AI Workflows and Publishing in Finance

HKUST PhD Short Course

J. Anthony Cookson

University of Colorado Boulder → Penn State Smeal College of Business

April 23, 2026

Wiki reference: velikov-mihail.github.io/ai-econ-wiki

3 hours · structured discussion with live demos

By the end of this session. . .

1. Set up a **functional AI workflow** for empirical research — persistent assistant, not a chatbot
2. Use AI to **compress the gap** between a dataset and a result
3. Understand LLMs as **measurement tools**, not just productivity aids
4. Reason clearly about how AI is **changing publishing** and what that means for careers starting now

Three blocks, 3 hours

Block 1 **50 min**

From Skeptic to Power User

Motivation · Blattman framework · Jagged frontier · Setup

Block 2 **70 min**

Empty Folder to Research Result

"Live" demo · Replication packages · Stress-testing · Literature

Block 3 **25 min**

LLMs as Measurement Tools

Market's Mirror as a research application

Coda **35 min**

AI and Publishing

Editorial view · Structural changes · Career implications

Block 1: From Skeptic to Power User

What changed, and why now?

The shift is **not** that AI got better at writing.

It's that AI can now act — read files, run code, manage memory, call external services — inside a project folder that persists across sessions.

Vivid example 1

A finance paper written in four days (*"vibe research"*)

Not an endorsement — a data point about the frontier.

Vivid example 2

Chris Blattman (UChicago) builds a full AI executive assistant — email triage, project dashboards, calendar, expenses — with no prior coding experience.

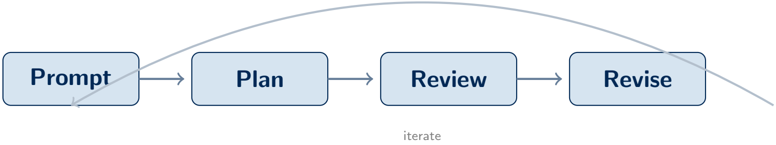
The right framing

“Become a complement to these tools,
not a substitute.”

- The goal is **not** to automate your research
- The goal is to **eliminate the friction** between your judgment and a result
- AI speeds up the mechanics; you supply the questions

The Blattman Framework

Site: claudeblattman.com — Blattman documented his entire workflow publicly as it evolved.



Prompt

Plan

Review

Revise

iterate

The key insight

Value is not in any single output. It's in *persistent context* — a `CLAUDE.md` that tells the assistant who you are, what your projects are, how you want to work.

Analogy

Like onboarding a very fast RA who never forgets what you've told them — and works 24 hours a day.

`CLAUDE.md` — persistent context

What goes in it:

- Who you are (role, domain expertise)
- Active research projects + co-authors
- How you want to communicate
- Data sources and file structure conventions
- Recurring workflows as named *skills*

What Blattman built:

- `/checkin` — daily inbox triage + calendar
- `/expenses` — reimbursement tracking
- `/editor` — editorial workload dashboard

Live demo: `/checkin` in action.

The goal: make the setup feel achievable, not magical.

Discussion

What's the most time-consuming non-research task in your week? Write it down. What would a `CLAUDE.md` entry for it look like?

The Jagged Frontier

AI is not uniformly better or worse than a human — it's jagged.

✓ Trust

- Boilerplate code
- Data cleaning
- Literature summaries (from docs you provide)
- First drafts of methods sections

🔍 Verify

- Any specific number or citation it generates
- Empirical claims
- Package documentation

✗ Don't delegate

- Identification strategy
- Economic interpretation
- Judgment about what's interesting
- Referee reports

The bottleneck is always human verification. AI speeds up drafts but creates new review work.

Three things that make AI useful vs. a toy

1. `CLAUDE.md` — persistent context file. Who you are, your projects, your preferences, your data sources. Without this, every session starts cold.
2. **Project folders** — AI works best inside a structured folder with data, code, and notes co-located. Think of it as a shared workspace.
3. **Skills / commands** — reusable workflows that encode recurring tasks. Write them once; invoke them with a single prompt.

Discussion

If you were writing a `CLAUDE.md` for yourself right now — a one-page brief for an AI assistant — what would be in it?

Block 2: From Empty Folder to Research Result (Research Scoping)

PGP's core insight

> *"AI dramatically shrinks the gap between*
> *a vague research idea and initial results."*
>
> Paul Goldsmith-Pinkham, BCF Princeton mini-series, March 2026

1. Start with a **question** and an empty folder
2. **Describe** the data you need and where to get it
3. Let AI **plan** the pipeline, then iterate
4. **Review** outputs at each step — your judgment is quality control

The researcher's job: describe the question clearly and review the output critically. The code is never the bottleneck — the questions are.

The demo dataset

The Social Signal

Cookson, Lu, Mullins & Niessner (2024), *JFE*

- 821,534 firm-day observations
- 1,500 firms × 10 years (2012–2021)
- Two variables:
    - `zee_sent_pc` — Sentiment PC1 (z-score)
    - `zee_attn_pc` — Attention PC1 (z-score)
- First principal component across StockTwits, Twitter, Seeking Alpha

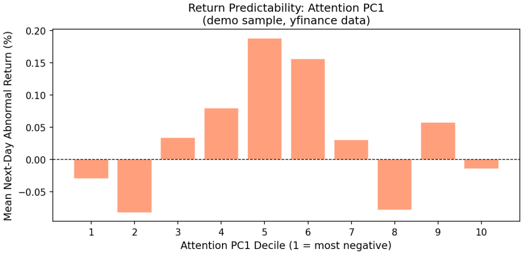
Publicly available:
data.mendeley.com/datasets/xffyybvw4j/1

The prompt that started it

"I have a firm-day social signal dataset. Help me understand what's in it and produce a time-series figure of average sentiment and attention. Then merge with Yahoo Finance returns and test whether sentiment predicts next-day abnormal returns."

Layer 1 — Firm-level basics

From one prompt: load data → describe → time-series figure → return predictability



What this figure shows

Return predictability by *attention* decile. Decile 1 = least attention; Decile 10 = most.

Higher attention → *lower* next-day abnormal return.

This is the classic overattention / contrarian pattern: when investors pile on, prices overshoot.

Sentiment shows the opposite sign (positive predictability).

Signs match Table 6 of the paper.

Layer 2 — Market-level analysis

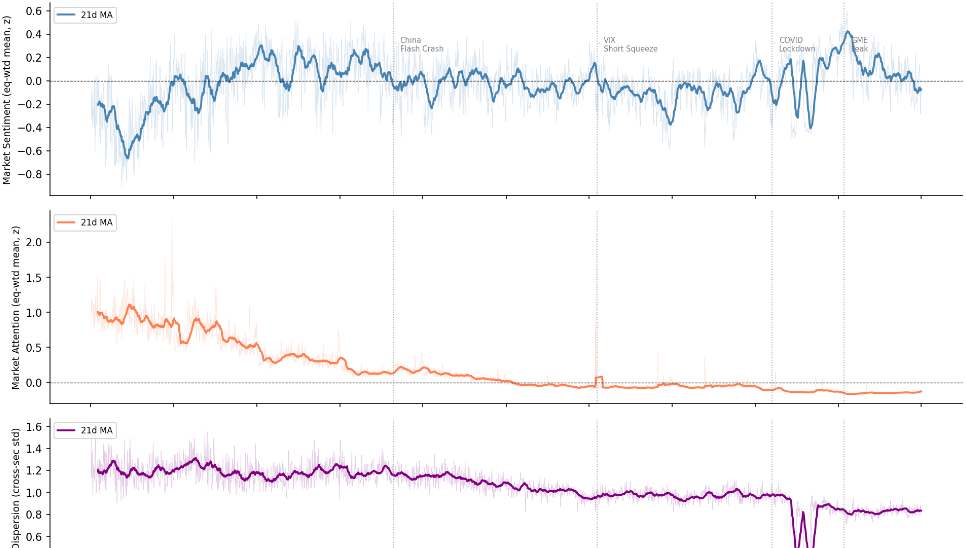
⌨ Prompt 2

"Can you do a demo of how you would produce an aggregate index and explore its market-level predictive properties for trading, volatility, and returns? Also consider the reverse relation — how do recent trading, vol and returns predict this aggregate index? Does this have anything to do with market news events like FOMC dates?"

What this produced:

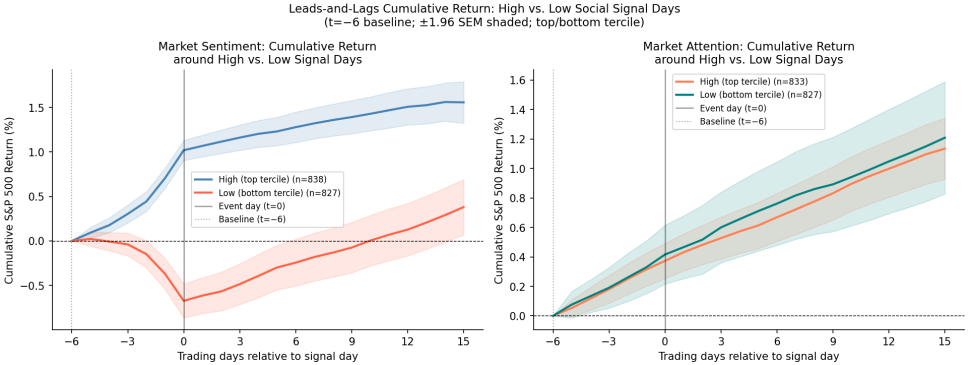
- Equal-weighted daily market sentiment & attention index
- Forward regressions: signal → next-day returns, VIX, volume
- Reverse regressions: market outcomes → signal (Newey-West HAC)
- FOMC event study (±15-day window, cumulative return)
- Leads-and-lags plot: high vs. low signal days

The market signal, 2012–2021



Market-Level Social Signal Index, 2012–2021
(aggregated from 1,500-firm social_signal_index.dta)

Leads-and-lags: what happens around high-signal days?



Event = day in top (blue) or bottom (red) tercile of signal. Cumulative S&P 500 return indexed to 0 at $t = -6$; shading = ±1.96 SEM.

Layer 3 — Cap-weighted robustness

⌨ Prompt 3

"Can you aggregate in a market value way and redo the analysis?"

Three indices compared:

Index	Coverage	Weight
EW-1500	All 1,500 firms	Equal
EW-8	8 large caps	Equal
CW-8	8 large caps	Market cap

Key finding:

EW-1500 and CW-8 sentiment correlation $r = 0.39$ — large caps diverge meaningfully from breadth-weighted sentiment.

Cap-weighting sharpens predictability ($\beta = -0.0007$, $t = -2.13^*$).

Interpretation: *what the mega-caps feel* is a different signal from *what the full market feels*.

What the demo illustrates

Researcher supplied	**Claude Code supplied**
Economic question	All Python code
Cap-weighting as a robustness check	Market cap from Yahoo Finance
FOMC event study as the right design	81 FOMC dates, event panel
“Leads and lags, not just coefficients”	Rewrote figure immediately
Benchmark date ($t = -6$) and window ($t+15$)	Cumulative indexing
Whether results make economic sense	All figures and tables

Four prompts. ∼15 minutes of real session time. The code was never the bottleneck. The questions were.

Larger pipelines: structure matters more

When data gets large or messy:

- Convert flat files to Parquet/DuckDB for fast querying PGP Episode 4
- Extract structured data from unstructured sources (EDGAR 10-K) PGP Episode 3
- **Planning mode**: ask AI to plan the full pipeline *before* writing any code
- **Sub-agents**: break a complex task into parallel components

PGP Episode 4 example

Mortgage market panel from HMDA data — county lender concentration, fintech/non-bank growth. The kind of data assembly task that used to take a RA a month.

Discussion

Describe a data assembly task in your current project you've been putting off because it's tedious. How would you describe it to an AI in plain English?

AI-augmented replication packages

The same capability that lets you *build* a new pipeline also lets you *navigate* an existing one — with no prior knowledge of the codebase.

Dickerson, Julliard & Mueller — JFE

github.com/Alexander-M-Dickerson/co-pricing-factor-zoo

Replication package ships with a `.claude/` directory and custom skills:

- `/onboard` — validates R, packages, data; auto-fixes gaps
- `/replicate-paper` — full pipeline with error recovery
- `/explain-paper` — explains any table or figure on demand

User prompt: *"Replicate the main text. If packages are missing, bootstrap them automatically first."*

No prior knowledge of the codebase required. This is a new standard for what a replication package can be.

AI as pre-mortem tool

Use AI to find the fatal flaw *before* referees do:

- Describe your identification strategy in a paragraph; ask AI to find the holes
- *"What would a skeptical referee say about this?"*
- *"What's the most obvious alternative explanation?"*
- Multiple AI agents auditing your DiD code for specification errors

Calibration note

AI will also find problems that aren't real. You need domain knowledge to filter. Use it to generate the list; use your judgment to prioritize.

Discussion

Describe your current identification strategy in two sentences. What's the hardest question you'd get at a seminar?

Literature, feedback, iteration

NotebookLM

Document-grounded AI — you feed it papers, it reasons over them. Much safer than asking AI to recall literature from memory. No hallucination risk on documents you provide.

AI referee reports

Useful as first-pass signal, not a substitute for judgment. Good at: missing robustness checks, unclear writing, obvious alternative hypotheses.

Feedback machines

The iterative loop: draft → AI critique → revise → repeat.

Key principle: use AI to **find weaknesses**, not to **write your paper**.

See also: Refine.ink

Block 3: LLMs as Measurement Tools

Beyond the assistant: LLMs as research instruments

LLMs are not just productivity tools.

They are instruments for measuring constructs —
beliefs, sentiment, preferences, reasoning —
that we couldn't measure at scale before.

Case study: The Market's Mirror

Cookson, Bhagwat, Dim & Niessner (working paper)

Question: How do demographics drive investor disagreement? Does demographic disagreement predict trading?

The approach:

- Imbue Llama 3.1 8B with **216 investor personas** (72 demographic combos × 3 political orientations, weighted by FINRA survey data)
- Ask all 216 personas for buy/hold/sell on **5.5 million** S&P 500 news headlines (2010–2025)
- Weighted SD of sentiment = LLM disagreement measure

Why this is methodologically novel:

- A *survey that could not be done with humans* (1.188 billion items)
- Allows revision: update the paper, re-query the same way
- Fixed training window → post-training validation (2024–25)

What the Market's Mirror finds

Key findings:

- Income and politics drive disagreement most
- Social/soft news generates more disagreement than hard financial news
- LLM disagreement validates against human survey data (Iliewa et al. 2025; Toubia et al. 2025 digital twins)
- Disagreement predicts abnormal trading volume, especially retail trading

The broader point:

This opens an entire new research agenda. Any latent construct — beliefs, preferences, reasoning styles — that you can elicit from a persona can now be measured at scale.

What questions become answerable that weren't before?

Discussion

What latent construct in your research area would you want to measure at scale if you could run 1 billion surveys? What would the right instrument look like?

The common thread

Three demonstrations, three layers of capability:

1. **AI accelerates the workflow** — from `CLAUDE.md` to `/checkin`, friction between judgment and result collapses
2. **AI builds and navigates research pipelines** — 821k firm-days to market-level analysis in four prompts, ∼15 minutes
3. **AI enables measurement at impossible scale** — 216 personas, 5.5 million headlines, 1.188 billion elicitations

In every case, research got **faster** *and* **better**.

Not faster *or* better. Both at once. That is worth pausing on before we ask what it means for publishing.

AI and the Future of Publishing

From the editor's desk

A view from ∼10 new manuscripts per month across Management Science and RCFS...

What's getting better

Writing is better, but this is mostly a Chatbot effect that has shown up already. I have seen more obvious AI-written reports on my own work than among my referees.

What's getting harder

More submissions. More polished-looking work.

Interesting tension: I evaluate more on the quality of the idea. This can be easier to sort out once everything is written passably.

True separation will come from using these tools in support of better ideas.
Skepticism: AI doesn't (yet?) ideate well. Will it ever?

Research and publishing are decoupling

"Research and publishing are now two different things."

- AI lowers the cost of a *plausible-looking* paper to near zero
- On the flip side, **truth-seekers** are enabled by these tools.
- How can we — as a profession — combat HARK'ing & encourage discovery?

The supply side

AI one-shot papers, vibe research, automated policy evaluation (Project APE) — these exist and will improve.

The quality filter

Replication and robustness become *easier* to demand. Editors will raise the bar. The threshold for "publishable" rises with the supply of cheap papers.

Replication packages as a new norm

Tie back to Dickerson et al. (Block 2): if AI can replicate a paper in one command, editors can *require* it.

For accountability:

Your identification strategy and code become legible to anyone with Claude Code — reviewers, competitors, editors. Methodological transparency is no longer optional.

For leverage:

You can build on others' work faster. A well-documented, AI-navigable package is a contribution that gets cited and used.

What this means for your career

The premium will shift toward:

- **Deep domain knowledge** — AI can't (yet?) supply the economic question (will it?)
- **Identification creativity** — insightful research (and refereeing) still requires human judgment
- **Real-world access** — proprietary data, field experiments, regulatory relationships
- **Measurement novelty** — like the LLM persona approach — as a source of edge

Discussion

What are you building that AI can't replicate? And what part of your current workflow are you doing manually that AI could handle tomorrow?

Genuine uncertainty — open questions

No settled answers on:

- How fast autonomous research agents arrive and how good they get, especially at ideating
- Whether LLM measurement tools will be accepted as identification or dismissed as fancy text mining
- Where the satisfaction in research comes from once discovery is cheap (Scott Cunningham perspective). Is discovery cheap?

To close

Discussion

What is one thing you'll do differently in your research workflow this week?

And what is the question about AI and research that you most want answered?

Course materials Wiki: velikov-mihail.github.io/ai-econ-wiki · claudeblattman.com · PGP mini-series (BCF Princeton)

Selected readings

Conceptual

- “The Shape of AI: Jaggedness, Bottlenecks and Salients” (Velikov wiki)
- “Research and Publishing Are Now Two Different Things” (Velikov wiki)
- “The Bitter Lesson” — Rich Sutton

Practical

- PGP Episode 1: paulgp.substack.com
- PGP Episode 2: From an empty folder to a figure
- claudeblattman.com — browse the workflows

For the ambitious

- Korinek, “Generative AI for Economic Research: Use Cases and Implications”
- “Feedback Machines: Writing and Editing Research Papers with AI”
- Dickerson et al. co-pricing replication package (AI-ready example)

The demo package

All code, figures, and prompt logs are in the course repository:

Market-Level Analysis demo

`demos/Market-Level Analysis/`

- `market_analysis.py` — equal-weighted baseline
- `market_analysis_capweighted.py` — cap-weighted comparison
- `demo_guide.md` — full prompt log + pedagogical notes

Run either script to reproduce all figures from scratch in ∼30 seconds.

Social Signal Demo

`demos/Social Signal Demo/demo_script.py`

Loads `social_signal_index.dta`, merges with Yahoo Finance, produces return predictability figures. No proprietary data needed.