Hawkish by Voice

Yueh-Heng Wu · Hui-Ching Chuang · Ju-Fang Yen

Department of Statistics, National Taipei University · hcchuang@gm.ntpu.edu.tw

Holding the words fixed, the way Q&A sounds contains market-relevant information.

15-MINUTE RESEARCH TALK

Do markets react to how the FOMC sounds?

Same words

“Inflation remains elevated.”

Text-based tone is held constant

Different delivery

Human voice − same-text TTS

The difference isolates non-textual delivery signals

Voice gap = net hawkishness(human) − net hawkishness(TTS)

Each sentence is paired with itself

FOMC audio
statement + Q&A

❯

ASR alignment
sentence timestamps

❯

Text labels
FOMC-RoBERTa

❯

WavLM
voice hawkishness

❯

Same-text TTS
counterfactual voice

What is held fixed

The words, sentence boundaries, and underlying semantic content.

What the gap can include

Delivery plus speaker composition in the interactive Q&A—not only the chair.

89 press conferences yield 35,619 paired sentences

89
press conferences

35,619
clean sentences

30,671
Q&A sentences

2011–25
sample window

Held-out speech classification

0.9125
pooled AUC

0.8362
accuracy

0.8362
macro F1

Human Q&A is modestly more hawkish than TTS

Sentence level

Gap = 0.0100

t = 9.94***

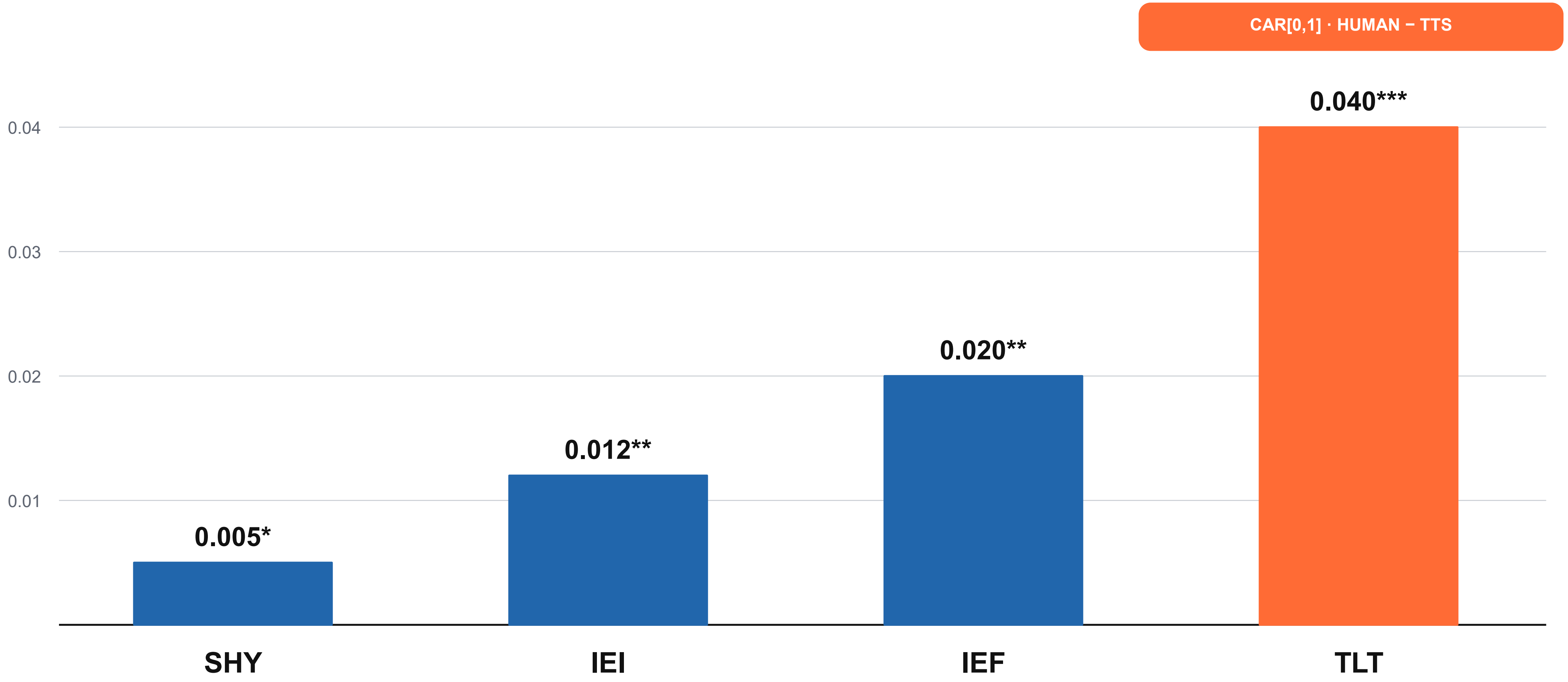
Meeting level

Gap = 0.0120

t = 2.24**

Paired Q&A estimates; ** $p<0.05$, *** $p<0.01$

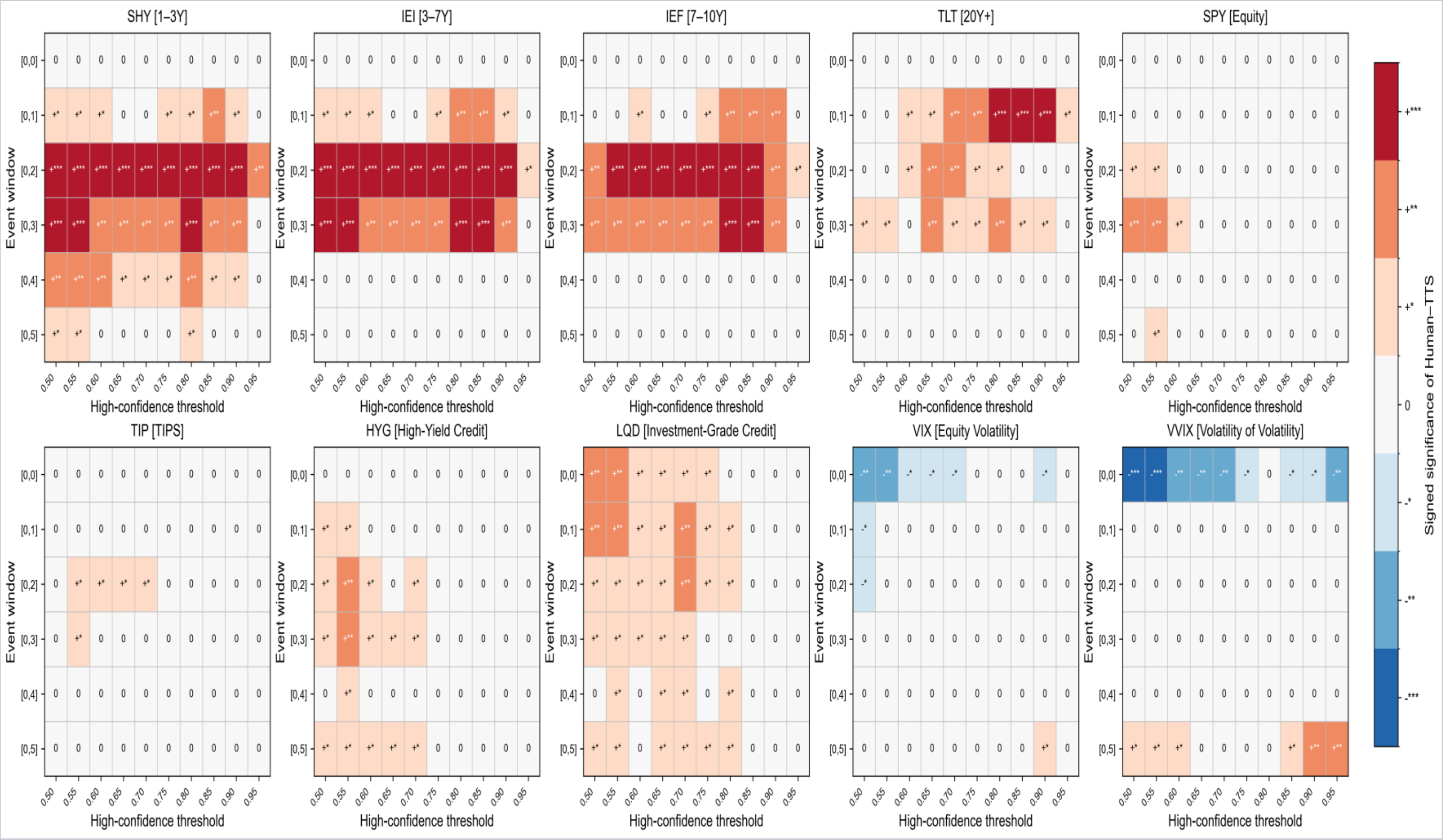
Treasury sensitivity rises with maturity



A 0.01 increase in the voice gap is associated with a larger two-day return at longer maturities.

Regression coefficients with standard controls; 89 meetings

The main effect is stable across design choices



Across thresholds

Signs persist across alternative high-confidence cutoffs.

Across windows

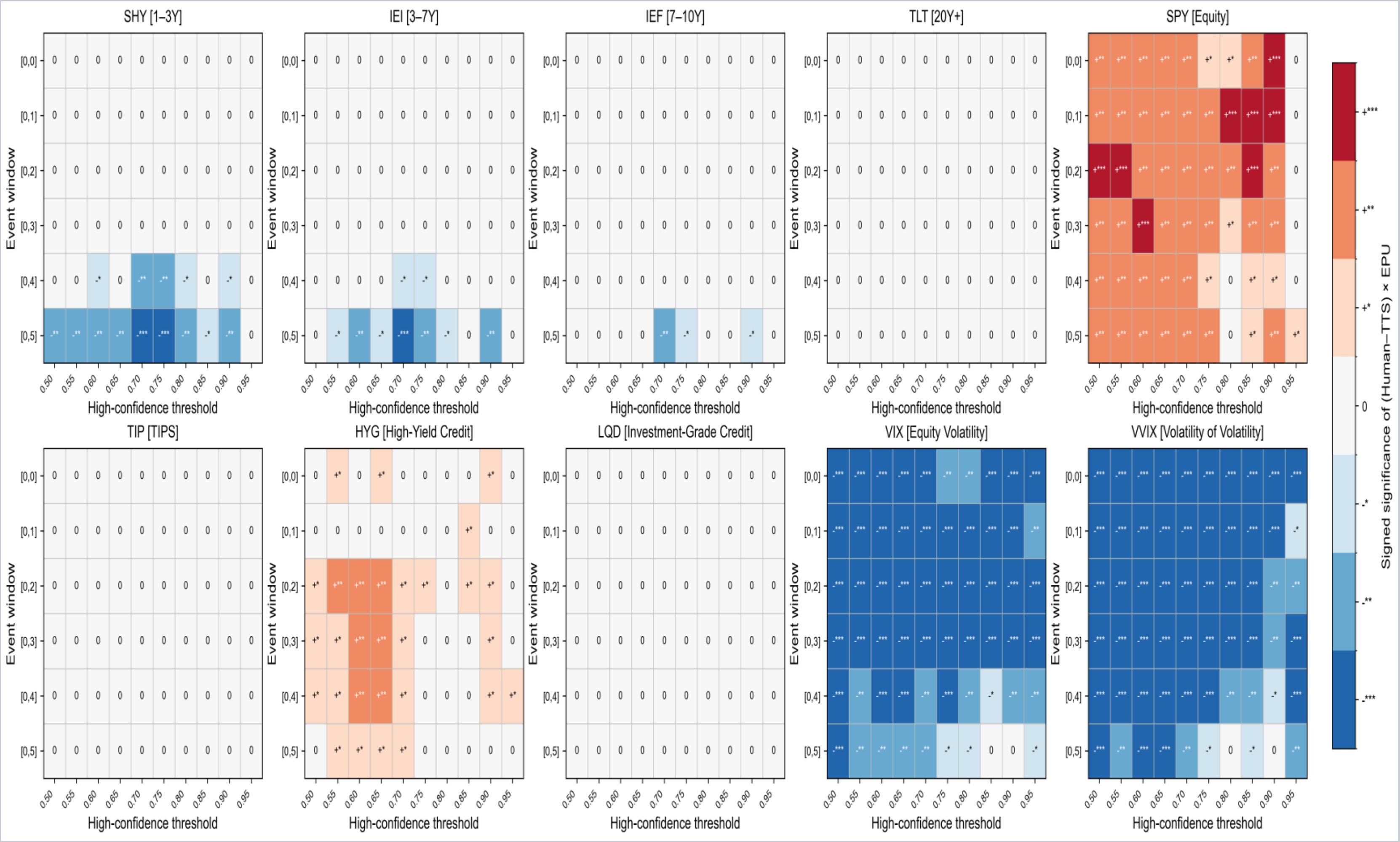
Results are visible beyond one event-window choice.

Across assets

The strongest unconditional pattern is in Treasuries.

Human–TTS coefficient heatmap across assets, thresholds, and windows

Policy uncertainty redirects the response



+0.0003

SPY × EPU

−0.058

VIX × EPU

−0.163

VVIX × EPU

Voice matters, but the estimand is broader than the chair

What we learn

Markets respond to non-textual information embedded in live FOMC Q&A, even after fixing the words.

What we do not claim

The coefficient is not a causal chair-only voice effect; reporters, moderator, and interaction matter.

Why it matters

Communication measurement should be multimodal—and evaluated against economic outcomes.

Bottom line: text and delivery are distinct channels of monetary-policy communication.