

Portfolio Oversight Lab: 60-Minute Interview Rehearsal

[Project](#) · [Field guide](#) · [Control methodology](#)

Goal: demonstrate that you can calculate, trace, challenge, and communicate the actual project—not merely repeat its summary. This workbook is preparation for portfolio research and analytics discussions, including JPMorgan Chase / 55ip requisition 210775966. It is not evidence of employment, endorsement, people-management experience, or production investment authority.

All numerical cases below are **invented teaching examples**. They are not additional historical results. The accounting case deliberately matches the transparent fixture in `tests/test_oversight_adversarial.py`, so the hand calculation can be checked against an executable example without licensed data.

Before starting

Use a clean sheet, a basic calculator, the repository, and a 60-minute timer. Complete questions A-G before opening the answer key. Notes are allowed only where the instructions explicitly permit source navigation. Say calculations and decisions aloud; this exposes unit confusion that a memorized paragraph can conceal.

Have the public environment installed using [the reproduction guide](#). Do not spend the timed exercise installing dependencies. No WRDS login is needed for the fixture tests. Do not run `run_oversight.py` expecting it to invent missing empirical data.

Record your attempt date, Git revision, and what you personally ran or inspected. A saved green receipt is prior evidence; it is not proof you executed the current checkout. Keep practice notes local unless you intentionally want to publish them.

The 60-minute sequence

Minutes	Task	Points	Completion evidence
0–5	A. Explain the project	5	A recorded or written 60-second answer
5–17	B. Funded accounting and linking	20	Calculation sheet with dollars, returns, and identities
17–27	C. Risk and economic judgment	20	TE calculation, feasible-budget decision, cost interpretation
27–35	D. SQL grain and missing data	15	Correct query shape and failure explanation
35–45	E. Navigate and challenge actual code	20	Symbols located, targeted tests run, limitations explained
45–53	F. Investigate an incident	10	Containment note, expected detector, recovery conditions
53–58	G. Make the PM recommendation	10	A 90-second decision and honest ownership statement
58–60	Score and select remediation	—	Score plus one specific next action

The timebox is an exercise constraint, not a reason to fake a result. If code takes longer to run, record that and finish the evidence check outside the timed session.

Questions—do not consult the answer key yet

A. Explain it in sixty seconds

Without notes, explain:

1. The business question and competing implementation choices.
2. One economic finding, with units and a qualification.
3. What the independent checker actually reconstructs.
4. One important common dependency and one missing production capability.
5. Your own contribution versus work generated with AI assistance.

Avoid the words “alpha,” “tax savings,” “production-grade,” and “independent audit” unless explaining precisely why the project does not establish them. Do not equate an agent-assisted workstream with managing employees.

B. Calculate the portfolio path

An account starts with \$100 in asset A and \$0 in asset B. Its frictionless benchmark starts \$50/\$50. On day 1, A earns +10% and B earns 0%. At that day's close, the account targets 50%/50%, paying **100 bp per dollar bought or sold**. The benchmark resets to 50%/50% without costs. On day 2, returns are A +10%, B 0%; on day 3, A −20%, B +5%. Neither portfolio trades after day 1. There are no external cash flows.

This intentionally large teaching fee is **not** the empirical study's 5-bp assumption.

1. Calculate the account's day-1 pretrade NAV, funded closing NAV, each closing sleeve, each signed trade, and each trade fee. Show $\text{sum}(\text{trades}) + \text{fees} = 0$.
2. Calculate gross return, net return, the exact cost contribution to return, and the cost/pretrade-NAV diagnostic. Explain why the last two are different.
3. Roll both portfolios to day 3. Calculate cumulative account return, benchmark return, and their difference in percentage points.
4. Calculate the account's linked A, B, and cost contributions using initial capital. Do the equivalent for the benchmark and reconcile cumulative active contributions.
5. A colleague sums the daily arithmetic active returns and reports 3.90%. Is that the cumulative return difference? Explain the numerical gap without calling it a rounding error.

C. Risk is not return

The proposed portfolio is 60% A / 40% B and policy is 50%/50%. Annualized volatilities are 20% and 10%, with correlation 0.25. There are no other assets.

1. Write the annualized covariance matrix and active-weight vector. Calculate estimated TE. State the units of variance and TE.

2. If correlation were 0.80 with unchanged volatilities, would this substitution's TE rise or fall? Calculate it and explain the direction.
3. The minimum feasible estimated TE is 1.20%, and the allowed increment is 25 bp. What is the budget? Is the original proposal feasible under that rule?
4. Does satisfying a proposal's estimated-risk budget guarantee that the full-history realized TE is below it? Name two distinct reasons beyond "markets change."
5. In a separate constant-NAV illustration, one-way turnover falls from 18% to 7%, the charge is 5 bp per traded dollar, and NAV is \$2 million. Estimate annual cost savings. Can you subtract that number from the extra TE to get a net performance benefit?
6. Name the missing data/model requirements before claiming country, sector, factor, or tax optimization from these ETF weights.

D. SQL: spot the attractive wrong answer

Suppose `daily` has one row per (date, client, method), while `trades` has one row per (date, client, method, ticker). The synthetic account above has two trade rows on day 1 and one daily row with `cost_dollars = 1.10`.

```
SELECT d.date, d.client, d.method,
       SUM(d.cost_dollars) AS reported_cost
FROM daily d
JOIN trades t
  ON d.date = t.date AND d.client = t.client AND d.method = t.method
GROUP BY d.date, d.client, d.method;
```

1. What cost does the query return for day 1, and why?
2. Sketch a corrected query comparing summed trade fees with the daily fee without multiplying it. What uniqueness checks must precede the comparison?
3. If an entire trade group or a daily row is missing, what does your query do? Explain why one directional left join is not a complete coverage control.
4. How can a saved benchmark and saved active-return series be changed together so SQL identities still pass? Identify which different layer should detect that corruption.
5. Why must NULL/NaN checks come before aggregation? Why is unconditional `COALESCE(..., 0)` unsafe for a missing daily match?

E. Trace four claims into real source

You may now open code, but not the answer key. Do not rely on line numbers; name the function, variable, condition, or assertion proving the claim.

1. **Timing:** locate the primary decision schedule and the checker's exact historical slice. Prove that execution-day returns cannot enter the covariance sample. Explain why a mere `signal_date < execution_date` check would be insufficient.
2. **Funding:** locate the primary funded rebalance solver, then locate the independent checker's cost and holdings arithmetic. Identify what they do differently and what input they still share.

3. **False reassurance:** locate the public test in which an internally consistent forged benchmark passes SQL consistency but fails source-driven reconstruction. Run it. Explain why the resulting test success does not mean the forged data passed validation.
4. **Incidents:** locate the runner's expected-detector map and its detected condition. Explain why `validator_execution = BLOCK` alone must not count as successful fault detection.

Run the two targeted tests relevant to questions B and E3 using their exact pytest node IDs. Record the command, observed test result, and one assertion you understand. Then identify the command for the full public test suite without claiming you ran it if you did not.

F. Write an incident note

A saved trade's recorded fee is increased by \$100, but its notional and holdings are unchanged. The control run blocks. A PM asks, "Did we lose \$100? Can you just correct the report so we can distribute it?"

Give a six-sentence response covering:

1. What is established, and what remains unknown.
2. Which output should be held, and what should not be inferred about trading.
3. The earliest identity to investigate and expected detector(s).
4. Scope and downstream reports to review.
5. Correction and dirty/clean regression evidence needed.
6. How this particular repository's in-memory experiment differs from a real incident resolution.

Then open `outputs/oversight/incident_summary.json`. Locate the appropriate case, its expected domain checks, observed failed checks, closure, and original-input status. Treat the file as **saved evidence**, not a new test execution.

G. Make a decision without overselling

Give a 90-second recommendation on the no-gold result:

- Lower turnover has about 60% less turnover, about 27 bp more realized gross TE, and about 1.10 bp less annualized modeled trading-cost rate than Minimum risk.
- Lower turnover happened to finish with more wealth in the examined history.
- The client asks if it should therefore become the default and whether it proves tax savings.

Include: your decision, why the historical wealth ordering does not establish alpha, one potentially relevant unmodeled client benefit, the evidence needed to change your recommendation, and a concrete reviewer/acceptance/rollback plan for a proposed research change. Explain your personal project work truthfully and identify one role requirement that the project alone cannot prove.

Answer key—open only after your attempt

A. What earns the five points

Award one point each for: a clear restrictions/implementation question; economic result in bp or dollars with its assumption; reconstruction of holdings/costs/NAV rather than just a checksum; common-source/specification limitations and an absent live capability; and accurate personal/AI-assisted ownership.

A strong answer need not match the field guide's script. It should make the listener understand the decision before hearing about libraries. "AI helped write and review the code; I personally verified these specified calculations" is defensible only after you actually perform that review. Use a narrower statement before then.

B. Worked accounting and linking solution

Day 1 funding. Pretrade holdings are \$110 and \$0. With equal closing sleeves, gross trading is \$110: the sale from A and purchase into B sum to that amount as long as both trades have the stated signs. Cost is $110 \times 0.01 = \$1.10$. Closing NAV is **\$108.90**, and each sleeve is **\$54.45**.

- A signed trade: $54.45 - 110 = -\$55.55$; fee $55.55 \times 0.01 = \$0.5555$.
- B signed trade: $54.45 - 0 = +\$54.45$; fee $54.45 \times 0.01 = \$0.5445$.
- Signed trades total $-\$1.10$; fees total $+\$1.10$; funding residual **\$0**.

This exercise permits fractional cents, matching the research abstraction. A real cash ledger would require a specified rounding/residual-cash policy.

Returns. Gross return is 10%. Net return is $108.90 / 100 - 1 = 8.90\%$. The exact cost contribution is $-1.10 / 100 = -1.10\%$, or **-110 bp**. Cost divided by pretrade NAV is $1.10 / 110 = 1.00\%$, or **100 bp**. Subtracting 100 bp from gross return would incorrectly give 9.00%. Subtracting either fee again from NAV-implied net return would double count it.

Wealth paths. These figures are calculated before any rounding for display:

Close	Account A	Account B	Account NAV	Benchmark A	Benchmark B	Benchmark NAV
Initial	100.0000	0.0000	100.0000	50.0000	50.0000	100.0000
Day 1 after rebalance	54.4500	54.4500	108.9000	52.5000	52.5000	105.0000
Day 2	59.8950	54.4500	114.3450	57.7500	52.5000	110.2500
Day 3	47.9160	57.1725	105.0885	46.2000	55.1250	101.3250

Account cumulative return is **5.0885%**. Benchmark cumulative return is **1.3250%**. Their difference is **3.7635 percentage points**, or 376.35 bp. The relative wealth ratio $105.0885 / 101.325 - 1$ is a different measure; this project's cumulative active difference is not that ratio.

Linking. Divide each sleeve's cumulative dollar gains by initial capital \$100:

- Account A: $(10 + 5.445 - 11.979) / 100 = 3.4660\%$.
- Account B: $(0 + 0 + 2.7225) / 100 = 2.7225\%$.
- Account costs: $-1.10 / 100 = -1.1000\%$.
- Sum: **5.0885%**.
- Benchmark A: $(5 + 5.25 - 11.55) / 100 = -1.3000\%$.
- Benchmark B: $(0 + 0 + 2.625) / 100 = 2.6250\%$.

- Benchmark sum: **1.3250%**.
- Linked active A: **4.7660 pp**; B: **0.0975 pp**; cost: **-1.1000 pp**; sum **3.7635 pp**.

Daily active return is 3.90% on day 1 and zero on days 2–3, because after the first rebalance the portfolios have identical weights. Summing it gives 3.90%, but the initial dollar advantage subsequently earns the same market path as the rest of each portfolio. Its terminal value is not its day-1 value. Wealth-scaled P&L, rather than unweighted daily addition, reconciles to the terminal difference.

Scoring (20): funded NAV/sleeves/trades/fees and identity 5; return denominator distinction 4; both terminal values and cumulative difference 4; exact linked components 5; daily-sum explanation 2. A dollars-versus-percentage-point error is not a rounding mistake.

C. Worked risk and economics solution

Annual covariance is:

```
S = [[0.0400, 0.0050],
      [0.0050, 0.0100]]
a = [0.10, -0.10]
active variance = 0.10^2 × (0.04 + 0.01 - 2 × 0.005)
                 = 0.0004
TE = sqrt(0.0004) = 0.02 = 2.00% annualized
```

The covariance entries are annualized decimal-return variance/covariance units; TE is annualized decimal-return volatility. Do not multiply 2% by sqrt(252) again.

At correlation 0.80, covariance is $0.20 \times 0.10 \times 0.80 = 0.016$. Active variance becomes $0.01 \times (0.04 + 0.01 - 0.032) = 0.00018$. TE is **1.3416408%**. More positively correlated sleeves make this long-A/underweight-B substitution less risky relative to the benchmark, holding volatilities fixed. Correlation does not change the restriction itself or forecast a return benefit.

Budget is $1.20\% + 0.25\% = 1.45\%$. The original 2.00% proposal is **infeasible** under this allowance. Its weights sum to one, but that does not establish risk feasibility. Do not silently enlarge the budget to accommodate it.

A point-in-time proposed allocation versus fixed policy is not the same estimand as full-history earned active returns versus a drifting monthly benchmark. Interim exposure drift, the legacy transition, repeated changing decisions, estimated covariance error, and future regime changes matter. The budget is not a guarantee about realized returns, loss, or full-history TE.

Turnover reduction is 11 percentage points one-way, hence approximately 22 percentage points gross traded notional. Cost-rate savings are $0.22 \times 0.0005 = 0.00011$, or **1.10 bp**. At constant \$2 million NAV the illustration is **\$220 annually**. TE is a volatility, whereas cost drag is a modeled return reduction; subtracting them as if they were commensurate expected returns has no defensible decision meaning. A risk preference or utility criterion would have to be specified separately.

Country/sector look-through requires effective-dated holdings and classification with coverage controls; factor risk requires exposures, factor covariance and specific risk with a declared model; tax decisions require lots, basis, tax rates, holding periods and household/wash-sale policy. A trading penalty alone provides none of those. State what you would build, not that those capabilities already exist.

Scoring (20): covariance/vector/2% TE 5; correlation calculation and intuition 3; 1.45% budget and rejection 3; forecast-versus-realized distinction 3; \$220 and noncommensurate-unit explanation 4; missing factor/tax inputs 2.

D. SQL solution and coverage limits

The query reports **\$2.20** because the one daily fee is copied onto each of two trade rows before summing. Correct arithmetic at the wrong grain remains wrong.

First check unique trade natural keys and unique daily keys, nonmissing fields, and the expected rebalance roster. Then aggregate trade rows before comparison:

```
WITH trade_fees AS (  
  SELECT date, client, method, SUM(cost_dollars) AS trade_cost  
  FROM trades  
  GROUP BY date, client, method  
)  
SELECT t.date, t.client, t.method,  
       t.trade_cost, d.cost_dollars AS daily_cost,  
       t.trade_cost - d.cost_dollars AS fee_residual,  
       CASE WHEN d.date IS NULL THEN 1 ELSE 0 END AS missing_daily  
FROM trade_fees t  
LEFT JOIN daily d  
  ON t.date = d.date AND t.client = d.client AND t.method = d.method;
```

This retains a trade group lacking a daily record. It does **not** find an expected rebalance with no trade group at all. Compare both tables with an independently derived expected calendar/account/method roster and use anti-joins or a full outer comparison at the relevant rebalance grain. Daily dates with no planned rebalance legitimately have no trade rows. The repository's structural coverage checks precede its SQL, so the query need not pretend it establishes complete coverage alone.

If both the benchmark table's return and every copied daily benchmark-return field are increased by 1 bp, while saved gross/net active returns are both decreased by 1 bp, their arithmetic relationships can still agree. A correspondingly forged benchmark NAV can preserve that table's internal wealth path too. The source-driven benchmark reconstruction in `_reconstruct` challenges the saved benchmark using asset returns and independently propagated benchmark holdings. Agreement among mutually altered saved columns is not independent economic evidence. The executable adversarial test uses a deliberately larger 1-percentage-point shift; the failure mechanism is the same.

SUM may ignore NULLs. Missing a fee or missing a matching row is not evidence of zero fees. A NaN can also interact differently with database types and comparisons. Validate domains first and retain explicit missing-match flags; only use zero when the business rule establishes an actual zero.

Scoring (15): \$2.20 and duplicate-grain cause 3; corrected aggregation and unique keys 4; bidirectional/expected-roster coverage 3; consistently forged benchmark explanation 3; NULL/NaN distinction 2.

E. Source map, commands, and what success means

The following links point to actual repository files. Search by symbol rather than memorizing line positions:

- [walkforward.py](#): `decision_schedule` selects prior-close history; `funded_rebalance` solves closing wealth after costs; `run_walkforward` records earned returns and closing trades.
- [oversight_core.py](#): `_risk_controls` uses `panel.iloc[loc-cfg["lookback"]:loc]`, an execution-exclusive slice. `decision_risk_window` reconciles the precise end, start and count. `_reconstruct` propagates previous holdings, recomputes $\text{abs}(\text{change}) \times \text{rate}$, and checks `trade_self_financing`; it does not call the primary solver.
- [tests/test_oversight_adversarial.py](#): `hand_case` constructs the independently hand-derived example; `test_internally_consistent_forged_benchmark_is_reconstructed_from_assets` tests shared-column false reassurance.
- [run_oversight.py](#): `expected_checks` maps cases to domain detectors; `domain_detected` requires an expected failed control; `detected` additionally requires an overall BLOCK.

Run from the repository root:

```
python -m pytest -q -W error tests/test_oversight_adversarial.py::test_high_return_fee_denominator_and_exact_multiperiod_linking tests/test_oversight_adversarial.py::test_internally_consistent_forged_benchmark_is_reconstructed_from_assets
```

The intended result is **two passing tests**, not two passing portfolio validations. The first checks the hand-calculated clean path. The second test succeeds because the economically false case is rejected by reconstruction despite SQL consistency. Read its individual assertions instead of inferring correctness from green text.

The full public suite command is:

```
python -m pytest -q -W error
```

Its count can grow with repository changes; quote your observed output and Git revision, not a memorized old count. It does not rerun the licensed 2014–2025 empirical pipeline. Source-data reconstruction and the primary engine still share returns and policy assumptions. Separate functions do not establish organizational independence or independently certify optimizer optimality.

If the generic exception handler produces `validator_execution = BLOCK`, the intended mutation may not have been diagnosed at all. It could simply have crashed before reaching the relevant checks. Requiring an expected domain failure prevents counting that crash as demonstrated detection.

Scoring (20): exact execution-exclusive slice and window metadata 5; primary versus challenger funding with shared-input limit 5; correct test/command/assertion interpretation 5; expected-domain detector and generic-error distinction 5. Award command-execution credit only when you actually ran it.

F. Incident response model

One concise response is:

“We have established a recorded-fee inconsistency, not an executed \$100 cash loss. I would hold the affected evidence package and avoid inferring a new trade from unreliable reporting. I would compare the recorded fee with absolute notional times the declared rate, then check funding and daily-fee reconciliations. I would identify the affected key, inspect other rows with the same source/transformation, and trace dependent aggregates, workbook and memo. Correction requires

the authoritative source or transformation fix, a test that still catches the dirty input, and a clean reconciliation—not editing a displayed total or widening tolerance. Here the input was deliberately mutated in memory; the original file was never damaged, so the recorded clean retest is not a claim of production remediation.”

In the saved receipt this is **INC-005**. Expected detector family includes `trade_cost_formula`, `sql_fee_error`, or `sql_funding_error`. Read the actual failed controls and closure instead of assuming every related identity must fail for every mutation. Closure additionally requires the clean retest and unchanged originals; an observed block alone is insufficient.

Illustrative functional owners, such as portfolio accounting or data engineering, are proposed responsibilities—not people assigned by the project. A live event would need an actual owner, decision rights and update cadence. Do not promise resolution by a particular clock time without evidence.

Scoring (10): distinguishes recorded discrepancy from loss 2; containment and no chart patch 2; right identity/detector/scope 2; dirty and clean evidence plus unchanged originals 2; test-versus-production distinction 2.

G. Recommendation and ownership model

A defensible decision is **not to adopt the lower-turnover method as a universal default based on this result alone**. Its direct modeled fee savings are small at the stated liquid-ETF assumption; added TE is a client/model-tracking choice, not a price that can mechanically be netted against fees. The higher realized wealth is an observed exposure outcome from inspected history, not proof that the optimizer predicts returns.

A client might value fewer trades, tax constraints or an operational trading budget. This experiment has not measured those benefits. Define the missing objective and data, then test it separately. For example, before shortening the covariance lookback, keep the clock, universe and restriction rules fixed; compare paired exposure/risk/turnover behavior; require all accounting and coverage controls to pass; identify a PM/research reviewer; and retain the previous configuration and evidence for rollback. A new historical rerun is not a pristine prospective test.

The ownership statement must identify what you personally specified, calculated, executed, reviewed, or changed. The project can demonstrate analytical understanding and a proposed review process; it cannot supply missing direct people-management experience, establish institutional approval rights, or prove prior JIRA/Barra/FactSet/Tableau use. Support those responsibilities with truthful separate work-history evidence.

Scoring (10): conditional PM decision 2; no alpha/tax inference 2; unmodeled client objective and required evidence 2; acceptance/reviewer/rollback specifics 2; honest personal ownership and role gap 2.

Scorecard and next action

Write A __/5, B __/20, C __/20, D __/15, E __/20, F __/10, G __/10. Total __/100. This is a self-assessment tool, **not a certification, hiring probability, or prediction of interview success**.

- **85-100:** ready to rehearse a live screen-share, provided none of the critical mistakes below occurred. Ask another person to challenge a claim and choose an unpracticed case.

- **70-84:** promising understanding with targeted gaps. Redo the weakest numerical or code-trace section without notes before polishing the pitch.
- **Below 70:** return to source and hand calculations. More fluent phrasing will not repair an accounting, data-grain, or risk-unit error.

Regardless of total, stop and repair any of these critical mistakes: funding fees twice or from nowhere; confusing prior and pretrade NAV; summing daily active returns as cumulative performance; accepting duplicated joins; treating forecast risk as guaranteed realized risk; claiming tax optimization, personal coding work, employee management, or production authority not established by your actual experience.

Use an error log with four fields: **my incorrect claim**, **source/identity disproving it**, **corrected calculation or explanation**, and **date I reproduced the correction unaided**. Do not record “reviewed risk” as a remediation; record a demonstrable outcome such as “recomputed 2.00% TE from covariance and rejected the 1.45% budget violation.”

A second-pass challenge once the basics are fluent

Do not add unreviewed features during an interview. Instead, explain how you would investigate one of these changes:

1. **External contribution arrives midperiod.** Why does initial-capital P&L linking no longer directly measure investment return? Specify flow timing, TWR versus MWR objectives, and separate unit tests before changing the identity.
2. **A cap must apply continuously.** Why is the current drift report insufficient? Distinguish monitoring frequency, trading authority, threshold policy, executable orders, and cost effects before promising compliance.
3. **Two independent paths agree on the wrong field.** How would vendor semantics, a distribution event fixture, a second authorized source, and human contract review address the common-mode risk?
4. **A clean public clone passes all fixture tests.** What is still missing before saying the saved empirical study was reproduced? Name licensed inputs, their manifest/vintage, the full run, current receipts, and report reconciliation.

These questions do not have a one-line production solution. The strong answer makes the new contract, evidence, limits, and decision owner explicit rather than stretching an old PASS into a new claim.