# A Seat at the Table: Turning AI Macroeconomic Views into Portfolios with a Factor Framework

Redouane Elkamhi  Jacky S.H. Lee*

July 2026

Redouane Elkamhi is Professor of Finance, Rotman School of Management at University of Toronto in Toronto, Ontario, Canada.

Jacky S.H. Lee is Senior Managing Director, Total Portfolio at Healthcare of Ontario Pension Plan in Toronto, Ontario, Canada. Editorial Advisory Board Member, Journal of Portfolio Management and Journal of Financial Data Science, Portfolio Management Research, UK.

*Corresponding Author

# A Seat at the Table: Turning AI Macroeconomic Views into Portfolios with a Factor Framework

July 2026


**Abstract**

Artificial intelligence is everywhere in investing—and yet, when it comes to the core task of deciding how capital is allocated, almost nowhere. This paper offers one concrete answer, worked out end to end and demonstrated with a working system. We keep the process investors already follow—form a view on the economy, then build the portfolio—and make the division of labor explicit. The authors set the analytical frame: AI operates within that frame through a pipeline of agents that gathers and organizes macroeconomic evidence and distills it into probabilistic views on growth and inflation regimes, honest about its own uncertainty. A transparent engine takes the second step, translating any view—machine-generated or human—into candidate portfolios inside hard constraints set by the CIO, the CRO, and the Board. The committee decides throughout: AI gets a seat at the table, not the chair. The engine also runs in an interactive portfolio construction tool developed with AI assistance from author-specified methodology and requirements.


**AI Use Disclosure.** This paper studies AI-assisted investment workflows. The research question, analytical framework, economic model, portfolio methodology, interpretation, and conclusions were developed by the human authors. AI-assisted tools were used within the author-designed research setting, to support prototype implementation from author-specified requirements, and for grammar, editing, and clarity. The authors reviewed and approved all outputs and take full responsibility for the paper.

**Key Takeaways**

- AI can be useful in asset allocation by helping form economic views, not by making decisions. In this paper, AI gathers and summarizes economic evidence each week, but only within rules set by the authors.
- Any view—whether it comes from AI or from people—is translated into a possible portfolio using the same transparent process. The portfolio must stay within limits set by the CIO, CRO, and Board.
- The examples show that governance still controls the outcome. Portfolio limits cap how far any view can move the allocation, and the no-view portfolio makes it clear what the view actually changed. At the limit, the mandate binds, not the machine.


*Keywords:* Artificial intelligence; Asset allocation; Macroeconomic regimes; Factor investing; Portfolio construction; Governance; Total portfolio approach.

*JEL Classification:* G11, G17, G23, E37

## 1 Introduction

AI is a buzz. Every institution is experimenting with it, every conference has a panel on it, and every board is asking about it. Yet ask the question that actually matters to an allocator—*how, concretely, should we use AI to decide how our capital is invested?*—and the conversation usually goes quiet, or goes vague. This paper answers it out loud.

We should be precise about which decision we mean. This paper is about **total-fund asset allocation for a multi-asset investor**—the highest-level decision of how capital is divided across equities, bonds, real assets, and the rest. It is not about alpha. Hedge funds and quantitative managers are already experimenting with AI for signal generation, security selection, and relative-value trading, with varying degrees of success; that is a different problem, at a different horizon, and it is not our subject. For the allocation decision itself—the one that determines most of a fund's risk and return—remarkably little has been said about how AI should participate; the recent proposal of Ang et al. (2026) is, to our knowledge, nearly alone. This paper is aimed squarely at that gap.

Consider the two answers usually on offer. The first is to wait: treat AI as a research curiosity while the allocation process stays what it has always been—a committee that meets periodically, digests what its members had time to read, and sets a view that must hold until the next meeting. Waiting feels prudent. The trouble with it is arithmetic. The information relevant to a macro view—data releases, surveys, market prices, positioning reports, central bank communications, earnings calls, expert commentary—arrives continuously and enormously, and human attention does not scale with it. However talented, a committee *samples* that flow. It cannot read it.

The second answer is to automate everything: let agents generate the assumptions, build the portfolios, vote on a winner, and write the board memo (for the fullest expression of this vision, see Ang et al., 2026). We understand the appeal. But we do not believe a fiduciary institution can—or should—hand the chair to a machine today, and we suspect most CIOs, CROs, and boards feel the same.

Our answer—for now—sits in between, and it is the thesis of this paper. In the age of AI, the opportunity is to rethink how investing gets done from the ground up—and to give AI the work it is actually best at. Today, that work is synthesis: reading vastly more than any team could, every week, without fatigue and without anchoring, and assessing what it reads—tone, sentiment, contradiction—with real competence. Turning that synthesis into an assessment of *probability* is a further step, and we are honest that it is a stretch: whether an AI's probabilistic read of the economy can stand beside a good macro strategist's or economist's is unproven, and the jury will be out for years. That is exactly what this paper attempts anyway. Success remains to be seen—but the uncertainty is a reason to build carefully and keep score, not a reason to sit out, and we suspect readers will value the attempt precisely because it is made with that humility. What humans are best at—and what only humans are accountable for—is judgment: deciding which risks the fund should truly take. In the end, the portfolio an institution chooses is a function of the investor's own

preferences and utility—its tolerance for risk, its liabilities, its comfort with uncertainty—and preference is something no AI can supply on the investor's behalf. That is the core of it. So we give AI a seat at the investment committee table, and we keep the chair human. The AI's view supplements and complements the views of the people around the table; it replaces no one. As the technology matures and live track records accumulate, institutions may choose to extend more trust. For now, we believe the industry moves forward through attempts like this one—built carefully, governed explicitly, and documented honestly.

The framework itself will feel familiar, because it follows what practitioners already do. Ask how asset allocation happens at almost any institution and you will hear the same two steps: form a view on the economy, then build the portfolio from the view. We keep those steps and change who does what. **Step 1**—the view—is performed by a scheduled pipeline of AI agents that runs regularly and unattended, producing probabilities on economic regimes along with its own confidence. **Step 2**—the portfolio—is performed by a deterministic engine that converts the view into one candidate portfolio, with the CIO's preferences and the risk appetite set by the CRO and the Board built into the mathematics as hard limits. The committee then does what committees are for: debate, challenge, decide. And because the engine does not care where a view came from, any committee member's view can be run through the identical machinery and land on the table as an equal candidate. We stress the word *candidate*: nothing this machinery produces is meant to be the final answer. Asset allocation weighs considerations that are qualitative and sometimes unquantifiable—a growing aversion to a deteriorating fiscal picture, unease about the future stance of central banks—none of which shows up in a covariance matrix. However good the machinery, it runs on assumptions, and assumptions are always incomplete against the full context of the decision. We respect, and expect, that investors will exercise their own judgment on the ultimate portfolio.

We should say clearly what this paper is not. It is not evidence that AI regime views make money. The track record of any AI's regime probabilities will be determined the slow way—by the industry, over years of live, dated forecasts—and we claim nothing about it here. But absence of proof is not a reason to sit out the era; it is a reason to build carefully. What protects the institution is not a backtest. It is the design: a human is the decision maker, the AI is an adviser, and that is the ultimate gate. Every view the system produces is dated, versioned, reproducible, and falsifiable, so the record that settles the question will accumulate on its own. And the record cuts both ways: an AI system can learn, be recalibrated, and improve, so what does not work today may well work tomorrow. With the technology advancing at its current pace, standing still is the one posture an investor cannot afford.

The paper tries to deliver three things of practical value. First, a **conceptual framework**: an architecture you can visualize working, in which the concerns that stall most institutions—an AI gone rogue, humans out of the loop, an unauditable black box—are answered by structure rather than by promise. Second, an **implementable methodology**: the complete mathematics for turning regime probabilities into a portfolio, reproducible with your own assumptions and constraints. Third, a small **proof of possibility**: the interactive

tool used in this paper was developed with AI assistance from a human specification. Sections 2 and 3 set the stage; Sections 4 and 5 describe the two steps; Sections 6 and 7 build and use the tool; Section 8 confronts the risks; Section 9 concludes.

**A note on terms, for the general reader**

A *regime* is a broad economic state—for example, growth rising while inflation falls; we use four, the combinations of rising or falling growth and inflation. A *capital market assumption* (CMA) is the industry's standard set of long-term return and risk forecasts for asset classes. A *factor* is a broad economic driver of returns that many assets share—here: growth, inflation, and the real interest rate. A *probabilistic view* says "rising growth is 60% likely," never "growth will rise." And *confidence* ($\gamma$) is a dial from 0 to 1 for how strongly the portfolio should act on a view. That is nearly all the vocabulary this paper requires.

## 2 The Ideas This Framework Stands On

Nothing in this framework is exotic. Its pieces come from ideas the industry already trusts, assembled in a new way—with AI supplying the one ingredient that was always missing: a disciplined view, produced on schedule.

The oldest piece is the observation that markets behave differently across economic regimes, and that allocation should respond (Ang and Bekaert, 2004; Kritzman et al., 2012). The obstacle was never the idea. It was that acting on it seemed to require forecasts nobody could reliably make—point estimates of returns, notoriously fragile things to build a portfolio on. In earlier work we showed a way out: express the view as *probabilities over regimes* instead. Portfolio tilts built from regime probabilities turn out to be stable, and their performance is remarkably forgiving of forecast error (Elkamhi et al., 2022). That forgiveness matters enormously here, because probabilities are exactly the honest output format of an AI system that knows it might be wrong.

The second piece is older still: the idea that a portfolio should move away from a neutral starting point in proportion to the strength of a view, descending from Black and Litterman (1992) and the confidence extensions that followed (Idzorek, 2007). Our view contract—a probability vector $\pi$ and a confidence $\gamma$—is Black–Litterman in spirit, translated into the language of regimes: two numbers a committee can debate in plain English.

The construction machinery rests on two further results from our earlier work. Capital market assumptions can be priced by a small set of macroeconomic factors, and portfolios that allocate to those factors are far more stable than portfolios built asset by asset (Elkamhi et al., 2021; Lee and Salerno, 2023). Together with the regime-tilting result, these give Step 2 its most practical property: if the CMA is reasonable, the portfolio that comes out is sensible—by construction, not by luck.

The newest piece is the technology itself. Large language models can now act

as agents—goal-directed programs that use tools, run scripts, and produce structured output—and ensembles of them can be made to check and challenge one another (Du et al., 2023). Applications in investment management are arriving quickly; the most complete to date is Ang et al. (2026), a pipeline in which dozens of agents generate capital market assumptions, construct candidate portfolios, and vote, through to an AI chief investment officer. Our framework makes different choices, and they are easy to state: the AI produces only the economic view; construction keys off the institution's single approved CMA through fixed, transparent mathematics; and the CIO, the CRO, and the Board are people, whose preferences and limits enter as hard constraints. The rest of this paper is about those choices.

One last influence frames the whole exercise. In our work on the total portfolio approach (Elkamhi and Lee, 2025) we argued that allocation is becoming an adaptive, whole-of-fund exercise in *preparedness* rather than the maintenance of a fixed plan. A disciplined, probabilistic, always-current view of the economy—with a transparent way to act on it, and to know what acting on it would change—is precisely the kind of tool that mindset calls for. This paper builds one. To be clear, we are not suggesting a CIO should act on every small move the framework registers; there sits yet another layer of judgment. Often the information is not in the view itself but in how the view *changes* over time: when the view shifts substantially—however the investment committee chooses to define substantially—that is a natural moment to re-examine the allocation. The framework offers exactly that kind of trigger point.

## 3 The Framework: Two Steps and a Human Boundary

Figure 1 is the whole framework on one page. Everything below the dashed line is machinery; everything above it is human; and the design guarantees that three things never cross the line downward—the analytical frame, the risk limits, and the decision.

The two steps mirror practice. **Step 1** produces the view: a pipeline of AI agents runs on a regular schedule, reads the week's evidence, and emits probabilities on economic regimes. The contract between the steps is deliberately minimal—a probability vector $\pi$ over the four combinations of rising and falling growth and inflation, and a confidence $\gamma \in [0, 1]$ that says how strongly the view should be expressed. **Step 2** produces the candidate: a deterministic engine tilts the institution's approved capital market assumptions by the view and maps the result to one or more feasible portfolios—each theoretically sound from a factor perspective, each inside every constraint. Alongside them, the engine always constructs a *no-view* companion—the same machinery with $\gamma = 0$—which gives the committee a clean point of reference.

Three design commitments run through everything that follows.

**The human sets the frame.** What the AI reads, how regimes are defined, which historical relationships map economic states to return factors—all of it is ratified by the researcher and pinned. The agents exercise judgment strictly inside that frame. The

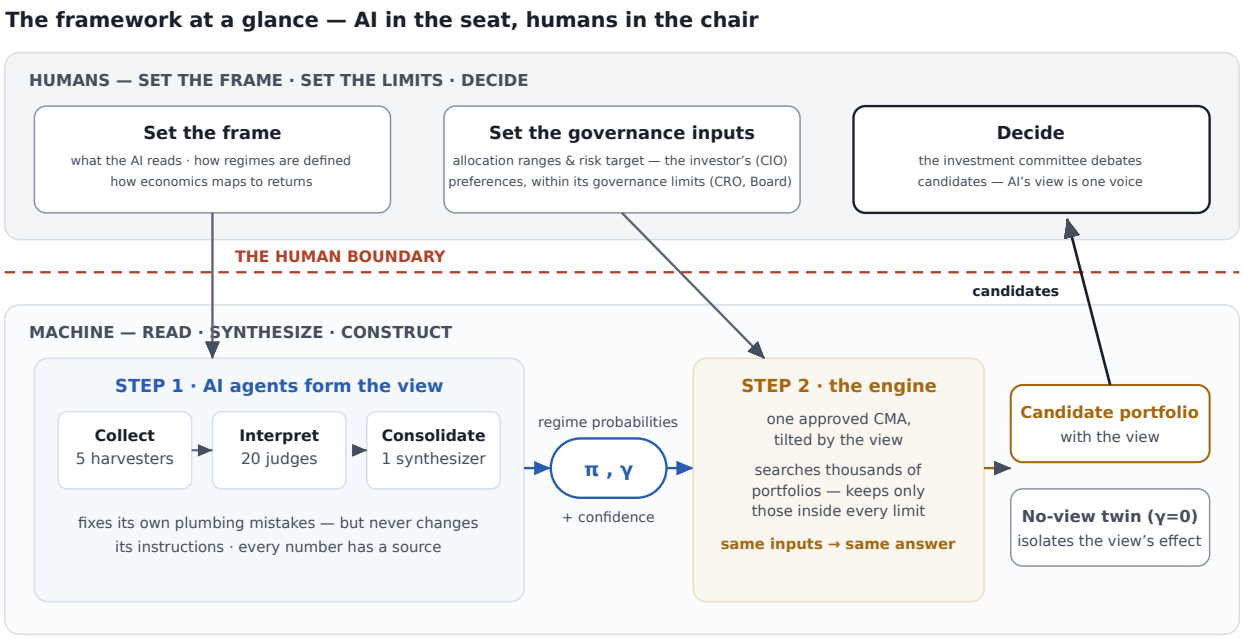



**Figure 1: The whole framework on one page.** Below the dashed line, AI agents read the current economic evidence—in our implementation, weekly—and condense it into a view of where the economy is heading: $\pi$, how likely each economic state is, and $\gamma$, how strongly to act on it. An engine then turns that view into a candidate portfolio, shown next to a "no-view twin" built with the view switched off—a point of reference that makes plain what the view changed. Above the line sit the people. They decide what the machine reads, they set the limits every portfolio must respect, and they make the decision. Nothing below the line can change the rules, move a limit, or invest a dollar.

system does maintain itself—it can fix its own operational mistakes, a broken data feed, a malformed file—but it is structurally forbidden from editing the frame it was given (Section 4.3 details how).

**Governance enters as inputs, not instructions.** The CIO's allocation preferences and the risk appetite set by the CRO and the Board are not prompts an agent promises to respect. They are hard constraints in the construction mathematics: minimum and maximum bands on every allocation group, and a portfolio volatility target the candidate must meet. An AI cannot talk its way past a constraint set. This is, we think, the practical difference between governance by design and governance by hope.

**The engine is view-source-agnostic.** Nothing in Step 2 knows who produced $\pi$ and $\gamma$. A human committee that disagrees with the machine can express its own regime probabilities—two numbers, debatable in plain language—and run the identical construction. The AI's view then lands on the table as *one* candidate among equals, which is what "a seat at the table" means operationally. It also means an institution can adopt the framework in stages: the engine is useful on day one with purely human views, before any AI is involved at all.

## 4 Step 1 — How AI Generates the Regime View

Here is the job description we would write for a macro research team if we could ignore human limits: read everything relevant, every week. Never anchor on last week's call. Treat your own conviction with suspicion, and argue against yourself before you conclude. Show your work—always. No human team could accept that job. This section describes the AI team that does, in enough detail to visualize it working and to copy its design; the algorithmic internals live in Appendices A and B.

Three principles shaped every choice, and each answers a specific failure mode of "ask an AI for a macro forecast." First, **separation of concerns**: collecting evidence, interpreting it, and aggregating interpretations are three different jobs done by three kinds of agent that never blur—so no agent can rationalize a number it wishes it had. Second, **the human sets the frame**: the machine exercises judgment only inside a researcher-ratified specification. Third, **the model is an input, not an oracle**: uncertainty is carried at every stage, disagreement between agents is information rather than noise, and the numeric core is deterministic, so a week-over-week change in the view is attributable to evidence, never to the machinery's mood.

### 4.1 What the system reads

The view rests on a wide, deliberately heterogeneous evidence base organized under four macroeconomic themes—**growth & activity**, **inflation & liquidity**, **policy rates & curve**, and **credit & financial conditions**—plus a shared cross-asset **positioning & sentiment** layer. The grouping is a choice, not a rule. It reflects what we consider matters most for an economy: growth and inflation themselves, and the forces that drive them—policy and rates, and credit conditions. The positioning and sentiment layer serves a different purpose: it informs conviction rather than direction. If growth is about to surprise to the upside and everyone already knows it, the information is likely already in the price—and the view deserves less weight, not more. Each theme is governed by a ratified specification whose input menu enumerates the exact signals to collect: the source, the lead time, the read direction, and whether the signal is a primary driver or a confirmatory cross-check. Across the four themes the menu totals on the order of 150 distinct signals, drawn from six classes of source (Table 1).

Two features are worth a practitioner's attention. Narrative is a first-class input: the system deliberately reads qualitative testimony alongside the hard series, precisely so that a story unsupported by numbers—or numbers unsupported by any story—surfaces as a contradiction. And consensus is structurally quarantined: what forecasters expect and how the market is positioned enter only as the benchmark against which surprise is scored and as a cap on conviction. They are never allowed to vote on the answer, so the system cannot simply reprint the crowd.

The word *quarantined* deserves one more beat, because a careful reader will notice that quarantine cannot mean irrelevance. Is consensus excluded from the answer, or is it

**Table 1: The six data classes the system ingests.** Consensus-type data—what forecasters predict, what market prices imply, how investors are positioned—is collected but ring-fenced. It serves two purposes only: it tells the system what is already expected, so the system knows what would count as a surprise; and it restrains overconfidence when a call agrees with a crowded consensus. What it never does is count as evidence when the regime probabilities are computed. The crowd sets the benchmark the answer is measured against; it casts no vote in the tallies the answer is computed from.

| Data class | Representative content | Role in the view |
|---|---|---|
| Official macro time-series | Inflation gauges (headline, core, trimmed, sticky); GDP and nowcasts; production; claims; curve, TIPS, term premium; bank credit; central bank balance-sheet and liquidity series | The quantitative backbone |
| Surveys & diffusion indices | Manufacturing and services PMIs; senior-loan-officer standards; consumer, small-business, and professional-forecaster expectations | Forward-looking; also supply the consensus benchmark |
| Market & prediction-market data | Curves, credit spreads, breakevens, FX, commodities, volatility; futures-implied policy paths; event contracts | The real-time market read and market-implied consensus |
| Positioning & flows | Futures positioning, investor-survey sentiment, put/call, fund flows, margin debt, manager surveys | Gauges how crowded the consensus is; sets the bar for what counts as a surprise—excluded from the probability calculation |
| Filings & company disclosure | Material-event filings, insider activity, earnings-call commentary of bellwether firms | Bottom-up corroboration of the top-down read |
| News, research & expert transcripts | Curated financial press, attributed house views, transcripts of expert macro commentary | Narrative and contradiction-hunting |

the benchmark that changes the answer? Both—and the two roles are different in kind. The quantities this system scores are *surprise* axes: the question each week is not "will growth be strong?" but "will growth beat what is already expected?"—and consensus is what defines *expected*. Move the consensus and the answer moves, the way a bond's excess return changes when its benchmark yield does: not because the benchmark voted, but because the outcome is measured from it. What the quarantine forbids is the second role. Consensus and positioning never enter the evidence tallies from which the probabilities are computed—the fact that a view is popular is never, by itself, a reason to believe it. Consensus defines the question; only evidence answers it. That is what "never a vote" means wherever it appears in this paper.

### 4.2 The agent society

The pipeline is a small society of specialized agents with sharply separated roles, run as three sequential waves on a schedule (Figure 2). **Five harvesters**—one per theme plus one shared positioning collector—*collect and never interpret*: a harvester may report that jobless claims printed 212k on Thursday, but it may not write "the labor market is weakening." **Twenty judges**—five analyst personas applied to each of the four themes—*interpret and never collect*: each judge independently scores its theme's regime components from the same evidence pack, producing probability vectors, an explicit conviction rating, and two to three quantified triggers that would prove it wrong. The five personas are archetypes any investment floor would recognize: a *hard-data* analyst who trusts the numbers over the narrative and demands that data corroborate any story; a *bottom-up* analyst who reads the economy as companies report it, through earnings, guidance, and revisions; a *sentiment-focused* analyst who works the news flow and positioning, hunting for gaps between the story and the data; a *global* analyst attentive to external demand, China, and worldwide linkages; and a *generalist* who applies the framework straight, with no tilt. The personas share an identical method and differ only in emphasis—the point is a controlled spread of plausible readings of one evidence base, not twenty different methodologies. **One synthesizer**—think of it as the aggregator—*consolidates and never re-judges*: it averages the judges, runs a pinned deterministic script, and writes the factor-level regime probabilities with a defended commentary. **Wave dispatchers**—the project managers of the pipeline—orchestrate, gate, and repair.

A word on how the workflow instructions were developed, because the process is part of the demonstration. The analytical design was set by the authors: the objective of the experiment, the macroeconomic framework, the agent roles, the evidence categories, the governance boundaries, and the portfolio-construction methodology were all specified by the researchers. AI-assisted tools were then used to help translate those author-defined requirements into operational instructions for the agent workflow, including drafting collection templates, orchestration steps, output formats, and quality-control checks. The process was iterative: the authors reviewed the draft workflow, tested sample outputs, identified gaps or ambiguities, and revised the specifications until the system behaved consistently with the intended design. AI tools also assisted in implementation tasks, such as proposing refinements to templates or flagging formatting issues, but the analytical frame, source selection criteria, scoring logic, and final workflow were reviewed, revised, and approved by the authors before being pinned for the demonstration. This division of labor is central to the paper's approach: humans set the research design, analytical boundaries, and governance rules; AI supports execution within that author-defined frame.

### 4.3 Gates, repairs, and a hard boundary on self-improvement

Every agent in the pipeline works from a written specification it must follow: the data sources it may read, the regime definitions and thresholds it scores against, the exact format its output must take, and the boundaries of its role. That specification is the contract that makes an unattended run checkable—and everything must be checkable, because no

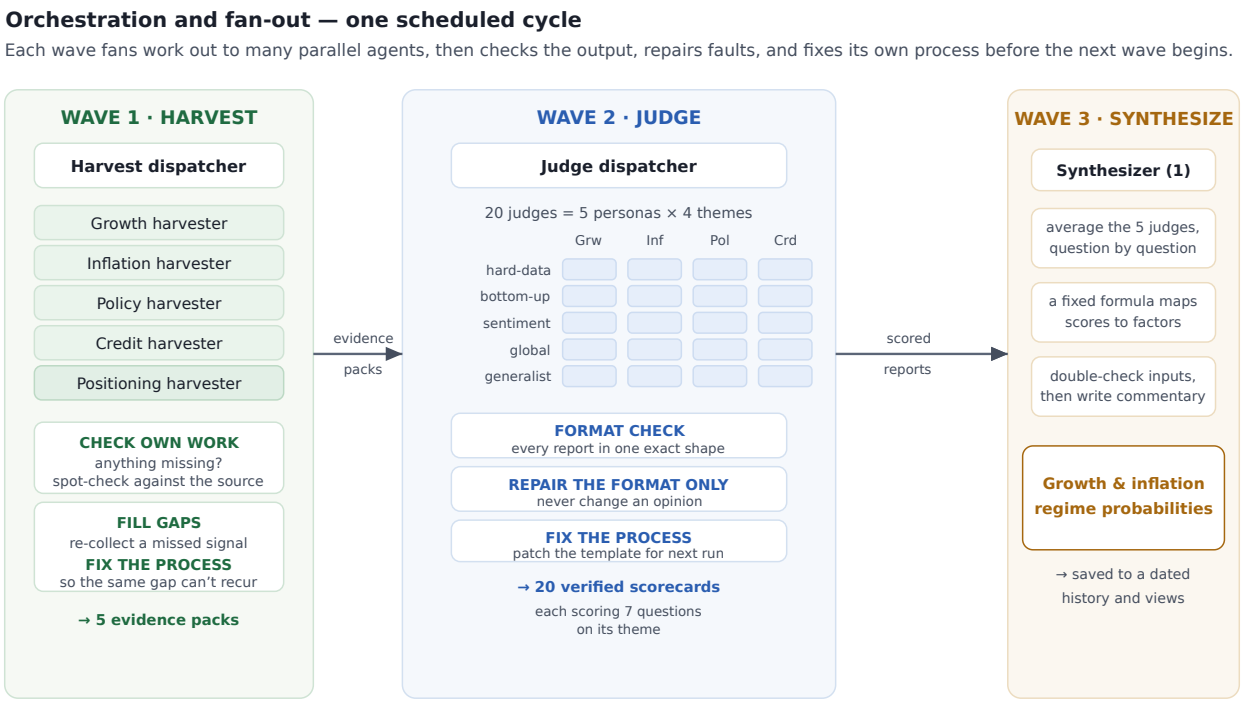



**Figure 2: One cycle of the pipeline, in three waves.** First the harvesters collect the week's evidence; then twenty judges independently interpret it; then one synthesizer averages the judges and turns the scores into factor probabilities. The jobs never mix—a collector cannot editorialize, a judge cannot go fetch friendlier data, the synthesizer cannot overrule a judge—and each wave leaves its work on disk, so any number can be traced back later. Note where the checking happens: the collected data is verified and repaired *before* any judge reads it, because that is the one place a single error would poison all twenty opinions at once. (In the judge grid, Grw, Inf, Pol, and Crd abbreviate the four themes: growth, inflation, policy, and credit.)

human is present when the pipeline runs. So every wave is written to gate its own output and fail loudly. After the harvesters return, the dispatcher checks that every primary signal is present or has a legitimate "not yet released" explanation, spot-checks readings against original sources, dispatches a targeted supplement to fill any gap—and then *self-improves*: it roots the gap to a cause and makes one durable edit to the responsible collection recipe so the same failure cannot recur. The judge wave has the analogous loop for format, with a repair agent that reconstructs malformed output from the judge's own report (re-ordering, never re-judging).

### The scope guard

Both loops are bounded by an explicit rule: agents may edit only **operational** logic—collection recipes, source endpoints, output formatting. They may **never** touch **analytical** content: regime definitions, thresholds, scoring rubrics, or the factor mapping. Anything that would change the analysis is flagged for the researcher and changed only by a logged, human-ratified decision. This is the mechanism that lets the plumbing improve itself every week while guaranteeing the view can never quietly drift to fit a number—and it is the sharpest single difference between our design and a self-rewriting pipeline.

*We should be honest about the line itself*: operational versus analytical is a judgment, not a law of nature. A swapped data source, a revised collection recipe, a missing-data rule—each is operational in intent, yet each can change what a judge sees, and a model-risk team could reasonably insist that some of these, too, pass through human approval. Nothing in the design resists that. Every automatic repair is logged and reviewable after the fact, and the boundary is a policy dial an institution can set as conservatively as its governance requires—including requiring sign-off on every change of any kind. This is our attempt at a workable default, not a claim that the line is drawn perfectly; what the architecture guarantees is more modest and more important: wherever the line is drawn, the analytical side of it cannot be crossed silently.

### 4.4 How a judge thinks

Each judge is memoryless—it receives no past reports, so it cannot anchor on a call it made last week—and executes a fixed procedure. It **counts clusters**: over a closed registry of leading and coincident indicator groups, it tallies how many read bullish, neutral, or bearish for its components on the three-to-six-month horizon; exact integers over a pinned registry, so a judge can neither invent nor drop evidence. It **scores the surprise, not the level**—and because this is the heart of the judgment, it deserves a careful description. Probabilities do not start from a blank slate: they start from a pinned base-rate table, roughly the historical frequency of each regime state, and the evidence must earn any distance from that anchor. The distance is judged against an explicit, quantified picture of what is already expected, which the evidence pack hands every judge in numbers: professional forecasters' medians, statistical nowcasts, the policy path implied by futures, market-based surprise indices, prediction-market odds, and—collected separately—the attributed house views of major institutions. An economy growing at trend is not a bullish surprise if trend is what everyone already predicts. We are candid about the epistemics here: nobody can truly know what is priced. In the end the judge—an LLM—makes a determination from these proxies, weighing the evidence in the pack against the measured consensus and against how crowded positioning already is, and that determination is precisely where we are pushing AI to test its usefulness. Whether it does this well is exactly the kind of question only a live, dated, scored track record can answer. The judge also reports **conviction as a separate axis** from probability, and a crowded consensus can cap that conviction but can never flip a direction. It must write an explicit **adversarial passage**—the strongest case against its own lean, from evidence already cited—and then rebut it or move its numbers. And it ends with **falsification triggers**: a series, a threshold, and a date that would prove the call wrong. Positioning data plays a specific, limited role in all of this: it tells the judge how crowded the trade already is. Positioning at an extreme raises the bar for calling a surprise in the same direction and caps the conviction the judge may claim, and it informs the discussion of risks—but it never enters the tallies from which the probabilities are computed.

**Design rules for an AI view pipeline (the transferable checklist)**

- **Collect $\neq$ interpret $\neq$ aggregate.** Three jobs, three classes of agent, structured artifacts between them.
- **Ensemble over oracle.** Many independent, memoryless judges; measure their disagreement and report it as uncertainty.
- **Consensus and positioning never count as evidence.** They define what is already expected—the benchmark surprise is scored against—and they cap conviction; they are structurally excluded from the probability calculation itself. Benchmark, yes; ballot, no (Section 4.1).
- **Score the surprise, not the level.** Anchor on base rates; distance must be earned.
- **Theory sets direction; data sets magnitude.** Economic theory fixes which way each regime state pushes each return factor; the week's evidence can scale that push—all the way to zero if the evidence is weak—but can never flip it. A noisy week can silence a channel; it cannot make the system claim the economics runs backward.
- **Uncertainty never collapses.** Carry judge dispersion, mapping error, and correlation jointly into a band; keep conviction a separate axis.
- **Deterministic numeric core.** Every calculation runs in fixed, version-pinned scripts with fixed random seeds—never in the language model's head. From the judges' recorded scorecards onward, a re-run reproduces the view digit for digit; so when the view changes, the cause is the evidence, never the machinery's mood. (The judges themselves are language models, and their layer is reproducible in a different sense—by archive, not by re-generation; Section 8.1.)
- **Self-improvement never touches the analysis.** Operational repairs are automatic; analytical change requires a logged human decision.
- **Demand falsifiability.** Every call ships with triggers that would prove it wrong; every view is dated and scoreable.

### 4.5 From twenty opinions to factor probabilities

The synthesizer consolidates in three moves; the full algorithm is in Appendix B. First, for each of the 28 regime components (seven per theme; Appendix A), it averages the five judges of the owning theme by regime-state *name*, keeping the dispersion:

$$\bar{p}_j = \frac{1}{n}\sum_{i=1}^{n} p_{ij}, \qquad \mathrm{SE}(\bar{p}_j) = \max\left(\frac{\hat{\sigma}_j}{\sqrt{n}},\ 1\,\mathrm{pp}\right), \tag{1}$$

where $p_{ij}$ is judge $i$'s probability for regime state $j$, $n$ is the number of judges averaged ($n = 5$), and $\hat{\sigma}_j$ is the sample standard deviation of those five probabilities; the one-percentage-point floor keeps a unanimous panel from claiming false precision. This way, genuine disagreement among the five readings survives into the output. It is worth being precise about what that dispersion measures: the five personas share one underlying model, so $\hat{\sigma}_j$ captures disagreement *within* that model's way of reading the world, not across models. The band it ultimately feeds is therefore best read as a lower bound on the true

uncertainty—an error all five judges share leaves no trace in the spread, and never widens the band. Section 8.1 returns to this squarely.

The second move turns those averaged probabilities into statements about **return factors**—the broad economic drivers of asset returns. The bridge is a lookup table we call the **factor-driver matrix**: one row for each regime state that has a genuine, economically grounded link to a factor. A row says, in effect, "if the economy is in *this* state, it pushes *that* factor in *this* direction, with *this* much force." Bank lending standards tightening pushes the growth factor down; consumer prices running above their trend push the inflation factor up. Each push is recorded as a signed loading $\beta$: the *sign* comes from economic theory, and the *magnitude* is estimated from decades of historical data—how strongly that state has actually moved the factor in the past. Two properties of this table matter for governance. It is built and ratified by the researcher, then pinned—no agent can add a row, drop a row, or change a loading. And because its magnitudes are estimated from history, they are estimates, not truths: every $\beta$ carries its own uncertainty, measured from the same historical record. A state's contribution to its factor is then simply

$$g = \bar{p} \cdot \beta, \tag{2}$$

the probability of being in the state times the push the state delivers. Note that *both* ingredients are uncertain—$\bar{p}$ because five judges disagree, $\beta$ because history is a noisy teacher—and the synthesizer does not pick one and ignore the other: in the simulation that follows, both are drawn jointly, so the final answer carries both sources of doubt at once.

The third move adds one more ingredient and then lets everything vary together. The 28 components are not independent readings of the world: labor momentum and cycle phase are often two windows on the same underlying economy, and if one is misread the other likely is too. So the researcher also ratifies and pins a correlation structure across the components—a fixed matrix saying which readings tend to move together—and the synthesizer runs a seeded Monte Carlo in which judge dispersion, loading uncertainty, and those cross-component correlations are all drawn jointly, ten thousand times. One rule disciplines every draw: *theory sets the direction, and noise is not allowed to overturn it*. If a draw hands a theoretically bullish state a negative contribution—pure sampling noise pointing against the economics—that contribution is **clamped**: set to zero rather than allowed to flip sign. The clamp is applied state by state, draw by draw: each component is floored at zero individually *before* anything is summed, so a noisy draw on one channel cannot be netted against the strength of another. There is no single clamp on a side's total; weak evidence silences channels one at a time, and it can never make the framework claim that tightening credit is good for growth. Each draw then produces the factor's probabilities by a simple ratio:

$$P(+) = \frac{\sum_{j \in \mathcal{S}_+} \max(g_j,\, 0)}{\sum_{j \in \mathcal{S}_+} \beta_j}, \qquad P(-) \text{ analogously}, \qquad P(\text{neutral}) = 1 - P(+) - P(-), \tag{3}$$

where $\mathcal{S}_+$ is the set of states that push the factor up (with $\mathcal{S}_-$ its downward counterpart), $g_j$

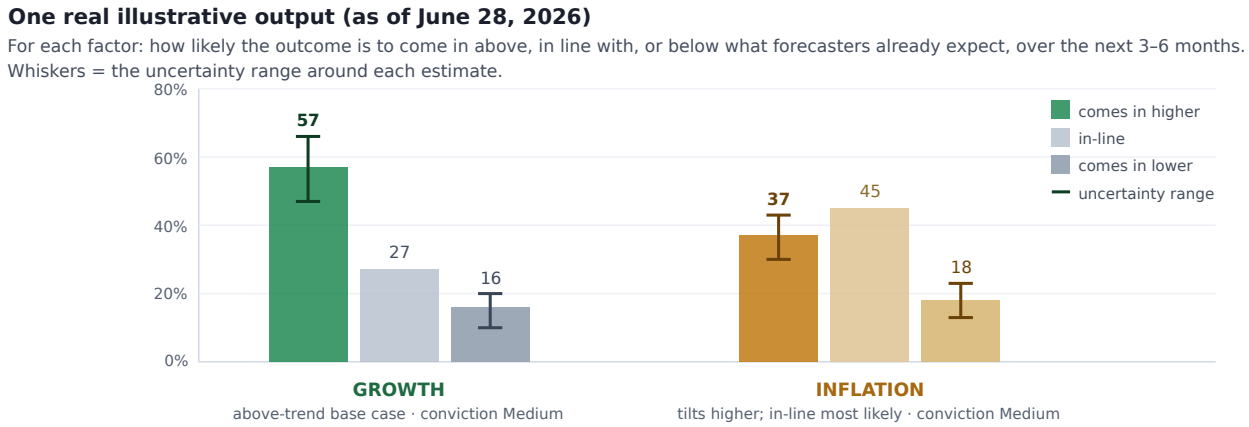



**Figure 3: A real weekly output (June 28, 2026)**—the view produced by the pipeline described above, and the one used in the illustration later in this paper. Read it like a weather forecast for the economy. Growth: the system puts a 57% chance on growth beating expectations, though the honest range on that number runs from 47 to 66, and its own conviction is only "Medium." Inflation is genuinely two-sided: coming in as expected is the single most likely outcome (45%), with the risk tilted modestly toward higher (37% against 18%). The ranges, not the single numbers, are the honest read—and they are what the confidence dial $\gamma$ responds to in the next step.

is state $j$'s drawn contribution—the max operator inside the sum is the clamp, applied to each state separately—and $\beta_j$ is its loading. The denominator is the strongest bullish push the framework recognizes: what the contributions would sum to if every bullish state were a certainty. The numerator is the bullish push the week's evidence actually delivered. So $P(+)$ has a plain-English reading: *the fraction of the maximum bullish case that the evidence showed up for*. If the loaded states deliver 57% of the strongest above-consensus case the matrix allows, the growth factor gets $P(+) = 57\%$; the bearish side is computed the same way from $\mathcal{S}_-$, and whatever force the evidence failed to deliver on either side is, fittingly, the probability of a neutral outcome.[1]

Before these numbers travel any further, we should say plainly what they are—and what they are not yet. A statement like "$P(+) = 57\%$" is *constructed* to behave like a probability: it lives on the right scale, it moves coherently with the evidence, and it carries an honest band. But whether 57% *means* 57%—whether, across the many weeks on which the system says 57%, the outcome actually arrives about 57% of the time—is a property called calibration, and it can only be earned by a live, dated, scored track record that does not yet exist (Section 8.2). Until it does, the honest reading of these outputs is **AI conviction scores expressed on a probability scale**: disciplined, evidence-weighted, uncertainty-banded—and unvalidated. We will keep calling them regime probabilities throughout, because that is the format and the ambition; the reader should hear the phrase with this caveat attached, everywhere it appears.

The reported view is the mean across the ten thousand draws with a 25th–75th percentile band, plus two diagnostics a committee actually wants: a **conviction** read (how

[1] In the rare draw where sampling noise pushes $P(+) + P(-)$ above one, the two sides are rescaled proportionally so the three probabilities sum to one; the neutral share in that draw is zero.

much the judges trusted the evidence on each side, distinct from the width of the band) and an **inflection** gauge that flags contrarian evidence building before a regime turns. The output is two vectors—growth and inflation—each with $P(+)$/neutral/$P(-)$ and its band, appended to a versioned time series. Figure 3 shows a real week.

What does "defended" mean in practice? It is easiest to show. Alongside the numbers, the synthesizer writes a rationalization of its own output—the case for the probabilities, argued from the underlying evidence, with the caveats stated rather than buried. Here is the commentary that accompanied the June 28 run—the same view shown in Figure 3—as a committee would receive it, lightly edited for length:

**Synthesizer commentary — the June 28, 2026 run (the view in Figure 3)**

"**Growth**—we see roughly a 57% chance the economy runs above its long-run trend over the next three to six months (versus about 16% below, with the rest in line), and we're highly confident in the direction even if not the magnitude. Nearly every indicator points the same way: the Atlanta Fed's real-time tracker has Q2 GDP around +2.5%, above the ∼2.2% trend; factory surveys are the strongest in years (S&P Global's gauge at 55.7, ISM new orders at 56.8, the Philly Fed survey up +27.3); banks are easing lending standards and business loan growth has swung from negative to about +8% year-on-year; financial conditions are among the loosest in years; analysts are raising earnings estimates broadly (about 73% of revisions are upgrades); and global activity is expanding (∼70%). We deliberately hold this at medium conviction, not high, for two reasons. First, the consumer is a real soft spot—incomes excluding government transfers have fallen three months running, and homebuilder sentiment (NAHB at 35) is deeply depressed. Second, almost everyone is already positioned for good news: investor bullishness is at a record (NAAIM at 98.6), the most crowded trade on record is being long semiconductors (80% of fund managers), and margin borrowing is at an all-time $1.42 trillion—so even a genuinely good number could be met by a flat or falling market.
**Inflation**—the most likely outcome is that it stays about in line rather than breaking higher (roughly a 37% chance of a meaningful upside surprise, 18% to the downside, and 'in line' the single biggest bucket). The picture is honestly mixed, which is why we don't lean hard either way. On the hot side: underlying prices are still running above target and not yet cooling (core PCE around 3.4%), cost pressures are building in the pipeline (producer and import prices up 6–7%), and demand signals from the yield curve and surveys are firm. On the cooler side: the monthly inflation readings have settled back into a moderate range rather than a hot one, longer-term inflation expectations remain well anchored, global money growth is subdued, and the recent drop in oil is dragging the headline number down. This is our lowest-conviction call.
**What to watch:** the two releases that decide which way each drifts are the July 1 manufacturing ISM (for growth) and the late-July core PCE inflation report (for inflation)."

Read it once and several of this section's design claims stop being abstract. Every probability names its evidence—actual readings, actual sources—so any single claim can be checked, or attacked, directly. The conviction cap is visible in action: the growth evidence is nearly one-way, and the system still refuses to claim more than medium conviction, partly because the consumer data disagrees and partly because positioning is so crowded that even good news might not pay—the never-a-vote rule and the surprise bar, both working in plain sight. The inflation call is two-sided out loud rather than hedged in a footnote, and

says so itself: honestly mixed, the lowest-conviction call on the page. And the closing line is the falsification discipline of Section 4.4 surfacing in plain language—two named releases, with dates, that would move the call. No committee member is obliged to agree with any of it. But because every claim names its evidence, disagreeing is cheap and precise—which is the point.

### 4.6 From regime probabilities to the view contract $(\pi, \gamma)$

The construction engine consumes a probability vector over the four growth×inflation regimes. The pipeline's growth and inflation outputs are three-state—$P(+)$, in-line, $P(-)$—so we convert and combine with two stated conventions. We **split the in-line mass equally** between the two sides of each axis—the subscripts $G$ and $I$ denote the growth and inflation axes, so $P_G(\uparrow) = P_G(+) + \frac{1}{2}P_G(0)$—and we **join the two axes as if independent** (in each $\pi$ subscript below, the first arrow is growth and the second is inflation):

$$\pi_{\uparrow\uparrow} = P_G(\uparrow)\,P_I(\uparrow), \quad \pi_{\uparrow\downarrow} = P_G(\uparrow)\,P_I(\downarrow), \quad \pi_{\downarrow\uparrow} = P_G(\downarrow)\,P_I(\uparrow), \quad \pi_{\downarrow\downarrow} = P_G(\downarrow)\,P_I(\downarrow). \tag{4}$$

These are choices, not theorems—growth and inflation surprises need not be independent, and in-line mass need not split evenly. The independence convention is, however, better supported than it may first appear: in our research, using GDP surprises and CPI surprises along with the correlation of their factor-mimicking portfolios, the two are only weakly correlated over the long run—even though the correlation can run high in selected periods. That evidence, we should add, concerns *contemporaneous* surprises at the horizon the view speaks to; over longer spans the two interact with leads and lags—inflation surprises shape policy, and policy shapes growth with delay—and the convention makes no claim about that richer dynamic structure. We state the conventions plainly because there is no uniquely right answer, and a future system can do better (for instance, estimating the joint distribution directly—an extension we return to in later sections). The confidence $\gamma$ is set from the system's own stated uncertainty—conviction, band width, and the strength of the lean—with a wide band or a Medium conviction pulling $\gamma$ down, so the portfolio leans in proportion to how much the *evidence*, not the machinery, supports the call.

## 5 Step 2 — From View to Candidate Portfolio

Now the view has to become a portfolio. This is, deliberately, the unglamorous half of the framework: no agents, no judgment—just transparent mathematics answering one question. *Given this view, held with this much confidence, what portfolio follows—inside this institution's limits?* This section gives the methodology at the level a team needs to reproduce it; derivations and implementation mechanics are in Appendix C. The universe in our implementation is 30 underlying assets priced by three macroeconomic return factors—economic growth, real rate, and inflation—and reported at the level of 14 asset classes, but nothing in the method depends on those counts. Throughout, $B$ is the $N \times M$ matrix of asset factor loadings ($N$ assets, $M$ factors), $\Sigma_a$ and $\Sigma_f$ the asset and factor covariance matrices, and $\mu_a$ the vector of expected excess returns. How the factor-mimicking portfolios are defined, and how the loadings in $B$ are estimated from them, is covered in detail in

our earlier work (Elkamhi et al., 2021; Lee and Salerno, 2023); readers who want that machinery can find it there, and readers who trust it can proceed.

### 5.1 Target in the space where estimation is kindest

The single most consequential choice is *where* to express the target. Following Lee and Salerno (2023), we work in **reverse-optimized factor-weight** space: for any portfolio $w$, the implied factor loading matrix

$$B_i = \Sigma_f^{-1}(B'B)^{-1}B'\,\Sigma_a \tag{5}$$

maps the portfolio to the factor weights $\theta(w) = B_i w$ that would rationalize it as optimal. Here $B$ is the matrix of asset factor loadings—one row per asset, one column per factor, each entry saying how much that asset moves when that factor moves—while $\Sigma_a$ and $\Sigma_f$ are the asset and factor covariance matrices and the prime denotes a transpose. The key result of that earlier paper is about what happens when expected returns are wrong—which, in practice, they always are to some degree. Feed slightly different expected returns into a classic optimizer and the asset weights can swing wildly; the optimal portfolio's *factor* weights, recovered by reverse optimization, barely move. A target expressed in this space is therefore robust to the errors we know our inputs contain—which is why every target and every comparison below lives in this coordinate system. Intuitively: asset weights fight over dozens of noisy, near-collinear instruments; reverse-optimized factor weights fight over three economically distinct bets.

### 5.2 One approved CMA, tilted by the view

Construction starts from a single set of capital market assumptions—in practice, the institution's approved CMA—and the view refines it rather than replacing it. Each regime $s$ carries a premia tilt $\delta_s$ and a covariance tilt $\Delta_s$ for the three factors, computed from the historical behavior of the factor-mimicking portfolios in each regime, and pinned. Given the view $(\pi, \gamma)$, the factor premia and covariance become confidence-scaled, probability-weighted mixtures, which propagate to the asset level through the factor structure:

$$\mu_a(\pi,\gamma) = m_{\text{cma}} + g\,B\lambda_f^0 + \gamma\,B\sum_s \pi_s\,\delta_s, \qquad \Sigma_a(\pi,\gamma) = \Sigma_a^0 + B\Big(\gamma\sum_s \pi_s\,\Delta_s\Big)B'. \tag{6}$$

Here $m_{\text{cma}}$ is the CMA's vector of expected excess returns, $\lambda_f^0$ the long-run factor premium implied by that CMA (its least-squares projection onto the loadings; Appendix C), $g$ a scalar that sets the *marginal* price of risk—it scales only the factor premium layered on top of what the CMA already implies, calibrating the expected returns used in this analysis (Section 5.3)—$\pi_s$ the view's probability of regime $s$, and $\Sigma_a^0$ the neutral asset covariance. Setting $\gamma = 0$ recovers the neutral inputs exactly. Only the systematic block of the covariance moves with the view; idiosyncratic risk is untouched. The CMA is the anchor throughout—the view can tilt it, never overwrite it—which is what keeps every candidate portfolio traceable to assumptions the institution has already debated and approved.

A careful reader may worry that Equation (6) counts the same premium twice—once inside $m_{\text{cma}}$ and again through $g\,B\lambda_f^0$, since $\lambda_f^0$ is itself extracted from the CMA. It does not, and the calibration of Section 5.3 is the reason. The scalar $g$ is solved so that the *total* expected return—CMA plus adjustment—prices the 60/40 calibration portfolio at exactly the Sharpe ratio the CIO chose. If that chosen Sharpe equals the one the CMA already implies, $g$ solves to zero and the expected returns are the approved CMA, untouched. If the house view is more optimistic than the CMA, $g$ comes out positive and adds the difference; more pessimistic, and $g$ comes out negative and removes it. The term $g\,B\lambda_f^0$ is therefore not a second helping of the CMA's premium but the *gap* between two prices of risk—the one the CMA embeds and the one the house holds—expressed along the factor structure. $m_{\text{cma}}$ is the raw CMA, always; the adjustment is marginal by construction, and zero when there is nothing to adjust.

Keying off a single CMA is a deliberate choice, and it earns its keep three ways. It keeps the whole institution on one baseline: the committee debates a tilt to numbers it has already ratified, not a fresh set of machine-generated assumptions each cycle. It keeps the candidate explainable: because the construction is fixed mathematics, "why did the portfolio move?" always traces back to the two view numbers. And it is forgiving: because the entire construction runs through the factor allocation approach, it inherits the stability results of Section 2, which neutralizes the usual concerns about building portfolios from a CMA—the framework does not depend on anyone, machine or human, getting expected returns exactly right. An investor who remains concerned can always elect to run the same construction under multiple CMAs; that adds a further layer of robustness to the analysis, and further confidence in the results.

### 5.3 The calibration portfolio: aligning the price of risk

The scalar $g$ in Equation (6) calibrates the overall level of factor risk premia layered on top of what the CMA already implies—the marginal price of risk embedded in the expected returns this analysis runs on. Rather than assert that level abstractly, we let the human calibrate it through their assumption on a portfolio everyone can reason about: a 60/40 equity/bond mix. The user supplies one number, $\hat{S}$—the Sharpe ratio a 60/40 portfolio should earn going forward—and $g$ solves in closed form so that the 60/40 earns exactly that (Appendix C):

$$g = \frac{\hat{S}\,\sigma_{60/40} - w'_{60/40}\,m_{\text{cma}}}{w'_{60/40}\,B\lambda_f^0}, \tag{7}$$

where $w_{60/40}$ is the weight vector of the 60/40 portfolio and $\sigma_{60/40}$ its volatility under the neutral covariance. We call the 60/40 a **calibration portfolio**, and the design deserves a word because it is one of the framework's most practical features. A CMA arrives with a market risk premium already baked in, and the CIO may or may not agree with it. Equation (7) gives the CIO a direct dial: if the house view is that a balanced portfolio earns a 0.35 Sharpe over the horizon, set $\hat{S} = 0.35$ and the entire premium structure scales consistently to match—a view on the market's price of risk, imposed in one interpretable number, without hand-editing thirty expected returns. We use a 60/40 for the dial because

most investors can *visualize* what a 60/40 Sharpe ratio is far more readily than that of any single asset class; its historical Sharpe is trivially computed from public data, so anyone can form a defensible forward-looking judgment about it. Because $g$ is calibrated on the neutral inputs, it is independent of the regime view: the view tilts the composition of returns, while the calibration sets their level.

### 5.4 The target, expressed as a risk budget

With the view-conditioned inputs in hand, the target factor exposure is the reverse-optimized factor weight of the optimal portfolio at those inputs, normalized so its factor-implied volatility equals the systematic share of the risk budget:

$$\theta^{\text{tgt}} = \phi\,\sigma^{\star}\,\frac{\theta^{\star}}{\|\theta^{\star}\|_{\Sigma_f}}, \qquad \theta^{\star} \propto \Sigma_f(\pi,\gamma)^{-1}(B'B)^{-1}B'\,\mu_a(\pi,\gamma), \tag{8}$$

where $\theta^{\star}$ is the reverse-optimized factor weight of the optimal portfolio at the view-conditioned inputs, $B$ is again the matrix of asset factor loadings, $\|\theta\|_{\Sigma_f}$ denotes the factor-implied volatility of a factor-weight vector, and $\sigma^{\star}$ is the portfolio volatility target—**the level of risk the CIO wants to run, which must sit within the risk appetite set by the CRO and the Board**. The scalar $\phi$ acknowledges that a real multi-asset portfolio carries some risk the factor set does not span; since the candidate portfolio is not known upfront, the share of its risk the macro factors will explain has to be estimated in advance, and past experience suggests $0.9$ is an informed guesstimate for a diversified long-only multi-asset portfolio. Equation (8) is where the risk appetite of the institution enters the mathematics as a hard scale on everything the view is allowed to do.

### 5.5 A feasible portfolio by constrained search

What remains is to find an actual portfolio—long-only sleeves, allocation bands, the CIO's minimums and maximums—whose reverse-optimized factor weights come as close as possible to the target. A closed-form solution exists in the unconstrained case (Appendix C), but real mandates make it inapplicable without distorting exactly the exposures it was built to hit. So the engine searches instead:

1. **Sample** a large number of candidate portfolios (tens of thousands) approximately uniformly over the constraint set $\mathcal{W}$—every draw respects every band by construction.

2. **Bracket the risk target:** pair sampled portfolios from just below and just above the volatility target and blend each pair so the blend lands on the target almost exactly—portfolio volatility is not linear in the weights, so a linear blend of a below-target and an above-target portfolio lands a shade under the target rather than exactly on it; the gap is negligible in practice.

3. **Score** every blend by its squared distance to the target in factor-weight space, $e^{(j)} = \|B_i w^{(j)} - \theta^{\text{tgt}}\|^2$.

4. **Average the best matches**—the top $N\%$ of blends with the least error, where $N$ is chosen by the user (we use the lowest-error few percent)—to form the candidate portfolio $\bar{w}$.

Averaging a neighborhood of near-optimal portfolios, rather than picking the single closest match, is deliberate: because the match is scored in the stable coordinate system of Equation (5), the portfolios in that neighborhood carry almost identical factor bets while differing in composition, so the average preserves the target exposure and buys diversification and noise-damping for free. The procedure is embarrassingly parallel, transparent to audit—a candidate's score involves nothing more than applying Equation (5) to its weights to read off its factor weights, subtracting the target factor weights, and squaring the difference—and indifferent to how gnarly the constraint set is, which is precisely where institutional reality lives.

### 5.6 The no-view twin

Every run also constructs a companion portfolio by the identical procedure with the view switched off—$\gamma = 0$, same CMA, same bands, same volatility target. The difference $\bar{w}^{\text{view}} - \bar{w}^{\text{no-view}}$ is the pure effect of the view: not the risk target, not the constraints, not the CMA—the view. In committee, this delta is the object of debate, and we would argue it is the single most useful exhibit the framework produces. It converts "what does the machine think?" into "here is precisely what the machine's view would change, in percentage points, within your limits."

### 5.7 Remarks

Three properties carry the practical weight. **Stability:** targeting in reverse-optimized factor space inherits the estimation robustness documented in the underlying papers—reasonable CMA in, sensible portfolio out. **Interpretable controls:** everything a human sets is a number a human can defend—two view numbers $(\pi, \gamma)$, one price-of-risk number $(\hat{S})$, one risk number $(\sigma^\star)$, and the bands. **Isolation:** the no-view twin makes the view's contribution visible and debatable. The two assumptions most deserving of scrutiny are the systematic-risk share $\phi$ and the sampling approximation itself; both are examined in Appendix C, and both matter less than they might because the objective is stated in the stable coordinate system.

## 6 Implementation: From Human Specification to Prototype

The methodology of Section 5 runs in an interactive application, and the way that application came to exist is itself evidence for this paper's thesis. The authors wrote the specification—the mathematics above, first prototyped as a single-scenario spreadsheet with every matrix and formula explicit. AI-assisted tools helped translate that author-specified methodology into software: a self-contained web application that freezes the same inputs, reproduces the spreadsheet's computation exactly (validated cell for cell at matched settings), and then extends it into the Monte-Carlo search of Section 5.5—tens of thousands

of feasible draws, risk-target bracketing, top-match averaging—with a live interface: a clickable regime map for setting $\pi$, steppers and sliders for the confidence, the risk target, and every allocation band, and a side-by-side display of the with-view and no-view portfolios with their differences and distribution whiskers. In the setting used throughout this paper, the engine generates 20,000 initial random portfolios (step 1 of Section 5.5), bootstraps them into roughly 100,000 risk-target blends (step 2), and averages the best 2,000 (step 4)—and on a modern laptop the whole run generally completes in under half a second. This setup is, in effect, portfolio construction in real time: change a probability, a band, or the confidence dial, and the candidate refreshes as fast as a committee can ask the question.

Two observations from the build, offered as experience rather than proof. First, the compression is real: what would have been a months-long internal development effort was a matter of days, and the binding constraint was never the coding—it was the completeness of the specification. Every hour spent making the spreadsheet prototype unambiguous repaid itself immediately; the AI implements what is specified, including any ambiguity. Second, the pattern is the same one that governs Step 1: **the human sets the frame, the machine works within it.** Specification-first AI development is the software version of ratify-the-parameters-then-let-the-agents-run—and in our experience it is the difference between AI as a demo and AI as infrastructure. The parallel is worth a CIO's attention: the skills an investment organization needs for this framework are mostly skills it already has—writing down what it believes, precisely.

## 7 Illustration: From an AI View to a Candidate Portfolio

This section runs the framework end to end. Nothing here is investment advice, and nothing here is a performance claim—the purpose is to let the reader *watch the machinery behave*: a real AI view becoming a portfolio tilt, two stress views probing the uncomfortable quadrants, a comparison of two different ways to target risk, and a confidence sweep showing exactly what $\gamma$ does. All numbers were produced by the application of Section 6, run headlessly from its own code with a fixed random seed, so every figure is exactly reproducible.

### 7.1 Setup

The illustration uses a 30-asset universe reported at 14 asset-class lines. Expected returns are a modified J.P. Morgan CMA viewed from a Canadian investor's perspective—a blend of the US, Canadian, and European assumption sets—with most assets currency-hedged except emerging-market equity. The governance inputs are deliberately generic (Table 2): broad 20–50% bands on the private, equity, and fixed-income groups, a tighter 0–10% band on the inflation-sensitive group, relative preferences within each group, a 0.35 Sharpe on the 60/40 calibration portfolio, and a 9% volatility target. It is worth noting that the sampled portfolios use a modest amount of leverage—total non-currency weights run somewhat above 100% (the no-view portfolio totals about 114%). We count non-currency weights because currency exposures are overlays: they are either implicit in the assets

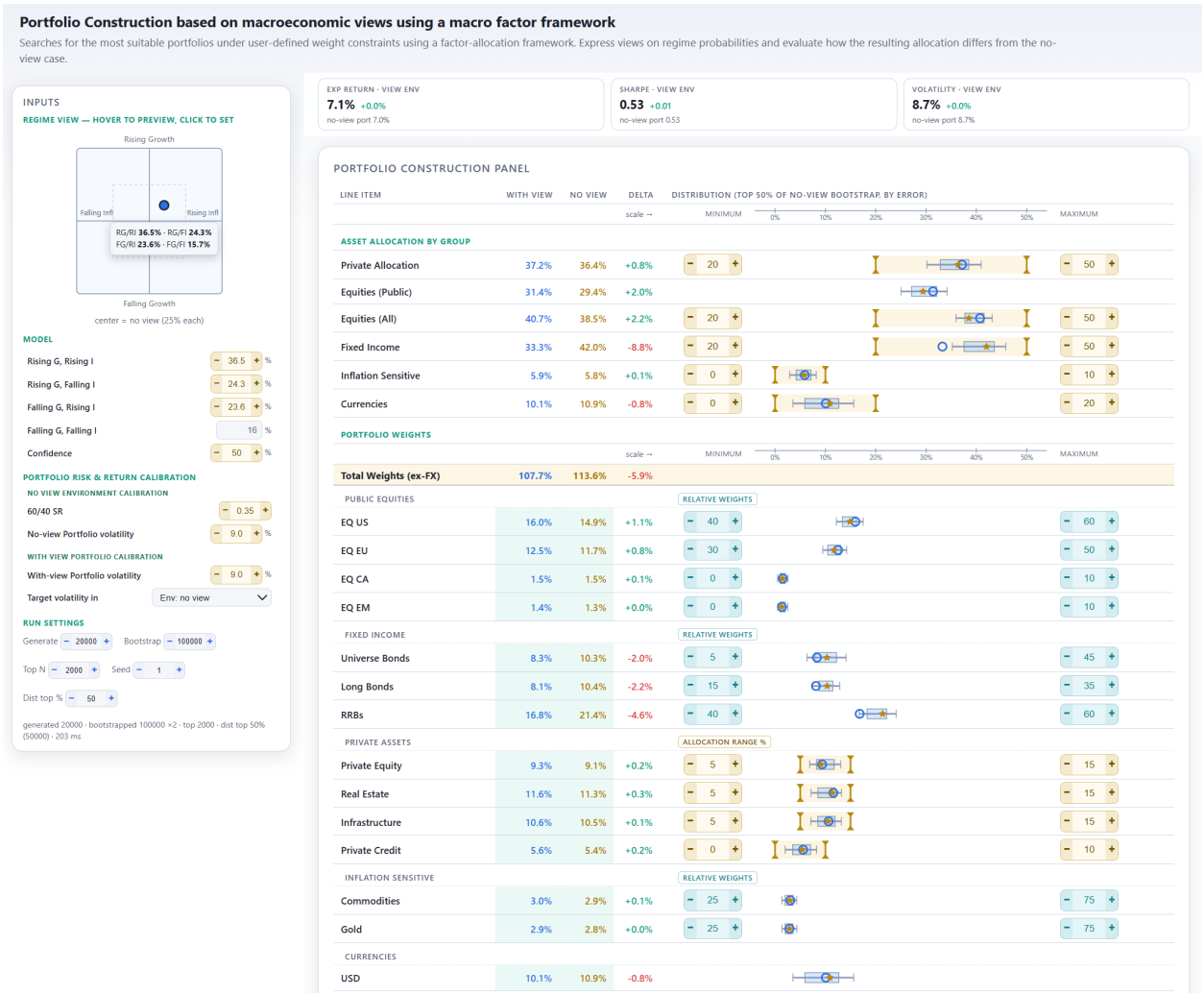



**Figure 4: The portfolio construction application.** The left panel holds the inputs: a regime map that sets $\pi$ by hovering and clicking, the four regime probabilities and the confidence $\gamma$, the 60/40 calibration Sharpe, the volatility targets with the environment selector, and the run settings; every allocation band in the construction panel is directly editable, and the bands shown are the illustration settings of Table 2. In this example the user has clicked into the rising-growth / rising-inflation quadrant—the callout reads RG/RI 36.5%, RG/FI 24.3%, FG/RI 23.6%, FG/FI 15.7% (RG/FG for rising and falling growth, RI/FI for rising and falling inflation)—at 50% confidence, and the tool prices what that view changes relative to the no-view portfolio: public equities rise to 31.4% from 29.4%, fixed income falls to 33.3% from 42.0%, leverage comes down (total non-currency weights 107.7% against 113.6%), and the cards across the top report both portfolios' expected return, Sharpe ratio, and volatility in the chosen environment. The whiskers beside each line show the distribution of near-optimal portfolios and are clickable and draggable, so a user can interrogate any allocation by hand. The no-view column differs from Table 3's Portfolio-"0" only in the last decimal, because the application runs a different random seed—a small live demonstration of how little the sampling noise matters. The run itself completed in about a fifth of a second. The application is self-contained and accompanies this paper as an interactive demo.

themselves or implemented with FX derivatives. A small amount of leverage of this kind is typical for most institutional investors.

One further choice deserves attention, because it turns out to matter: *where the 9% is*

**Table 2: Illustration settings.** The bands express the CIO's preferences, set within the risk appetite of the CRO and the Board; the 60/40 Sharpe is the house view on the market price of risk. Stress-scenario probabilities are round numbers by design.

| Setting | Value |
|---|---|
| Asset class group bands | Private 20–50%; All equities 20–50%; Fixed income 20–50%; Inflation-sensitive 0–10%; Currencies 0–20% |
| Equity preferences (relative) | US 40–60%; Europe 30–50%; Canada 0–10%; EM 0–10% |
| Bond preferences (relative) | Universe 5–45%; Long 15–35%; Real-return 40–60% |
| Inflation-sensitive preferences (relative) | Commodities 25–75%; Gold 25–75% |
| Private assets (individual) | PE 5–15%; Real estate 5–15%; Infrastructure 5–15%; Private credit 0–10% |
| Calibration & risk | 60/40 Sharpe $\hat{S} = 0.35$; volatility target $\sigma^\star = 9\%$, applied two ways: measured in the no-view environment, and measured in the view environment |
| Scenario A — the AI view (June 28, 2026) | From Figure 3 via Equation (4): $\pi = (42.0, 28.6, 17.6, 11.9)\%$ over (G↑I↑, G↑I↓, G↓I↑, G↓I↓); $\gamma = 0.5$ for Medium conviction (an illustrative mapping, stated as such) |
| Scenario B — stagflation | $\pi = (15, 5, 65, 15)\%$; $\gamma = 0.8$ |
| Scenario C — recession | $\pi = (5, 15, 15, 65)\%$; $\gamma = 0.8$ |

*measured*. A view does not just move expected returns—it moves the covariance too, so the same portfolio carries a different volatility depending on whether risk is computed under the neutral assumptions or under the view-tilted ones. We therefore run every scenario twice: once targeting 9% measured in the **no-view environment** (the neutral covariance), and once targeting 9% measured in the **view environment** (the covariance the view itself implies). The committee is really making two decisions, not one—what to hold, and how much risk to run against which measuring stick—and we will see below that the second decision can change the answer to the first.

### 7.2 What the views do

The results live in Tables 3 through 6 and Figure 5, and they are easiest to read with a short map in hand—so here it is. There are four portfolios throughout: Portfolio-"0", built with no view at all, and one candidate for each of the three views. Table 3 shows what each portfolio holds—and how it differs from Portfolio-"0"—when the 9% risk target is measured in the no-view environment. Table 4 then prices those same four portfolios under different assumed worlds. Portfolio-"0" is evaluated under all four environments—the neutral one and each of the three views—because the same holdings carry a different volatility, expected return, and Sharpe ratio depending on which world you assume. Each candidate is evaluated twice: once under the neutral environment, which speaks the language of the mandate ("what does this portfolio look like under our approved assumptions?"), and once under its own view, which speaks the language of the bet ("what does it look like if the

**Table 3: Allocations of the four portfolios, with the 9% risk target measured in the no-view environment** (percent of portfolio). Portfolio-"0" is identical across scenarios by construction—same CMA, bands, risk target, and seed—so every difference from it is attributable to the view alone. Totals exceed 100% ex-currency because the portfolios carry modest leverage. Deltas are computed before rounding. The bars read like a spreadsheet's data bars: blue scales with the weight (a full bar would be 50%), green and red scale with the size of each positive and negative tilt.

| | Portfolio "0" (no view) | Portfolio "view-A" (AI view) | Δ | Portfolio "view-B" (Stagflation) | Δ | Portfolio "view-C" (Recession) | Δ |
|---|---|---|---|---|---|---|---|
| **Equities (public)** | 29.3 | 36.3 | +7.0 | 22.2 | −7.1 | 31.2 | +1.8 |
| US | 14.9 | 18.6 | +3.7 | 11.1 | −3.8 | 15.7 | +0.8 |
| Europe | 11.7 | 14.7 | +3.0 | 8.6 | −3.1 | 12.1 | +0.4 |
| Canada | 1.4 | 1.5 | +0.1 | 1.3 | −0.2 | 1.6 | +0.2 |
| Emerging markets | 1.4 | 1.6 | +0.2 | 1.2 | −0.1 | 1.8 | +0.4 |
| **Fixed income** | 42.0 | 28.6 | −13.5 | 43.6 | +1.5 | 47.7 | +5.7 |
| Universe bonds | 10.2 | 7.4 | −2.7 | 10.5 | +0.4 | 11.6 | +1.4 |
| Long bonds | 10.4 | 6.8 | −3.6 | 10.8 | +0.4 | 13.1 | +2.7 |
| Real-return bonds | 21.5 | 14.3 | −7.2 | 22.3 | +0.8 | 23.0 | +1.5 |
| **Private allocation** | 36.7 | 35.3 | −1.4 | 40.3 | +3.6 | 31.6 | −5.1 |
| Private equity | 9.2 | 9.1 | −0.1 | 11.1 | +1.8 | 8.4 | −0.9 |
| Real estate | 11.3 | 11.0 | −0.3 | 11.4 | 0.0 | 8.5 | −2.8 |
| Infrastructure | 10.5 | 9.8 | −0.7 | 11.7 | +1.3 | 9.7 | −0.7 |
| Private credit | 5.6 | 5.3 | −0.3 | 6.1 | +0.5 | 4.9 | −0.7 |
| **Inflation-sensitive** | 5.9 | 2.3 | −3.6 | 8.4 | +2.5 | 5.5 | −0.4 |
| Commodities | 3.0 | 1.2 | −1.8 | 4.1 | +1.1 | 2.6 | −0.4 |
| Gold | 2.9 | 1.1 | −1.8 | 4.3 | +1.4 | 2.9 | 0.0 |
| **Currencies (USD)** | 11.6 | 11.6 | 0.0 | 5.5 | −6.0 | 12.1 | +0.5 |
| **Total ex-currency** | 114.0 | 102.5 | | 114.5 | | 116.0 | |

view is right?"). Tables 5 and 6 repeat the entire exercise with the 9% measured in each view's own environment. With that map in hand, on to the views themselves.

**Scenario A — the real view.** The June 2026 pipeline output (Figure 3) reads: growth likely above trend, inflation modestly tilted higher but truly two-sided, conviction Medium—so $\gamma = 0.5$, a deliberate half-expression. The engine's response (Table 3, Figure 5) is the moderate pro-growth tilt a practitioner would sketch by hand, executed with discipline: public equities up 7.0 points, funded from duration—fixed income down 13.5 points, spread across universe bonds, long bonds, and real-return bonds—with the small inflation-sensitive sleeve trimmed as well and the private allocation left nearly where it was. Measured in the view's own environment, expected return improves by about 9 basis points at essentially the same risk. The tilt is decisive but not heroic—which is the correct translation of a Medium-conviction, two-sided view. The machine did not manufacture certainty it did not have.

**Scenario B — stagflation.** Throughout this scenario, the comparison is between portfolios sized to the same volatility measured in the no-view environment (Table 3). Falling growth with rising inflation is the allocator's least comfortable quadrant: equities

**Table 4: Portfolio characteristics for the four portfolios of Table 3.** Each portfolio is priced under different assumed worlds: Portfolio-"0" under all four environments, and each candidate under the neutral environment and under its own view. Blank cells are combinations not relevant to that portfolio.

| **Characteristic** | **Portfolio "0"** (no view) | **Portfolio "view-A"** (AI view) | **Portfolio "view-B"** (Stagflation) | **Portfolio "view-C"** (Recession) |
|---|---|---|---|---|
| *Volatility (%)* | | | | |
| under no view | 8.9 | 9.0 | 8.9 | 8.9 |
| under view A | 8.6 | 8.6 | | |
| under view B | 9.0 | | 9.0 | |
| under view C | 9.9 | | | 10.0 |
| *Expected return (%)* | | | | |
| under no view | 6.4 | 6.4 | 6.4 | 6.3 |
| under view A | 7.5 | 7.6 | | |
| under view B | 4.2 | | 4.6 | |
| under view C | 3.7 | | | 3.8 |
| *Sharpe ratio* | | | | |
| under no view | 0.45 | 0.44 | 0.45 | 0.43 |
| under view A | 0.59 | 0.60 | | |
| under view B | 0.20 | | 0.24 | |
| under view C | 0.14 | | | 0.15 |

and nominal bonds tend to fail together. Expressed at $\gamma = 0.8$, the engine's equity move is best read from the top down. Overall equities—public plus private equity—decrease by about five points. Within them, the mix shifts toward privates: public equities are sold hard ($-7.1$) while the private allocation is added to ($+3.6$). Only about half of that private add goes to private equity ($+1.8$, so only slightly); the rest goes to infrastructure ($+1.3$) and private credit ($+0.5$), with real estate left unchanged—and infrastructure in particular is the kind of real asset whose inflation pass-through makes it a better stagflation holding than equities. The engine also adds to the inflation-sensitive sleeve ($+2.5$, split across gold and commodities) and sells the US dollar ($-6.0$). Inside fixed income, a critic could note that the weights look symmetric—nominal bonds add $+0.4$ apiece for universe and long bonds, and real-return bonds add $+0.8$—but the weights understate what actually moved. The real-return bonds in this exercise are long-term instruments, with duration far above universe bonds and slightly above long bonds, so the same weight buys considerably more rate exposure: on a risk basis, duration is being shifted from nominal to real—exactly what the quadrant calls for, since stagflation is the environment in which nominal duration disappoints and inflation-linked duration is the better defense. The more interesting exhibit is what the engine *could not* do. The unconstrained factor target called for outright negative growth exposure and a large positive inflation exposure; the candidate keeps its growth exposure positive and captures only a fraction of the inflation exposure—because the long-only sleeves, the group minimums, and the 0–10% cap on the inflation-sensitive group are hard constraints. **The bands visibly cap the tilt.** A committee reading this sees governance working in the numbers themselves: conviction proposes, the mandate disposes.

**Scenario C — recession.** As in Scenario B, the comparison is between portfolios sized to the same volatility measured in the no-view environment (Table 3). Falling growth

**Table 5: Allocations for the same three views, with the 9% risk target measured in the view environment instead** (percent of portfolio). Portfolio-"0" is unchanged from Table 3; the only thing that has moved is the measuring stick for each candidate's risk. Scenario B barely notices; Scenario C changes materially. Bars read as in Table 3.

| | Portfolio "0" (no view) | Portfolio "view-A" (AI view) | Δ | Portfolio "view-B" (Stagflation) | Δ | Portfolio "view-C" (Recession) | Δ |
|---|---|---|---|---|---|---|---|
| **Equities (public)** | 29.3 | 37.0 | +7.7 | 21.6 | −7.7 | 24.8 | −4.5 |
| US | 14.9 | 18.9 | +4.1 | 10.9 | −4.0 | 12.4 | −2.5 |
| Europe | 11.7 | 14.9 | +3.3 | 8.4 | −3.3 | 9.9 | −1.8 |
| Canada | 1.4 | 1.6 | +0.2 | 1.2 | −0.2 | 1.3 | −0.2 |
| Emerging markets | 1.4 | 1.6 | +0.2 | 1.2 | −0.2 | 1.2 | −0.1 |
| **Fixed income** | 42.0 | 29.7 | −12.3 | 43.9 | +1.8 | 47.3 | +5.3 |
| Universe bonds | 10.2 | 7.5 | −2.6 | 10.4 | +0.3 | 11.4 | +1.2 |
| Long bonds | 10.4 | 7.3 | −3.1 | 11.0 | +0.7 | 13.0 | +2.6 |
| Real-return bonds | 21.5 | 14.8 | −6.6 | 22.4 | +0.9 | 23.0 | +1.5 |
| **Private allocation** | 36.7 | 37.1 | +0.5 | 40.2 | +3.6 | 30.8 | −5.9 |
| Private equity | 9.2 | 8.8 | −0.4 | 11.1 | +1.9 | 8.3 | −0.9 |
| Real estate | 11.3 | 11.8 | +0.5 | 11.3 | −0.1 | 8.5 | −2.9 |
| Infrastructure | 10.5 | 10.3 | −0.2 | 11.8 | +1.3 | 9.6 | −0.9 |
| Private credit | 5.6 | 6.2 | +0.5 | 6.1 | +0.5 | 4.4 | −1.2 |
| **Inflation-sensitive** | 5.9 | 2.6 | −3.3 | 8.4 | +2.4 | 4.7 | −1.3 |
| Commodities | 3.0 | 1.3 | −1.7 | 4.1 | +1.1 | 2.2 | −0.8 |
| Gold | 2.9 | 1.3 | −1.6 | 4.2 | +1.3 | 2.5 | −0.5 |
| **Currencies (USD)** | 11.6 | 10.8 | −0.8 | 5.7 | −5.8 | 12.1 | +0.5 |
| **Total ex-currency** | 114.0 | 106.4 | | 114.1 | | 107.5 | |

and falling inflation at $\gamma = 0.8$ is a duration story, and the engine tells it: fixed income rises 5.7 points across the curve—long bonds +2.7, universe bonds +1.4, real-return bonds +1.5. On the equity side, start from the top: equities in the broad sense—public together with the private allocation—go down by about three points, because the engine simply does not want growth risk in a recession. Within that smaller block, it also lowers the proportion held in privates. The net result of those two decisions, when portfolios are sized to the same volatility in the no-view environment, is that the public equity line ends up slightly higher (+1.8) while the reduction falls mostly on the other risky assets—real estate, infrastructure, and private credit—with private equity itself only trimmed (−0.9). Target the 9% in the view environment instead (Table 5) and both public and private equities fall—because in a recession environment, risk rises. Table 5 is therefore doing more than reshuffling allocations within asset classes: it is also changing the portfolio's overall risk deployment. Portfolio-"view-C" in Table 5 de facto de-risks the whole portfolio, evident too in its gross exposure, which falls to about 108% of capital against Portfolio-"0"'s 114%. (For readers puzzled by that +1.8 public equity line: Appendix D re-runs this scenario with the private book pinned, and the entry disappears—public equities are then sold outright, confirming that the oddity belonged to the free private book, not to the view's opinion of equities.)

Now read Table 4, because it explains everything in Table 3. Start with Portfolio-"0"'s

**Table 6: Portfolio characteristics for the portfolios of Table 5**, priced the same way as in Table 4: Portfolio-"0" under all four environments, each candidate under the neutral environment and under its own view.

| **Characteristic** | **Portfolio "0"** (no view) | **Portfolio "view-A"** (AI view) | **Portfolio "view-B"** (Stagflation) | **Portfolio "view-C"** (Recession) |
|---|---|---|---|---|
| *Volatility (%)* | | | | |
| under no view | 8.9 | 9.3 | 8.9 | 8.0 |
| under view A | 8.6 | 9.0 | | |
| under view B | 9.0 | | 8.9 | |
| under view C | 9.9 | | | 8.9 |
| *Expected return (%)* | | | | |
| under no view | 6.4 | 6.5 | 6.4 | 6.0 |
| under view A | 7.5 | 7.8 | | |
| under view B | 4.2 | | 4.6 | |
| under view C | 3.7 | | | 3.9 |
| *Sharpe ratio* | | | | |
| under no view | 0.45 | 0.44 | 0.45 | 0.44 |
| under view A | 0.59 | 0.60 | | |
| under view B | 0.20 | | 0.24 | |
| under view C | 0.14 | | | 0.17 |

volatility column: the same holdings measure 8.9% in the neutral world, 8.6% under view A, 9.0% under view B, and 9.9% under view C. Nothing about the portfolio changed—only the assumed world did—and yet under the recession view it is a full point riskier than the mandate believes it to be. That one column is the reason the whole exercise is run twice. A committee acting on a view is making two decisions, not one: what to hold, and how much risk to run—and the second decision includes choosing the world in which risk is measured. Where the worlds agree, the choice is a technicality: under view B the two measurements nearly coincide, and the stagflation candidate comes out almost the same either way. Where the worlds disagree, the choice has teeth: under view C, insisting on 9% in the recession world (Tables 5 and 6) produces a visibly more defensive, less levered portfolio than the same view measured neutrally (Tables 3 and 4).

The Sharpe ratios of Tables 4 and 6 carry a quieter message worth pausing on. Each candidate beats Portfolio-"0" in its own world—0.60 against 0.59 under view A, 0.24 against 0.20 under B, 0.15 or 0.17 against 0.14 under C. That much should be expected: a portfolio tilted toward a world ought to win in that world. The more valuable reading is the other row. Under the neutral lens, the candidates sit at or slightly below Portfolio-"0"—0.43 to 0.45 against 0.45—and that small gap deserves a name: it is the *opportunity cost* of the tilt, what the portfolio stands to give up if the view is wrong and the neutral assumptions turn out to be the truth. This is one of the more useful things the tool does: it prices both sides of the bet at once—the expected gain in the world the view describes, and the expected cost in the world the mandate assumes. Here the cost happens to be small; it will not always be, and a committee should insist on seeing both numbers before it acts. Notice, too, what the levels themselves say: no portfolio makes stagflation or recession pleasant—the best available Sharpe in those worlds is a fraction of the neutral one—the candidates simply

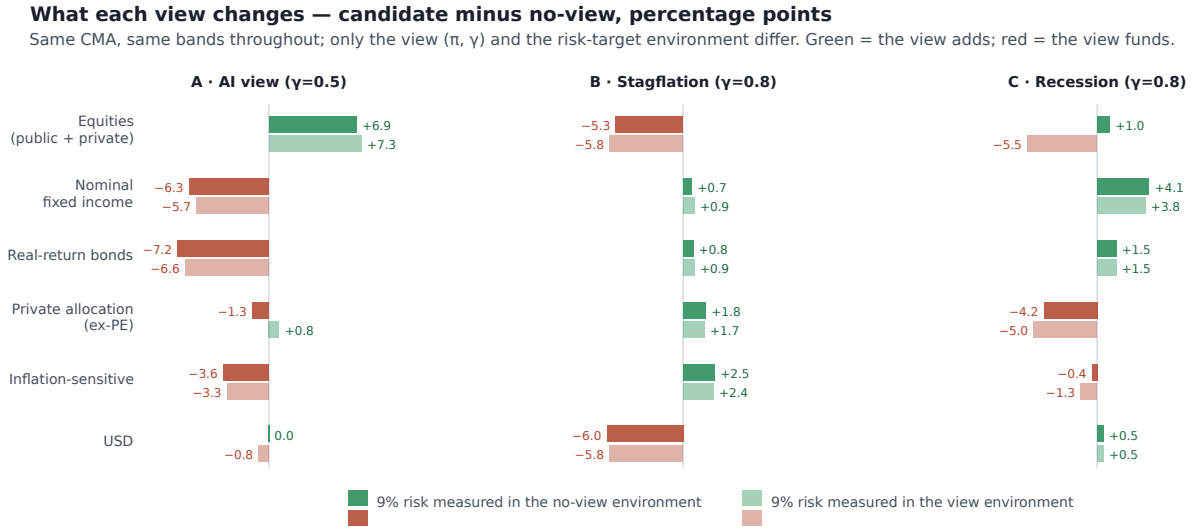



**Figure 5: What each view would change, under both risk-targeting choices.** Every bar compares a candidate portfolio against the same no-view portfolio—so each bar is precisely what that view, and nothing else, does to the allocation. The groupings: "Equities (public + private)" combines public equities with private equity; "nominal fixed income" is universe plus long bonds; "private allocation (ex-PE)" is the private allocation excluding private equity—real estate, infrastructure, and private credit. The darker bar targets 9% volatility measured under the neutral assumptions; the lighter bar targets 9% measured under the view's own assumptions. Where the two bars agree (A and B), the choice of measuring stick is a technicality; where they diverge (C, equities), it is a real decision. This, we would argue, is the right format for an AI's macro opinion to arrive at a committee: a set of tilts that are debatable, bounded by the mandate, and reversible.

lose less and defend better. None of this is a flaw in either risk-targeting convention; it is the framework making implicit decisions explicit. An institution that tilts on a view of the world should decide, deliberately, whether its risk budget is stated in the world it has, or in the world it believes is coming.

A final caution belongs here, precisely because the tables are so specific. The model produces exact-looking numbers, and they should never be taken as anything more than a *suggestion*. Whatever truth the model has to offer lives around each number, not at it. Nor will the weights always feel perfectly intuitive—Scenario C's public-versus-private split is a good example—and that is to be expected from any construction machinery working inside real constraints. The test a committee should apply is not "does every line match my instincts?" but "are the results generally reasonable, in magnitude and in the direction of the tilts?" When they are, the candidate has done its job: informing a decision that remains, as ever, the committee's to make.

### 7.3 What confidence does — and where it stops

The confidence sweep (Figure 6) re-runs Scenario A's probabilities with $\gamma$ from 0 to 1, holding everything else fixed (risk is measured in the no-view environment throughout). Three behaviors are worth reading off the chart. At $\gamma = 0$ the candidate coincides with

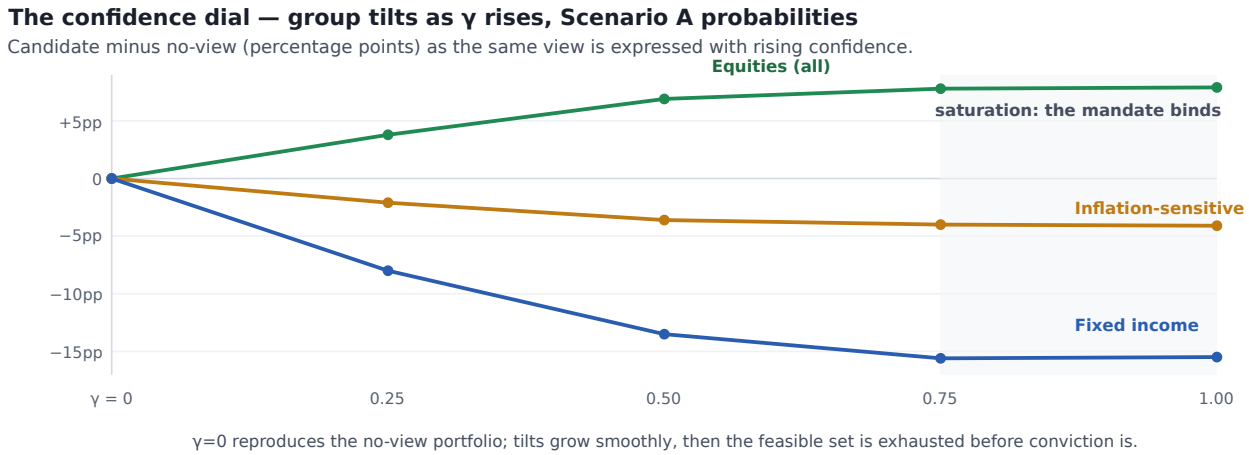



**Figure 6: What the confidence dial does.** The view is held fixed at the June 2026 probabilities; the only thing that changes along the horizontal axis is how strongly the committee chooses to act on it. At zero, the candidate simply *is* the no-view portfolio. As confidence rises, the engine leans further into the view—more equities, less duration—smoothly and predictably. And then the lines go flat: somewhere past $\gamma \approx 0.75$ the portfolio runs out of room, because the bands and the risk target will not let it lean any further. That flat stretch is worth showing to a board, because it is the picture of a mandate doing its job: no matter how sure the machine becomes, there is a maximum tilt it can buy, and this chart shows exactly where it sits. (Every point is a seeded run and exactly reproducible.)

the no-view portfolio—the anchor is exact, by construction. Tilts then grow smoothly with confidence: at $\gamma = 0.25$ the portfolio has sold 8.0 points of fixed income; at $\gamma = 0.5$, 13.5 points. And then, past $\gamma \approx 0.75$, **the tilts stop growing**: between $\gamma = 0.75$ and $\gamma = 1.0$ the equity add moves from +7.8 to +7.9 points and fixed income barely moves at all, because the portfolio has reached the edge of what the bands, the long-only sleeves, and the risk target jointly permit. The feasible set is exhausted before conviction is.

We did not design this saturation as a feature, but it is one. It means that within this framework there is a maximum tilt *no level of machine conviction can exceed*—the mandate, not the model, is the binding constraint at the extremes. For an investor asking "what is the worst the AI's view could do to the portfolio?", that question has a concrete, computable answer.

### 7.4 How the tool should be used in practice

The scenarios above let every asset class move, because the point was to watch the machinery. In practice, investors do not move everything when they tilt. The most obvious case is private assets: commitments take quarters to build and longer to unwind, so over any horizon a regime view speaks to, the private book is what it is. The tool accommodates this without any new machinery—set an asset class's minimum equal to its maximum, and it is *pinned*. The engine then treats the private allocation as given and reallocates only what can actually move, and the question changes from "what is the ideal portfolio under this view?" to the question a committee actually faces: *given our privates, how much equity, bonds, and commodities should we hold?* The result is more constrained, and for exactly that reason more realistic and more useful. The same trick handles exclusions: an investor

not permitted to hold gold or commodities pins them at a 0% relative weight, or pins the whole inflation-sensitive group at zero. This is worth pausing on, because it is a design point and not an accident: constraints are the engine's native language, so every practical restriction—illiquidity, mandate exclusions, transition limits—enters through the same dial the CIO's preferences already use, with no change to the mathematics.

Pinning also earns its keep as a diagnostic. Run the same view twice, once with the private book pinned and once with it free, and the difference between the two candidates is the price of immobility: what the locked allocation costs, in expected return and in the tilts the view could not express. A committee does not need to act on that number, but it is a useful thing to know—it tells the institution how much its illiquidity is constraining its ability to respond to the economy it believes is coming. We do not leave this as advice: Appendix D runs exactly this exercise, repeating the full illustration of Section 7 with the private book pinned at Portfolio-"0"'s holdings—the configuration we would expect an adopting institution to run first.

The second habit we would urge is to evaluate candidates under *both* risk-targeting conventions, because "how much risk to take" is a real question and the dial that answers it should not be turned by default. Targeting risk in the no-view environment is the natural governance choice: most institutional risk systems measure risk over long lookback windows, so the no-view environment is the closest analogy to what the fund's own risk report will say. Targeting the same risk in the view environment carries real investment content: it says the fund wants to run the same *level* of risk regardless of regime—in effect, risk-targeting the portfolio by regime rather than by calendar. We are not suggesting investors religiously target risk. We are suggesting they see clearly what Section 7.2 demonstrated: a portfolio choice under a view is two decisions, the mix and the risk, and an investor who examines only one of them has let the machinery make the other. Used this way, the tool is less a portfolio generator than a disciplined way of asking what-if questions—and the committee that asks both questions is the committee the framework was built for.

### 7.5 The committee frame

Picture the Monday this framework is built for. The weekly view arrives before the meeting does—Figure 3, with its probabilities, its bands, its stated conviction, and a written case that already carries the strongest argument against itself. The engine has turned it into the Scenario A column of Table 3: one candidate portfolio, at mandate risk, inside every band, sitting beside its no-view twin so everyone can see precisely what the view would change—a few points of equity here, a few points of duration there. Around the table, reactions differ, as they should. One member finds the growth call too sunny; she reads the same data as recession-shaded. Fine: her probabilities go into the same engine, and a few seconds later Scenario C's column sits beside Scenario A's—same mandate, same mathematics, her judgment alongside the machine's, as equals on the page. Now the committee is debating the only things worth debating: whose regime probabilities are better supported, and how much confidence the evidence has earned. The machine has done what machines do well—read everything, price the alternatives consistently, show its work. The

people do what only people can: weigh what the numbers cannot see, and choose. That meeting is the whole proposal of this paper, acted out in an hour.

## 8 Risks, Limits, and the Shape of Adoption

### 8.1 What can go wrong

Let us start with the failure modes, because pretending they do not exist is how trust gets lost. Large language models misread numbers, over-weight vivid stories, and occasionally state things that are simply not so. Our defenses are procedural rather than magical, and it helps to see how they stack. Harvesters report and never conclude, so a hallucinated interpretation has no doorway into the evidence. A sample of every evidence pack is spot-checked against original sources before any judge reads it. Twenty independent judges dilute any single misreading. The arithmetic runs in a pinned script rather than in a model's head. And conformance gates catch malformed output before it can propagate. These controls reduce model error; they do not eliminate it. Two deeper limits deserve the same candor. The pinned factor loadings and cross-variable correlations are estimated from history—they are the framework's most consequential modeling assumptions, and they can simply be wrong, which is one reason the output is always a band and never a point. And all twenty judges currently share one base model: the personas diversify emphasis, not the machinery underneath, so errors can correlate in ways the ensemble does not fully wash out—which also means the reported bands, built from judge dispersion, inherit the same blind spot. They are honest about the disagreement the ensemble can see, and silent about the error it cannot; a band is best read as a floor on the uncertainty, not a ceiling. Mixing base models across the judges is the obvious mitigation, and it is on our list. A related honesty concerns *where* determinism lives. It begins at the judges' recorded scorecards: everything downstream—the averaging, the mapping, the simulation—is pinned and seeded, and reproduces digit for digit. The judges themselves are large language models, and re-running one is not guaranteed to reproduce the same scorecard unless the model version, the prompts, and the sampling settings are all frozen—and even a frozen configuration is hostage to vendor behavior over time. This is why every run archives its evidence packs and scorecards: a past view is re-derived from the record, never by asking a model to answer twice. Finally, a distinction we repeat because it is easy to blur: determinism is not accuracy. Reproducibility guarantees that the same recorded inputs yield the same view; it says nothing about whether the view is right. We do not claim any of this makes the system bulletproof; no design does. It is our attempt—built with these failure modes in view, honest about the ones that remain, and arranged so that when something does slip through, it surfaces as a wider band or a flagged gap rather than a confident wrong answer.

### 8.2 How the view will be judged

How will anyone know whether the views are any good? Not by backtest—and it is worth being clear about why. A model trained on an internet-scale corpus has, in effect, already read history; ask it to "forecast" 2019 and it is reciting, not predicting, so

any historical simulation is contaminated by lookahead. The only clean test is the slow one: live, dated, versioned forecasts, scored as the outcomes arrive. The system is built for exactly that discipline—every view ships with falsification triggers and is written to a time series with fields waiting for the realized outcome—and whether AI regime views carry skill is a question the industry will settle over the coming years, with this system among the entrants. In the meantime, the framework's protection does not depend on the answer: the human gate, the bands, and the risk target bound what any view, skilled or not, can do to the portfolio.

### 8.3 The questions investors ask, and the design's answers

Raise this framework with any investment organization and the same handful of concerns surfaces, almost word for word—we have heard each of the questions in Table 7 more than once, and every one of them deserves a real answer. The table pairs each question with the specific design mechanism that addresses it. The point, in every row, is that the answer is built into the architecture rather than promised in a policy: a structure that cannot misbehave in a particular way is worth more than an assurance that it will not.

### 8.4 Why we keep the chair human

Keeping the chair human is a choice, and we want to make the case for it affirmatively rather than timidly. Accountability for an institution's allocation is legally and culturally human, and no board memo written by an agent changes that. Regime calls are low-frequency, high-stakes judgments where context, humility, and a willingness to be persuaded still matter—these are the parts of the job people are for. And the marginal value of AI today lies overwhelmingly in breadth of evidence-processing: the pipeline reads more in a week than a macro team reads in a quarter, and that is where the gains are—not in delegating the final synthesis. So: AI at the table, not in the chair, for now. We will also say plainly what would move us. A multi-year live record of well-calibrated regime views would justify extending trust—perhaps first by letting the view adjust its own $\gamma$ within a ceiling the committee sets. The boundary is a policy dial, not a wall, and it should move the way everything in this framework moves: on evidence.

### 8.5 What it takes to build this — and what it pays

What does it actually take to build one of these? Less than an institution might fear—and the order of construction matters more than the budget. **Build Step 2 first.** The construction engine is transparent mathematics over things the institution already owns—an approved CMA, allocation bands, a risk target—and it is useful the day it is finished, with purely human views and no AI anywhere in sight. It is also the cheapest way to find out whether the framework fits the institution's decision culture. **Add the view pipeline second**, starting from the design rules of Section 4. They are model-agnostic, and in our experience the architecture matters more than the model behind it. The real intellectual property is the frame—the input menus, the regime definitions, the factor mapping—and writing it down precisely is mostly work the institution's economists have

**Table 7: The questions investors ask about AI in the allocation process, and the design mechanisms that answer them.** Every answer is structural—a property of how the system is built, not a promise about how it will behave.

| The question investors will ask | The design answer |
|---|---|
| "What if the AI goes rogue—changes its own rules?" | The scope guard: self-repair may touch only operational logic (collection recipes, formatting). Regime definitions, thresholds, rubrics, and the factor mapping are pinned; changing them requires a logged, human-ratified decision. The analysis cannot drift silently. |
| "What about hallucination?" | Collectors never interpret; readings are spot-checked against original sources before use; twenty independent judges score the same evidence; all arithmetic runs in pinned scripts, not in a model's head. Residual error shows up as dispersion—which widens the reported band rather than hiding inside a confident point estimate. |
| "Who is accountable for the allocation?" | The same people as today. The AI produces a view; the engine produces a candidate; the human committee decides. Nothing in the framework can move a dollar. |
| "Can we audit it?" | Every number traces through the synthesizer, the judges, and the evidence packs to a dated, cited source. Every run archives its evidence packs and judge scorecards, and the numeric layer is seeded and deterministic—so any past view re-derives exactly from its archived artifacts. Audit replays the record; it does not re-ask the model. |
| "How do we keep humans genuinely in the loop, not rubber-stamping?" | The artifacts are designed for challenge: a no-view twin that isolates the view's effect, bands that cap it, an adversarial passage in every judge report, and a human-view path through the same engine so disagreement is cheap to express. (On rubber-stamping as an automation failure mode, see Parasuraman et al., 2000.) |
| "What is the worst the AI's view can do to the portfolio?" | A computable answer: the maximum tilt inside the bands at the risk target (Section 7.3). Beyond a confidence level the mandate binds—no conviction can exceed it. |

already half-done in their heads; the exercise is getting it onto paper. One researcher who owns that frame, plus commodity AI agents, is a sufficient team. The tooling, as Section 6 showed, can largely be developed with AI assistance from a sufficiently precise human specification.

And the payoff? We would frame it as productivity, not headcount. The pipeline does not replace an investment team; it takes the lowest-value hours of the team's week—the

reading, the tracking, the collating of a firehose no team was ever going to cover—and hands them back as judgment hours. The same professionals cover more ground, notice contradictions sooner, and walk into committee with a disciplined, already-benchmarked view waiting to be challenged. The team produces more insight, not fewer people.

### 8.6 Ideas for real-world implementation, and future work

Where could an adopter take this next? A few directions seem natural to us, and we offer them as ideas rather than prescriptions. The cadence is free: the pipeline can run weekly, monthly, or on whatever rhythm the committee actually meets, with the view arriving as one voice at the table. Even under a monthly decision cadence, it is worth letting the pipeline run weekly and tracking the candidates through time—presenting a running average rather than a single snapshot insulates decisions from any one run, and surfaces model drift early. Once a history of weekly probabilities has accumulated, a second-stage model—a hidden-Markov or Kalman-type filter—can estimate the smoother path beneath the weekly reads and feed that to the engine instead. The independence and equal-split conventions of Equation (4) are improvable: a future system can estimate the joint growth–inflation surprise distribution directly. Mixing base models across the judges would blunt the monoculture risk of Section 8.1. And the simplest extension of all: put humans through the same pipe—use the tool to elicit and discipline the committee's own views, with or without any AI upstream.

And yet the most important item on this list is none of the above. It is simply to try. We said in the introduction that standing still is the one posture an investor cannot afford; Lauren Bacall is said to have put it better— *"standing still is the fastest way of moving backwards in a rapidly changing world."* Inaction is not an option; the way forward is to try, to keep score, and to learn from what the track record says. This paper is a blueprint for starting.

## 9 Conclusion

AI is a buzz, and buzz is not an allocation process. This paper proposed one concrete way to put AI to work in asset allocation, shaped by how allocation is actually practiced: form a view, then build the portfolio. We let AI do the step it is genuinely better at—reading everything, every week, and distilling it into probabilistic regime views that carry their own uncertainty—and we kept construction as transparent mathematics under explicitly human constraints, with a no-view twin that shows a committee exactly what the view would change and bands that bound what any view can do. The AI holds a seat at the table. The chair, the risk limits, and the decision remain where accountability lives: with people.

We have claimed no track record, because none can honestly be claimed yet: the skill of AI regime views will be settled by live, dated forecasts over years, and this system is built to be scored, not believed. What we do claim is that the architecture is ready—every piece an institution needs is described in these pages, and none of it waits on a research breakthrough. The two-number contract between view and portfolio. The human boundary

no agent can cross. The rule that consensus and positioning set the benchmark but never count as evidence. The scope guard that lets the plumbing improve itself without ever touching the analysis. The calibration portfolio that turns a philosophical argument about the price of risk into one debatable number. The no-view twin that shows a committee exactly what a view would change, and nothing else. All of it is reproducible from this paper, and the construction engine earns its keep from the day it is built, with or without an AI upstream. The natural next steps—running at the committee's own cadence, averaging candidates through time, filtering the weekly probabilities into their smoother path—are engineering, not research.

Like it or not, every institution will have to answer the question this paper started with, because everyone else is already trying. Our answer is neither to look away from AI nor to hand it the wheel. It is to give the machine the reading and the discipline, keep the judgment and the accountability, and let the track record—honestly dated, honestly scored—tell us when to move the boundary. The goal, for now, is not a machine at the wheel. It is a driver with a better view of the road.

## DISCLAIMER

The views and opinions expressed in this study are solely those of the authors and do not necessarily reflect the official policy or position of any other agency, organization, employer, or company, including the authors' employer in the asset management industry. The information provided in this study is for general informational purposes only and should not be construed as legal, tax, investment, financial, or other professional advice.

## ACKNOWLEDGMENT

The authors would like to thank Michael Wissell for his support on innovation and research.

## DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES

This paper examines the use of artificial intelligence in institutional asset allocation. The research question, conceptual contribution, economic regime framework, portfolio-construction methodology, governance design, interpretation of results, and conclusions were developed by the human authors.

Generative AI and AI-assisted technologies were used in three limited and disclosed ways.

First, AI agents were used as part of the research setting described in the paper. The authors designed the AI experiment, including the analytical objective, workflow architecture, economic framing, role definitions, governance boundaries, and portfolio methodology. Within that author-designed framework, AI agents were used to demonstrate

how macroeconomic evidence could be gathered, organized, and synthesized into structured regime views. These AI-generated outputs are treated as experimental inputs within the methodology, not as independent authorial conclusions, investment recommendations, or substitutes for human judgment.

Second, AI-assisted tools were used to support implementation. The authors specified the mathematical prototype, workflow requirements, portfolio-construction logic, and user-interface requirements. AI-assisted tools were then used to help translate those author-defined specifications into a working software prototype. The resulting tool was reviewed and tested by the authors against the intended methodology.

Third, AI-assisted tools were used during manuscript preparation to improve grammar, readability, formatting, and editorial clarity. They were not used to originate the paper's research question, economic model, portfolio methodology, analytical conclusions, or investment interpretation.

All analytical design choices, model assumptions, workflow specifications, methodological decisions, empirical illustrations, interpretation, and conclusions were reviewed, revised, and approved by the human authors. The authors take full responsibility for the content of the paper. No AI system is listed as an author, cited as an author, or assigned responsibility for any part of the research.

## REFERENCES


Ang, A., N. Azimbayev, and A. Kim. 2026. "The self-driving portfolio: Agentic architecture for institutional asset management". SSRN Working Paper No. 6504386.

Ang, A. and G. Bekaert. 2004. "How regimes affect asset allocation" *Financial Analysts Journal* 60(2): 86–99.

Black, F. and R. Litterman. 1992. "Global portfolio optimization" *Financial Analysts Journal* 48(5): 28–43.

Du, Y., S. Li, A. Torralba, J. B. Tenenbaum, and I. Mordatch. 2023. "Improving factuality and reasoning in language models through multiagent debate". arXiv preprint arXiv:2305.14325.

Elkamhi, R. and J. S. H. Lee. 2025. "Total portfolio approach" *The Journal of Portfolio Management* 51(8): 109–127.

Elkamhi, R., J. S. H. Lee, and M. Salerno. 2021. "Factor investing using capital market assumptions" *The Journal of Portfolio Management* 48(2): 119–143.

Elkamhi, R., J. S. H. Lee, and M. Salerno. 2022. "Portfolio tilts using views on macroeconomic regimes" *The Journal of Portfolio Management* 49(3): 7–24.

Idzorek, T. 2007. "A step-by-step guide to the Black–Litterman model". In S. Satchell (Ed.), *Forecasting Expected Returns in the Financial Markets* pp. 17–38. Academic Press.

Kritzman, M., S. Page, and D. Turkington. 2012. "Regime shifts: Implications for dynamic strategies" *Financial Analysts Journal* 68(3): 22–39.

Lee, J. S. H. and M. Salerno. 2023. "Factor-targeted asset allocation: A reverse optimization approach" *Financial Analysts Journal* 79(3): 75–94.

Parasuraman, R., T. B. Sheridan, and C. D. Wickens. 2000. "A model for types and levels of human interaction with automation" *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans* 30(3): 286–297.

## Appendix A Step 1 Detail — Components, Personas, and Judged Structure

This appendix fills in the structure that Section 4 sketched, for the reader who wants to see the machine at working resolution. Each of the four themes is scored on seven regime components—the 28 components of Table A.1—and each component is a probability vector over its own ratified regime states (cycle phase over recession/early/mid/late, for example, or GDP trend over above/below). Five judge personas (Table A.2) score every component of their theme independently each cycle, so a full run produces $28 \times 5 = 140$ scored vectors. The judges are memoryless, receive identical evidence, and differ only in stated emphasis.

For the reader who wants the judge's procedure end to end, here it is: (i) count bullish/neutral/bearish readings over a fixed registry of ten indicator clusters (seven leading, three coincident; consensus excluded), exact integers, primary clusters driving the call and confirmatory clusters adjusting only conviction; (ii) anchor on a pinned base-rate table and earn any distance from it by scoring the surprise versus consensus, not the level; (iii) report conviction (low/medium/high) on a rubric independent of the probabilities, with consensus extremes capping conviction but never flipping direction; (iv) write a mandatory adversarial passage arguing against the lean from evidence already cited, then rebut it or move the numbers; (v) close with two to three quantified falsification triggers (series, threshold, date). The shared positioning pack is readable by all judges but is structurally excluded from cluster counts and probabilities.

**What a harvester actually produces.** Each theme's evidence pack is a structured table, one row per signal, and every row carries the same fields: the source and the exact reading; the release vintage, so a judge knows what was knowable when; the signal's lead time relative to the cycle; the ratified read-direction—which way the number pushes which component; and its tier, primary or confirmatory. A missing signal is logged as a gap with a stated reason rather than silently dropped, and a missing *primary* signal triggers the supplement-and-repair loop of Section 4.3. Four harvesters map one-to-one onto the themes. The fifth—the shared positioning harvester—collects the cross-asset positioning, sentiment, flow, and volatility picture once per cycle: futures positioning across equity, rate, and commodity legs; professional and retail survey sentiment; options put/call; fund flows and margin balances; the monthly manager survey. It is read by all twenty judges but kept physically separate from the four theme packs, and that separation is what enforces never-a-vote at the data layer: a crowding reading cannot leak into a regime probability, because it never enters a theme's evidence in the first place.

**The cluster registry.** Judges do not weigh individual series one by one; they count *clusters*—small, pinned groups of related indicators that move together. Each theme's registry holds ten clusters: seven leading and three coincident, with consensus-type data excluded by construction. A judge tallies, per component, how many clusters read bullish, neutral, or bearish on the three-to-six-month horizon—exact integers over a closed list, so evidence can be neither invented nor conveniently forgotten. Primary clusters drive the call; confirmatory clusters move only conviction; and evidence is weighted by how well its lead time matches the horizon, so a long-leading signal is not asked to explain this quarter.

**Table A.1: The 28 regime components scored each cycle—seven per theme.** Bracketed entries list a component's regime states (Rec = recession); components without brackets carry their own ratified state lists in the specification.

| Growth & Activity | Inflation & Liquidity | Policy Rates & Curve | Credit & Fin. Conditions |
|---|---|---|---|
| Cycle phase [Rec/Early/Mid/Late] | Global money growth | Policy stance | Bank lending standards |
| GDP trend [Above/Below] | CPI level band | Path gap vs market | Credit impulse |
| Manufacturing survey | CPI vs trend | Curve regime | Commercial & industrial loans |
| Services survey | Underlying internals | Term premium | Financial conditions index |
| Labor momentum | Inflation expectations | Real-rate stance | Credit spreads |
| Global / China | Price pipeline | Auctions & flows | Household credit |
| Earnings momentum | Central bank & net liquidity | Global policy | Funding & plumbing |

**Table A.2: The five judge personas.** All apply the identical method; the lens changes emphasis, not procedure—producing a controlled spread of interpretations of one evidence pack.

| Persona | Emphasis |
|---|---|
| Hard-data | Hard and leading data clusters; skeptical of narrative; demands the numbers corroborate any story. |
| Bottom-up | Bellwether earnings commentary, guidance, and revision breadth—the economy as firms report it. |
| Sentiment | News, narrative, and positioning; hunts contradictions between story and data, and consensus extremes. |
| Global | External and global clusters, commodity demand, foreign-revenue exposure of domestic firms. |
| Generalist | Textbook-balanced application of the framework, no tilt. |

## Appendix B Step 1 Detail — The Synthesis Algorithm

This appendix walks the pinned synthesis script end to end—what it computes, in what order, and why each design choice was made. It is written for the reader who wants to build one, so alongside the general recipe we follow a single concrete thread: a hypothetical week in which the evidence points to a credit crunch. Wherever we describe our inputs, substitute your own—the algorithm does not care what the components are or where the judges' probabilities come from, only that everything carrying economic meaning is fixed before the run and only the week's evidence changes. Figure B.1 at the end draws the whole path in one picture.

**What the algorithm consumes.** Four objects. Three are pinned in advance and identical every week: (i) a *component inventory*—a list of variables, each with a small set of mutually exclusive states. Ours has 28 components; one of them is bank lending standards, read from the Fed's loan officer survey, with three states: aggressive tightening, moderate tightening, accommodative. (ii) A *factor-driver matrix* mapping states to return factors, described below. (iii) A *cross-variable correlation matrix* saying which components tend to be misread together. The fourth object arrives weekly: the judge ensemble's probability and standard error for every state. Add a seed—ours is the run date—and the run is exactly reproducible. A reader with a different model needs nothing else; these four objects are the entire interface.

**Per-variable estimates.** For each component-state $j$, the five owning-theme judges are averaged by regime-state *name*, never by list position—a validation gate reconstructs each input from the judge ledgers and asserts value and name-binding before the script runs, because a silent misalignment between a probability and the state it describes is the one bug no downstream step can catch. In our credit week, suppose the five credit-theme judges put aggressive tightening at 55, 60, 60, 65, and 70%: the ensemble carries forward $\bar{p} \approx 62\%$ with a standard error reflecting that 15-point spread, as in Equation (1). The spread is not a nuisance to be averaged away; it is the first of the three uncertainties the simulation will carry to the output. This is the only place judge output enters; everything after is arithmetic on $(\bar{p}_j, \mathrm{SE}_j)$ pairs.

**The factor-driver matrix.** Structural links from regime states to the three return factors (e.g., the factor mimicking portfolios for growth, real rate, and inflation) are ratified and pinned. Only states passing a validation bar—minimum historical hit-rate and effect size, with the economically sensible sign—carry a loading $\beta$: direction from theory, magnitude in discrete strength grades calibrated from history, and a reliability grade attached that scales the loading's own error.[2] In the example: aggressive tightening loads *negative* on growth—theory says banks pulling back starves activity—at, say, a medium

[2]Our pinned calibration, for concreteness: three strength grades mapping to $|\beta|$ of 0.45, 0.825, and 1.25 (in units of return-on-volatility), with loading error $\sigma_\beta = \max(0.40, |\beta|/2)$ multiplied by a reliability factor of 1.00, 1.15, or 1.30 for strong, medium, or weak historical evidence. The floor of 0.40 caps any single loading's $t$-statistic near 2—history is never allowed to claim certainty. An implementer should recalibrate all five numbers to their own factor definitions and sample.

strength grade, because history shows the link is real but not the strongest in the matrix. The accommodative state loads positive; the moderate state fails the validation bar and carries no loading at all, which is the matrix's way of saying *history cannot distinguish this state's growth signal from noise, so it is not allowed to speak*.

**The derived layer.** Not every informative component was part of the researcher's own historical validation, and this is where a second, "derived" layer comes in. Consider corporate credit spreads. Suppose the pinned matrix carries no hand-calibrated spread loading—the original research never covered it—yet a week in which spreads are widening plainly says something about growth. The derived layer answers with a transitive argument rather than a new opinion. History shows that the spreads-widening state travels with a particular weighted bundle of already-ratified anchor states—GDP below trend, credit impulse negative, bear flattening—so it is allowed to inherit the loading that bundle implies, and no more: a calibrated linkage regression measures the co-movement, the candidate's loading is the coefficient-weighted sum of its anchors' loadings, and its uncertainty is propagated through the same coefficients, so a state admitted secondhand carries secondhand error. The logic is the same one an analyst uses informally—"spreads widening usually means what GDP-below-trend means"—made explicit, calibrated, and pinned. Two safeguards keep the layer honest. First, each derived link still faces a human ratification bar of minimum historical hit-rate and effect size, exactly as the anchors did; in our matrix, spreads-widening passed for growth and inflation, but its real-rate link fell below the floor and was rejected—inheritance is per link, not per state. Second, the share of total loading weight carried by the whole derived layer is reported each run, so a matrix quietly filling up with secondhand evidence would be visible immediately. The synthesizer can never edit either layer: if the week's evidence seems to demand a new link, that is a research finding for the human, not a runtime decision for the machine.

**Contribution uncertainty.** A loaded state's contribution is $g = \bar{p}\,\beta$—in the example, a medium-strength bearish push on growth, scaled by the 62% the judges gave it—with

$$\mathrm{Var}(g) = \beta^2\sigma_p^2 + \left(\tfrac{1}{n_{\text{states}}}\right)^2\sigma_\beta^2 + \sigma_p^2\sigma_\beta^2,$$

where $\sigma_p$ is the ensemble SE, $\sigma_\beta$ the loading's empirical error, and $n_{\text{states}}$ the number of states in the variable. The three terms have plain readings: the judges may be wrong about the state (that 15-point spread), history may be wrong about how much the state matters, and both can be wrong at once. The one non-obvious choice sits in the middle term, and it is deliberate: the loading error is weighted by the *no-view* baseline $1/n_{\text{states}}$, not by the current $\bar{p}$. Suppose next month the judges put aggressive tightening at only 10%. Weighting by $\bar{p}$ would then shrink the mapping-error term to almost nothing—as if history's uncertainty about what tightening *does* had improved because tightening became unlikely. It has not, so the weight stays at the uniform baseline.

**Correlated simulation—why there is a PCA in the middle.** Here is the problem the middle of the algorithm exists to solve. In our credit week, three separate components go bearish together: lending standards tighten aggressively, the credit impulse turns negative,

and business loan growth contracts. Are those three independent pieces of bad news, or one piece of bad news measured three ways? The honest answer is mostly the latter—and if the simulation treated their uncertainties as independent, the three would average against each other and the final band would come out dishonestly narrow, manufacturing confidence out of redundancy. So, per factor, the script assembles the covariance of all loaded contributions, $\Omega = \mathrm{diag}(\sigma_g)\, R\, \mathrm{diag}(\sigma_g)$, where $\sigma_g$ collects the per-state contribution uncertainties and $R$ is built by lookup from the pinned correlation matrix: two states inherit the correlation of their parent variables, and states of the *same* variable get zero, because they are mutually exclusive readings, not co-moving ones. The script then eigendecomposes $\Omega$ (clipping any numerically negative eigenvalue to zero) and keeps the leading principal components—the independent directions in which this week's evidence actually co-moves—until they explain 90% of the contribution variance. In the example, the three credit readings collapse into essentially one component: *the credit story*. The retained count $k$ is reported as a diagnostic in its own right, because it answers, in one integer, a question a committee might actually ask: *how many independent macro stories does this week contain?* A week in which dozens of states collapse into five components is a week with five stories—perhaps the credit crunch, a still-resilient labor market, sticky services inflation, and two more.

**The draw loop.** Ten thousand draws, seeded by the run date. Each draw adds correlated noise—generated through the retained components, so that when the credit story draws pessimistic, all three credit states move together rather than offsetting one another—to the contribution *means held raw*: never smoothed, never projected. The asymmetry is deliberate, and the example shows why. Suppose that amid the credit gloom one component dissents: global money growth is elevated, a bullish reading. Projecting the means onto the components would average that lone testimony into the dominant bearish theme—the dissent would simply vanish into the consensus. Holding the means raw preserves it, while running the noise through the components keeps the uncertainty honest about correlation. Within each draw, every state's contribution is then clamped toward zero on its home-sign side, individually, before any summation—theory fixes direction, evidence scales magnitude. Concretely: if a draw's noise hands the bullish manufacturing-PMI state a *negative* contribution, that state is set to zero for that draw; it is not allowed to testify against growth, and its bad draw cannot be netted against another bullish state's good one, because each state is clamped before the sum. A state's side is set by the original sign of its $\beta$, never re-assigned by the draw. Equation (3) converts the clamped contributions into the draw's probabilities, with the denominators $\sum \beta$ fixed per side across all draws; in the rare draw where $P(+) + P(-)$ exceeds one, both sides are rescaled proportionally. The projection the means are protected from is, however, used deliberately elsewhere: the inflection gauge of Section 4.5 reruns the identical loop with component-projected means, and a widening gap between raw and projected probabilities is the money-growth dissent *accumulating*—contrarian evidence building against the dominant theme, often the earliest sign of a turn.

**What is reported.** Each factor's probabilities are the means across draws, banded by the 25th–75th percentiles of the draws—a percentile interval, because a symmetric standard error is misleading for a probability bounded in $[0, 1]$. In the example week the growth

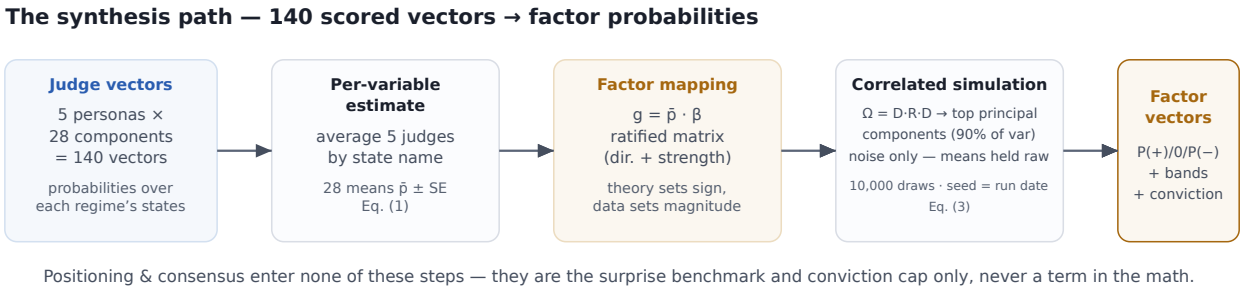



**Figure B.1: The consolidation path, left to right:** 140 judge scorecards are averaged question by question; a ratified matrix converts each averaged score into a signed push on a return factor; and a correlated Monte Carlo carries the three sources of uncertainty—judge disagreement, mapping error, and cross-variable correlation—into the final probability ranges. The middle of that last step deserves a sentence in words: the covariance of the pushes is reduced to a handful of principal components, which are the week's independent macro stories; random noise is drawn through those components so correlated states share their uncertainty; and the means themselves are never smoothed or projected—so a lone dissenting signal survives to the output instead of being averaged away by its correlated neighbors.

factor might come out around $P(-) = 55\%$ with a band of roughly 48–62%: a clearly bearish call whose width the reader can trace back to named sources—judges spread fifteen points on the key state, the loading is medium-reliability, and the three credit readings are one story, not three. Around the band sit the diagnostics, kept deliberately distinct because they answer different questions: the signed lean $P(+) - P(-)$ with its $t$-statistic against the simulated dispersion and the share of draws in which the direction survives (three answers to "how sure are we?"); bull-side and bear-side conviction, the contribution-weighted averages of the judges' *stated* conviction (how much the machine trusted its inputs—an axis independent of both the band and the $t$); the inflection gaps; the retained component count $k$; the derived layer's weight share; and the unclamped sum of contributions, which passes to the downstream state-estimation layer where the clamp would discard information the smoother can use. The factor sign convention is simple throughout: growth and inflation are surprise axes, so positive always means above consensus.

That is the whole machine. An implementer who has pinned an inventory, a loading matrix, and a correlation matrix, and who receives weekly $(\bar{p}, \mathrm{SE})$ pairs, can reproduce every step of it from this appendix. The discipline that makes it trustworthy is not in any single formula but in the separation the steps enforce: pin everything that carries meaning, and randomize only what carries noise.

## Appendix C Step 2 Detail — Derivations and the Sampler

This appendix holds what an implementer needs and a general reader can safely skim: the exact universe, the derivations behind Equations (5)–(8), and the mechanics of the sampling search. Nothing here changes the story told in Section 5—it makes that story reproducible.

**Universe and mapping.** The implementation prices 30 underlying assets on the three macro factors and reports results at 14 asset-class lines through a fixed mapping. All expected returns are excess of cash. For convenience, the working matrices augment the 30 assets with three synthetic "factor assets" carrying identity loadings, so factor-space quantities—the view factor covariance in particular—can be read directly off blocks of the augmented covariance rather than re-derived. Nothing methodological depends on these counts or on the augmentation; they are stated so an implementer can reproduce the arithmetic exactly.

**The reverse-optimization target.** Lee and Salerno (2023) show that when $w$ is the mean–variance tangency portfolio, its reverse-optimized factor weights are

$$\theta^\star = \kappa^{-1}\Sigma_f^{-1}(B'B)^{-1}B'\mu_a,$$

and that these weights are materially more stable to estimation error in $\mu_a$ than either asset weights or explicit exposures $Bw$—the motivating result for stating the target, Equation (8), in this space. The risk-aversion scalar $\kappa$ drops out under the risk-budget normalization. Note that the factor covariance enters the target only through the leading $\Sigma_f^{-1}$; the premium-recovery step $(B'B)^{-1}B'\mu_a$ is ordinary least squares and does not depend on $\Sigma_f$, so the harder-to-estimate covariance tilt influences the target through a single, well-understood inversion.

**The view at the factor level.** Equation (6) in the main text states the view-conditioned inputs at the asset level; the underlying factor-level objects are the confidence-scaled, probability-weighted mixtures

$$\lambda_f(\pi,\gamma) = \lambda_f^0 + \gamma\sum_{s=1}^{4}\pi_s\,\delta_s, \qquad \Sigma_f(\pi,\gamma) = \Sigma_f^0 + \gamma\sum_{s=1}^{4}\pi_s\,\Delta_s,$$

which propagate to assets through the loadings. Setting $\gamma = 0$ recovers the neutral inputs exactly; only the systematic block of the asset covariance moves with the view. The per-regime tilts $\delta_s$ (a 3-vector per regime) and $\Delta_s$ (a $3\times 3$ matrix per regime) are estimated from historical regime episodes and pinned; they are inputs to the engine, not outputs of the AI pipeline, and revising them is a researcher decision like any other change to the frame.

**The calibration closed form.** The baseline expected-return vector $\mu_a^0 = m_{\text{cma}} + g\,B\lambda_f^0$ is affine in $g$, so the 60/40's expected excess return is $a_0 + g\,a_1$ with $a_0 = w'_{60/40}m_{\text{cma}}$,

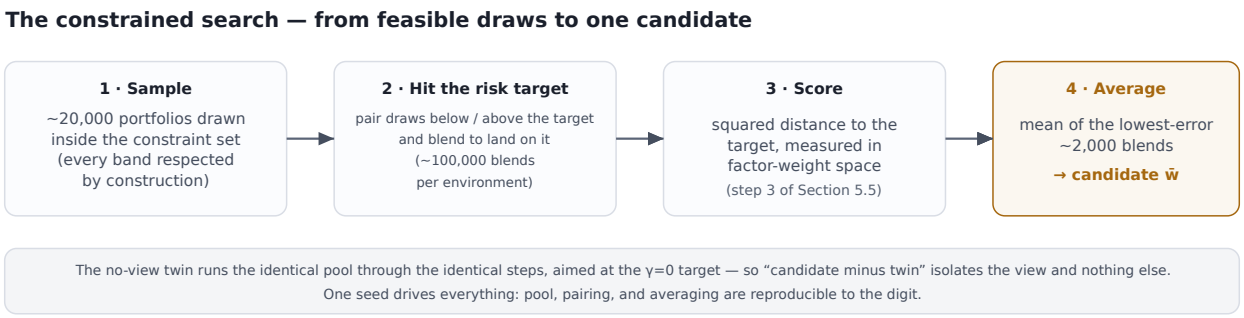



**Figure C.1: The sampling-and-averaging search of Section 5.5, as implemented.** Steps 1–2 guarantee feasibility and the exact risk target; step 3 enforces the factor target in the stable coordinate system; step 4 supplies diversification and damps sampling noise.

$a_1 = w'_{60/40} B \lambda_f^0$; equating the 60/40 Sharpe to $\hat{S}$ (on a geometric basis) and solving gives Equation (7). Because the calibration uses neutral inputs, $g$ is independent of $(\pi, \gamma)$: the view never changes the assumed price of risk, only the composition of returns around it. The long-run factor premium itself is the least-squares projection $\lambda_f^0 = (B'B)^{-1}B'\mu_a$ of the CMA onto the loadings.

**Why not the closed form.** Unconstrained, the target-matching problem has a ridge solution $w = [c\, B_i' B_i + (1-c)I]^{-1}[c\, B_i' \theta^{\text{tgt}} + (1-c)w_0]$, where $c \in [0,1]$ trades fidelity to the factor target against proximity to an anchor portfolio $w_0$ and $I$ is the identity matrix. Institutional constraints—long-only sleeves, group bands, budget conditions—cannot be imposed on this inverse without projections that distort precisely the exposures being targeted. The sampling procedure of Section 5.5 is the constrained analogue of its $c \to 1$ limit: restricting draws to the feasible polytope and averaging the closest matches plays the role of the anchor term, pulling the solution into the well-diversified interior while the error metric enforces the target.

**The sampler.** The sampler is the one piece with real moving parts, so here it is in the order it runs. Candidate portfolios are drawn by a hit-and-run scheme over the constraint set: each group total is drawn from the midpoint of its band displaced by an asymmetric random offset scaled to the room on each side; relative weights within equities, bonds, and inflation-sensitive sleeves are drawn the same way and demeaned so they sum to the group total; the four private sleeves are drawn individually against their own minimums and maximums, with the aggregate pulled into the all-private band by a scaler bounded at one. The all-private band is auto-validated against the private assets' own limits before every generation—held inside $[\,\Sigma\text{min} + 1\%,\ \Sigma\text{max} - 1\%\,]$ straddling the midpoint sum—which guarantees the scaler never amplifies and no single private asset can be pushed outside its individual range. Risk-target matching brackets each candidate pair below and above the volatility target and blends them onto the target (risk being convex, the blend lands marginally under, never over—the near-exactness noted in Section 5.5). The candidate portfolio is the average of the lowest-error few percent of roughly one hundred thousand such blends; the no-view twin repeats everything at $\gamma = 0$ against the regime-neutral target. All draws are seeded.

**Table C.1: Symbol-to-control map for implementers:** every quantity in the paper and where it lives in the application.

| Quantity (paper) | Control in the application |
|---|---|
| Regime probabilities $\pi$ | The clickable regime map, or the four model inputs (Rising G/Rising I, Rising G/Falling I, Falling G/Rising I; the fourth is the residual) |
| Confidence $\gamma$ | "Confidence" (percent) |
| Calibration Sharpe $\hat{S}$ | "60/40 SR" under no-view environment calibration |
| Risk target $\sigma^\star$ | "No-view portfolio volatility" and "With-view portfolio volatility" (both 9% in our illustration) |
| Risk-measurement environment | "Target volatility in" selector—the paper uses both settings: Tables 3–4 measure risk under the no-view covariance, Tables 5–6 under the view covariance |
| Constraint set $\mathcal{W}$ | Group bands, relative-weight ranges, and private-asset min/max steppers, all editable in the table |
| Pool, blends, averaged set | "Generate" ($\sim$20,000), "Bootstrap" ($\sim$100,000), "Top N" ($\sim$2,000) |
| Reproducibility | "Seed" (fixes pool and bootstrap) |

**Run sizes and the blend.** Our settings, for concreteness: a pool of roughly 20,000 feasible portfolios; about 100,000 bracketed blends per environment; the lowest-error 2,000 averaged into the candidate. The blend itself is linear: pairing a below-target draw $w_{lo}$ (volatility $\sigma_{lo}$) with an above-target draw $w_{hi}$, the mixing weight $\lambda = (\sigma_{hi} - \sigma^\star)/(\sigma_{hi} - \sigma_{lo})$ is applied to weights and to factor exposures (both linear), and the blend's risk is then recomputed—by convexity it lands marginally under the target, never over. Expected return, volatility, and Sharpe are reported on a geometric basis (the Sharpe adds back $\frac{1}{2}\sigma^2$), each computed under the stated environment's covariance. One seed drives the pool and the bootstrap alike, so the entire candidate—not merely the target—is reproducible to the digit. Figure C.1 summarizes the search; Table C.1 maps every symbol in this paper to its control in the application, for anyone reproducing the work.

**Sensitivities.** The two modeling choices most deserving of scrutiny are the systematic-risk share $\phi$ and the sampling-and-averaging approximation. Because the objective is stated in reverse-optimized factor-weight space, the recommended allocations are robust to reasonable variation in both: perturbing $\phi$ rescales the target without rotating it, and widening or narrowing the averaged set moves diversification, not the factor bets. The practical failure mode to monitor is a risk target near the edge of what the bands permit, where the bracketing step can fail to find portfolios on both sides—the application warns rather than extrapolates.

## Appendix D The Illustration Re-Run with the Private Book Pinned

The scenarios of Section 7 let every asset class move, because the point there was to watch the machinery. But as Section 7.4 argued, that is not how an institution would actually run a regime tilt: private commitments take quarters to build and longer to unwind, so over any horizon a view speaks to, the private book is what it is. A CIO's first request will therefore be the version of the exercise this appendix delivers: the full illustration repeated with the four private sleeves *pinned*—each minimum set equal to its maximum at Portfolio-"0"'s holdings from Table 3 (private equity 9.2%, real estate 11.3%, infrastructure 10.5%, private credit 5.6%). Everything else is unchanged: same CMA, same bands elsewhere, same 0.35 calibration Sharpe, same 9% risk target applied both ways, same seed. One consistency check is worth reporting before the results: rebuilt under the pinned bands, the no-view twin reproduces the original Portfolio-"0" to within 0.2 percentage points on every line—the pin merely formalizes holdings the neutral construction had already chosen, and the small residue is sampling noise doing its usual, negligible work.

Table D.1 shows the result with risk measured in the no-view environment, and reading it against Table 3 answers a question each scenario poses differently: *what does the view do when its private lever is taken away?* Scenario A barely notices—public equities +6.6 points against +7.0 before, funded from the same 13.6 points of duration—because the June view never wanted to move the private book much in the first place; pinning costs it almost nothing. Scenario B loses its favorite move. Free, the stagflation view added 3.6 points to the private allocation, infrastructure above all; pinned, that door is closed, and the engine responds by selling *less* public equity (−4.5 against −7.1)—the private add and the deeper public cut were two halves of one rebalancing it can no longer perform—while leaning harder on the levers that still move: the inflation-sensitive sleeve (+2.7), the dollar (−6.4), and the same duration shift from nominal toward real. Scenario C is the most instructive of the three. In the main text, the recession candidate's public equity line rose 1.8 points—the counterintuitive entry Section 7.2 paused to explain as the byproduct of shrinking the private block while holding volatility at target. Pin the privates and that byproduct vanishes: public equities are now sold outright (−1.2), exactly the move intuition expects from a recession view. The oddity was never the view's opinion of equities; it was the free private book expressing the same opinion through a different door.

Table D.3 repeats the exercise with each candidate's 9% measured in its own view environment, and the two-decisions lesson of Section 7.2 returns with the volume turned up. Measured neutrally, the pinned recession candidate grosses *up*—118.4% of capital against 114.2%—because with equities cut and privates immobile, the only way to hold 9% of neutral-world risk is to buy more of the low-risk asset, duration. Measured in the recession world, where that duration is doing real defensive work and equity risk runs hot, the same view de-levers the fund to 110.5% gross and sells public equities 7.7 points. Neither answer is wrong; they are answers to different questions about the risk budget, and the gap between them is precisely the decision Section 7.4 urged committees to make deliberately rather than by default. The characteristics in Tables D.2 and D.4 carry the same quiet message as their main-text counterparts: each candidate beats Portfolio-"0" in its own

world (0.60 against 0.59 under A, 0.22–0.23 against 0.20 under B, 0.16–0.18 against 0.14 under C), the opportunity cost under the neutral lens stays small, and no portfolio makes stagflation or recession pleasant—the candidates simply defend better with the levers they are permitted to pull.

Two readings of this appendix, and we intend both. Narrowly, it is the robustness check a skeptical reader would ask for: the qualitative story of Section 7—disciplined tilts, visible constraints, small opportunity cost, the mandate binding before conviction does—survives intact when the least liquid third of the portfolio is frozen. More broadly, it is the configuration we would expect an adopting institution to run *first*: the free-privates tables of the main text demonstrate the machinery; the pinned tables here are the deployment setting, where the committee's real question—given our privates, how should the liquid book respond to the economy we believe is coming?—gets its direct answer. And the two runs together price what Section 7.4 called the cost of immobility: the difference between them is what the locked private book keeps the view from expressing.

**Table D.1: Allocations with the private book pinned, 9% risk target measured in the no-view environment** (percent of portfolio). The four private lines are frozen at Portfolio-"0"'s holdings—the wall of grey zeros down the private block is the constraint doing its job—so every tilt must now be expressed through public equities, bonds, the inflation-sensitive sleeve, and currency. Bars read as in Table 3. Deltas are computed before rounding.

| | Portfolio "0" (no view) | Portfolio "view-A" (AI view) | Δ | Portfolio "view-B" (Stagflation) | Δ | Portfolio "view-C" (Recession) | Δ |
|---|---|---|---|---|---|---|---|
| **Equities (public)** | **29.5** | **36.1** | **+6.6** | **24.9** | **−4.5** | **28.3** | **−1.2** |
| US | 15.0 | 18.4 | +3.4 | 12.6 | −2.4 | 14.2 | −0.8 |
| Europe | 11.7 | 14.5 | +2.8 | 9.7 | −2.1 | 11.1 | −0.7 |
| Canada | 1.4 | 1.6 | +0.2 | 1.3 | −0.1 | 1.4 | 0.0 |
| Emerging markets | 1.3 | 1.5 | +0.2 | 1.4 | 0.0 | 1.6 | +0.2 |
| **Fixed income** | **42.1** | **28.6** | **−13.6** | **44.2** | **+2.1** | **48.1** | **+5.9** |
| Universe bonds | 10.2 | 7.4 | −2.8 | 10.4 | +0.1 | 11.5 | +1.3 |
| Long bonds | 10.3 | 7.0 | −3.3 | 11.2 | +0.9 | 13.3 | +3.1 |
| Real-return bonds | 21.6 | 14.2 | −7.4 | 22.6 | +1.0 | 23.2 | +1.6 |
| **Private allocation** | **36.6** | **36.6** | **0.0** | **36.6** | **0.0** | **36.6** | **0.0** |
| Private equity | 9.2 | 9.2 | 0.0 | 9.2 | 0.0 | 9.2 | 0.0 |
| Real estate | 11.3 | 11.3 | 0.0 | 11.3 | 0.0 | 11.3 | 0.0 |
| Infrastructure | 10.5 | 10.5 | 0.0 | 10.5 | 0.0 | 10.5 | 0.0 |
| Private credit | 5.6 | 5.6 | 0.0 | 5.6 | 0.0 | 5.6 | 0.0 |
| **Inflation-sensitive** | **6.0** | **2.1** | **−3.9** | **8.7** | **+2.7** | **5.4** | **−0.5** |
| Commodities | 3.0 | 1.1 | −2.0 | 4.3 | +1.3 | 2.5 | −0.5 |
| Gold | 2.9 | 1.0 | −1.9 | 4.4 | +1.4 | 2.9 | 0.0 |
| **Currencies (USD)** | **11.7** | **12.2** | **+0.6** | **5.3** | **−6.4** | **12.6** | **+0.9** |
| **Total ex-currency** | **114.2** | **103.4** | | **114.5** | | **118.4** | |

**Table D.2: Portfolio characteristics for the pinned-privates portfolios of Table D.1**, priced as in Table 4: Portfolio-"0" under all four environments, each candidate under the neutral environment and under its own view.

| Characteristic | Portfolio "0" (no view) | Portfolio "view-A" (AI view) | Portfolio "view-B" (Stagflation) | Portfolio "view-C" (Recession) |
|---|---|---|---|---|
| *Volatility (%)* | | | | |
| under no view | 8.9 | 9.0 | 9.0 | 9.0 |
| under view A | 8.6 | 8.7 | | |
| under view B | 9.0 | | 9.0 | |
| under view C | 9.9 | | | 10.0 |
| *Expected return (%)* | | | | |
| under no view | 6.4 | 6.4 | 6.3 | 6.4 |
| under view A | 7.5 | 7.7 | | |
| under view B | 4.2 | | 4.4 | |
| under view C | 3.7 | | | 3.9 |
| *Sharpe ratio* | | | | |
| under no view | 0.45 | 0.45 | 0.44 | 0.45 |
| under view A | 0.59 | 0.60 | | |
| under view B | 0.20 | | 0.22 | |
| under view C | 0.14 | | | 0.16 |

**Table D.3: Allocations with the private book pinned, 9% risk target measured in the view environment instead** (percent of portfolio). Portfolio-"0" is unchanged; only each candidate's measuring stick has moved. As in the main text, Scenario C is where the choice has teeth: measured in its own recessionary world, the candidate sells public equities hard and brings gross exposure down to 110.5% of capital against Portfolio-"0"'s 114.2%.

| | Portfolio "0" (no view) | Portfolio "view-A" (AI view) | Δ | Portfolio "view-B" (Stagflation) | Δ | Portfolio "view-C" (Recession) | Δ |
|---|---|---|---|---|---|---|---|
| **Equities (public)** | **29.5** | **37.1** | **+7.6** | **24.2** | **−5.2** | **21.8** | **−7.7** |
| US | 15.0 | 19.0 | +4.0 | 12.2 | −2.8 | 11.0 | −4.0 |
| Europe | 11.7 | 15.0 | +3.3 | 9.4 | −2.3 | 8.6 | −3.1 |
| Canada | 1.4 | 1.6 | +0.2 | 1.3 | −0.1 | 1.1 | −0.4 |
| Emerging markets | 1.3 | 1.5 | +0.2 | 1.3 | 0.0 | 1.1 | −0.2 |
| **Fixed income** | **42.1** | **29.7** | **−12.5** | **44.5** | **+2.3** | **47.8** | **+5.6** |
| Universe bonds | 10.2 | 7.7 | −2.6 | 10.6 | +0.3 | 11.5 | +1.3 |
| Long bonds | 10.3 | 6.9 | −3.3 | 11.3 | +1.0 | 13.2 | +2.9 |
| Real-return bonds | 21.6 | 15.1 | −6.6 | 22.6 | +1.0 | 23.1 | +1.4 |
| **Private allocation** | **36.6** | **36.6** | **0.0** | **36.6** | **0.0** | **36.6** | **0.0** |
| Private equity | 9.2 | 9.2 | 0.0 | 9.2 | 0.0 | 9.2 | 0.0 |
| Real estate | 11.3 | 11.3 | 0.0 | 11.3 | 0.0 | 11.3 | 0.0 |
| Infrastructure | 10.5 | 10.5 | 0.0 | 10.5 | 0.0 | 10.5 | 0.0 |
| Private credit | 5.6 | 5.6 | 0.0 | 5.6 | 0.0 | 5.6 | 0.0 |
| **Inflation-sensitive** | **6.0** | **2.6** | **−3.4** | **8.7** | **+2.7** | **4.4** | **−1.6** |
| Commodities | 3.0 | 1.3 | −1.7 | 4.3 | +1.2 | 2.1 | −1.0 |
| Gold | 2.9 | 1.3 | −1.6 | 4.4 | +1.5 | 2.3 | −0.6 |
| **Currencies (USD)** | **11.7** | **10.2** | **−1.4** | **5.2** | **−6.5** | **13.1** | **+1.5** |
| **Total ex-currency** | **114.2** | **106.0** | | **114.0** | | **110.5** | |

**Table D.4: Portfolio characteristics for the pinned-privates portfolios of Table D.3**, priced the same way: Portfolio-"0" under all four environments, each candidate under the neutral environment and under its own view.

| Characteristic | Portfolio "0" (no view) | Portfolio "view-A" (AI view) | Portfolio "view-B" (Stagflation) | Portfolio "view-C" (Recession) |
|---|---|---|---|---|
| *Volatility (%)* | | | | |
| under no view | 8.9 | 9.3 | 8.9 | 8.1 |
| under view A | 8.6 | 9.0 | | |
| under view B | 9.0 | | 9.0 | |
| under view C | 9.9 | | | 9.0 |
| *Expected return (%)* | | | | |
| under no view | 6.4 | 6.5 | 6.3 | 6.1 |
| under view A | 7.5 | 7.8 | | |
| under view B | 4.2 | | 4.4 | |
| under view C | 3.7 | | | 4.0 |
| *Sharpe ratio* | | | | |
| under no view | 0.45 | 0.44 | 0.44 | 0.45 |
| under view A | 0.59 | 0.60 | | |
| under view B | 0.20 | | 0.23 | |
| under view C | 0.14 | | | 0.18 |