Regime Detection: BOCPD vs MMD

Bayesian Statistics — Final Project

Elijah Whitham-Powell

Recap: Bayesian Online Changepoint Detection

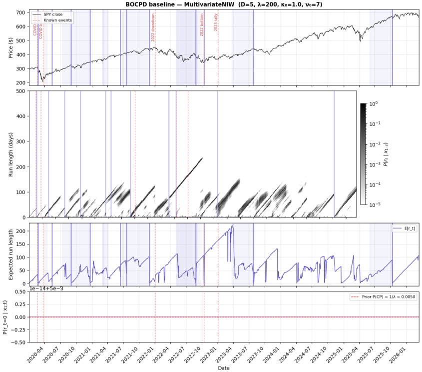
From the preview presentation (Adams & MacKay, 2007):

Maintain a posterior over **run length** r_t — time since the last changepoint.

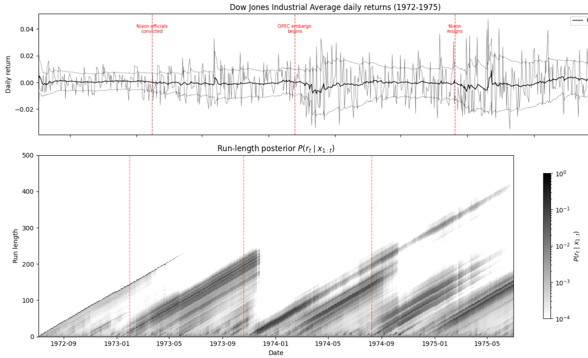
At each observation, update recursively:

$$P(r_t, x_{1:t}) = \sum_{r_{t-1}} P(r_t \mid r_{t-1})\, P(x_t \mid r_{t-1})\, P(r_{t-1}, x_{1:t-1})$$

Extension: I generalized from univariate (Normal-Inverse-Gamma) to **multivariate** data using Normal-Inverse-Wishart priors.



Validating the Implementation: Reproducing Figure 3



Using the paper's exact parameters:

Data	DJIA daily returns Jul 1972 – Jun 1975
μ_0	0
κ_0	1
α_0	1
β_0	10^{-4}
λ	250

What I Built

Three packages, shared data pipeline, all open source.

`bocpd`

Pluggable observation models:
— Univariate NIG
— Multivariate NIW

Three hazard functions:
— Constant, Increasing, Decreasing

Run-length truncation (r_{max}: $>100\times$ speedup)

`mmd_regime_detection`

Sliding window MMD with permutation testing

RBF kernel via `kta`

Boundary extraction with z-score thresholding

No distributional assumptions

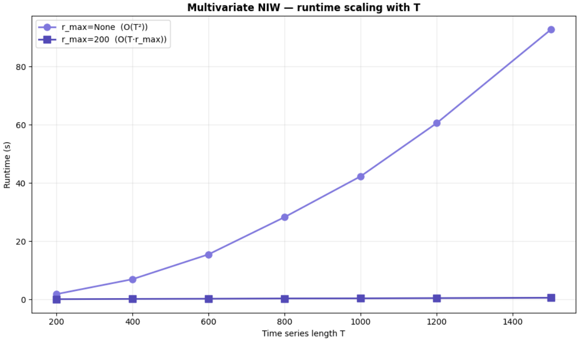
`finfeatures`

Yahoo Finance data source

Feature pipeline:
— Log-OHLCV transform
— 5 features per day

Both methods consume the same data through the same pipeline

Making It Practical: Run-Length Truncation



Without truncation: $O(T^2)$
NIW on 1500 obs: $\approx 103\,\mathrm{s}$

With $r_{\max} = 200$: $O(T \cdot r_{\max})$
Same data: $\approx 0.6\,\mathrm{s}$

170× speedup

Comparison Design: 2×2

Separate the **framework effect** from the **feature effect**.

	Univariate (log returns)	**Multivariate** (log OHLCV, $D = 5$)
BOCPD	NIG, $\lambda = 100$	NIW, $\lambda = 100$, $r_{\max} = 600$
MMD	RBF, window=30, step=5	RBF, window=30, step=5

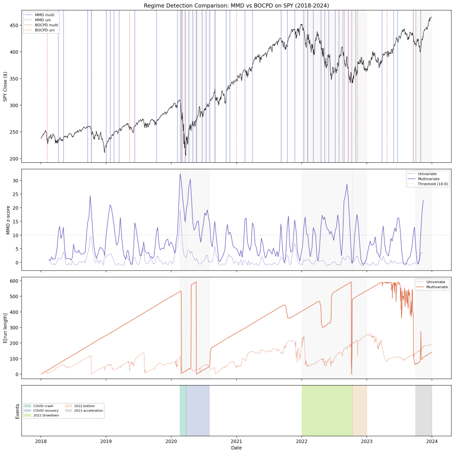
Same data: SPY daily prices, 2018–2024

Same validation: 5 known market events

Key question: Does adding features help both methods equally?

(Spoiler: no — opposite effects.)

SPY Results: Detection Overview



BOCPD multivariate: 4 boundaries MMD multivariate: 43 boundaries

Full-resolution notebook: `https://github.com/whitham-powell/regime-detection-comparison/blob/master/notebooks/comparison.md`

Validation Against Known Events

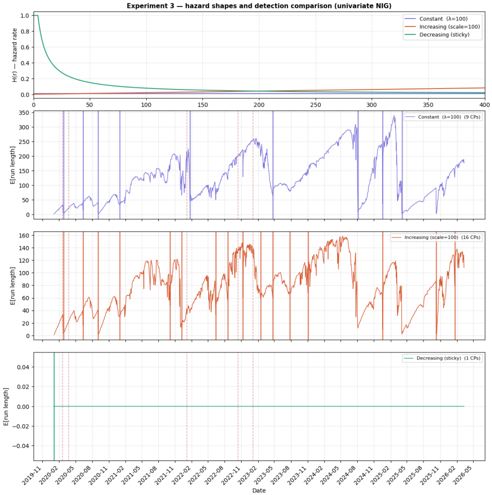
	BOCPD		MMD	
Event	Uni	Multi	Uni	Multi
COVID crash	N	Y (+12d)	Y (+1d)	Y (+8d)
COVID recovery	N	Y (−22d)	N	Y (−5d)
2022 drawdown	Y (+31d)	Y (+277d)	N	Y (+10d)
2022 bottom	Y (+29d)	N	N	Y (+19d)
2023 acceleration	N	Y (+1d)	N	Y (−13d)
Detected	**2/5**	**4/5**	**1/5**	**5/5**

Feature effect asymmetry: multivariate features make BOCPD *more conservative* (9→4 boundaries) but make MMD *more sensitive* (1→43 boundaries).

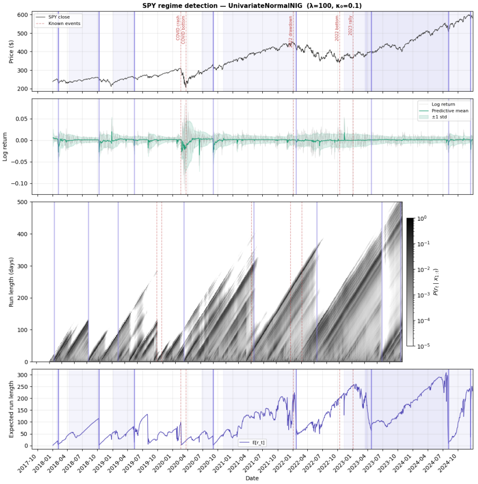
BOCPD Sensitivity: Hazard Function Shape

Your **prior belief about how regimes age** directly controls detection.

Hazard	CPs
Constant	9
Increasing	16
Decreasing	1

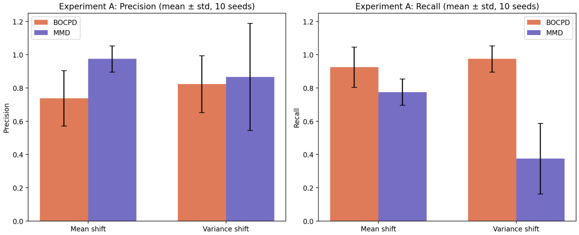


Bayesian Updating in Action: Predictive Envelope



Envelope **widens** at changepoints (fresh prior), **tightens** as evidence accumulates.

Synthetic Validation: Correctly Specified Model



5 Gaussian segments, 200 obs each, 4 planted CPs.

Mean shift (A1):

BOCPD recall **0.93**, MMD recall 0.78

Both high precision

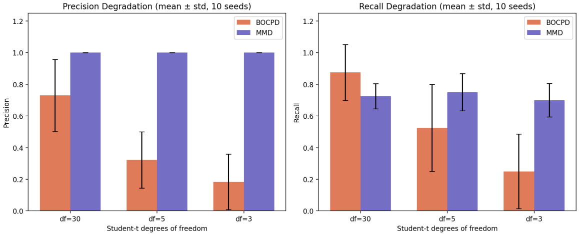
Variance shift (A2):

BOCPD recall **0.97**, MMD recall **0.38**

Full-resolution notebook:

`https://github.com/whitham-powell/regime-detection-comparison/blob/master/notebooks/comparison_synthetic.md`

Model Misspecification: Heavy-Tailed Data



Same mean-shift task, but Student-t observations.

	df=30	df=5	df=3
Prec.	0.73	0.32	**0.18**
Recall	0.88	0.53	**0.25**
FP/run	1.5	4.2	**4.6**
CI cov.	0.93	0.88	**0.62**
Prec.	1.00	1.00	**1.00**
Recall	0.72	0.75	**0.70**

Framework Comparison and Practical Takeaways

Aspect	**BOCPD**	**MMD**
Distribution	Gaussian (NIG / NIW)	None (characteristic kernel)
Online	Yes (by design)	No (requires full windows)
Output	$P(\text{CP})$ with credible intervals	z-score from permutation null
Key tuning	Hazard λ, prior hyperparams	Window size, bandwidth, threshold
Multivariate	NIW (full covariance)	Native (kernel on feature vectors)
Cost	Fast (conjugate updates)	Expensive (permutation test)

Use BOCPD for:

Real-time monitoring. Calibrated uncertainty. Streaming univariate signals.

Use MMD for:

Offline review. Robustness to non-Gaussian data. High-dimensional features.

Best strategy:

Combine both — BOCPD for alerting, MMD for periodic review.

Limitations and Future Work

Limitations

BOCPD: Gaussian assumption — CI coverage degrades to 0.62 under heavy tails (df=3).

MMD: 43 boundaries suggests over-sensitivity. Window and threshold require tuning.

Both: No ground truth for real market regime boundaries.

Future Work

Student-t observation model for BOCPD — robustness to fat tails while preserving Bayesian uncertainty.

Warm-up and state persistence — enable true online deployment without reprocessing history.

Questions?

References

Adams, R. P. & MacKay, D. J. C. (2007). Bayesian Online Changepoint Detection.
`https://arxiv.org/abs/0710.3742`

Gretton, A., Borgwardt, K., Rasch, M., Schölkopf, B., & Smola, A. (2012). A Kernel Two-Sample Test. *JMLR*, 13:723–773.

Code

`bocpd`: `https://github.com/whitham-powell/bocpd`
`mmd-regime-detection`: `https://github.com/whitham-powell/mmd-regime-detection`
`finfeatures`: `https://github.com/whitham-powell/finfeatures`
`regime-detection-comparison`:
`https://github.com/whitham-powell/regime-detection-comparison`

Rendered Notebooks (Full-Resolution Figures)

All experiments are available as rendered Markdown on GitHub.

BOCPD Experiments (`bocpd-regime-detection`)

https://github.com/whitham-powell/bocpd-regime-detection/blob/master/examples/rendered/bocpd_experiments.md
https://github.com/whitham-powell/bocpd-regime-detection/blob/master/examples/rendered/bocpd_experiments_univariate.md
https://github.com/whitham-powell/bocpd-regime-detection/blob/master/examples/rendered/rmax_runtime_comparison.md
https://github.com/whitham-powell/bocpd-regime-detection/blob/master/examples/rendered/demo_spy_regime_detection.md

https://github.com/whitham-powell/bocpd-regime-detection/blob/master/examples/rendered/replicate_adams_mackay_fig3.md

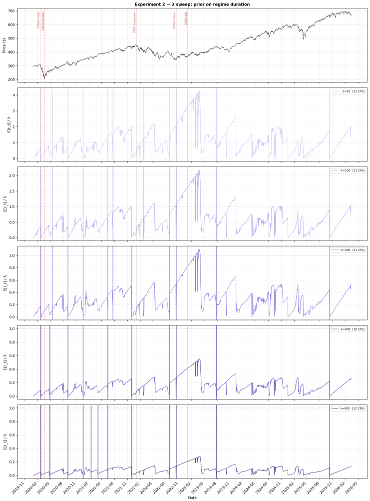
Comparison Experiments (`regime-detection-comparison`)

https://github.com/whitham-powell/regime-detection-comparison/blob/master/notebooks/comparison.md

https://github.com/whitham-powell/regime-detection-comparison/blob/master/notebooks/comparison_synthetic.md

Backup Slides

Backup: λ Sweep (Multivariate NIW on SPY)

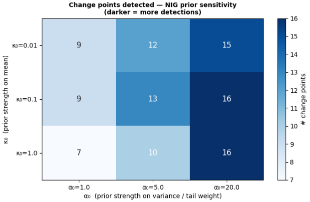


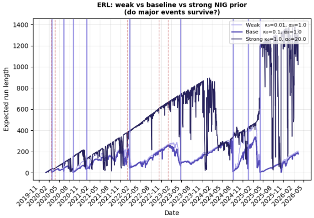
λ	Boundaries
50	11
100	11
200	11
400	10
800	12

Multivariate NIW is **insensitive to λ** in this range — the posterior updates dominate the hazard prior.

Univariate NIG shows more variation ($10 \rightarrow 7$ across the same range) because the 1D signal provides less evidence per observation.

Backup: NIG Prior Sensitivity ($\kappa_0 \times \alpha_0$ Grid)





α_0 **dominates**: controls Student-t degrees of freedom ($df = 2\alpha_0$).

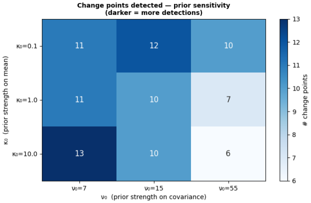
$\alpha_0 = 1$: very heavy tails ($df = 2$), tolerant of outliers $\to$ 7–9 CPs.

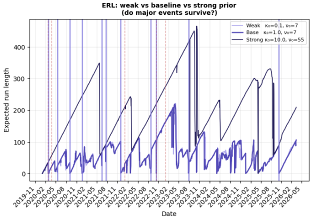
$\alpha_0 = 20$: light tails ($df = 40$), flags smaller deviations $\to$ 15–16 CPs.

κ_0 **has a weak effect**: controls how quickly the mean adapts. Varying κ_0 by $100\times$ changes CP count by only 1–2.

Takeaway: α_0 is the prior knob that matters most in 1D.

Backup: NIW Prior Sensitivity ($\kappa_0 \times \nu_0$ Grid)





ν_0 **dominates**: controls how quickly the covariance estimate adapts.

$\nu_0 = 7$ (weak): 11–13 CPs.
$\nu_0 = 55$ (strong): 6–10 CPs.

Opposite direction from univariate α_0: stronger covariance prior makes the model *less* responsive, requiring larger distributional shifts to trigger a detection.

κ_0 has a weaker but non-negligible effect in the multivariate case: counts vary by 1–3 per row.

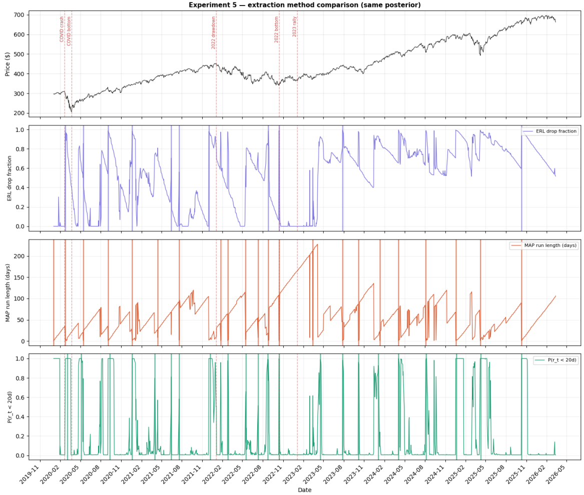
Backup: MMD Window Size Sensitivity

Window	Boundaries	Mean |Offset| (days)	Runtime (s)
20	25	17.2	6.6
30	42	11.0	8.8
40	53	11.8	10.7
60	74	27.0	29.4

Boundary count scales roughly linearly with window size. Smaller windows detect more boundaries but at higher computational cost. Window=30 balances sensitivity and offset accuracy (11 days mean offset from known events).

The z-score threshold of 10.0 was chosen to avoid the zero-p-value floor of the permutation test (500 permutations). A higher threshold reduces boundary count; the sensitivity sweep in the comparison notebook explores this.

Backup: Extraction Method Comparison (Same Posterior)



Three ways to summarize the same posterior:

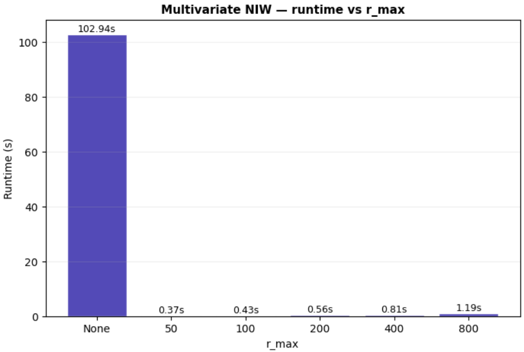
Method	CPs
ERL drops	11
MAP r_t	22
$P(r_t < 20)$	22
All three agree	5

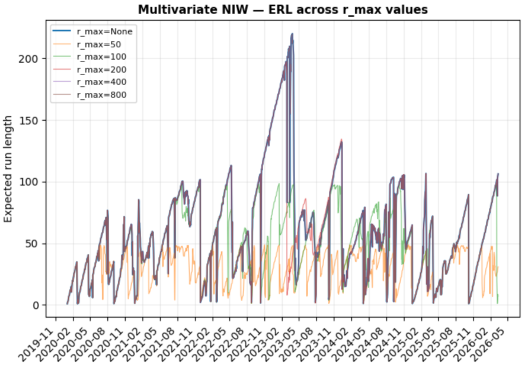
ERL is the most conservative — integrates over all run lengths, smoothing out transient fluctuations.

MAP is noisiest — can snap to $r = 0$ on a single unusual observation.

All experiments in this presentation use ERL extraction.

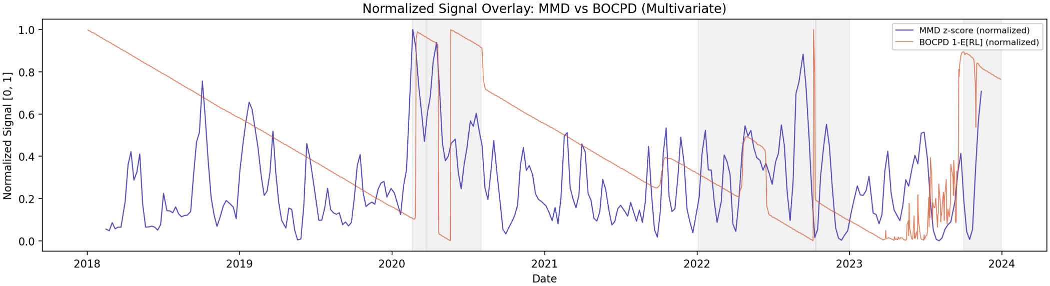
Backup: $r_{\max}$ Accuracy — ERL Overlay (Multivariate NIW)





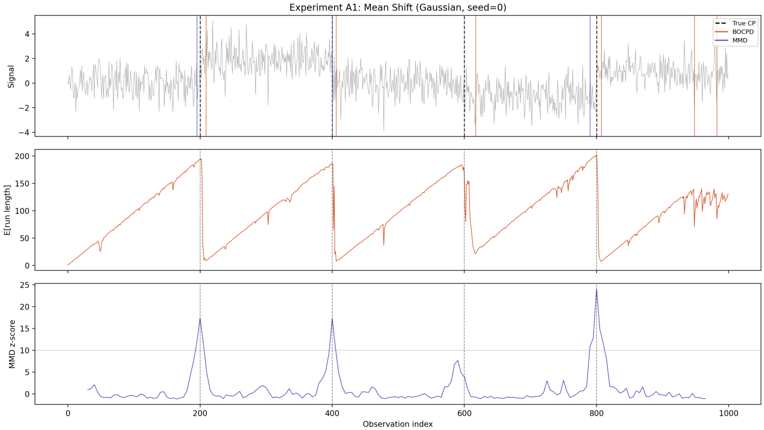
$r_{\max} = 50$: visibly diverges (cannot represent regimes > 50 days). $r_{\max} = 200+$: qualitatively identical to untruncated. Choosing $r_{\max}$ at 2–4$\times$ expected regime duration balances speed and fidelity.

Backup: Normalized Signal Overlay (MMD vs BOCPD)



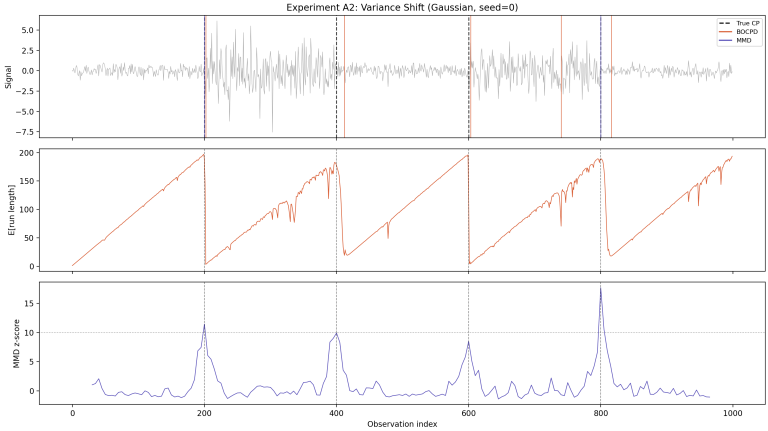
Both signals peak during the same major events (COVID crash, 2022 drawdown) but differ in dynamics: BOCPD reacts sharply to individual observations while MMD builds gradually across a window. The signals are complementary rather than redundant.

Backup: Synthetic A1 — Mean Shift (Single Seed)



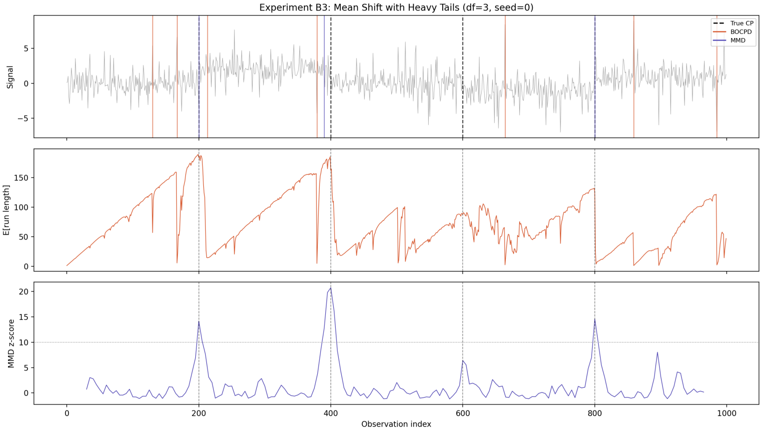
BOCPD detects all 4 CPs with slight latency (9–17 obs late). MMD detects 3/4, missing the $0 \to -1$ transition. Both methods' continuous signals show clear responses at the true change point locations (dashed lines).

Backup: Synthetic A2 — Variance Shift (Single Seed)



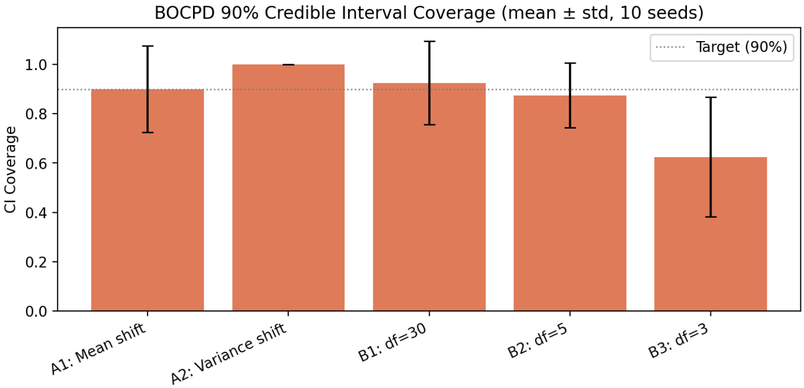
BOCPD detects 5 boundaries (4 true + 1 FP) — the NIG model's explicit variance tracking catches all scale changes. MMD detects only 2/4: the $0.5 \to 2.0$ and $0.5 \to 0.5$ transitions (largest contrasts). The $2.0 \to 0.5$ and $1.5 \to 0.5$ shifts fall below MMD's z-score threshold.

Backup: Synthetic B3 — Heavy Tails df=3 (Single Seed)



BOCPD fires 7 detections — the first two (t=130, 167) are false positives from outliers within the first segment, *before* any true change point. The Gaussian model interprets heavy-tailed draws as evidence of a new regime. MMD detects 3/4 true CPs with zero false positives.

Backup: BOCPD 90% Credible Interval Coverage



Under correct specification (A1, A2, B1): coverage at or above the 90% target — credible intervals are well-calibrated. Under misspecification (B2, B3): coverage degrades to 0.88 and 0.62 respectively, confirming that Bayesian uncertainty estimates require a correctly specified model.