Fama–MacBeth (1973) Tables 1–3 — Replication Plan

(PhD team: Peter Yin& Bing Han; Duration: 10 days)

0) Purpose & scope

This document records how we are going to replicate Table 1 (period definitions & counts), Table 2 (estimation-period portfolio statistics), and Table 3 (monthly cross-section regressions). **This is a plan only and the formats and names of the output datasets and deliverables might be different in the final code submission.**

1) Data, sources, and extraction

1.1 Universe & sample

- Listings: NYSE common stocks only (SHRCD ∈ {10, 11}; EXCHCD = 1).
- Sample start: 1926-01.
- Sample end: 1968-06.
- Security identifiers: use PERMNO as the stable key (carry PERMCO for diagnostics).

1.2 Tables & pulls

- CRSP monthly returns (crsp.msf) joined to a filtered security set from mse — Task: "Crsp monthly return data task" — Output: data_ret (monthly ret, later mkt, plus year, month).
- Exchange code SHRCD and EXCHCD comes from crsp.msenames.
- CRSP delisting file (crsp.msedelist) for dlstdt, dlret — Task: "crsp delist data task" — Output: data_delist.
- Risk-free rate (for $\gamma_0 - rf$ diagnostics only), ff.factors_monthly — Task: "rf data task" — Output: data_factor (with year, month, rf).
- Equal-weighted market proxy (Fisher-style) — Task: "Market return (Fisher, 1996) task" — Output: mkt = mean(ret) by calendar month merged back to data_ret. Data comes from crsp.msi.

Ownership (co-leads): Peter & Bing — both author and run the WRDS queries and extraction scripts; ensure row counts, date ranges, and filters match specs; create checkpoint CSVs/RDS.

2) Data cleaning & construction rules

2.1 Filters & joins

- Keep NYSE, common shares; drop rows with missing ret.
- Build calendar keys year, month from date.
- Generate monthly equal-weighted market mkt via month-wise mean of individual ret.

2.2 Delisting treatment in test months

- For each test month (year = i, month = m), left-join dlret and replace ret with (1 + ret) × (1 + coalesce(dlret, 0)) − 1 when dlstdt equals the test month.
- Exclude names that delisted before the current month via a delist_set.

Ownership (co-leads): Peter & Bing — implement and validate delisting adjustment; produce QA logs for counts before/after adjustment.

3) Period design (Table 1 backbone)

- Nine period blocks with formation (F), estimation (E), and testing (T) years defined in the provided periods list.
- Eligibility thresholds used during replication: Formation β estimation (F): per-stock min observations ≥ 48 months; Estimation β/σ (E): per-stock min observations ≥ 60 months.
- Available vs. meets-requirement counts (Table 1): Available: names with a return in the first month of T; Meets data requirement: names satisfying both F and E thresholds.

Ownership (co-leads): Peter & Bing — materialize the nine blocks from periods, compute per-block counts (available vs. meets-requirement), and export the Table 1 grid + counts.

4) Functions to build (as implemented)

- estimate_beta(data, min_obs): OLS time-series fit ret ~ mkt per unit (stock or portfolio), returning beta, se_beta, R^2, sd_ret, sd_resid.
- assign_portfolios(beta_df, n_port = 20): sort by β and allocate to 20 portfolios with tails receiving leftovers; output adds portfolio ∈ {1,…,20}.
- Period grid helper (implicit via periods list): extract fstart…fend, estart…eend, tstart…tend for P1–P9.
- Formation-stage wrapper: filter data_ret to F; compute stock-level beta_f with min_obs = 48; assign 20 portfolios.
- Estimation-stage wrapper: re-estimate stock-level beta_e and sd_resid in rolling E windows with min_obs = 60; construct month-level portfolio panels for testing by averaging over portfolio constituents.
- run_fmb(data, formula_str): for each (year, month) in test data, run cross-section OLS on 20 portfolios; returns tidy rows with coefficients estimate by term and r_squared.
- fmb_coef_stats(fmb_model, rf_data): aggregates monthly γ's to mean, sd, t-stat; reports acf1 (γ autocorrelation), and γ adjusted by rf if needed; includes mean/sd of monthly r_squared.

Ownership of implementation: Peter — implements and unit-tests (1)–(5); Bing — implements and unit-tests (6)–(7).

5) Portfolio formation, estimation & testing

5.1 Formation (by block)

- Input: formation_data = data_ret in F years.
- Compute beta_f per stock via estimate_beta(…, min_obs = 48).
- Apply assign_portfolios(beta_f, 20).
- Output: stock → portfolio map for the block.

5.2 Estimation (rolling, aligned to T)

- For each test calendar year in T (loop i = tstart…tend):
- Estimation sample: estimation_data = data_ret in [estart, eend + n], where n = i − tstart (rolling extension).
- Compute beta_e per stock with min_obs = 60.
- Build month panels: join beta_f ∩ beta_e ∩ current month returns (post-delist adjustment & pre-exclusion of prior delists).
- Aggregate to portfolio-level means: beta, sd_resid, ret; append to result_all with keys (year, month, portfolio).

5.3 Estimation-period portfolio regressions (Table 2)

- Construct beta_p_est by re-running estimate_beta on portfolio monthly series within E.
- Compute Table 2 stats: beta, se_beta, R^2, sd_resid, and the ratio sd_resid / sd_resid_i (baseline from first test month).

5.4 Testing (Table 3 models)

- Prepare result_all, then add beta2 = beta^2.
- Define subsample slices: Entire 1935–1968; decades/half-decades; 1935–45, 46–55, 56–68.
- Cross-section OLS models (monthly, 20 portfolios):
- Model 1 (Panel A): ret ~ beta
- Model 2 (Panel B): ret ~ beta + beta^2
- Model 3 (Panel C): ret ~ beta + sd_resid
- Model 4 (Panel D): ret ~ beta + beta^2 + sd_resid
- For each slice, run run_fmb(…, formula); then summarize with fmb_coef_stats() to produce mean γ's, standard deviations, t-statistics, acf(1) of γ (and optionally γ − rf), and mean/sd of R^2.

6) Model & regression details

- Time-series step (per stock / per portfolio): OLS ret ~ mkt using monthly data in F and E windows; store beta, se_beta, R^2, sd_ret, sd_resid.
- Cross-section step (per month, 20 portfolios): OLS across 20 observations with the four Table-3 formulas; intercept is $\gamma_0 t$ (no restrictions); capture coefficients and r_squared monthly.
- Aggregation step: over target months in each slice, report mean γ, sd γ, t = mean_gamma / (sd_gamma / √T), acf(1) of the γ series; optionally the same for γ − rf; report mean/sd R^2.

7) Dedicated Table-visualisation track

We reserve explicit time and modules for styling so the tables match the originals as close as we can. We will use kableExtra (LaTeX/HTML with booktabs and spanners) and/or gt (spanners/footnotes) depending on fidelity.

Visualisation modules (optional: we might just create the tables within our main codes without creating separate files)

- viz/theme_fm_jpe.R — helpers: fmt_num(dp), fmt_t(t), fmt_r2(dp), fmt_pct(dp); header utilities; rules/spacing utilities; caption_note().
- viz/table_fm_t1.R — renders Table 1 (period grid + counts; two count columns).
- viz/table_fm_t2.R — renders Table 2 (β, seβ, R^2, sdε, ratio).
- viz/table_fm_t3.R — renders Table 3 Panels A–D (γ mean, sd, t, acf(1), R^2).
- Deliverables: LaTeX/PDF and PNG exports for each table; side-by-side visual diff versus the original PDF.

8) Outputs & acceptance criteria

Table 1

- Deliverables: period grid (P1–P9) + two counts per block ("available" and "meets requirements").
- Checks: period edges match; counts plausible across time; reproducible from a single script.

Table 2

- Deliverables: per-portfolio estimation-period statistics; include ratio sd_resid / sd_resid_i.
- Checks: monotone pattern in beta across portfolios; reasonable R^2, sd_resid, ratios.

Table 3

- Deliverables: Panels A–D for each slice with: mean γ's, sd γ's, t-stats, acf(1), mean/sd R^2; store the monthly γ time series.
- Checks: 20 obs each month; stable coefficient signs/magnitudes; acf(1) in [−1, 1].

Visual fidelity (tables)

- Headers, spanners, panel labels exactly positioned.
- Rules (top/bottom double; mid single) and panel separators match.
- Decimals & signs as in original (coefficients, t-stats, R^2).
- Footnotes/captions and slice labels present and placed identically.

9) Roles & handoffs (equal contribution)

Data engineering & portfolio construction — Peter & Bing (co-leads)

- WRDS extracts; filters; mkt construction; delist adjustments.
- Formation (beta_f) & portfolio assignment; Estimation (beta_e) rolling fits.
- Build full testing panel result_all + beta2.

- Produce Table 1 counts & Table 2 estimation statistics.

Cross-section regressions & inference — Bing (lead), Peter (support)

- Implement/run run_fmb() for Panels A–D across all slices; compute γ series; aggregate with fmb_coef_stats() (including acf(1)).
- Package per-slice Panel results for visualisation modules; sensitivity runs; summary memo.

Visualisation & aesthetics — Peter (lead), Bing (review)

- Implement theme and table modules (viz/ files).
- Recreate Tables 1–3 to match the original look: fonts, lines, alignment, decimals, panel notes; export LaTeX/PDF and PNG snapshots.
- Maintain a style guide and ensure consistent theming.

10) 10-day timeline & milestones

Day	Peter Yin	Bing Han	Milestones
1	Re-run pulls; rebuild data_ret, data_delist, data_factor; construct mkt; QC	Mirror run; cross-check counts; set slice helpers	Clean extracts + monthly coverage report
2	Implement/validate delist rule; add year, month; finalise filters	Implement estimate_beta() tests & perf checks	Data foundation locked
3	Build P1–P9 grid; compute Table-1 counts	Independently recompute spot counts; reconcile	Table 1 (raw) ready
4	Formation: compute beta_f (min_obs = 48); assign 20 portfolios	Estimation rolling fits beta_e (min_obs = 60) smoke test	Formation & Estimation logic verified
5	Assemble monthly testing panel (result_all) with delist handling	Add beta2; run one month across Panels A–D (smoke test)	Full panel sample validated
6	Start Viz Track: create theme_fm_jpe.R; scaffolds table_fm_t1/t2/t3 (kableExtra/gt)	Build run_fmb() & fmb_coef_stats(); verify with one slice	Viz scaffolds compile; model helpers green
7	Table 2: compute stats; implement exact layout (headers/spanners/rules/decimals); export PDF/PNG	Run Models 1–4 for all slices; produce γ series	Table 2 (styled v1); γ series complete
8	Table 3: implement Panels A–D layout; add footnotes & spanners; export	Validate Panel values vs aggregates; add acf(1)	Table 3 (styled v1)
9	Table 1: final aesthetics; side-by-side visual-diff, tweak spacing/lines/decimals	Sensitivity passes; finalize reporting tables	Tables 1–3 (styled v2)
10	Final visual-diff pass; README & differences/limitations; package tables/figures	Clean-machine replicate; checksum outputs	Release candidate delivered

11) Reproducibility, structure, and tooling

```
01_extract/           # WRDS pulls (co-owned: Peter & Bing)
02_build/             # formation, estimation, testing panel (co-owned)
03_models/            # FMB runs + aggregates (Bing lead)
viz/                  # theme + table builders (Peter lead)
outputs/              # PDFs/PNGs/CSVs of tables & γ series
utils/                # helpers (formatting, checks)
README.md             # instructions (session info, package versions)
```

- Version control & logs: commit at each Day's milestone; save intermediate RDS (beta_f, beta_e, result_all, and Tables 1–3).
- Quality gates: deterministic portfolio assignment; every monthly cross-section has 20 obs; documented delist thresholds (48/60 months); visual consistency (fonts, number formats, captions).

12) Quick checklist (deliverables)

- Table 1 grid + counts (P1–P9), styled
- Table 2 portfolio statistics (β, se, R^2, sd_resid, ratio), styled
- Table 3 Panels A–D across slices (γ means, sds, t, acf(1), R^2)
- γ time-series CSV for each model & slice (optional)
- Figures: β monotonicity, γ-series, residual σ distributions, Gantt timeline
- README + run scripts + QA logs

Github Repo Address: https://github.com/bing-han-dk/maf900-a3-team