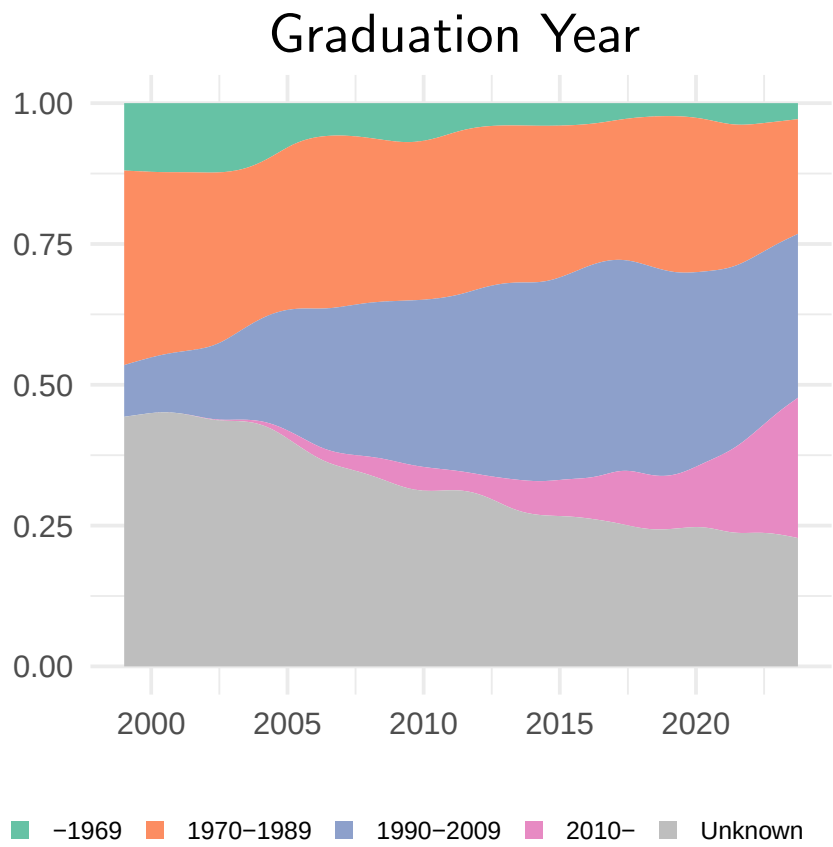
Synthetic Forecasters

We collect publicly available data (e.g., LinkedIn, personal websites) to:

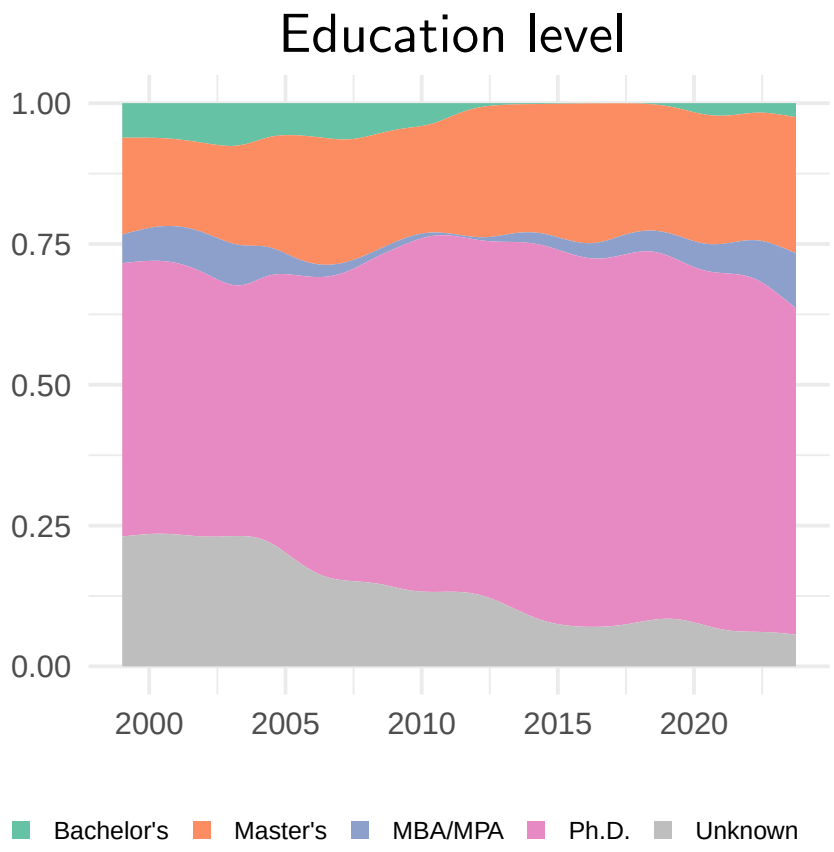
- Create a set of **synthetic forecasters** by endowing them with:
  - Education, job title, affiliation, company location
  - Experience and possible geographic or sector biases
  - Social media presence, interviews, etc.
- These features vary widely across actual SPF participants individuals

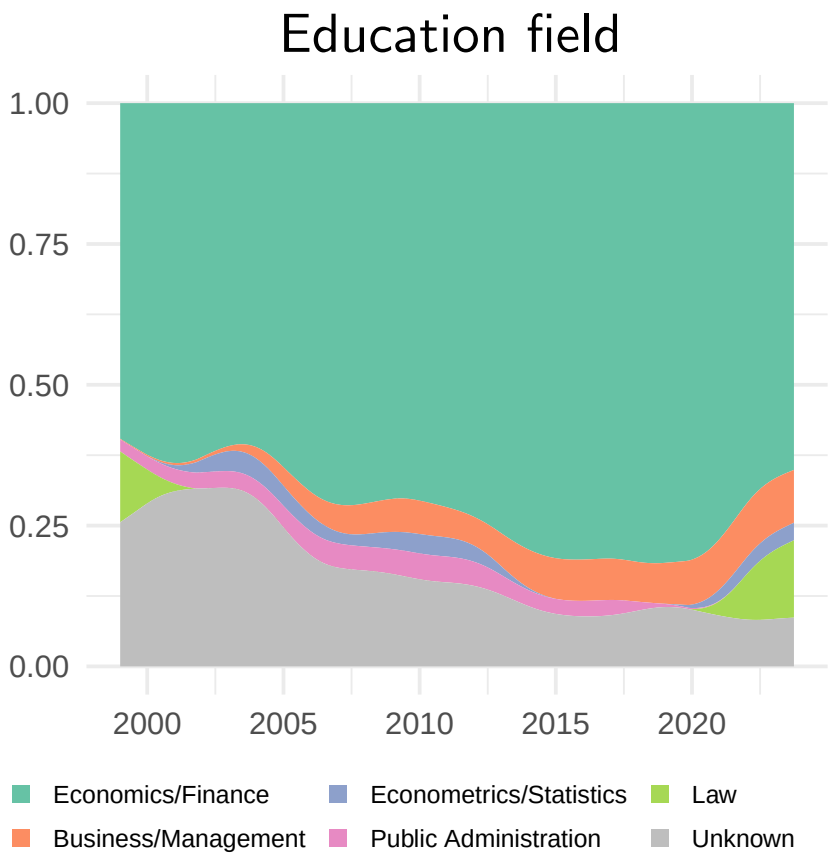
Personal Characteristics I



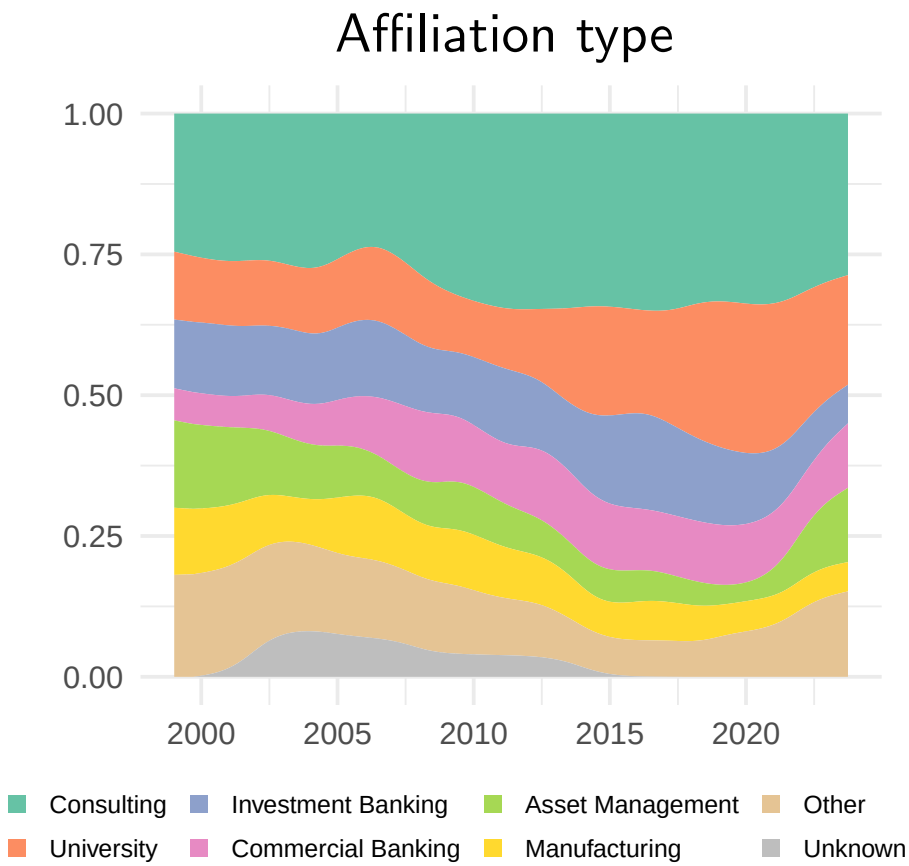


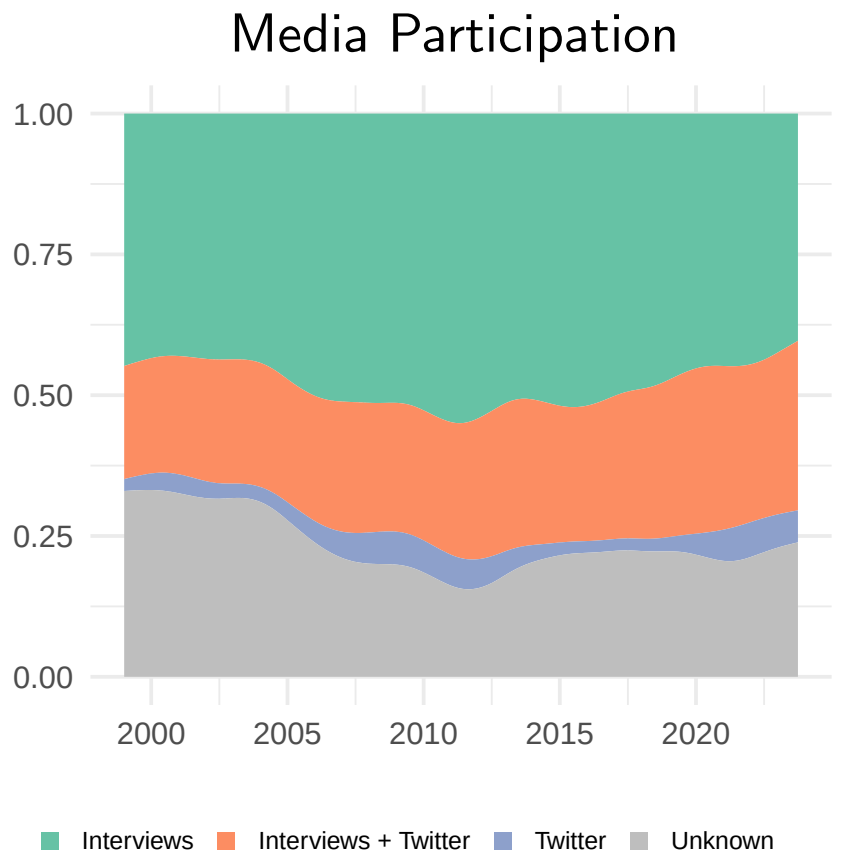
Personal Characteristics II





Personal Characteristics III





Method

1. We use a set of LLMs (e.g., GPT-4o mini) and prompt them with:
   - **Synthetic forecaster personas** (i)
   - **Real-time data** (up to quarter t)
   - **Past SPF median forecasts**

$$m_{i,t}^{\text{AI}} = \mathbb{E}^{\text{AI}}\left[f\left(x_t, z_t\right) \mid x_t,\ f\left(x_{t-1}, z_{t-1}\right) + \bar{\Delta}_{t-1},\ w_{i,t} \right].$$

2. The model is then instructed to forecast the same variables over the same horizons as human SPF forecasters
3. Evaluate LLM forecasts *versus* actual SPF and realized outcomes

Prompt

You are a participant on a panel of Survey of Professional Forecasters. Your name is [name], you graduated from [*alma mater*] with a [education] around [graduation year]. Today, you work as [title] at [affiliation]. It's [affiliation types] organization.
Your organization is based in [company location].
You are originally from [country of origin]. [social media status].
We are in [quarterly date]. You are about to fill out the forecast form for [quarterly date].
Using only the information available to you as of [quarterly date], please provide your best numeric forecasts for the following variables: [variables].
Do this for the following quarters: t (current quarter), t+1, t+2, t+3, and t+4, as well as annual forecasts for this and next year (annual averages). You have the most recent real-time data on key macroeconomics variables available to you as of today: [real-time data].
The forecasts made by the SPF panel during the previous quarter were as follows (for t-1, t, t+1, t+2, t+3, t+4; where t is previous quarter: [past median forecasts].
Do not incorporate any data that was not available to you beyond the current date in your forecasts. Do consider all relevant information on the broad economic conditions and current Federal Reserve actions (up to, but not beyond [release date]).
Use available information, and your professional judgment and experience. Your forecast is anonymous. Provide the forecasts as a sequence of numerical values only. Please only provide your forecasts in the format: (t, t+1, t+2, t+3, t+4, this year's average, next year's average).

Prompt

You are a participant on a panel of Survey of Professional Forecasters. Your name is [name], you graduated from [*alma mater*] with a [education] around [graduation year]. Today, you work as [title] at [affiliation]. It's a [affiliation types] organization. Your organization is based in [company location].
You are originally from [country of origin]. [social media status].

Prompt

You are a participant on a panel of Survey of Professional Forecasters...

We are in [quarterly date]. You are about to fill out the forecast form for [quarterly date]. Using only the information available to you as of [quarterly date], please provide your best numeric forecasts for the following variables: [variables].

Prompt

You are a participant on a panel of Survey of Professional Forecasters...
We are in [quarterly date]...

Do this for the following quarters: t (current quarter), t+1, t+2, t+3, and t+4, as well as annual forecasts for this and next year (annual averages). You have the most recent real-time data on key macroeconomics variables available to you as of today: [real-time data].

Prompt

You are a participant on a panel of Survey of Professional Forecasters...
We are in [quarterly date]...
Do this for the following quarters...

The forecasts made by the SPF panel during the previous quarter were as follows (for t-1, t, t+1, t+2, t+3, t+4; where t is previous quarter): [past median forecasts].

Prompt

You are a participant on a panel of Survey of Professional Forecasters...
We are in [quarterly date]...
Do this for the following quarters...
The forecasts made by the SPF panel during the previous quarter...

Do not incorporate any data that was not available to you beyond the current date in your forecasts. Do consider all relevant information on the broad economic conditions and current Federal Reserve actions (up to, but not beyond [survey release date]).

Prompt

You are a participant on a panel of Survey of Professional Forecasters...
We are in [quarterly date]...
Do this for the following quarters...
The forecasts made by the SPF panel during the previous quarter...
Do not incorporate any data that was not available...

Use available information, and your professional judgment and experience. Your forecast is anonymous. Provide the forecasts as a sequence of numerical values only. Please only provide your forecasts in the format: (t, t+1, t+2, t+3, t+4, this year's average, next year's average).

Prompt

You are a participant on a panel of Survey of Professional Forecasters. Your name is [name], you graduated from [*alma mater*] with a [education] around [graduation year]. Today, you work as [title] at [affiliation]. It's [affiliation types] organization.
Your organization is based in [company location].
You are originally from [country of origin]. [social media status].
We are in [quarterly date]. You are about to fill out the forecast form for [quarterly date].
Using only the information available to you as of [quarterly date], please provide your best numeric forecasts for the following variables: [variables].
Do this for the following quarters: t (current quarter), t+1, t+2, t+3, and t+4, as well as annual forecasts for this and next year (annual averages). You have the most recent real-time data on key macroeconomics variables available to you as of today: [real-time data].
The forecasts made by the SPF panel during the previous quarter were as follows (for t-1, t, t+1, t+2, t+3, t+4; where t is previous quarter: [past median forecasts].
Do not incorporate any data that was not available to you beyond the current date in your forecasts. Do consider all relevant information on the broad economic conditions and current Federal Reserve actions (up to, but not beyond [release date]).
Use available information, and your professional judgment and experience. Your forecast is anonymous. Provide the forecasts as a sequence of numerical values only. Please only provide your forecasts in the format: (t, t+1, t+2, t+3, t+4, this year's average, next year's average).

Results

Results

- Large data set comprising point forecasts for 20+ variables at different horizons for both human and AI forecasters
- Focus here is on most relevant policy variables:
  - CPI inflation rate
  - Real GDP
  - Unemployment rate
  - 3-month Treasury bill rate

Results

- Large data set comprising point forecasts for 20+ variables at different horizons for both human and AI forecasters
- Focus here is on most relevant policy variables:
  - CPI inflation rate
  - Real GDP
  - Unemployment rate
  - 3-month Treasury bill rate
- **Three main take-aways:**

  #1 AI $\approx$ humans: While AI and human forecasts are qualitatively similar, there are quantitative differences

  #2 AI $\succ$ humans: AI often achieves lower forecasting errors

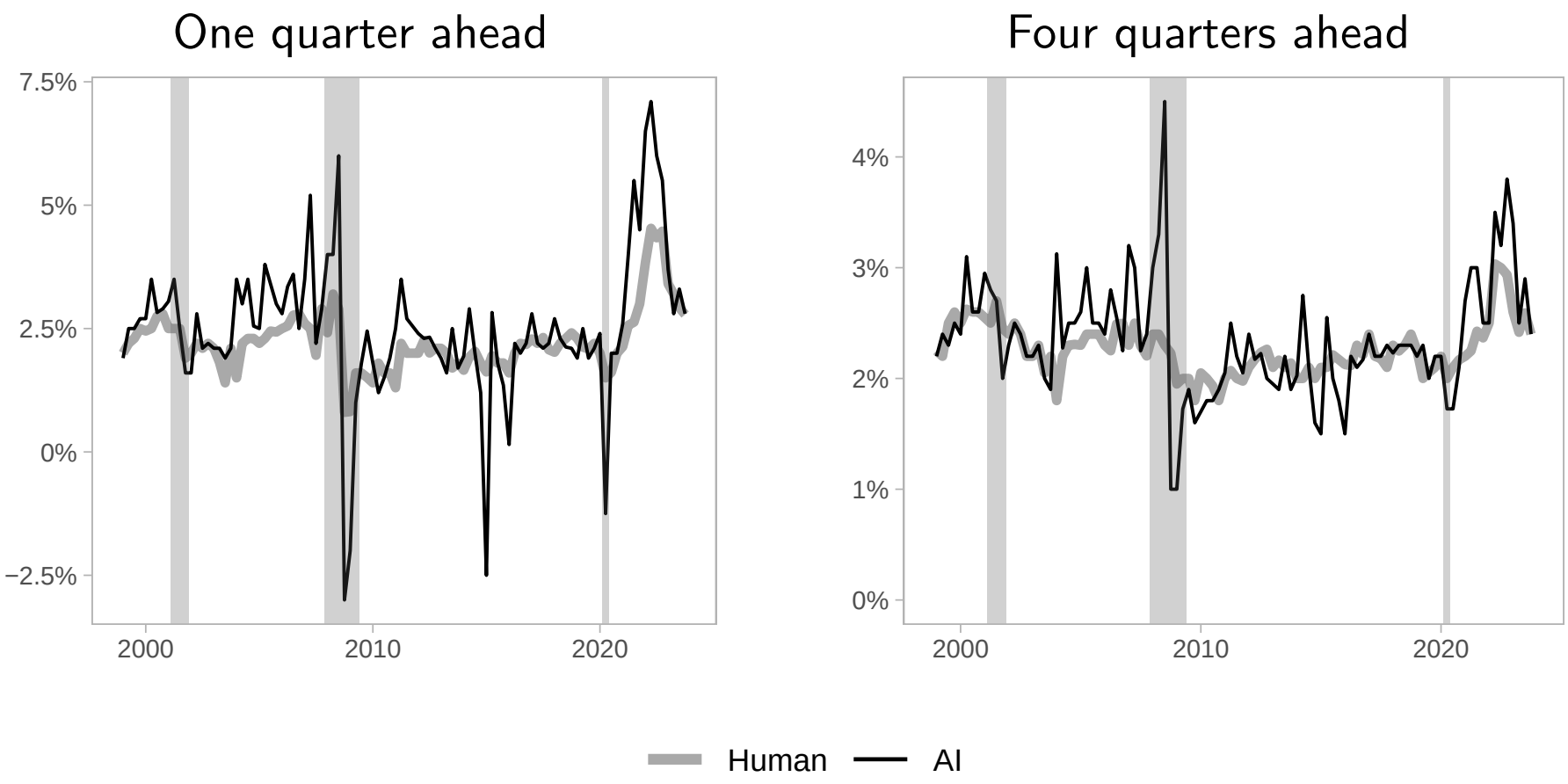
  #3 AI $\succ$ humans $|$ human input: Accuracy of AI hinges on human input in prompt

Result #1: AI $\approx$ humans

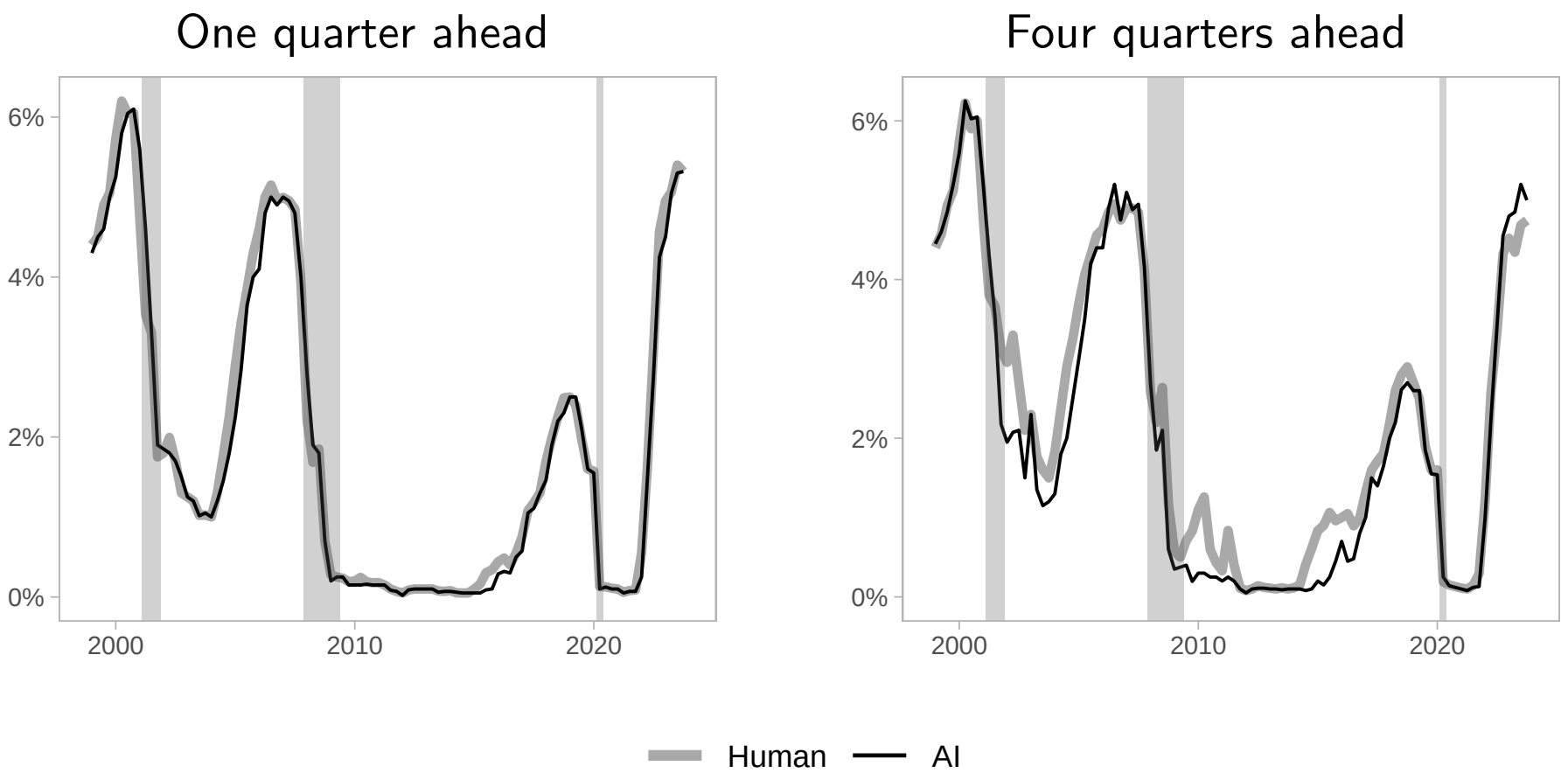
Forecast Distributions

- Median AI and human forecasts often align quite closely
- Significant distributional differences in tails and skewness
- At times, AI is more volatile or more "reactive" to data changes than human forecasts

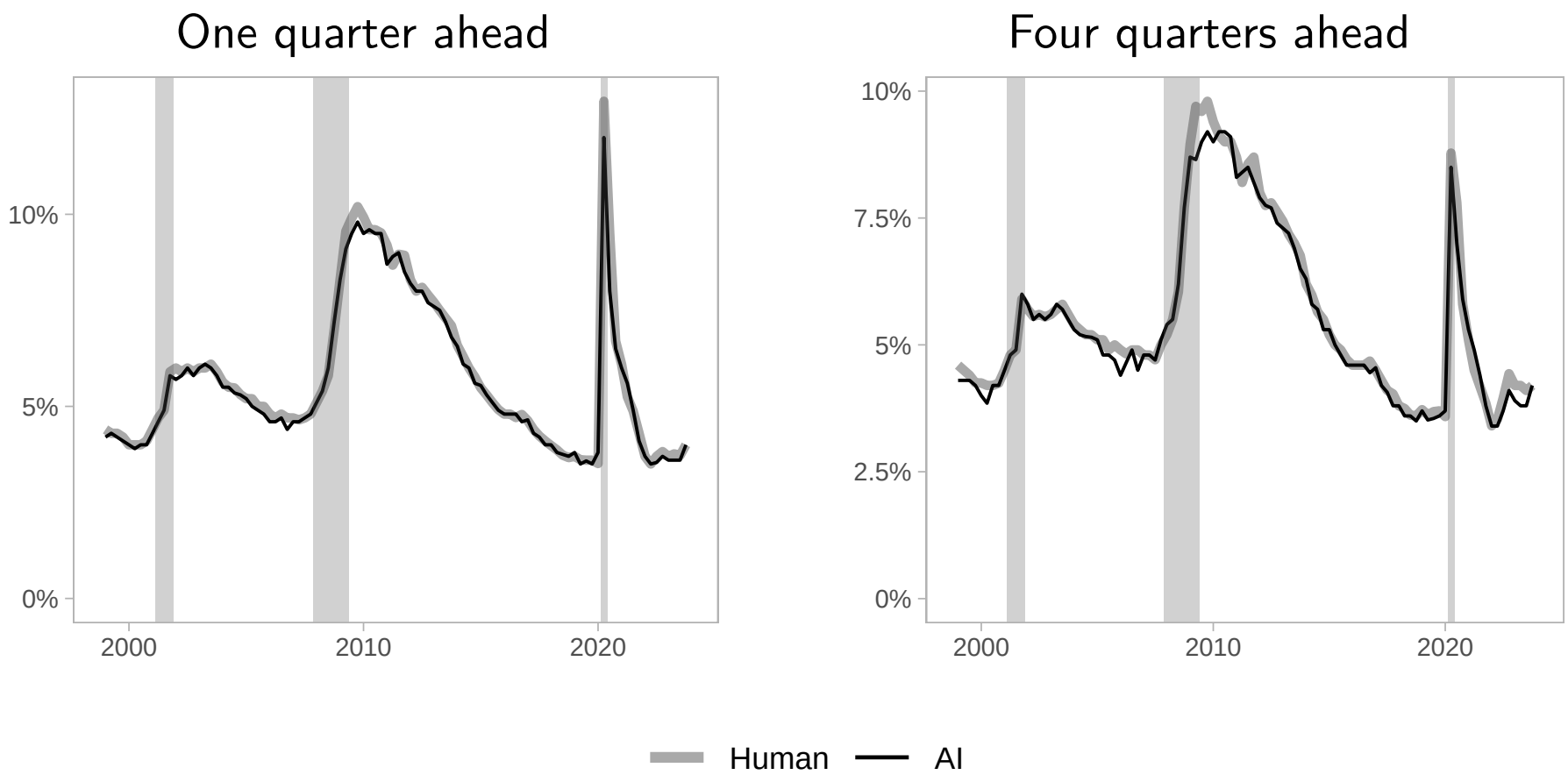
Median Forecasts: CPI Inflation



Median Forecasts: T-Bill Rate (3-month)

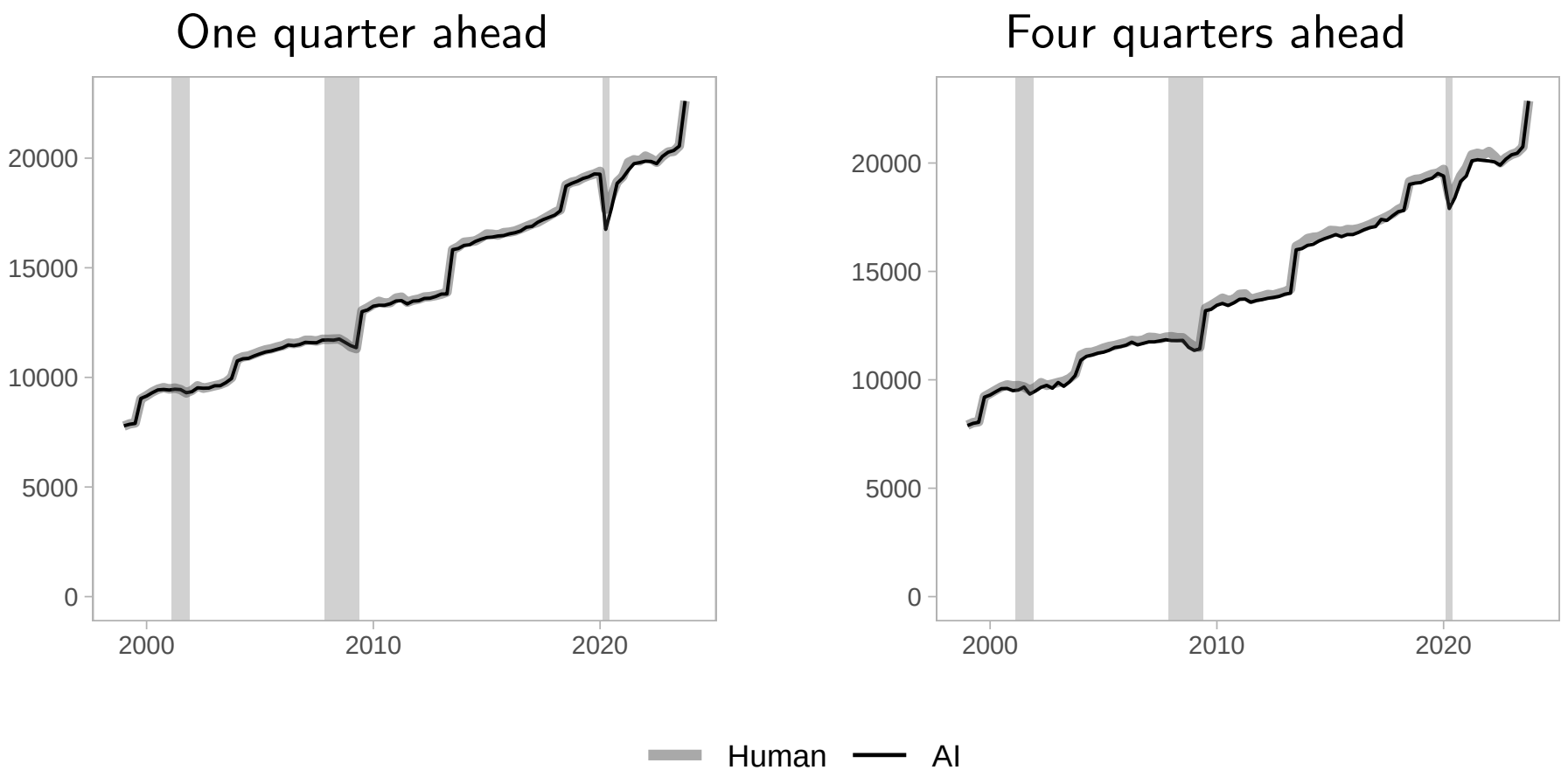


Median Forecasts: Unemployment



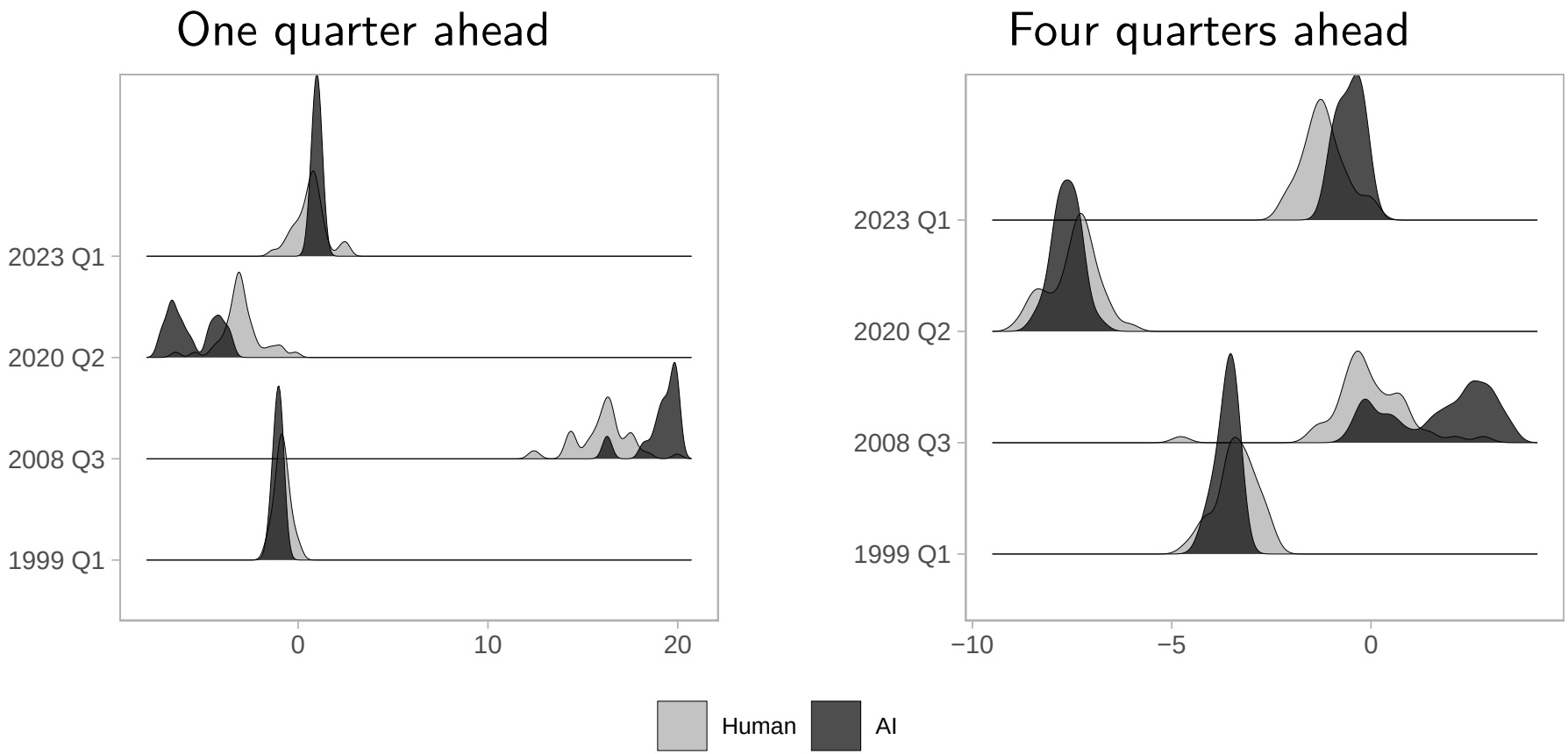
Shaded areas are NBER recessions

Median Forecasts: Real GDP



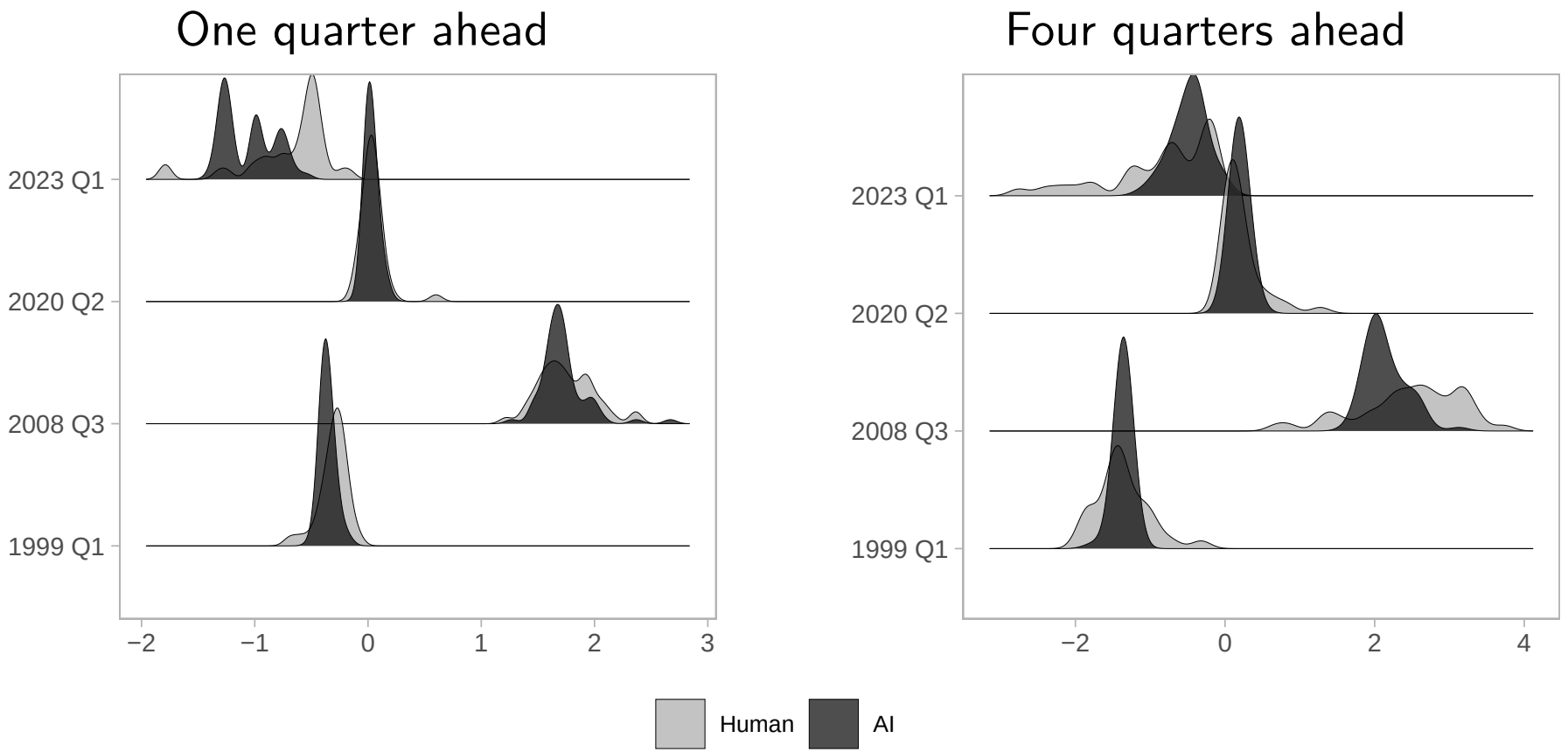
Shaded areas are NBER recessions

Densities of Individual Forecasts: CPI Inflation



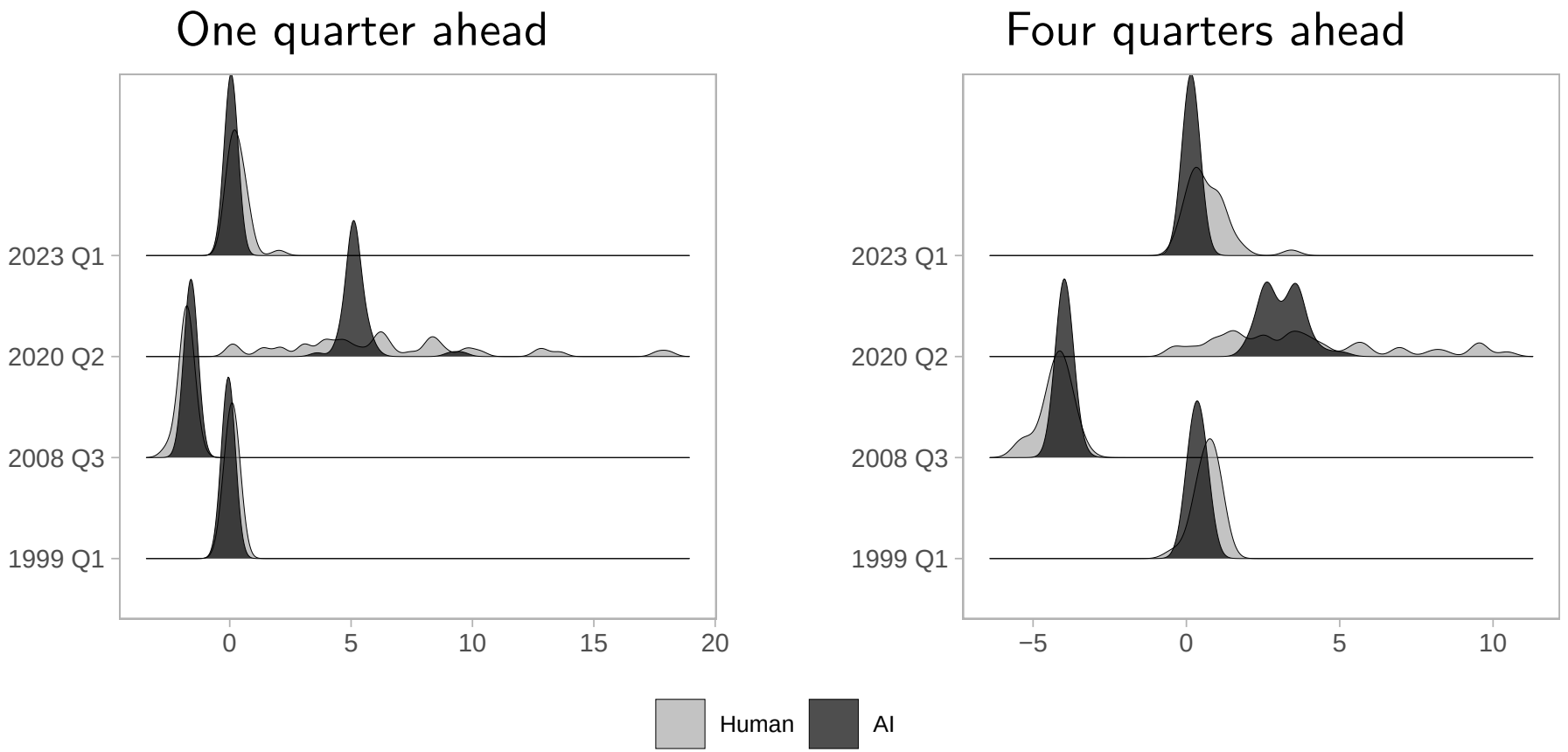
Densities of forecast errors (forecast - realized value). 1999 Q1: Earliest survey; 2008 Q3: GFC; 2020 Q2: COVID-19; 2023 Q1: Latest survey with four-quarter-ahead realization.

Densities of Individual Forecasts: T-Bill Rate (3-month)



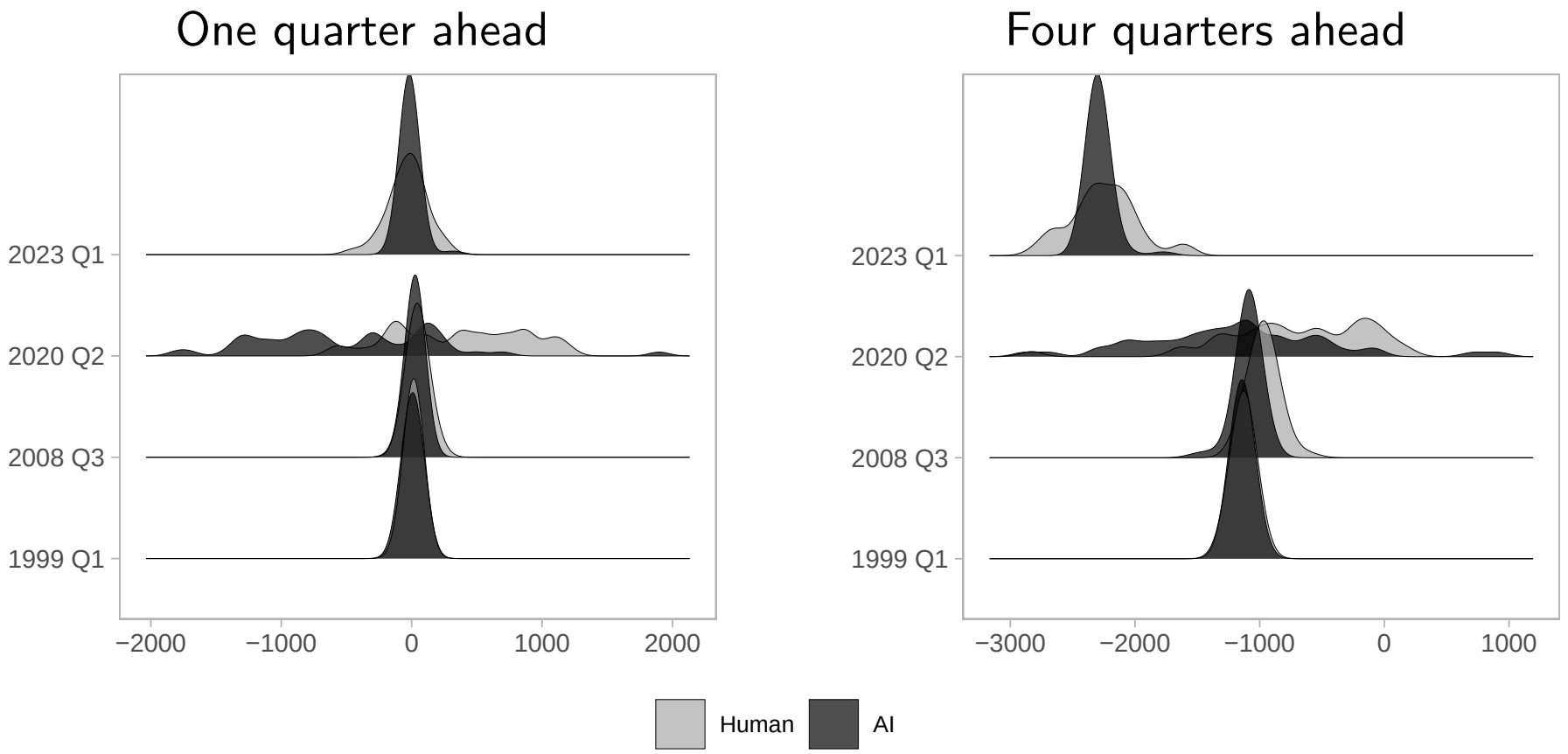
Densities of forecast errors (forecast - realized value). 1999 Q1: Earliest survey; 2008 Q3: GFC; 2020 Q2: COVID-19; 2023 Q1: Latest survey with four-quarter-ahead realization.

Densities of Individual Forecasts: Unemployment



Densities of forecast errors (forecast - realized value). 1999 Q1: Earliest survey; 2008 Q3: GFC; 2020 Q2: COVID-19; 2023 Q1: Latest survey with four-quarter-ahead realization.

Densities of Individual Forecasts: Real GDP



Densities of forecast errors (forecast - realized value). 1999 Q1: Earliest survey; 2008 Q3: GFC; 2020 Q2: COVID-19; 2023 Q1: Latest survey with four-quarter-ahead realization.

Result #2: AI $\succ$ humans

Forecast Accuracy (MAE)

- AI forecasts often outperform human forecasts, especially at longer horizons
- Gains are most pronounced for variables like real GDP and unemployment rate
- Including past SPF data is essential for strong performance (otherwise forecasts degrade)

"LLMs extract latent (z_t) information from human forecasts while also processing x_t more effectively."

Proportion of Quarters Where AI is More Accurate

Horizon (quarters)	0		1		4	
	Pct	P-val	Pct	P-val	Pct	P-val
CPI Inflation Rate	**0.69**	0.01***	0.47	0.74	**0.55**	0.78
T-bill	**0.51**	1.00	0.47	0.97	**0.60**	0.08*
Unemp	**0.81**	0.00***	**0.74**	0.01***	**0.63**	0.22
Real GDP	**0.70**	0.00***	**0.75**	0.00***	**0.63**	0.01**

Boldfaced values are ≥ 0.5. P-val reports significance of randomized tests of Pct$= 0.5$.

Result #3: AI $\succ$ humans | human input

AI Forecast Accuracy without Human Input

	Generic		Generic, w/o real-time data		Generic, w/o real-time data, w/o past SPF data	
Horzion	0	4	0	4	0	4
T-bill	**1.09**	**1.01**	0.74	**1.03**	**1.07*****	**1.08*****
Unemp	**1.02**	**1.02*****	**1.20**	**1.02**	**1.12**	**1.10*****
Real GDP	**1.15**	**1.04**	**1.37**	**1.08**	**7.57*****	**1.53*****
CPI	0.90	**1.02**	**1.09**	**1.02**	**1.09**	**1.13****
Average	**1.14**	**1.06**	**1.31**	**1.06**	**8.88**	**2.52**

Values are MAEs relative to MAEs of baseline AI forecasts. Boldfaced values are ≥ 1.
P-val reports significance of randomized tests of Pct$= 1$.

Addressing Temporal Leakage

- LLM might recall future data from its training set
- Mitigation:
  - Strict instructions to use only data *up to t*
  - Real-time "dated" data sets (no future info)
  - Out-of-sample test (e.g., 2024 data) outside model's training window
- Recall test: Ask the model to recall past realized values from the data set. On average, errors are 16x larger than our baseline nowcasting results.

Discussion

- Humans have access to unobservable insights but can suffer systematic biases
- LLMs see only structured data and historical patterns, but can approximate the "latent" aspects by:
  - reading past human forecasts,
  - adjusting to persona-specific biases
- Hybrid approach: AI + human signals can exceed pure human or purely data-driven ML forecasts
- Demonstrates the viability of AI-augmented macroeconomic surveys, potentially powerful for policy or research: "virtual forecasting lab"

Tools

- `Expected Parrot`: a set of tools for running experiments with many AI agents and models at once

Thank you!

Feedback is appreciated:

kazinnik [at] stanford.edu

Appendix

The Federal Reserve Bank of Philadelphia thanks the following forecasters for their participation in recent surveys:

Lewis Alexander, Nomura Securities; **Scott Anderson**, Bank of the West (BNP Paribas Group); **Robert J. Barbera**, Johns Hopkins University Center for Financial Economics; **Peter Bernstein**, RCF Economic and Financial Consulting, Inc.; **Wayne Best** and **Michael Brown**, Visa, Inc.; **Jay Bryson**, Wells Fargo; **J. Burton**, **G. Ehrlich**, **D. Manaenkov**, and **T. Ranoso**, RSQE, University of Michigan; **Christine Chmura, Ph.D.**, and **Xiaobing Shuai, Ph.D.**, Chmura Economics & Analytics; **Gary Ciminero**, **CFA**, GLC Financial Economics; **Gregory Daco**, Oxford Economics USA, Inc.; **Rajeev Dhawan**, Georgia State University; **Bill Diviney**, ABN AMRO Bank NV; **Michael R. Englund**, Action Economics, LLC; **Sacha Gelfer**, Bentley University; **James Glassman**, JPMorgan Chase & Co.; **Jan Hatzius**, Goldman Sachs; **Brian Higginbotham**, U.S. Chamber of Commerce; **Fred Joutz**, Benchmark Forecasts; **Sam Kahan**, Kahan Consulting Ltd. (ACT Research LLC); **N. Karp**, BBVA Research USA; **Walter Kemmsies** and **Ryan Severino**, Jones Lang LaSalle; **Jack Kleinhenz**, Kleinhenz & Associates, Inc.; **Rohan Kumar**, Decision Economics, Inc.; **Thomas Lam**, Sim Kee Boon Institute, Singapore Management University; **John Lonski**, Moody's Capital Markets Group; **Matthew Luzzetti**, Deutsche Bank Securities; **IHS Markit**; **Robert McNab**, Old Dominion University; **R. Anthony Metz**, Pareto Optimal Economics; **R. M. Monaco**, TitanRM; **Michael Moran**, Daiwa Capital Markets America; **Joel L. Naroff**, Naroff Economic Advisors; **Brendon Ogmundson**, BC Real Estate Association; **Perc Pineda, Ph.D.**, Plastics Industry Association; **Philip Rothman**, East Carolina University; **Chris Rupkey**, MUFG Union Bank; **Sean M. Snaith**, **Ph.D.**, University of Central Florida; **Constantine G. Soras, Ph.D.**, CGS Economic Consulting, Inc.; **Stephen Stanley**, Amherst Pierpont Securities; **Charles Steindel**, Ramapo College of New Jersey; **Susan M. Sterne**, Economic Analysis Associates, Inc.; **James Sweeney**, Credit Suisse; **Thomas Kevin Swift**, American Chemistry Council; **Maira Trimble**, Eaton Corporation; **Gary Wagner**, University of Louisiana at Lafayette; **Mark Zandi**, Moody's Analytics; **Ellen Zentner**, Morgan Stanley.

This is a partial list of participants. We also thank those who wish to remain anonymous.

▸ Return

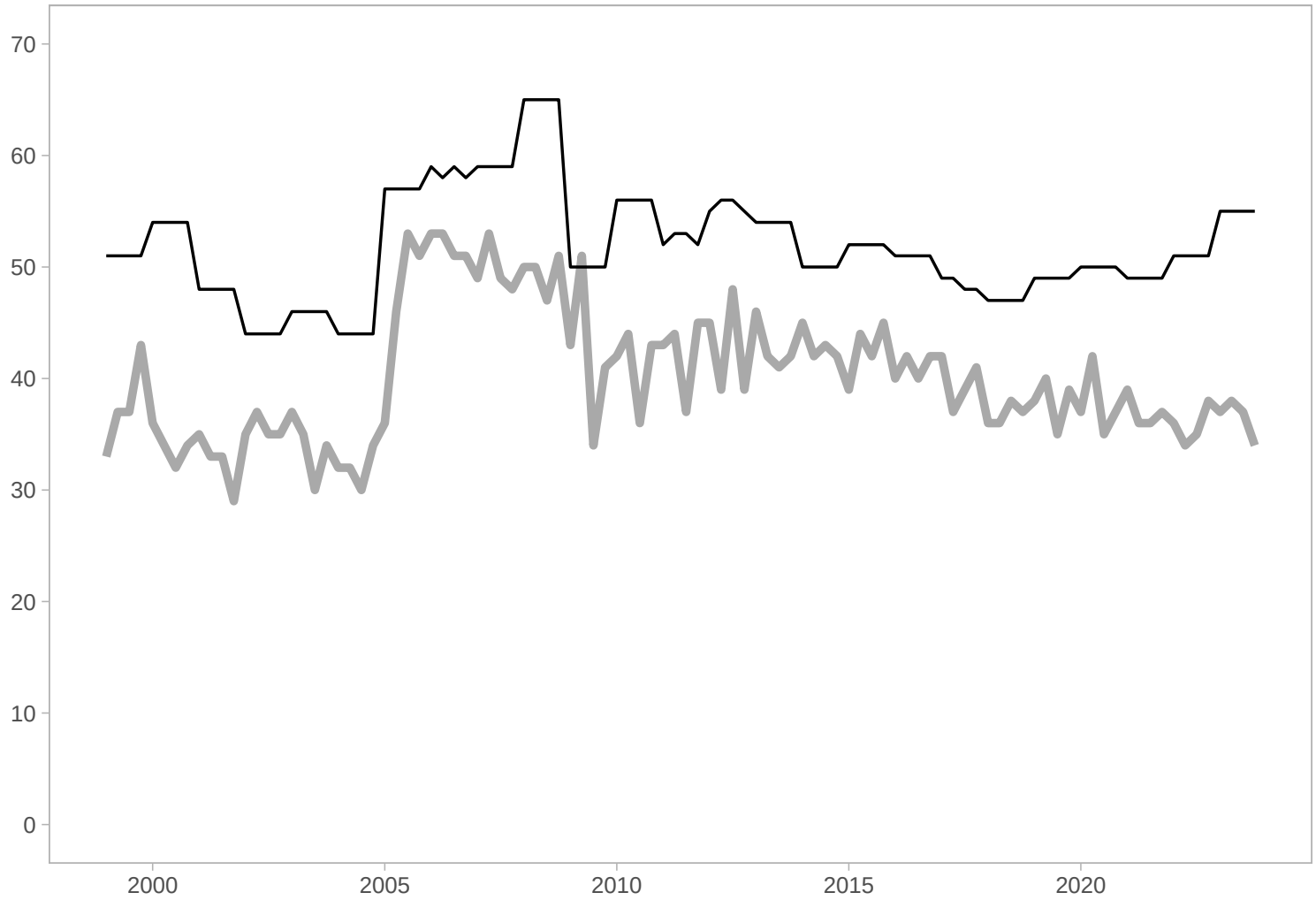
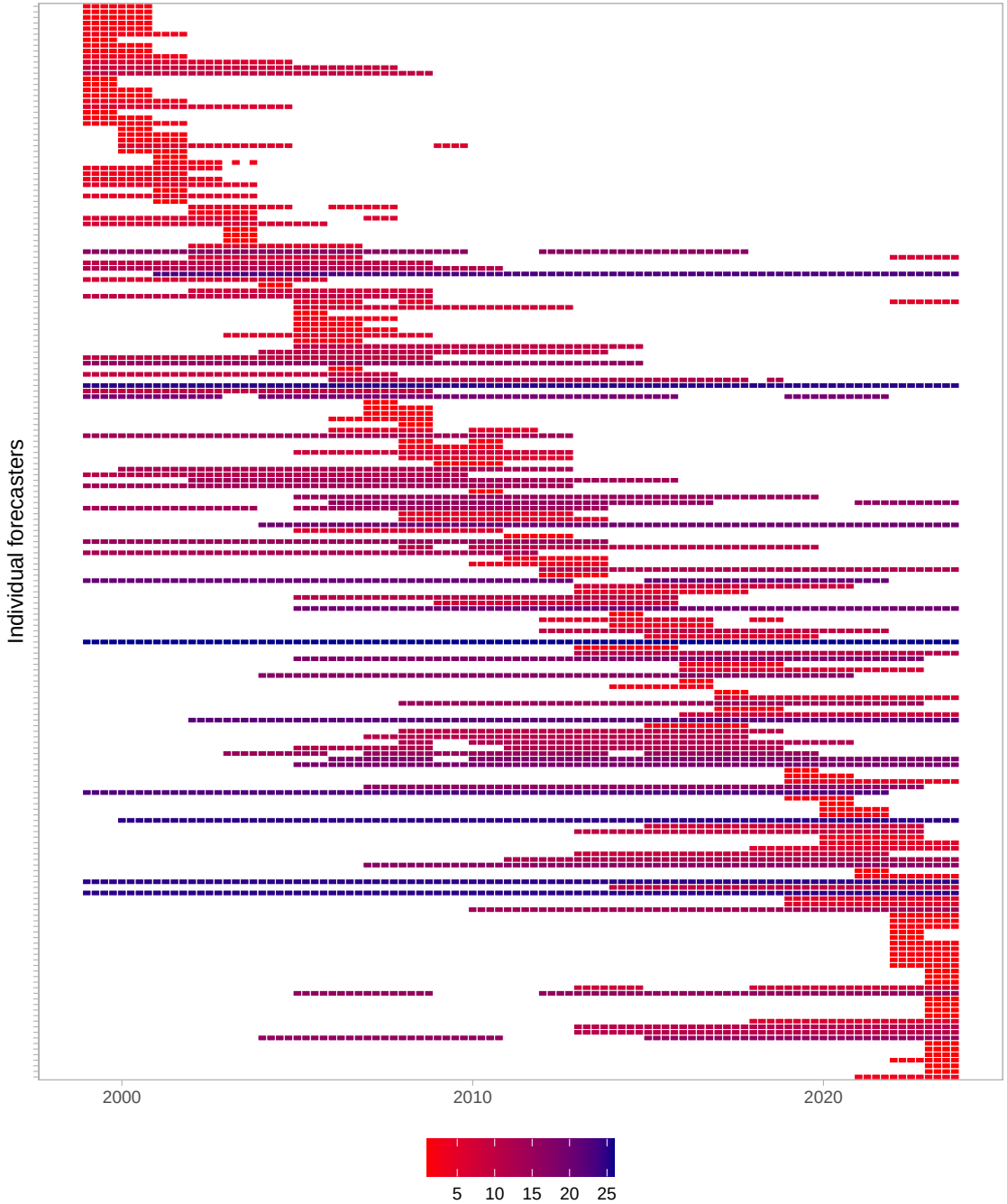


Figure: **Number of forecasters in the SPF panel over time**


Individual forecasters
2000
2010
2020
5
10
15
20
25