

Supplementary Information

When Validation Defines Abnormality: Prediction Targets in Bayesian Accrual Measurement
Quality & Quantity – Online Resource 1

S1. Benchmark catalogue and Bayesian configuration

Table S1. Candidate normal-accrual benchmark catalogue

ID	Short name	Status	Main normal-accrual feature
M01	Jones	Core	Revenue change and PPE capacity
M02	Modified Jones	Core	Revenue change net of receivables plus PPE
M03	Performance Mod. Jones	Core	Modified Jones with lagged operating performance
M04	Dechow–Dichev	Core	Cash-flow mapping across adjacent periods
M05	McNichols	Core	Cash-flow mapping integrated with Jones predictors
M06	McNichols performance	Core	McNichols mapping with performance control
M07	Ball–Shivakumar style	Core	Asymmetric response to adverse cash-flow states
M08	Performance/volatility ext.	Secondary	Performance plus size/volatility controls
M09	No-lookahead cash-flow/Jones	Core	Cash-flow/Jones mapping excluding lead CFO
M10	Operating-cycle extension	Secondary	Performance-matched mapping with operating-cycle proxy

Each admissible family is estimated under pooled and firm-random-intercept structures where feasible. The common Bayesian configuration uses a Student- t likelihood, four chains, 12,000 iterations per chain, 4,000 warmup iterations, `adapt_delta`=0.99, and `max_treedepth`=15. Standardized slopes and the intercept use $N(0, 0.10)$ priors; scale parameters use exponential priors and the Student- t degrees of freedom uses a Gamma prior. Prior predictive simulation is used to assess prior plausibility. Primary exact-K-fold inclusion is governed by $\widehat{R}$, bulk and tail effective-sample-size, divergence, and treedepth diagnostics; posterior predictive checks are retained as secondary model-adequacy diagnostics.

S2. Simulation evidence for the heterogeneity mechanism

Table S2. Row-minus-grouped Firm-RE weight premium under simulated structure

Simulation block	σ_{firm}	Row Firm-RE	Group Firm-RE	Premium	$Pr(\text{premium} > 0)$
Static LMER	0.0	0.1764	0.1797	-0.0033	0.4533
Static LMER	0.1	0.9288	0.0665	0.8622	1.0000
Static LMER	0.3	0.9733	0.0168	0.9565	1.0000
Temporal-dependence LMER	0.0	0.2090	0.1958	0.0132	0.4953
Temporal-dependence LMER	0.1	0.9191	0.0855	0.8336	1.0000
Temporal-dependence LMER	0.3	0.9700	0.0197	0.9503	1.0000

Notes. The static block varies panel length $T \in \{3, 7, 15\}$ with 1,500 replications per heterogeneity level. The temporal block additionally varies same-firm AR(1) persistence over $\rho \in \{-0.10, -0.05, 0, 0.05, 0.10\}$ and contains 7,500 replications per heterogeneity level. A small Bayesian **brms** confirmation pilot produces the same qualitative positive-premium pattern once heterogeneity is introduced, but its low replication count and a high- $\hat{R}$ boundary cell make it diagnostic rather than primary evidence. The simulation is used to discriminate mechanisms, not to calibrate the empirical magnitude.

S3. New-firm scoring and pooled-only mechanism checks

Table S3. Sensitivities that isolate firm-level heterogeneity

Panel A: Fully marginal new-firm scoring				
Space	Population-level Firm-RE	Marginal Firm-RE	Grouped leading benchmark	
No lookahead	0.0368	0.0000	M09 no-lookahead, pooled	
Ex post	0.0405	0.0000	M06 McNichols performance, pooled	

Panel B: Pooled-only stacking				
Space	L1 family shift	Max family shift	Leading family under both targets	
No lookahead	0.0151	0.0075	No-lookahead family	
Ex post	0.0141	0.0070	McNichols performance	

Notes. Marginal new-firm scoring integrates over a newly drawn firm random effect rather than evaluating the Firm-RE candidate at the population mean. Pooled-only stacking removes all Firm-RE candidates and compares accounting-family allocations under the two held-out-unit targets. The maximum pooled-family weight shifts are 0.0070 ex post and 0.0075 no lookahead.

S4. Denominator and fold-local preprocessing boundaries

Table S4. Boundary diagnostics for benchmark allocation and abnormal-accrual classification

Space	Diagnostic	Baseline	Sensitivity	Interpretation
No lookahead	Firm-RE row-group shift	0.5559	0.4161	Remains positive; top-cell identity less stable
Ex post	Firm-RE row-group shift	0.5801	0.9568	Remains positive under fold-local preprocessing
No lookahead	Raw top-5% Jaccard	0.6917	0.6854	Material turnover persists
Ex post	Raw top-5% Jaccard	0.6485	0.6148	Material turnover persists
No lookahead	z_{est} top-5% Jaccard	0.4196	0.4331	Material turnover persists
Ex post	z_{est} top-5% Jaccard	0.3776	0.4173	Material turnover persists

Table S4 continued. Denominator sensitivity of uncertainty-scaled tails

Space	Original z_{est} Jaccard	Floor/cap range	z_{est} vs. z_{pred} range
No lookahead	0.4196	0.4196–0.6917	0.2329–0.3196
Ex post	0.3776	0.3776–0.6485	0.2548–0.3133

Notes. Fold-local values re-estimate winsorization and standardization constants using the training fold. Denominator diagnostics show that the larger divergence of the uncertainty-scaled score is partly denominator-dependent; the raw-residual result does not rely on this scaling choice.

S5. Subsequent accounting composition of target-specific tails

Table S5. Descriptive subsequent-outcome coefficients by top-tail membership

Score	Tail membership	Future CFO	Future earnings	Future ROA
Raw	Common to both	-0.0238	-0.0160	-0.0220
Raw	Row-only	-0.0383	-0.0280	-0.0249
Raw	Grouped-only	0.0562	0.0341	0.0396
z_{est}	Common to both	-0.0275	-0.0216	-0.0287
z_{est}	Row-only	-0.0214	-0.0164	-0.0185
z_{est}	Grouped-only	0.0064	0.0066	0.0005

Notes. Coefficients are descriptive discriminating observables, not causal effects and not direct evidence of managerial intent. The underlying regressions use 3,936 firm-years (3,918 for accrual reversal, not shown here) and 568 firms (567 for accrual reversal). Full source exports supplied with the reproducibility package include standard errors, p -values, Benjamini–Hochberg adjusted q -values, effect-size transforms, and the corresponding predictive-uncertainty-scaled specification.

Source and reproducibility note

The tables above are condensed from analysis-pipeline exports retained with the project records. Raw FiinPro/FiinPro-X observations cannot be redistributed under the data licence. Non-proprietary aggregate outputs, code identifiers, model manifests, fold assignments, diagnostics, and source-to-manuscript mappings can be supplied in anonymized form during peer review where needed to audit the reported measurement pipeline without treating the licensed raw data as public.