

# Audit Report: Workflow 34c88b2bf0194f5b9b72793845290a52

## Overview

- Source JSON: reports\workflow-report-34c88b2bf0194f5b9b72793845290a52.json
- Run ID: 34c88b2bf0194f5b9b72793845290a52
- Papers input: papers\Use\_of\_Cumulative\_Sums\_of\_Squares\_for\_Re.pdf
- Started at: 2026-02-15T21:46:06.642379+00:00
- Finished at: 2026-02-15T21:52:34.727145+00:00
- Duration: 0:06:28.084766

## Effective Configuration

- Chat model: gpt-5-nano
- Embedding model: text-embedding-3-large
- Top K: 10
- Chunk words / overlap: 350 / 75
- Batch size: 64
- Database URL configured: True

## Step Outcomes

- prep: completed
- ingest: num\_pdfs=1, num\_papers=1
- enrich: openalex=0, citec=0
- econ\_data: fetched
- agentic: completed
- index: indexed
- report\_store: pending

## Agentic Summary

- Status: completed
- Main question: What is the key contribution?
- Report question set: structured
- Structured questions generated: 83
- Confidence mean/median: 0.2724441085027672 / 0.26825243553753414
- Confidence labels: low=77, medium=6, high=0

## Final Answer

- Key contribution
  - Introduces ICSS (Cumulative Sums of Squares) as a fast, retrospective detector of variance changes in time series, using the  $D_k$  statistic to flag potential change points. (Inclán & Tiao, page 3, words 0-349)

- Core mechanics
  - For a fixed candidate change point  $k$ , the two-sample F statistic is  $F_{\{k,T-k\}} = [(\sum_{i=1}^k \hat{a}_i^2)/k] / [(\sum_{i=k+1}^T \hat{a}_i^2)/(T-k)]$  and with  $C_k = \sum_{i=1}^k \hat{a}_i^2$ , equivalently  $F_{\{k,T-k\}} = [(C_T - C_k)/(T - k)] / (C_k/k)$ .
  - $D_k$  is a monotone function of  $F_{\{k,T-k\}}$ ; the changepoint is identified by maximizing  $|D_k|$  over  $k$ . (Inclán & Tiao, page 3, words 0-349)
- Change-point localization
  - Compute  $D_k$  for  $k = 1, \dots, T$ ; set  $k^* = \arg \max_k |D_k|$ . If the maximum exceeds a critical boundary, infer a variance change at or near  $k^*$ . (Inclán & Tiao, page 3, words 0-349)
- Comparative framework in the study
  - ICSS as fast, retrospective detector; LR tests for one/more changes; Bayesian posterior odds for selecting the number of changes. (Inclán & Tiao, pages 7–9, various)
- One-change-point findings (Part I)
  - Small variance change ( $A=2$ ): hard to detect, especially near the start.
  - Larger change ( $A=3$ ): correct detections  $>80\%$  when change is mid-series; performance improves with longer series and larger  $A$ . (Inclán & Tiao, page 9, words 550-904)
- Two-change-points findings (Part II)
  - ICSS tends to miss small changes; LR and Bayesian approaches improve detection when changes are larger or favorably located. (Inclán & Tiao, page 9–10, words 550-904; page 10, words 0-84)
- Computational considerations
  - ICSS is computationally lighter and avoids full likelihood surface searches; LR tests are heavier, especially with multiple change points. (Inclán & Tiao, page 9, words simple; page 11, words 275-507)
- Quantitative takeaway
  - For short series ( $T = 100$ ), small variance changes are hard to detect; detection improves with longer  $T$  or larger changes; Bayesian and LR methods offer higher power at greater computational cost. (Inclán & Tiao, pages 9–11, various)
- If you want exact equations or precise typographical forms for a item, I can pull them out. (Context: Inclán & Tiao, various sections)

## Sub-Answers

### Sub-answer 1

- Question: ResponseTextConfig(format=ResponseFormatText(type='text'), verbosity='medium')
- Question tokens estimate: 19
- Answer:
- What ICSS does (retrospective variance-change detection)
  - Uses Cumulative Sums of Squares to detect variance changes in a time series; assesses potential change points by examining  $D_k$  statistics across  $k = 1, \dots, T$ . If the process is Normal with mean 0, variance changes yield a likelihood-ratio shape in the  $D_k$  plot. (Inclán & Tiao, page 3, words 0-349)
- Key quantity:  $D_k$  and its relation to the usual F-test
  - For a fixed candidate change point  $k$ , the two-sample F statistic testing  $\sigma_1^2 = \sigma_2^2$  is  $F_{\{k, T-k\}} = [(\sum_{i=1}^k a_i^2)/k] / [(\sum_{i=k+1}^T a_i^2)/(T-k)]$
  - Let  $C_k = \sum_{i=1}^k a_i^2$  be the cumulative sum of squares. Then  $F_{\{k, T-k\}} = [(C_T - C_k)/(T - k)] / (C_k/k)$
  - $D_k$  is a monotone function of  $F_{\{k, T-k\}}$ ; thus maximizing  $|D_k|$  over  $k$  and comparing to a boundary identifies a change point. (Inclán & Tiao, page 3, words 0-349)
- How to locate a change point from  $D_k$ 
  - Compute  $D_k$  for  $k = 1, \dots, T$  and locate  $k^* = \arg \max_k |D_k|$ .
  - Compare  $\max |D_k|$  to its boundary; if it exceeds the boundary, you infer a variance change at or near  $k^*$ . (Inclán & Tiao, page 3, words 0-349)
- Methods compared in the study
  - ICSS algorithm (Cumulative Sums of Squares) as a fast, retrospective detector.
  - Likelihood ratio (LR) tests for one or more changes, with simulation-based critical values.
  - Bayesian posterior odds approach for deciding the number of change points. (Inclán & Tiao, pages 7–9, various)
- Part I: one changepoint simulation setup
  - 1,000 replicates per design point; residuals derived from AR(1) processes; LG tests and ICSS applied to residuals after fitting AR(1).
  - Scenarios include variance ratio  $A = 2$  and  $A = 3$ , with change location  $K/T = 0.25, 0.50, 0.75$ . (Inclán & Tiao, page 7, words 0-349)

- Part I: one changepoint results (from the study)
  - With a small variance change ( $A = 2$ ): hard to detect, especially when the change is near the beginning.
  - With a larger change ( $A = 3$ ): correct identifications occur more than 80% of the time when the change is in the middle; performance improves with longer series and larger variance ratios. (Inclán & Tiao, page 9, words 550-904)
- Part II: two changepoints simulation setup
  - Similar design, but two changepoints at locations  $K1/T$  and  $K2/T$ ; compare ICSS, LR tests ( $n-1$  to  $n$  changes), and Bayesian posterior odds for  $NT = 0,1,2$  changes. (Inclán & Tiao, page 8–9, words 0-329 and 550-904)
- Part II: two changepoints results
  - Table-like summaries show how often the procedure correctly identifies  $N\_T$  (the number of changepoints) across configurations; ICSS tends to miss small changes, while LR and Bayesian approaches improve detection under larger shifts and favorable locations. (Inclán & Tiao, page 9, words 550-904; page 10, words 0-84)
- CPU time and computational considerations
  - ICSS: computational burden is relatively modest; its CPU time scales with series length but is generally favorable versus LR-based approaches.
  - Likelihood ratio tests: heavy computational burden, especially with multiple changepoints; max-likelihood evaluation becomes costly for long series and many potential change points. (Inclán & Tiao, page 9, words simple; page 11, words 275-507)
  - Exact statements: LR tests consume substantial CPU time; ICSS avoids searching the full likelihood surface, trading some power for speed. (Inclán & Tiao, page 11, words 275-507)
- Quantitative takeaway
  - For short series (e.g.,  $T = 100$ ), small variance changes are hard to detect; increasing  $T$  or the magnitude of the change improves detection rates, particularly for the middle of the series. Bayesian and LR methods can offer higher detection power at the cost of greater computation. (Inclán & Tiao, pages 9–11, various)

If you want, I can pull out specific equations or reproduce the exact forms used in the paper (with precise typographical corrections) for a given item.

## Structured Q&A Appendix

This section mirrors `agentic.report_questions` for audit traceability.

**A01: What is the main research question of the paper?**

- Category: A) Research question / contribution
- Confidence: low (0.2574489296458133)
- Retrieval method: local
- Evidence type: title/overview
- Answer:

The main research question is how to retrospectively detect changes in variance in a time series using cumulative sums of squares (the ICSS approach) and evaluate its effectiveness compared with alternative methods.

- Quote snippet:  
Use of Cumulative Sums of Squares for Retrospective Detection of Changes of Variance
- Citation anchors:
  - page=1, words=0-230, section=title, note=Paper title indicating main research question.

**A02: What is the paper's primary contribution relative to the existing literature?**

- Category: A) Research question / contribution
- Confidence: low (0.2743771636600953)
- Retrieval method: local
- Evidence type: methodological contribution
- Answer:

The paper introduces the ICSS algorithm for retrospective, multiple-change detection in variance, using centered cumulative sums of squares to identify variance-change points, and shows it achieves competitive performance with substantially lower computational burden than Bayesian or likelihood-ratio methods.

- Quote snippet:  
We propose a procedure to detect variance changes based on an iterated cumulative sums of squares (ICSS) algorithm.
- Citation anchors:
  - page=2, words=0-40, section=Introduction, note=States the contribution: proposes a procedure to detect variance changes based on an iterated cumulative sums of squares (ICSS) algorithm.
  - page=2, words=275-624, section=ICSS Algorithm, note=Describes the ICSS approach to search for multiple variance change points using centered cumulative sums of squares.

**A03: What is the central hypothesis being tested?**

- Category: A) Research question / contribution
- Confidence: low (0.25458346383585695)
- Retrieval method: local
- Evidence type: text
- Data source: Inclán and Tiao: Retrospective Detection of Changes in Variance
- Answer:

The central hypothesis is that there is no change in variance ( $NT = 0$ ) versus there being a change ( $NT = 1$ ) in the time series, i.e., testing for a variance change point.

- Quote snippet:

Let  $NT = 0$  represent the case of no change of some specified boundaries with high probability. These and let  $NT = 1$  represent the case of one change.

- Citation anchors:
  - page=3, words=275-624, section=2.2, note=Defines  $NT = 0$  (no change) and  $NT = 1$  (one change) and discusses testing for a change in variance.
  - page=3, words=0-349, section=2.2, note=Mentions the likelihood ratio for testing one change against no change in variance.

**A04: What are the main outcomes of interest (dependent variables)?**

- Category: A) Research question / contribution
- Confidence: low (0.26044345018932097)
- Retrieval method: local
- Evidence type: text
- Data source: Inclán and Tiao (1994). Retrospective Detection of Changes in Variance, JASA
- Answer:

The main dependent variables are the locations of the change points (the number of change points  $NT$  and their positions  $K$ ) and the variances in each segment (the segment variances  $a_i$ ).

- Quote snippet:

The maximum likelihood estimator of  $K$  is obtained by maximizing this function with respect to  $K$ .

- Citation anchors:
  - page=3, words=If-KZ, section=Two Change-Point Scenario, note=Describes two changepoints and corresponding variances; indicates  $E[D_k]$  depends on changepoint locations.

- page=3, words=The-K, section=Estimation of K, note=States that the MLE of K is obtained by maximizing the likelihood with respect to K.
- page=6, words=We-variance., section=ICSS and changepoint tuning, note=Discusses obtaining a homogeneous-variance series after identifying change points (cp) and updating cp locations.

**A05: What are the key treatment/exposure variables (independent variables)?**

- Category: A) Research question / contribution
- Confidence: low (0.21784745458251123)
- Retrieval method: local
- Evidence type: text
- Answer:

There are no treatment/exposure variables in this analysis. The study examines changes in variance over time, with time (the index  $t$  or  $k$ ) serving as the ordering variable and potential change points  $K1$ ,  $K2$  defining segments. The key parameters relate to segmentation ( $NT$ ,  $K1$ ,  $K2$ ) and the segment variances before/after the change.

- Quote snippet:  
For a fixed  $k$ , the value of  $D_k$  can be written as a function of the usual  $F$  statistic for testing equality of variances between two independent samples.
- Citation anchors:
  - page=3, words=0-349, section=2.1 Relationship of  $D_k$  to the  $F$  Statistic and the Likelihood Ratio, note=Describes splitting a series into two samples with variances and using  $F$  statistics; time index  $k$  defines the potential change point.

**A06: What setting/context does the paper study (country, market, period)?**

- Category: A) Research question / contribution
- Confidence: low (0.29801931529493225)
- Retrieval method: local
- Evidence type: text
- Data source: IBM stock price data as reported by Box and Jenkins (1976)
- Table/Figure: Figure 4 (Sequence of  $D_k$  from ICSS applied to IBM stock prices)
- Answer:

United States stock market; IBM stock prices; period May 17, 1961 to November 2, 1962.

- Quote snippet:  
The series analyzed is the first difference of the logarithm of the IBM stock closing prices from May 17, 1961, to November 2, 1962, as reported by Box and Jenkins (1976).
- Citation anchors:
  - page=6, words=0-349, section=3.1 Example: IBM Stock Prices, note=IBM stock prices data used as a real-data example; May 17, 1961 to November 2, 1962.

**A07: What is the main mechanism proposed by the authors?**

- Category: A) Research question / contribution
- Confidence: low (0.2356826157787199)
- Retrieval method: local
- Evidence type: textual
- Answer:

The main mechanism is the Iterated Cumulative Sums of Squares (ICSS) algorithm, which uses a centered cumulative sum of squares to locate multiple variance-change points retrospectively.

- Quote snippet:  
We propose a procedure to detect variance changes based on an iterated cumulative sums of squares (ICSS) algorithm.
- Citation anchors:
  - page=2, words=0-349, section=Introduction/ICSS, note=Authors propose detecting variance changes using an iterated cumulative sums of squares (ICSS) algorithm.
  - page=5, words=825-1174, section=Iterated Cumulative Sums of Squares (ICSS) Algorithm, note=Detailed steps and description of the ICSS algorithm.

**A08: What alternative mechanisms are discussed?**

- Category: A) Research question / contribution
- Confidence: low (0.24258059419127234)
- Retrieval method: local
- Evidence type: Comparative discussion of alternative methods (Bayesian approach; likelihood ratio tests).
- Answer:

A Bayesian approach and likelihood ratio tests.

- Quote snippet:  
the ICSS algorithm offers results comparable to those obtained by a Bayesian approach or by likelihood ratio tests



- Citation anchors:
  - page=2, words=0-349, section=Introduction, note=Mentions Bayesian approach and likelihood ratio tests as alternatives to ICSS.

**A09: What are the main policy implications claimed by the paper?**

- Category: A) Research question / contribution
- Confidence: low (0.2661868704730924)
- Retrieval method: local
- Evidence type: methodology\_focus
- Answer:

The paper does not articulate any policy implications; it is focused on methodological issues of detecting variance changes in time series (comparisons of ICSS, LR, and Bayesian approaches), including performance and computational cost, with guidance for practitioners on applying these methods.

- Quote snippet:
 

The simulation experiment to compare the performance of the ICSS algorithm with the likelihood ratio and a Bayesian approach.
- Citation anchors:
  - page=1, words=0-230
  - page=2, words=550-899
  - page=4, words=0-349
  - page=4, words=275-624
  - page=6, words=550-899
  - page=7, words=0-349
  - page=9, words=275-624
  - page=9, words=550-899
  - page=11, words=275-507
  - page=12, words=0-349

**A10: What is the welfare interpretation (if any) of the results?**

- Category: A) Research question / contribution
- Confidence: low (0.29126722563578344)
- Retrieval method: local
- Evidence type: scope and purpose of study (methodological)
- Answer:

No welfare interpretation is provided; the results are methodological, focusing on retrospective detection of variance changes in time series using the ICSS algorithm and related statistics, not on welfare implications.

- Quote snippet:
 

Incldn and Tiao: Retrospective Detection of Changes in Variance

- Citation anchors:
  - page=3, words=0-349, section=Two Variance Changes: Brownian bridge asymptotically, note=Discusses variance-change detection and  $D_k$  as function of F statistic; centers on methodology.
  - page=4, words=0-349, section=Incln and Tiao: Retrospective Detection of Changes in Variance, note=Shows title indicating focus on variance-change detection; no welfare discussion.

**A11: What are the main limitations acknowledged by the authors?**

- Category: A) Research question / contribution
- Confidence: low (0.24853943459540173)
- Retrieval method: local
- Evidence type: Limitation discussion in CPU time and detection power sections
- Answer:

The authors acknowledge multiple limitations: (1) a heavy computational burden and high CPU time required, especially for posterior-odds/bayesian approaches and when handling multiple changepoints; (2) reduced ability to detect small variance changes in short series (e.g., variance ratio  $A = 2$  in a 100-observation series), particularly when the change occurs early in the sequence; (3) performance deteriorates with shorter series and more challenging changepoint configurations, though longer series and larger variance changes improve detection; (4) feasibility concerns for the posterior odds method (e.g., 1,358 CPU hours for  $T = 500$ ) which limits its practicality in some scenarios.

- Quote snippet:
 

One concern in ICSS design is the heavy computational burden of calculating posterior odds, and it's hard to detect a small variance change ( $A=2$ ) at the beginning.
- Citation anchors:
  - page=9, words=275-624, section=4.6 CPU Time Requirements, note=Notes the heavy computational burden of calculating posterior odds.
  - page=9, words=550-899, section=4.5 The ICSS Algorithm Applied to Residuals, note=States difficulty in detecting a small variance change for series of 100 observations ( $A=2$ ), especially if it occurs at the beginning.
  - page=9, words=825-904, section=4.6 CPU Time Requirements, note=Mentions the feasibility issue of posterior odds, e.g., 1,358 CPU hours for  $T = 500$ .

**A12: What does the paper claim is novel about its data or identification?**

- Category: A) Research question / contribution
- Confidence: low (0.28669586887752774)
- Retrieval method: local
- Evidence type: textual\_claim
- Answer:

The paper claims novelty in retrospective detection of multiple variance-change points in a sequence of independent observations using the iterated cumulative sums of squares (ICSS) algorithm.

- Quote snippet:

This article studies the problem of multiple change points in the variance of a sequence of independent observations. We propose a procedure to detect variance changes based on an iterated cumulative sums of squares (ICSS) algorithm.

- Citation anchors:
  - page=2, words=0-349, section=Introduction, note=Proposes ICSS to detect multiple variance change points; novelty in handling multiple changes.
  - page=2, words=0-349, section=Introduction/ICSS algorithm, note=Notes retrospective, systematic search for change points in variance using cumulative sums.

**B01: What is the identification strategy (in one sentence)?**

- Category: B) Identification strategy / causal design
- Confidence: low (0.27731002864874177)
- Retrieval method: local
- Evidence type: algorithm\_description
- Answer:

An iterative ICSS procedure: compute  $D_k$  on the current segment, identify the point with the maximum absolute  $D_k$  as a changepoint, split the series at that point, and repeat on the subseries until convergence or a maximum number of iterations.

- Quote snippet:

Step 0. Let  $t_l = 1$ . Step 1. Calculate  $D_k(a[t_l : T])$ . Let  $k^*(a[t_l : T])$  be the point at which  $\max_k |D_k(a[t_l : T])|$  is obtained.

- Citation anchors:
  - page=5, words=825-1174, section=Iterated Cumulative Sums of Squares (ICSS) Algorithm, note=Describes Step 0-Step 1 and iterative splitting to locate changepoints

- page=6, words=275-624, section=Iterative ICSS procedure, note=Discusses convergence and continuing the process on subseries

**B02: Is the design experimental, quasi-experimental, or observational?**

- Category: B) Identification strategy / causal design
- Confidence: low (0.26824989134262495)
- Retrieval method: local
- Evidence type: simulation study / experimental design (synthetic data)
- Data source: Simulated data generated for methodology evaluation of ICSS algorithm.
- Answer:

Experimental (simulation study).

- Quote snippet:  
The simulation experiment has two separate parts, the first for one changepoint and the second for two changepoints in the generated series.
- Citation anchors:  
– page=7, words=0-20, section=Simulation Experiment, note=Statement describing the simulation experiment with two parts (one changepoint and two changepoints).

**B03: What is the source of exogenous variation used for identification?**

- Category: B) Identification strategy / causal design
- Confidence: low (0.3249980422368381)
- Retrieval method: local
- Evidence type: conceptual
- Answer:

Variance changes in the error term (i.e., changes in the variance of the observations) serve as the exogenous variation used for identification.

- Quote snippet:  
This article studies the problem of multiple change points in the variance of a sequence of independent observations.
- Citation anchors:  
– page=2, words=0-349, section=Introduction, note=Source of exogenous variation identified as changes in variance of independent observations.

**B04: What is the treatment definition and timing?**

- Category: B) Identification strategy / causal design
- Confidence: low (0.11992656712503762)
- Retrieval method: local
- Evidence type: not\_present\_in\_context
- Data source: Inclan, C. and Tiao, G.C. (1994). Use of Cumulative Sums of Squares for Retrospective Detection of Changes of Variance. Journal of the American Statistical Association.
- Answer:

The provided context does not define a treatment or its timing; it discusses detecting changes in variance (variance change points) using the Iterated Cumulative Sums of Squares (ICSS) algorithm.

- Quote snippet:  
Iterated Cumulative Sums of Squares (ICSS) Algorithm
- Citation anchors:
  - page=1, words=0-230
  - page=2, words=0-349
  - page=4, words=0-349
  - page=5, words=550-899
  - page=5, words=825-1174
  - page=5, words=1100-1198
  - page=9, words=275-624
  - page=11, words=0-349
  - page=12, words=0-349
  - page=12, words=275-624

**B05: What is the control/comparison group definition?**

- Category: B) Identification strategy / causal design
- Confidence: low (0.19701504363095051)
- Retrieval method: local
- Evidence type: definition
- Answer:

The control/comparison group is the case of no change in variance (homogeneous variance), i.e.,  $NT = 0$ , against the alternative of one change ( $NT = 1$ ).

- Quote snippet:  
Let  $NT = 0$  represent the case of no change of some specified boundaries with high probability. These and let  $NT = 1$  represent the case of one change.
- Citation anchors:

- page=3, words=0-349, section=Definitions of change points, note=Defines NT = 0 as no change and NT = 1 as one change.

**B06: What is the estimating equation / baseline regression specification?**

- Category: B) Identification strategy / causal design
- Confidence: low (0.2855057810106444)
- Retrieval method: local
- Evidence type: Model specification / baseline regression (variance-change framework)
- Answer:

The estimating equation is the likelihood-ratio test for a change in variance, assuming the innovations  $\{a_i\}$  are  $\text{Normal}(0, \sigma_i^2)$ . Specifically, if the  $\{a_i\}$  are Normal with mean 0 and variances, one obtains the likelihood ratio for testing one change against no change in variance, and the statistic  $D_k$  is related to an F statistic.

- Quote snippet:  
If the  $\{a_t\}$  are Normally distributed, with mean 0 and variances  $\sigma_t^2$ ,  $t = 1, \dots, T$ , then we can obtain the likelihood ratio for testing one change against no change in variance, the plot of  $D_k$  will exhibit
- Citation anchors:
  - page=3, words=0-349, section=Model setup /  $D_k$  and F statistic, note=Baseline model assumes normal errors with potential variance changes; estimating equation via likelihood ratio for one change in variance.

**B07: What fixed effects are included (unit, time, two-way, higher dimensional)?**

- Category: B) Identification strategy / causal design
- Confidence: low (0.27610975909446506)
- Retrieval method: local
- Evidence type: not\_discussed\_in\_context
- Answer:

No fixed effects are mentioned in the provided context.

- Quote snippet:  
Incldn and Tiao: Retrospective Detection of Changes in Variance
- Citation anchors:
  - page=2, words=825-916
  - page=3, words=275-624
  - page=3, words=550-882

- page=4, words=0-349
- page=4, words=275-624
- page=6, words=275-624
- page=6, words=550-899
- page=10, words=0-84
- page=11, words=275-507
- page=12, words=0-349

**B08: What standard errors are used (robust, clustered; at what level)?**

- Category: B) Identification strategy / causal design
- Confidence: low (0.33150099288163026)
- Retrieval method: local
- Evidence type: textual
- Answer:

Standard errors are computed for a proportion from 10,000 simulation replicates:  $SE = \sqrt{\hat{p}(1 - \hat{p}) / 10,000}$ , i.e., at the proportion level across replicates (not described as robust or clustered).

- Quote snippet:  
The standard errors were obtained as  $SE = \sqrt{\hat{p}(1 - \hat{p}) / 10,000}$ , where  $\hat{p}$  is the proportion of series for which  $klast = tl - 1$ .
- Citation anchors:
  - page=5, words=275-624, section=Step 2b, note=Contains the line describing the standard errors, including the formula
  - page=9, words=0-349, section=Table 6 / Sampling Distribution, note=Also references standard errors computed as  $SE = \sqrt{\hat{p}(1 - \hat{p}) / 10,000}$

**B09: What is the key identifying assumption (parallel trends, exclusion restriction, ignorability)?**

- Category: B) Identification strategy / causal design
- Confidence: low (0.2723274028030592)
- Retrieval method: local
- Evidence type: absence\_of\_identifying\_assumption
- Assumption notes: No explicit causal-identifying assumption is stated in the provided text.
- Answer:

The provided context does not specify any identifying assumption (parallel trends, exclusion restriction, ignorability).

- Quote snippet:

For a fixed  $k$ , the value of  $D_k$  can be written as a function of the usual  $F$  statistic for testing equality of variances

- Citation anchors:
  - page=3, words=0-349, section=2.1 Relationship of  $D_k$  to the  $F$  Statistic and the Likelihood Ratio, note=No identifying assumptions discussed in this statistical methods paper.

**B10: What evidence is provided to support the identifying assumption?**

- Category: B) Identification strategy / causal design
- Confidence: low (0.2664307225007051)
- Retrieval method: local
- Evidence type: Simulation results (Tables 4-8) and theoretical justification of the ICSS method, including the property that the mode of  $k^*$  is at the change point.
- Data source: Inclhn and Tiao (1994)
- Table/Figure: Table 4
- Assumption flag: True
- Assumption notes: Variance changes are discrete events that can be detected by the ICSS procedure; simulation results demonstrate correct identifications under various series lengths and variance ratios, supporting the identifying assumption.
- Answer:

The identifying assumption—that variance changes occur at discrete points that can be detected via the ICSS method—is supported by simulation evidence showing that the ICSS algorithm correctly identifies changepoints and that the mode of the change-point locator  $k^*$  aligns with the true change point.

- Quote snippet:
 

”the mode of  $k^*$  is exactly at the change point where the change in variance occurs.”
- Citation anchors:
  - page=9, words=290-320, section=Results and interpretation of mode of  $k^*$ , note=State that the mode of  $k^*$  is exactly at the change point where the change in variance occurs.
  - page=8, words=0-40, section=Results of Part One, note=Table 4: frequency distribution of NT obtained with ICSS, LR tests, and log posterior odds.

**B11: Are there event-study or pre-trend tests? What do they show?**

- Category: B) Identification strategy / causal design
- Confidence: medium (0.3887858064534689)
- Retrieval method: local



- Evidence type: **text**
- Data source: IBM stock price data (first difference of log closing prices, May 17, 1961 to November 2, 1962)
- Table/Figure: Figure 4
- Answer:

No. The provided context does not discuss event-study or pre-trend tests. The paper develops and applies the ICSS algorithm for retrospective detection of variance changes in time series, and the IBM stock-price example demonstrates identifying variance-change points rather than evaluating pre-event trends or event-study effects.

- Quote snippet:  
There is no substantial autocorrelation for this series.
- Citation anchors:
  - page=6, words=275-624, section=IBM stock prices example, note=ICSS applied; multiple variance-change points detected; no event-study/pre-trend tests discussed.
  - page=7, words=0-349, section=IBM example continuation, note=No discussion of event-study tests; illustrates variance-change points in IBM data.

**B12: What instruments are used (if IV)? Define instrument and first stage.**

- Category: B) Identification strategy / causal design
- Confidence: low (0.17732312317760468)
- Retrieval method: local
- Evidence type: No discussion of instrumental variables; content is about variance-change detection (ICSS).
- Answer:

The provided context does not discuss any instruments or an instrumental-variables framework; it focuses on the ICSS algorithm for detecting variance changes in time series, not on IVs or a first stage.

- Quote snippet:  
"Iterated Cumulative Sums of Squares (ICSS) Algorithm Step 0. Let  $t_1 = 1$ ."
- Citation anchors:
  - page=5, words=Iterated-Algorithm, section=DETECTIONS OF MULTIPLE CHANGES: THE ITERATED CUMULATIVE SUMS OF SQUARES ALGORITHM, note=ICSS algorithm description; no IV discussion present in this fragment.

**B13: What is the first-stage strength (F-stat, partial  $R^2$ , relevance evidence)?**

- Category: B) Identification strategy / causal design
- Confidence: low (0.31210124437097825)
- Retrieval method: local
- Evidence type: Definition and relationship between  $D_k$  and F statistic from the article
- Data source: Includn and Tiao (1994), Journal of the American Statistical Association
- Answer:

The first-stage strength is captured by the F statistic for testing equality of variances between two subsamples, defined as  $F_{\{k,T-k\}} = ((C_T - C_k)/(T - k)) / (C_k/k)$ . The  $D_k$  statistic is then a monotone function of this F statistic.

- Quote snippet:  
For a fixed  $k$ , the value of  $D_k$  can be written as a function of the usual F statistic for testing equality of variances between two independent samples.
- Citation anchors:
  - page=3, words=0-349, section=Relationship of  $D_k$  to the F statistic, note=Definition of F-statistic for testing equality of variances and its relation to  $D_k$

**B14: If RDD: what is the running variable and cutoff? bandwidth choice?**

- Category: B) Identification strategy / causal design
- Confidence: low (0.3425596506801698)
- Retrieval method: local
- Evidence type: not\_discussed\_in\_context
- Answer:

Not available in the provided context; there is no discussion of regression discontinuity design, running variable, cutoff, or bandwidth.

- Quote snippet:  
The ICSS Algorithm Applied to Residuals
- Citation anchors:
  - page=3, words=0-349
  - page=3, words=275-624
  - page=4, words=0-349
  - page=4, words=275-624
  - page=4, words=550-869
  - page=6, words=275-624

- page=9, words=0-349
- page=9, words=275-624
- page=9, words=825-904
- page=11, words=275-507

**B15: If DiD: what is the timing variation (staggered adoption)? estimator used?**

- Category: B) Identification strategy / causal design
- Confidence: medium (0.41047338929895166)
- Retrieval method: local
- Evidence type: not\_covered\_in\_context
- Data source: Incln and Tiao (1994) ICSS variance-change detection paper
- Answer:

The provided context does not discuss a difference-in-differences timing variation estimator (staggered adoption); it describes methods for detecting variance changes (ICSS).

- Quote snippet:  
Incln and Tiao: Retrospective Detection of Changes in Variance
- Citation anchors:
  - page=2, words=825-916
  - page=3, words=550-882
  - page=4, words=0-349
  - page=4, words=550-869
  - page=5, words=0-349
  - page=5, words=1100-1198
  - page=6, words=0-349
  - page=6, words=275-624
  - page=6, words=550-899
  - page=11, words=275-507

**C01: What dataset(s) are used? (name sources explicitly)**

- Category: C) Data, sample, and measurement
- Confidence: low (0.2668669299082501)
- Retrieval method: local
- Evidence type: text
- Data source: IBM stock closing prices dataset, as reported by Box and Jenkins (1976).
- Table/Figure: Figure 4
- Answer:

IBM stock prices dataset: the first difference of the logarithm of IBM stock closing prices from May 17, 1961 to November 2, 1962, as reported by Box and

Jenkins (1976).

- Quote snippet:

The series analyzed is the first difference of the logarithm of the IBM stock closing prices from May 17, 1961, to November 2, 1962, as reported by Box and Jenkins (1976).

- Citation anchors:
  - page=6, words=0-349, section=3.1 Example: IBM Stock Prices, note=Describes the dataset used: IBM stock prices and time frame; cites Box and Jenkins (1976).

**C02: What is the unit of observation (individual, household, firm, county, transaction, product)?**

- Category: C) Data, sample, and measurement
- Confidence: low (0.24619288704307682)
- Retrieval method: local
- Evidence type: textual
- Data source: IBM stock closing prices (daily), May 17, 1961 to November 2, 1962
- Answer:

IBM stock closing prices (daily observations).

- Quote snippet:

The series analyzed is the first difference of the logarithm of the IBM stock closing prices from May 17, 1961, to November 2, 1962.

- Citation anchors:
  - page=6, words=0-349, section=3.1 Example: IBM Stock Prices, note=Shows unit of observation as daily IBM stock price series (first difference of log prices).

**C03: What is the sample period and geographic coverage?**

- Category: C) Data, sample, and measurement
- Confidence: low (0.26825243553753414)
- Retrieval method: local
- Evidence type: textual evidence from the article
- Data source: IBM stock closing prices; source Box and Jenkins (1976)
- Answer:

Sample period: May 17, 1961 to November 2, 1962. Geographic coverage: not specified in the provided context.

- Quote snippet:

The series analyzed is the first difference of the logarithm of the IBM stock closing prices from May 17, 1961, to November 2, 1962.

- Citation anchors:
  - page=6, words=0-349, section=3.1 Example: IBM Stock Prices, note=Contains the IBM stock prices example and the period May 17, 1961 to November 2, 1962.

**C04: What are the sample restrictions / inclusion criteria?**

- Category: C) Data, sample, and measurement
- Confidence: low (0.23026758427151148)
- Retrieval method: local
- Evidence type: simulation design description and inclusion criteria
- Data source: Simulation study; synthetic time series (ICSS algorithm performance).
- Answer:

In the simulation study, the authors used fixed, predefined sample configurations rather than observational data: three series lengths ( $T = 100, 200, 500$ ); Part One changepoint locations at fractions  $0.25T, 0.50T$ , and  $0.75T$ ; variance-change magnitudes  $A = 2$  or  $3$ ; Part Two with the same three lengths but three two-change-point location sets  $(0.33T, 0.66T)$ ,  $(0.20T, 0.80T)$ ,  $(0.58T, 0.80T)$  and six variance configurations (eliminating symmetric permutations). Additionally, replicates differ by configuration (e.g., 10,000 replicates for  $A = 1$ , otherwise 1,000).

- Quote snippet:
 

For Part One, we used three series lengths (  $T = 100, 200$ , and  $500$  ), three locations of the changepoint (  $K = .25T, .50T$ , and  $.75T$  ), and two values for the variance ratio (  $A = 2$  and  $3$  ).
- Citation anchors:
  - page=7, words=0-349, section=4. SIMULATION EXPERIMENT; Part One & Part Two, note=Describes series lengths, changepoint locations, variance ratios for Part One and Part Two; symmetry elimination.
  - page=9, words=0-349, section=Notes on sampling/replicates, note=Mentions replicates count (10,000 for  $A = 1$ ; 1,000 otherwise).

**C05: What is the sample size (N) in the main analysis?**

- Category: C) Data, sample, and measurement
- Confidence: low (0.29223383662827496)
- Retrieval method: local
- Evidence type: textual evidence from ICSS example showing  $D_k$  is evaluated up to 700 observations

- Answer:

700

- Quote snippet:  
Dk plot Dk plot k=1, ....700 k=1,....700
- Citation anchors:
  - page=3, words=0-349, section=Dk statistics and ICSS algorithm, note=The Dk plot is described for k = 1,...,700, implying a sample size of 700 observations in the main example

**C06: How is the key outcome measured? Any transformations (logs, z-scores, indices)?**

- Category: C) Data, sample, and measurement
- Confidence: low (0.29366393060057905)
- Retrieval method: local
- Evidence type: text
- Answer:

The key outcome is measured with the Dk statistic, defined as a function of the F statistic for testing equality of variances between the two samples split at k. The change point location is estimated as  $k^* = \arg \max_k |Dk|$ , and a change is declared if  $\max_k |Dk|$  exceeds a boundary derived from the asymptotic distribution of Dk under constant variance (i.e., no change).

- Quote snippet:  
for a fixed k, the value of Dk can be written as a function of the usual F statistic for testing equality of variances between two independent samples.
- Citation anchors:
  - page=3, words=0-349, section=Dk and F statistic, note=Dk is a function of the F statistic for testing equality of variances; boundaries derived from the asymptotic distribution.
  - page=3, words=275-624, section=Change-point detection, note= $\max_k |Dk|$  determines the location of the change point ( $k^*$ ).
  - page=2, words=825-916, section=Behavior of Dk, note=The plot of Dk against k will oscillate around 0 for series with homogeneous variance; boundaries are used for decision.
  - page=3, words=0-349, section=Asymptotic boundaries, note=Boundaries can be obtained from the asymptotic distribution of Dk assuming constant variance.

**C07: How is treatment/exposure measured? Any constructed variables?**

- Category: C) Data, sample, and measurement
- Confidence: low (0.2465336639861547)
- Retrieval method: local
- Evidence type: methodology\_summary
- Data source: A sequence of independent observations with possibly changing variance (modeled as normally distributed  $a_i$  with piecewise-constant variances).
- Answer:

There is no treatment/exposure measurement discussed in this article. The work analyzes a sequence of independent observations that may experience changes in variance over time. Constructed variables used for detection include the cumulative sums of squares ( $C_k$ ) and the centered cumulative sum of squares ( $D_k$ ), which are used to identify potential change points. The method treats the data as a sequence with possible variance changes (not a treatment/exposure variable) and uses these constructed statistics to detect changes.

- Quote snippet:

This article studies the problem of multiple change points in the variance of a sequence of independent observations.

- Citation anchors:
  - page=2, words=0-60, section=Introduction, note=Overview: problem of multiple change points in the variance of a sequence of independent observations.
  - page=2, words=61-140, section=Introduction, note=ICSS algorithm proposed for detecting variance changes.
  - page=3, words=0-50, section=Methods, note=Construction of  $D_k$  and its relation to the cumulative sum of squares; linkage to F statistic.

**C08: Are there key covariates/controls? Which ones are always included?**

- Category: C) Data, sample, and measurement
- Confidence: low (0.2841094694486537)
- Retrieval method: local
- Evidence type: text
- Data source: ICSS paper: Use of Cumulative Sums of Squares for Retrospective Detection of Changes of Variance (Tiao & Inl  n, 1994)
- Assumption flag: True
- Assumption notes: No covariates or controls are included; model assumes independent observations with mean zero and possibly changing variances.
- Answer:

No. The article does not discuss key covariates or controls. It treats the data as a sequence of independent observations with mean zero and potentially changing variances; there are no covariates or external controls included in the model.

- Quote snippet:

The main interest is to study the variance of a given sequence of observations retrospectively, so we can use all the information on the series to indicate the points of variance change.

- Citation anchors:
  - page=2, words=0-349, section=I. INTRODUCTION, note=States problem of multiple change points in variance of a sequence of independent observations and describes ICSS approach.
  - page=2, words=825-916, section=I. INTRODUCTION, note=Notes the main interest is retrospective variance detection using all information in the series.
  - page=3, words=275-624, section=2. CENTERED CUMULATIVE SUMS OF SQUARES, note=Presents assumption that  $a_t$  are Normal with mean 0 and variances  $a_t^2$ ; discusses likelihood for variance change.

**C09: How are missing data handled (dropping, imputation, weighting)?**

- Category: C) Data, sample, and measurement
- Confidence: low (0.24362276088827278)
- Retrieval method: local
- Evidence type: not addressed in context
- Data source: Incln and Tiao, 1994, Journal of the American Statistical Association (JASA)
- Answer:

The provided context does not specify how missing data should be handled (dropping, imputation, or weighting).

- Quote snippet:

4.5 The ICSS Algorithm Applied to Residuals

- Citation anchors:
  - page=2, words=825-916
  - page=5, words=275-624
  - page=6, words=0-349
  - page=6, words=275-624
  - page=6, words=550-899
  - page=7, words=0-349
  - page=8, words=0-329
  - page=9, words=0-349



- page=9, words=275-624
- page=10, words=0-84

**C10: Are weights used (survey weights, propensity weights)? How?**

- Category: C) Data, sample, and measurement
- Confidence: low (0.2805437626211552)
- Retrieval method: local
- Evidence type: No weights discussed; context is about variance-change detection with ICSS.
- Data source: Simulation study with simulated series and residuals from AR(1) processes.
- Answer:

No. The article focuses on detecting variance changes in a sequence of observations using the ICSS algorithm and does not mention or use survey weights or propensity weights.

- Quote snippet:

This article studies the problem of multiple change points in the variance of a sequence of independent observations. We propose a procedure to detect variance changes based on an iterated cumulative sums of squares (ICSS) algorithm.

- Citation anchors:
  - page=2, words=0-349, section=I. INTRODUCTION, note=ICSS algorithm for variance changes; no weights discussed.

**C11: Are data linked/merged across sources? How is linkage performed?**

- Category: C) Data, sample, and measurement
- Confidence: low (0.2025617110628378)
- Retrieval method: local
- Evidence type: not described in context
- Answer:

The provided excerpts do not describe linking or merging data across sources; there is no information on data linkage or linkage methods.

- Quote snippet:

The ICSS Algorithm Applied to Residuals

- Citation anchors:
  - page=3, words=0-349, section=Introduction to ICSS/Variance Change Detection, note=Content discusses variance change detection in time series using the ICSS algorithm; no data linkage or merging is described.

**C12: What summary statistics are reported for main variables?**

- Category: C) Data, sample, and measurement
- Confidence: medium (0.3528184759025107)
- Retrieval method: local
- Evidence type: table
- Answer:

Mode, mean, and standard deviation of  $k^*$  are reported.

- Quote snippet:

Table 6. Summaries from the Sampling Distribution of  $k'$  for Series in bold type... Mode of  $k^*$  Mean of  $k^*$  SD of  $k^*$

- Citation anchors:
  - page=page 9, words=Table 6-SDof $k^*$ , section=4.4 Sampling Distribution of  $k^*$ , note=Table shows summaries for  $k^*$  including Mode, Mean, and SD.

**C13: Are there descriptive figures/maps that establish baseline patterns?**

- Category: C) Data, sample, and measurement
- Confidence: low (0.28962113769275105)
- Retrieval method: local
- Evidence type: figure\_description
- Data source: Inclan, C. & Tiao, G. (1994). Use of Cumulative Sums of Squares for Retrospective Detection of Changes in Variance. JASA.
- Table/Figure: Figure 1 (a and b)
- Answer:

Yes. The paper includes descriptive figures that establish baseline patterns for variance: Figure 1a shows a constant-variance series (white noise) where  $C_k$  is roughly a straight line, and Figure 1b shows the same series with two variance changes, illustrating how  $D_k$  captures changes.

- Quote snippet:

For the series with homogeneous variance,  $C_k$  presents roughly a straight line with slope  $a_2 = 1$ .

- Citation anchors:
  - page=3, words=0-349, section=Figure 1 description (baseline and variance changes), note=Figure 1a shows constant variance; Figure 1b shows two changes; baseline  $C_k$  is straight

**D01: What is the headline main effect estimate (sign and magnitude)?**

- Category: D) Results, magnitudes, heterogeneity, robustness
- Confidence: low (0.2764263296484324)

- Retrieval method: `local`
- Evidence type: `not_stated_in_context`
- Answer:

There is no headline main effect estimate provided in the given context.

- Quote snippet:
 

the centered (and normalized) cumulative sum of squares. improves on Wichern et al. by reducing the computational The plot of  $D_k$  against  $k$  will oscillate around 0 for series with homogeneous variance.
- Citation anchors:
  - page=2, words=825-916
  - page=3, words=0-349
  - page=3, words=275-624
  - page=3, words=550-882
  - page=4, words=0-349
  - page=4, words=275-624
  - page=4, words=550-869
  - page=5, words=0-349
  - page=5, words=275-624
  - page=9, words=0-349

#### **D02: What is the preferred specification and why is it preferred?**

- Category: D) Results, magnitudes, heterogeneity, robustness
- Confidence: low (0.22499800338784004)
- Retrieval method: `local`
- Evidence type: `recommendation`
- Data source: Simulation results in Section 4 comparing ICSS with likelihood-ratio tests and Bayesian approaches
- Table/Figure: Tables 4-8
- Answer:

Use the ICSS algorithm (the ICSS procedure) for detecting variance changes; it is preferred for long series with multiple change points because it identifies changepoints iteratively without exhaustive searches, offering lower CPU time and robust performance compared with likelihood-ratio tests or Bayesian methods, especially as series length grows or multiple changes occur.

- Quote snippet:
 

lend support to the recommendation to use the ICSS algorithm when we need to analyze long series with multiple change points
- Citation anchors:
  - page=11, words=0-349, section=4.6 CPU Time Requirements, note=Statement recommending ICSS for long series with multiple

change points

**D03: How economically meaningful is the effect (percent change, elasticity, dollars)?**

- Category: D) Results, magnitudes, heterogeneity, robustness
- Confidence: low (0.30573986846011725)
- Retrieval method: local
- Evidence type: lack\_of\_economic\_significance\_in\_context
- Data source: IBM stock prices (Box and Jenkins data), 1961-1962
- Answer:

The provided context does not discuss any economic meaning such as percent change, elasticity, or dollar amounts; it focuses on detecting variance changes in time series using the ICSS algorithm and provides examples (e.g., IBM stock prices) but does not quantify economic impact.

- Quote snippet:

The ICSS algorithm to other approaches and illustrates the variance shift at an unknown point in a sequence of independent observations, focusing on the detection of points of change
- Citation anchors:
  - page=2, words=the-change, section=CENTERED CUMULATIVE SUMS OF SQUARES, note=ICSS algorithm focuses on variance change detection; no economic effect sizes reported.

**D04: What are the key robustness checks and do results survive them?**

- Category: D) Results, magnitudes, heterogeneity, robustness
- Confidence: low (0.3380853248769257)
- Retrieval method: local
- Evidence type: simulation study with method comparisons
- Data source: Simulated data; ICSS algorithm performance study
- Answer:

Robustness checks included: (1) comparing the ICSS algorithm with likelihood-ratio tests and with a Bayesian approach; (2) varying series length ( $T = 100, 200, 500$ ), changepoint locations (beginning, middle), and variance-change magnitudes ( $A = 2$  and  $3$ ); (3) testing both one- and two-changepoint scenarios; (4) using simulated replicates to assess performance; (5) evaluating results on residuals and noting CPU-time considerations. Overall, detection is weaker for short series and small variance changes ( $A = 2$ , cp near the start), but improves with longer series or larger  $A$ ; the Bayesian approach performs best for  $A = 2$  and is comparable for  $A = 3$ , indicating the conclusions are robust to method but with caveats about power in small samples.

- Quote snippet:  
The simulation experiment has two separate parts, the first for one changepoint and the second for two changepoints.
- Citation anchors:
  - page=7, words=0-349, section=Simulation design, note=Two-part simulation to assess robustness across one vs two changepoints; varying T, changepoint location, variance ratio.
  - page=9, words=550-899, section=Results of Part Two, note=Short series (T=100) hard to detect small variance change (A=2); performance improves with A and location.
  - page=9, words=550-899, section=CPU Time and robustness, note=ICSS algorithm efficiency improves with longer series and larger variance ratios; Bayesian method robust across settings.

#### **D05: What placebo tests are run and what do they show?**

- Category: D) Results, magnitudes, heterogeneity, robustness
- Confidence: low (0.20043731555329525)
- Retrieval method: local
- Evidence type: not described in context
- Answer:

Placebo tests are not described in the provided context.

- Quote snippet:  
The simulation experiment has two separate parts, the first for one changepoint and the second for two changepoints in the generated series.
- Citation anchors:
  - page=page 7, words=275-556, section=Simulation experiments for ICSS algorithm, note=Notes that the simulation has two parts; no mention of placebo (placebo tests) is made.

#### **D06: What falsification outcomes are tested (unaffected outcomes)?**

- Category: D) Results, magnitudes, heterogeneity, robustness
- Confidence: low (0.2874098242390074)
- Retrieval method: local
- Evidence type: topic\_not\_discussed
- Data source: Incln and Tiao (1994) Retrospective Detection of Changes in Variance
- Answer:

The provided text does not discuss falsification outcomes or unaffected (unrelated) outcomes.

- Quote snippet:

The performance of different procedures used to determine the number of variance changes in a series can be measured in several ways.

- Citation anchors:
  - page=3, words=0-349, section=Background/Variance Changes, note=Discusses Dk, F-statistic, and change-point detection; no mention of falsification/unaffected outcomes.
  - page=7, words=0-349, section=LR tests and changepoint identification, note=Focuses on evaluating methods to determine the number of variance changes; no falsification outcomes discussed.
  - page=9, words=275-624, section=Results and methods, note=Describes simulation results, posterior odds, and sampling distributions; no falsification outcomes mentioned.

**D07: What heterogeneity results are reported (by income, size, baseline exposure, region)?**

- Category: D) Results, magnitudes, heterogeneity, robustness
- Confidence: low (0.32050029240926176)
- Retrieval method: local
- Evidence type: textual content review
- Data source: JASA 1994 paper: Two Variance Changes; ICSS algorithm simulation study.
- Answer:

The provided material does not report heterogeneity by income, size, baseline exposure, or region.

- Quote snippet:

The simulation experiment has two separate parts, the first for one changepoint and the second for two changepoints in the generated series.

- Citation anchors:
  - page=3, words=0-349, section=Variance change detection, note=Covers Dk, F-statistic relationship, and variance-change detection; no subgroup heterogeneity by socio-economic or regional factors.
  - page=9, words=0-349, section=Simulation study, note=Describes two-part simulation (one changepoint and two changepoints) focusing on variance changes, not heterogeneity across income/size/baseline exposure/region.

**D08: What mechanism tests are performed and what do they imply?**

- Category: D) Results, magnitudes, heterogeneity, robustness

- Confidence: low (0.20558593617213333)
- Retrieval method: local
- Evidence type: textual-analytic
- Answer:

They perform likelihood-ratio tests to determine the number of variance change points (NT) and a Bayesian posterior-odds approach to decide the most likely number of changepoints, complemented by the ICSS algorithm to retrospectively detect variance changes. The LR tests assess whether there is evidence to reject  $NT = n - 1$  in favor of  $NT = n$ ; the posterior odds quantify the relative likelihood of different numbers of changepoints given the data; and ICSS identifies and locates changepoints, with performance depending on series length and variance ratios.

- Quote snippet:  
Let  $LR_{n-1,n}$  denote the likelihood ratio statistic for testing  $H_0: NT = m$  against  $H_a: NT = n$ .
- Citation anchors:
  - page=8, words=0-329, section=4.1. Assessing the Evidence with Respect to the Number of Change Points Using Likelihood Ratio Tests and the Posterior Odds Ratio, note=Describes LR tests and posterior odds for number of changepoints; mentions  $NT = 0,1,2$

#### **D09: How sensitive are results to alternative samples/bandwidths/controls?**

- Category: D) Results, magnitudes, heterogeneity, robustness
- Confidence: medium (0.41549035980986854)
- Retrieval method: local
- Evidence type: experimental results from ICSS simulation and application
- Table/Figure: Table 3
- Answer:

Results are highly sensitive to sample size and the magnitude/location of variance changes. Small samples (e.g.,  $T = 100$ ) make detecting a small variance change ( $A = 2$ ) difficult, especially if the change occurs early in the series. Performance improves with longer series (200+ observations) and larger variance ratios ( $A = 3$ ). In almost every instance with 200+ observations, ICSS outperforms likelihood-ratio tests. The mode of the change-point location estimator  $k^*$  tends to align with the true change point as sample size grows, and the best detectability occurs when the larger variance is in the middle of the series. Computational choices (e.g., number of replicates) and boundary decisions (e.g., 1.358 quantile) also affect results, but bandwidth-like parameters are not extensively analyzed.

- Quote snippet:

In almost every instance with 200 observations or more, the ICSS algorithm gives better results than the likelihood ratio tests.

- Citation anchors:
  - page=9, words=1-20, section=4.3 Results of Part Two: Series With Two Changepoints, note=Small-sample difficulty: hard to detect a small variance change when  $T = 100$  ( $A = 2$ )
  - page=6, words=280-320, section=4.1 Assessing the Evidence with Respect to the Number of Change Points, note=ICSS often yields better results than LR tests with larger samples
  - page=9, words=320-360, section=4.4 Sampling Distribution of  $k^*$ , note=Mode of  $k^*$  aligns with the change point; converges as sample size grows
  - page=7, words=20-60, section=4. SIMULATION EXPERIMENT, note=Best results across change-point locations when the larger variance is in the middle (.33, .67)
  - page=9, words=710-730, section=4.6 CPU Time Requirements, note=ICSS performance improves with longer series and larger variance ratios

**D10: What are the main takeaways in the conclusion (bullet summary)?**

- Category: D) Results, magnitudes, heterogeneity, robustness
- Confidence: low (0.2047301701363368)
- Retrieval method: local
- Evidence type: Conclusion/summary statements from the ICSS paper
- Answer:
  - ICSS is effective for detecting variance changes, especially in longer series and when there are multiple changepoints.
  - For short series (around 100 observations), detecting a small variance change ( $A = 2$ ) is hard, particularly if the change occurs at the start.
  - With around 200 observations or a larger variance ratio ( $A = 3$ ), correct identifications improve, and if the changepoint is in the middle, identifications exceed 80%.
  - Best performance across locations occurs when changepoints are near 0.33 and 0.67 of the series; detecting two changepoints may require around 500 observations in monotone-variance scenarios.
  - The ICSS algorithm uses an iterative scheme to locate changepoints and avoids evaluating all possible locations; convergence typically happens in a few iterations, with a practical cap to avoid cycling.



- Compared methods show that Bayes can be best for  $A = 2$  but is very CPU-intensive; ICSS generally offers competitive performance with far lower computational burden, especially for long series with multiple change-points; results support recommending ICSS for such tasks.
- CPU-time considerations: ICSS has costs that grow with series length; in practice, the method balances accuracy and efficiency, and longer series and larger variance changes improve reliability.

- Quote snippet:

We conclude that for series of 100 observations, it is hard to detect a small variance change (variance ratio  $A = 2$ ) - particularly when it appears at the beginning of the series.

- Citation anchors:
  - page=page 9, words=0-349, section=4.2. Results of Part One: Series With One Changepoint, note=Difficulty detecting small variance change for 100-observation series with  $A=2$ ; start point worsens detection.
  - page=page 9, words=0-349, section=4.2. Results of Part One: Series With One Changepoint, note=For  $A=3$ , identifications exceed 80% when the changepoint is in the middle ( $n \sim 100$ ).
  - page=page 9, words=550-899, section=4.2. Results of Part One: Series With One Changepoint, note=ICSS performance improves with longer series and larger variance ratios.
  - page=page 9, words=825-904, section=4.5 The ICSS Algorithm Applied to Residuals, note=ICSS uses an iterative scheme for locating changepoints; avoids exhaustive search.
  - page=page 11, words=0-349, section=4.6 CPU Time Requirements, note=CPU time costs grow with series length; practical considerations noted.

**E01: What are the most important prior papers cited and why are they central here?**

- Category: E) Citations and related literature
- Confidence: low (0.2772812212214367)
- Retrieval method: local
- Evidence type: literature-citation analysis
- Answer:

The most important prior papers cited are those that established the change-point literature for variance, provided early detection tests, and offered Bayesian or likelihood-based identification methods. Key works include Hsu, Miller, and Wichern (1974) on modeling variance changes in stock returns; Hinkley (1971) on inference about change points from cumulative sums; Hsu (1977, 1979, 1982) on variance shifts at unknown points; Wichern, Miller, and Hsu (1976) on au-

toregressive models with variance change; Brown, Durbin, and Evans (1975) on testing constancy of regression relationships over time; Baufays and Rasson (1985) on estimating variances and change points (simultaneously); Booth and Smith (1982) on a Bayesian approach to retrospective identification of change-points; and Haccou, Meelis, and van de Geer (1988) on the number of change points. These works are central because they (a) formalize the problem of detecting changes in variance, (b) develop testing procedures (CUSUM/f-statistics and likelihood/Bayesian approaches), and (c) motivate the ICSS framework’s focus on identifying multiple variance-change points with cumulative sums of squares.

- Quote snippet:

The statistical literature on changes of variance started with Hsu, Miller, and Wichern (1974), who offered this formulation as an alternative to the Pareto distribution to model stock returns.

- Citation anchors:
  - page=page 2, words=0-349, section=Introduction, note=Cites Hsu, Miller, and Wichern (1974) as foundational to variance-change literature; mentions using cumulative sums to model stock returns.
  - page=page 2, words=40-100, section=Introduction, note=Mentions Hinkley (1971) and Smith (1975, 1980) among early change-point references.
  - page=page 12, words=0-349, section=Appendix/References, note=Cites Booth and Smith (1982) Bayesian approach to identification of change-points.
  - page=page 12, words=275-624, section=References, note=Cites Hsu (1977, 1979, 1982) and Baufays & Rasson (1985) among key variance-change methodologies.

## **E02: Which papers does this work most directly build on or extend?**

- Category: E) Citations and related literature
- Confidence: low (0.2504107580870808)
- Retrieval method: local
- Evidence type: Literature/background references
- Answer:

It directly builds on and extends the variance-change literature initiated by Hsu, Miller, and Wichern (1974) and the cumulative-sums framework of Brown, Durbin, and Evans (1975).

- Quote snippet:

”The statistical literature on changes of variance started with Hsu, Miller, and Wichern (1974), who offered this formulation as an alternative to the Pareto distribution to model stock returns.”

- Citation anchors:
  - page=2, words=0-349, section=Background/Introduction, note=Mentions foundational variance-change work: Hsu, Miller, and Wichern (1974). Also cites Brown, Durbin, Evans (1975) for the cumulative sum of squares framework.
  - page=2, words=0-349, section=Background/Introduction, note=Explicitly states the approach uses a centered version of the cumulative sum of squares presented by Brown, Durbin, Evans (1975).

**E03: Which papers are used as benchmarks or comparisons in the results?**

- Category: E) Citations and related literature
- Confidence: low (0.30018478436149454)
- Retrieval method: local
- Evidence type: textual evidence from article describing comparisons and references
- Answer:

The results are benchmarked against traditional comparison methods and standard time-series references: likelihood ratio tests and the log posterior odds (Bayesian approach) are used for comparison, with Box, G. E. P., and Jenkins (1976); Billingsley (1968); and Bratley, P., Fox, B. L., and Schrage, L. E. (1987) cited as benchmark papers.

- Quote snippet:

The simulation experiment has two separate parts, the first for one changepoint and the second for two changepoints in the generated series.

- Citation anchors:
  - page=page 7, words=0-349, section=4.1. Assessing the Evidence with Respect to the Number of Change Points Using Likelihood Ratio Tests and the Posterior Odds Ratio, note=Describes comparing ICSS with LR tests and posterior odds (Bayesian) methods.
  - page=page 12, words=0-120, section=APPENDIX A: APPROXIMATE EXPECTED VALUE OF  $D_k$ , note=Benchmarks cited within references include Box, Jenkins (1976); Bratley, Fox, Schrage (1987); Billingsley (1968).
  - page=page 12, words=120-240, section=APPENDIX A: APPROXIMATE EXPECTED VALUE OF  $D_k$ , note=Cited standard references for time-series and simulation used as benchmarks.
  - page=page 12, words=240-349, section=APPENDIX A: APPROXIMATE VALUE OF  $D_k$ , note=Billingsley (1968) referenced in context of foundational probability results.

**E04: What data sources or datasets are cited and how are they used?**

- Category: E) Citations and related literature
- Confidence: low (0.24775436052606295)
- Retrieval method: local
- Evidence type: Simulation study using synthetic time-series data (AR processes) to evaluate variance-change detection methods.
- Data source: Simulated time-series data (synthetic autoregressive processes) used to test ICSS performance; no external datasets cited.
- Answer:

The study uses simulated time-series data (synthetic AR processes) rather than real datasets. These synthetic series, including AR(1) residuals and series with one or two variance-change points, of lengths  $T = 100, 200$ , and  $500$ , are generated to evaluate the ICSS algorithm against likelihood ratio and Bayesian approaches.

- Quote snippet:

The simulation experiment has two separate parts, the main way is by the number of 'correct identifications'.

- Citation anchors:
  - page=2, words=0-349, section=Introduction, note=Mentions the problem of multiple change points in variance and that simulation results compare ICSS to Bayesian or LR approaches.
  - page=6, words=550-899, section=4. SIMULATION EXPERIMENT, note=Describes the ICSS procedure and systematic search for change points; discusses simulation setup.
  - page=9, words=0-349, section=4. SIMULATION EXPERIMENT, note=The same series used in the first part of the simulation experiment were used to obtain an autoregressive process.
  - page=9, words=550-899, section=4.2 Results of Part One, note=Discusses results for residuals and applying ICSS to AR(1) residuals.
  - page=9, words=275-624, section=4. SIMULATION EXPERIMENT, note=References to tables and results from the simulation (Tables 4-8) evaluating ICSS against other methods.

**E05: What methodological or econometric references are cited (e.g., DiD, IV, RDD methods)?**

- Category: E) Citations and related literature
- Confidence: low (0.3465951953234926)
- Retrieval method: local
- Evidence type: textual reference list illustrating time-series change-point/variance-change econometrics

- Answer:

The methodological/econometric references cited are predominantly time-series change-point and variance-change literature (e.g., Bayesian and likelihood-based change-point methods, variance-change detection) rather than DiD, IV, or RDD approaches. Notable sources include Box & Jenkins (Timeseries Analysis, Forecasting and Control), Hinkley (1971), Hsu (1977, 1979, 1982), Booth & Smith (1982), Inclán (1991), Tsay (1988), Wichern, Miller & Hsu (1976), Menzefricke (1981), Smith (1975, 1980), Worsley (1986), and related references in the paper’s bibliography.

- Quote snippet:

Box, G. E. P., and Jenkins, G. M. (1976), Timeseries Analysis, Forecasting and Control, San Francisco: Holden Day.

- Citation anchors:

– page=12, words=550-858, section=References, note=Contains discussion of change-point/time-series econometric references (e.g., Hinkley, Hsu, Booth & Smith, Inclán, Tsay, Wichern et al., Box & Jenkins).

#### **E06: Are there any seminal or classic references the paper positions itself against?**

- Category: E) Citations and related literature
- Confidence: low (0.23734506593616342)
- Retrieval method: local
- Evidence type: literature\_reference
- Answer:

Yes. The paper cites and contrasts with seminal literature on variance change and change-point detection, notably Hsu, Miller, and Wichern (1974) and subsequent foundational works such as Hinkley (1971), Brown, Durbin, and Evans (1975), Smith (1975, 1980), Billingsley (1968), and Bayesian/likelihood approaches like Booth and Smith (1982) and Abraham & Wei (1984).

- Quote snippet:

The statistical literature on changes of variance started with Hsu, Miller, and Wichern (1974), who offered this formulation as an alternative to the Pareto distribution to model stock returns.

- Citation anchors:

– page=2, words=0-349, section=Introduction, note=Mentions Hsu, Miller, and Wichern (1974) as early literature on variance changes.  
 – page=12, words=550-858, section=References, note=References include Hinkley (1971); Brown, Durbin, Evans (1975); Billingsley (1968); Booth & Smith (1982); Abraham & Wei (1984); Hsu (1977, 1979, 1982).

**E07: Are there citations to code, data repositories, or appendices that are essential to the claims?**

- Category: E) Citations and related literature
- Confidence: low (0.24801875970379558)
- Retrieval method: local
- Evidence type: code and appendices
- Answer:

Yes. The article cites a Fortran program of the ICSS algorithm that is available by request, and it references appendices (Appendix A and Appendix 6) containing methodological details and proofs.

- Quote snippet:

The Fortran program of the ICSS algorithm is available from us upon request; send an electronic mail message to [inclan@guvax.georgetown.edu](mailto:inclan@guvax.georgetown.edu).

- Citation anchors:
  - page=6, words=838-898, section=Code availability, note=Fortran ICSS algorithm program available by request ([inclan@guvax.georgetown.edu](mailto:inclan@guvax.georgetown.edu)).
  - page=11, words=0-40, section=APPENDIX A, note=Approximate expected value of  $D_k$ ; regression models.
  - page=12, words=0-25, section=APPENDIX 6, note=Proof of Theorem 1.

**E08: What gaps in the literature do the authors say these citations leave open?**

- Category: E) Citations and related literature
- Confidence: low (0.23079481257921125)
- Retrieval method: local
- Evidence type: literature\_gap
- Answer:

The citations leave open the need for efficient, systematic detection of multiple variance changes. Much of the prior literature focuses on single changepoints or relies on heavy computational Bayesian/likelihood approaches, making simultaneous detection of several variance changes impractical.

- Quote snippet:

One of the concerns that motivated the design of the ICSS algorithm is the heavy computational burden involved in the calculation of posterior odds.

- Citation anchors:

- page=2, words=60-90, section=Introduction, note=Mentions prior work and the search for multiple change points in an iterative way, implying limited prior methods for multiple changes.
- page=9, words=560-585, section=4.6 CPU Time Requirements, note=States the heavy computational burden of posterior odds and the burden in multi-change detection, motivating a need for a simpler method.

**F01: Are replication files or code provided? If so, where?**

- Category: F) Replication and transparency
- Confidence: low (0.1940297762178587)
- Retrieval method: local
- Evidence type: textual
- Data source: ICSS replication materials: Fortran program available on request from Inclán (inclan@guvax.georgetown.edu).
- Answer:

Yes. The Fortran program of the ICSS algorithm is available from the authors upon request; contact inclan@guvax.georgetown.edu.

- Quote snippet:  
The Fortran program of the ICSS algorithm is available from us upon request; send an electronic mail message to inclan@guvax.georgetown.edu.
- Citation anchors:
  - page=6, words=825-968, section=ICSS implementation availability, note=Fortran program of the ICSS algorithm available on request; email provided.

**F02: Is there a pre-analysis plan or registered trial? Provide details if mentioned.**

- Category: F) Replication and transparency
- Confidence: low (0.22318956605936977)
- Retrieval method: local
- Evidence type: no\_pre\_analysis\_plan\_mentioned
- Data source: Inclán and Tiao (1994), 'Use of Cumulative Sums of Squares for Retrospective Detection of Changes of Variance', JASA
- Answer:

No. The provided context does not mention any pre-analysis plan or registered trial.

- Quote snippet:  
The simulation experiment has two separate parts, the first for one changepoint and the second for two changepoints

- Citation anchors:
  - page=7, words=275-556, section=4. SIMULATION EXPERIMENT, note=Contains a statement about the simulation experiment having two separate parts (one and two changepoints), indicating no pre-registered design is described.

**F03: Are data access constraints disclosed (restricted access, proprietary data, NDAs)?**

- Category: F) Replication and transparency
- Confidence: low (0.1822168798303961)
- Retrieval method: local
- Evidence type: Access restrictions in publication repository terms
- Data source: JSTOR Terms and Conditions of Use (publication access constraints) and article data references.
- Answer:

Yes. The context discloses data access constraints via JSTOR's Terms and Conditions of Use, including restrictions on downloading entire issues and the requirement to use content for personal, non-commercial use.

- Quote snippet:
 

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use.
- Citation anchors:
  - page=1, words=0-230, section=JSTOR access terms, note=Cites JSTOR Terms and Conditions of Use; restricts downloading whole issues and non-commercial use.

**F04: Are key steps in data cleaning and construction documented?**

- Category: F) Replication and transparency
- Confidence: low (0.24536972844114352)
- Retrieval method: local
- Evidence type: data\_construction\_and\_cleaning\_documentation
- Data source: Box and Jenkins (1976)
- Table/Figure: Figure 4
- Answer:

Yes. The article documents data construction and cleaning steps (e.g., transforming data by taking the first difference of the logarithm of stock prices) and procedures for diagnosing and correcting variance changes/outliers as part of the ICSS method.

- Quote snippet:



The series analyzed is the first difference of the logarithm of the IBM stock closing prices from May 17, 1961, to November 2, 1962, as reported by Box and Jenkins (1976).

- Citation anchors:
  - page=6, words=550-899, section=3.1 Example: IBM Stock Prices, note=Data construction and cleaning illustrated by IBM stock price series; first difference of log prices used.

**F05: Are robustness and sensitivity analyses fully reported or partially omitted?**

- Category: F) Replication and transparency
- Confidence: low (0.3097132235620471)
- Retrieval method: local
- Evidence type: Assessment of robustness/sensitivity reporting in the article's simulation study.
- Data source: Simulation study described in the article.
- Table/Figure: Table 6 and Table 7 summarize sampling distributions; Table 8 shows residual AR(1) results.
- Answer:

Partially reported. The study conducts robustness/sensitivity checks across several scenarios (varying series length, variance changes, and change-point locations) but some analyses are not fully reported due to computational limits; for example, posterior odds results for large T (T=500) were not obtained, and additional results are promised elsewhere.

- Quote snippet:
 

The posterior odds ratio was not obtained for series with T = 500, because it would have required a total of 1,358 CPU hours.
- Citation anchors:
  - page=9, words=0-349, section=4.3, note=Posterior odds for series with T=500 not obtained due to CPU hours; indicates partial reporting due to computation.
  - page=9, words=275-624, section=4.3-4.6, note=Discussion of heavy computational burden and robustness across scenarios; CPU time concerns.
  - page=9, words=550-899, section=4.6, note=CPU Time Requirements section detailing computational burden that limits reporting.

**G01: What populations or settings are most likely to generalize from this study?**

- Category: G) External validity and generalization
- Confidence: low (0.21155484744527037)
- Retrieval method: local

- Evidence type: **Introduction/Overview**
- Data source: Simulated data ( $N(0,1)$ ); residuals from AR models discussed in simulations
- Assumption notes: Assumes independent observations with possible variance changes; applicable to residuals and financial time series.
- Answer:

The study generalizes to time series or sequential data where variance can change over time (variance change points) in sequences of independent observations, including financial time series that do not have constant variance; the ICSS method is demonstrated on simulated data and residuals from models, suggesting applicability to settings with variance shifts in practice.

- Quote snippet:  
This article studies the problem of multiple change points in the variance of a sequence of independent observations.
- Citation anchors:
  - page=2, words=0-349, section=Introduction, note=This article studies the problem of multiple change points in the variance of a sequence of independent observations; mentions finance context.

## **G02: What populations or settings are least likely to generalize?**

- Category: G) External validity and generalization
- Confidence: low (0.20619964347148861)
- Retrieval method: local
- Evidence type: assumptions about data distribution and independence (Normality of the innovations) used to derive likelihood-ratio testing and Dk behavior
- Data source: Theoretical derivations and simulation study in Inclhn and Tiao (1994); Gaussian residuals used in simulations.
- Assumption flag: **True**
- Assumption notes: Assumes Gaussian (Normal) and independent observations; results may not generalize to non-normal or dependent data.
- Answer:

Populations or settings with non-normal, non-iid errors (e.g., non-Gaussian residuals, autoregressive or dependent data, heavy tails), since the method and its properties in the cited work rely on Gaussian, independent observations.

- Quote snippet:  
If the  $\{a_i\}$  are Normally distributed, with mean 0 and variances  $a_i$ ,  $t = 1, \dots, T$ , then we can obtain the likelihood ratio for testing one change against no change in variance
- Citation anchors:

- page=3, words=0-349, section=2.1, note=The method derivation assumes normally distributed  $a_t$  and enables likelihood-ratio testing for variance changes.

**G03: Do the authors discuss boundary conditions or scope limits?**

- Category: G) External validity and generalization
- Confidence: low (0.26157004968814634)
- Retrieval method: local
- Evidence type: textual evidence from algorithm description and implementation notes
- Answer:

Yes. The authors discuss boundary conditions or scope limits in their ICSS change-point method. They describe using a threshold  $D^*$  (boundaries) to decide when a change point exists, explain that boundaries can be obtained from the asymptotic distribution, and note practical limits on the iterative procedure (an iteration cap of 20) and convergence criteria (stop when the number of points stops changing by more than a specified amount).

- Quote snippet:  
"If  $M(t_1 : t_2) > D^*$ , max, 1 D, I can be obtained."
- Citation anchors:
  - page=3, words=275-624, section=ICSS boundary definition, note=Boundaries can be obtained from the asymptotic distribution; NT=0/1 change hypothesis.
  - page=5, words=0-349, section=ICSS Step 1/Thresholds, note=Threshold  $D^*$  used: If  $M(t_1:t_2) > D^*$ , a changepoint can be obtained.
  - page=5, words=825-1174, section=ICSS Implementation details, note=Iteration limit of 20 to avoid cycling; convergence criterion.
  - page=6, words=550-899, section=Convergence criteria, note=Convergence when number of points does not change by more than a specified amount.

**G04: How might the results change in different time periods or markets?**

- Category: G) External validity and generalization
- Confidence: medium (0.37509148185008667)
- Retrieval method: local
- Evidence type: textual (simulation results and IBM stock example from the ICSS methodology)
- Data source: Inclán and Tiao (1994), ICSS algorithm; IBM stock prices example
- Answer:

Results may differ across time periods or markets due to sample size, the number and location of variance changes, and the magnitude and pattern of shifts. In small samples (e.g.,  $T=100$ ), the ICSS algorithm tends to detect only one changepoint; detection improves with longer series and favorable change locations (best when changepoints are around  $1/3$  and  $2/3$  of the series). Monotone increases in variance can make detection harder; real-data examples (IBM stock) show multiple changepoints depending on data, and outliers can mislead the  $D_k$  peak.

- Quote snippet:

For series of 100 observations, we find only one changepoint most of the time.

- Citation anchors:
  - page=9, words=0-349, section=4.4 Sampling Distribution of  $k^*$ , note=For series with 100 observations, we find only one changepoint most of the time.
  - page=9, words=0-349, section=4.5 The ICSS Algorithm Applied to Residuals, note=The mode of  $k^*$  is exactly at the change point where the change in variance occurs; as the sample size increases,  $k^*$  concentrates around the true change-point.
  - page=6, words=0-349, section=Two change-points configuration, note=The best results across location of the points of change are at about  $0.33T$  and  $0.67T$ .
  - page=2, words=275-624, section=IBM Stock Prices, note=ICSS applied to IBM stock prices shows multiple potential changepoints ( $cp1=342$ ,  $cp2=397$ ,  $cp3=526$ ) and illustrates the procedure on real data.
  - page=9, words=275-624, section=4.4 Sampling Distribution of  $k^*$ , note=The distribution is skewed;  $k^*$  is biased towards the middle of the series.

#### **H01: Are key variables measured directly or via proxies?**

- Category: H) Measurement validity
- Confidence: low (0.278265772321319)
- Retrieval method: local
- Evidence type: data\_description
- Data source: Box and Jenkins (1976) IBM stock prices dataset
- Table/Figure: Figure 4
- Answer:

They are measured directly from the observed data; the analysis uses the time series of IBM stock prices (specifically, the first difference of the log prices) as the data, and the statistics ( $D_k$ ,  $C_k$ , etc.) are computed from that series rather than proxies for latent variables.

- Quote snippet:  
The series analyzed is the first difference of the logarithm of the IBM stock closing prices from May 17, 1961, to November 2, 1962, as reported by Box and Jenkins (1976).
- Citation anchors:
  - page=6, words=275-624, section=3.1 Example: IBM Stock Prices, note=Describes using IBM stock prices; data are the first difference of the logarithm of closing prices.

## H02: What measurement error risks are acknowledged or likely?

- Category: H) Measurement validity
- Confidence: low (0.26744468580487696)
- Retrieval method: local
- Evidence type: textual\_inference\_from\_methodology
- Assumption flag: True
- Assumption notes: Assumes iid  $\text{Normal}(0, \hat{a}^2)$  errors for observations; measurement error risk arises if normality or independence is violated (affecting  $D_k$  thresholds and change-point detection).
- Answer:

The study acknowledges measurement error via the assumption that the observed series are driven by independent, identically distributed  $\text{Normal}(0, \hat{a}^2)$  errors (the  $a_i$ ). This implies the results rely on normally distributed, independent measurement noise; violations (e.g., non-normality, autocorrelation, or non-constant variance) could affect the distribution of  $D_k$  and the reliability of variance-change detection.

- Quote snippet:  
Let  $a_1, a_2, \dots$  be a sequence of independent, identically distributed  $\text{Normal}(0, \hat{a}^2)$  random variables.
- Citation anchors:
  - page=5, words=0-60, section=Theorem 1 / Model assumptions, note=States independent, identically distributed  $\text{Normal}(0, \hat{a}^2)$  errors for  $a_i$
  - page=3, words=0-40, section=Model/setup, note=Mentions  $N(0, 1)$  white noise and Normal assumptions for  $a_i$

## H03: Are there validation checks for key measures?

- Category: H) Measurement validity
- Confidence: low (0.25965959834216273)
- Retrieval method: local
- Evidence type: validation\_checks
- Data source: ICSS algorithm from Inclhn and Tiao (1994), JASA

- Assumption flag: **True**
- Assumption notes: Assumes the  $a_i$  are normally distributed with mean 0 and variances  $a_i$ ; normality underpins the likelihood ratio and Dk-based tests
- Answer:

Yes. The article describes validation checks for key measures (Dk and change points) by comparing the maximum  $|Dk|$  to a predetermined boundary  $D^*$ , using asymptotic and empirical quantiles to decide if a change point exists; it repeats steps 2a–2c and may apply a Bayesian check (Step 3) to refine change points, then uses a corrected series for residual diagnostics (e.g., the Q statistic).

- Quote snippet:  
If this maximum absolute value exceeds a pre- determined boundary, then we may conclude that there is a change point near  $k^*$  and take  $k^*$  as an estimate of the change observations.
- Citation anchors:
  - page=3, words=0-349, section=Step 2a, note=Dk boundary validation: maximum exceeding boundary implies a change point near  $k^*$
  - page=5, words=0-349, section=Step 2a, note=Step 2a describes computing  $t_2$  and evaluating Dk
  - page=5, words=275-624, section=Step 2b, note=Step 2b describes starting the search toward the end from the first change
  - page=6, words=550-899, section=Step 3, note=Step 3: Bayesian analysis to refine change points
  - page=7, words=0-349, section=Residual diagnostics, note=After correction, residual diagnostics like the Q statistic are used

#### **H04: Do the authors discuss construct validity for core outcomes?**

- Category: H) Measurement validity
- Confidence: low (0.2123228301267673)
- Retrieval method: local
- Evidence type: Negative evidence: the article focuses on detecting variance changes using ICSS; there is no discussion of construct validity for core outcomes.
- Answer:

No.

- Quote snippet:  
This article studies the problem of multiple change points in the variance of a sequence of independent observations.
- Citation anchors:
  - page=page 2, words=0-349, section=Introduction, note=Contains the problem statement and scope: detection of changes in variance;

no mention of construct validity.

**I01: What policy counterfactuals are considered or implied?**

- Category: I) Policy counterfactuals and welfare
- Confidence: low (0.2852497082884739)
- Retrieval method: local
- Evidence type: no\_policy\_counterfactuals\_present\_in\_text
- Data source: IBM stock prices data (May 17, 1961 to November 2, 1962) used in ICSS example; references Box and Jenkins (1976).
- Answer:

No policy counterfactuals are discussed or implied in the provided context.

- Quote snippet:

The main interest is to study the variance of a given sequence of observations retrospectively, so we can use all the information on the series to indicate the points of variance change.

- Citation anchors:
  - page=2, words=550-899, section=2. CENTERED CUMULATIVE SUMS OF SQUARES, note=ICSS algorithm context; text discusses variance change detection rather than policy counterfactuals.

**I02: What are the main welfare tradeoffs or distributional impacts discussed?**

- Category: I) Policy counterfactuals and welfare
- Confidence: low (0.2454828695112901)
- Retrieval method: local
- Evidence type: textual summary
- Answer:

The main tradeoffs discussed are (1) computational burden versus detection accuracy for variance changepoints (ICSS is computationally intensive, especially when searching for several change points, compared with likelihood- or Bayesian-based approaches), (2) how detection performance depends on data length, the location of changepoints, and the variance ratio (longer series and larger variance changes improve detectability), and (3) distributional aspects that affect detection, such as how  $D_k$  behaves asymptotically (its distribution resembles Brownian motion under homogeneity), the piecewise-linear shape of  $E[D_k]$  with slope changes at changepoints, and the bias of the estimated changepoint location toward the series center, which together influence the difficulty of identifying multiple changepoints.

- Quote snippet:

heavy computational burden involved in looking for several points of change simultaneously.

- Citation anchors:
  - page=2, words=550-899, section=Introduction, note=cost of using the ICSS algorithm in comparison to other approaches; heavy computational burden involved in looking for several points of change simultaneously.
  - page=4, words=0-349, section=Asymptotic Behavior of  $D_k$ , note=The asymptotic distribution of  $D_k$  and related discussion; two changepoints; maximum likelihood estimator of  $K$ .
  - page=4, words=550-869, section=Two Change Points, note=Two changepoints and the piecewise linear behavior of  $E[D_k]$  with slope changes at changepoints; symmetry considerations.
  - page=9, words=550-899, section=Part One/Two Results, note=The ICSS algorithm performance improves notably as series length increases and variance ratios grow.
  - page=9, words=825-904, section=Algorithm Properties, note=The ICSS algorithm avoids calculating a function at all possible locations of the change-point.
  - page=9, words=275-624, section=Simulation Results, note=Masking effect with multiple changepoints and the iterative approach to mitigate it.

### **I03: Are cost-benefit or incidence analyses provided?**

- Category: I) Policy counterfactuals and welfare
- Confidence: low (0.28148167429525717)
- Retrieval method: local
- Evidence type: textual analysis of the document's methods and results
- Answer:

No; cost-benefit or incidence analyses are not provided. The text analyzes computational cost using CPU time and simulation results, not economic incidence analyses.

- Quote snippet:
 

The ICSS algorithm avoids calculating a function at all possible locations of the change-
- Citation anchors:
  - page=page 9, words=The ICSS algorithm avoids calculating a function at all possible locations of the change--change-, section=4.6 CPU Time Requirements, note=Describes computational burden and iterative changepoint search; indicates focus on CPU time rather than cost-benefit or incidence analyses.
  - page=page 11, words=On average, after cutting and analyzing the pieces, we need to perform-operations., section=CPU Time Analysis



- (Appendix A), note=Plain statement about CPU operations, reinforcing that the discussion centers on computational cost.
- page=page 11, words=This section presents actual CPU times obtained in the simulations described earlier.-earlier., section=CPU Time Analysis, note=Explicitly references CPU time measurements rather than economic cost/benefit or incidence analysis.

#### **I04: What policy recommendations are stated or implied?**

- Category: I) Policy counterfactuals and welfare
- Confidence: low (0.21516643219529108)
- Retrieval method: local
- Evidence type: none (policy recommendations not present; methodological content only)
- Data source: Inclan & Tiao (1994), ICSS algorithm; Journal of the American Statistical Association
- Answer:

No policy recommendations are stated or implied in the provided excerpts; the content is technical and methodological, describing the ICSS algorithm for retrospective detection of variance changes and related statistical analyses.

- Quote snippet:

The ICSS algorithm avoids calculating a function at all

- Citation anchors:
  - page=page 5, words=825-1174, section=Iterated Cumulative Sums of Squares (ICSS) Algorithm, note=Describes the ICSS procedure and convergence criteria for change-point detection.
  - page=page 6, words=550-899, section=Algorithmic details and Bayesian analysis, note=Discusses variance changes, algorithm performance, and Bayesian context.
  - page=page 8, words=0-329, section=ICSS Algorithm Applied to Residuals; Part Two Results, note=Residuals analysis and comparison with LR tests; methodological results.
  - page=page 9, words=0-349, section=Results: ICSS vs LR; posterior odds; changepoint identification, note=Discussion of results and identifications of changepoints; methodological evaluation.
  - page=page 9, words=550-899, section=Results: ICSS algorithm performance; series with two changepoints, note=Further results and interpretation of ICSS performance.
  - page=page 12, words=0-349, section=Appendix and References; Proof of Theorem 1, note=Appendix materials and references; not policy-related.

**J01: What missingness or attrition patterns are reported?**

- Category: J) Data quality and integrity
- Confidence: low (0.3050894031058159)
- Retrieval method: local
- Evidence type: not\_reported\_in\_context
- Answer:

No missingness or attrition patterns are reported in the provided context.

- Quote snippet:

This procedure performs a systematic search for possible change points of the series.

- Citation anchors:
  - page=3, words=550-600, section=Variance change detection, note=Content discusses variance change points; no missingness/attrition discussion.
  - page=4, words=0-60, section=Intro/Methods, note=Paper focuses on retrospective detection of changes in variance; no missing data discussion.

**J02: How are outliers handled (winsorization, trimming, exclusions)?**

- Category: J) Data quality and integrity
- Confidence: low (0.3496550751118726)
- Retrieval method: local
- Evidence type: text
- Answer:

Outliers are handled by applying a dedicated outlier-detection procedure as part of the residual diagnostics; the text does not specify methods like winsorization, trimming, or exclusions. It notes that outliers can affect the Dk plot and recommends complementing the search for variance changes with an outlier-detection step (see Chang, Tiao, and Chen 1988).

- Quote snippet:

for variance changes with a procedure for outlier detection (see Chang, Tiao, and Chen 1988).

- Citation anchors:
  - page=6, words=0-349, section=Residual diagnostics and outlier detection in ICSS algorithm, note=Mentions 'a procedure for outlier detection' to complement search for variance changes.

**J03: Are there data audits or validation steps described?**

- Category: J) Data quality and integrity
- Confidence: low (0.24508686605825253)

- Retrieval method: `local`
- Evidence type: `simulation study and diagnostic validation`
- Data source: IBM stock prices (May 17, 1961 to November 2, 1962) as reported by Box and Jenkins (1976); first difference of logarithm of closing prices.
- Table/Figure: Table 1-4; Figures 4a-4g
- Answer:

Yes. The article describes validation/validation steps including simulation studies comparing the ICSS algorithm to likelihood-ratio and Bayesian approaches, residual diagnostics, outlier checks, convergence criteria, and cross-validation on real data (IBM stock prices).

- Quote snippet:  
The simulation experiment to compare the performance of the ICSS algorithm with the likelihood ratio and a Bayesian approach.
- Citation anchors:
  - page=page 9 words 275-624, words=275-624, section=4. SIMULATION EXPERIMENT, note=Simulation comparing ICSS algorithm with likelihood ratio and Bayesian approaches; discusses validation via simulation and posterior odds.
  - page=page 6 words 275-624, words=275-624, section=3.1 Example: IBM Stock Prices, note=IBM stock price example used to illustrate validation/diagnostics of ICSS; includes residuals and convergence checks.

#### **J04: Is there evidence of reporting bias or selective sample inclusion?**

- Category: J) Data quality and integrity
- Confidence: `low` (0.2815055471289827)
- Retrieval method: `local`
- Evidence type: `disclosure of computational constraints affecting reporting`
- Answer:

No evidence of reporting bias or selective sample inclusion is evident in the provided material. The authors explicitly acknowledge computational constraints that limited reporting (e.g., not obtaining the posterior odds for  $T=500$  due to CPU hours), rather than omitting results selectively.

- Quote snippet:  
The posterior odds ratio was not obtained for series with  $T = 500$ , because it would have required a total of 1,358 CPU hours.
- Citation anchors:
  - page=7, words=0-349, section=4.1 Assessing the Evidence with Respect to the Number of Change Points Using Likelihood Ratio Tests

- and the Posterior Odds Ratio, note=Notes that the simulation used 1,000 replicates per design point.
- page=9, words=0-349, section=4.6 CPU Time Requirements, note=States heavy computational burden; posterior odds for T=500 not obtained due to 1,358 CPU hours; mentions 10,000 replicates for ICSS vs 1,000 for other cases.
- page=9, words=550-899, section=4.4 Sampling Distribution of  $k^*$ , note=Discusses distribution of  $k^*$  and relates replicates to computational limits; notes variability due to limited runs.

**K01: What goodness-of-fit or diagnostic metrics are reported?**

- Category: K) Model fit and diagnostics
- Confidence: low (0.3192011172563224)
- Retrieval method: local
- Evidence type: Statistical metrics and simulation results (LR tests, posterior odds,  $N\_T$  distributions,  $k^*$  distribution, CPU time).
- Table/Figure: Table 4 and Table 5
- Answer:

They report likelihood-ratio tests for the number of variance changes, log posterior odds with posterior probabilities for the number of changepoints, frequency distributions of detected changepoints ( $N\_T$ ) from the ICSS algorithm, the sampling distribution of the changepoint location  $k^*$ , and CPU-time as a computational diagnostic.

- Quote snippet:
 

Table 4. Frequency Distribution of  $N$ , Obtained with the ICSS Algorithm, Likelihood Ratio Tests and Log Posterior Odds for Series with One Variance Change
- Citation anchors:
  - page=8, words=0-329, section=Table 4, note=Reports frequency distribution of  $N\_T$  obtained with the ICSS algorithm, likelihood ratio tests and log posterior odds for series with one variance change.
  - page=8, words=0-329, section=Table 5, note=Reports frequency distribution of  $N\_T$  with ICSS, LR tests, and log posterior odds for series with two variance changes.
  - page=9, words=275-624, section=4.4 Sampling Distribution of  $k^*$ , note=Presents the sampling distribution of  $k^*$ , including mode, mean, and SD, for the location where  $\max_k |D_k|$  is attained.
  - page=9, words=550-899, section=4.6 CPU Time Requirements, note=Discusses CPU time requirements and computational burden of the ICSS and posterior-odds approaches.

**K02: Are functional form choices tested (logs, levels, nonlinearities)?**

- Category: K) Model fit and diagnostics
- Confidence: low (0.3409648803258408)
- Retrieval method: local
- Evidence type: text
- Data source: IBM stock prices; first difference of the logarithm of prices (ICSS example).
- Assumption flag: True
- Assumption notes: Assumes normality of the increments  $a_i$ ; changes in variance are modeled via changepoints and tested with likelihood ratio methods.
- Answer:

No. The provided context does not indicate testing alternative functional forms (such as logs vs levels or nonlinear specifications). It reports using a log transform of stock prices as part of preprocessing (first-difference of the log prices) and then tests variance changes under a normality assumption using likelihood- and ICSS-based methods, not explicit functional-form comparisons.

- Quote snippet:

The series analyzed is the first difference of the logarithm of the IBM stock closing prices from May 17, 1961, to November 2, 1962.

- Citation anchors:
  - page=6, words=275-624, section=3.1 Example: IBM Stock Prices, note=Exemplar uses the first difference of the logarithm of IBM stock closing prices; describes preprocessing with logs rather than testing different functional forms.

**K03: Are residual checks or specification tests reported?**

- Category: K) Model fit and diagnostics
- Confidence: low (0.3297158477161626)
- Retrieval method: local
- Evidence type: quote
- Answer:

Yes. The article discusses residual checks and states that after using the ICSS algorithm to obtain a corrected, variance-homogeneous series, one should perform residual diagnostics (e.g., the Q statistic) on that corrected series because using the original variance-changing series distorts these checks.

- Quote snippet:

"This 'corrected cpl series' should then be used to obtain other residual diagnostics, like the Q statistic, which would be distorted if calculated with the series that has variance changes (see Inclan 1991)."

- Citation anchors:
  - page=6, words=825-968, section=Residual diagnostics after ICSS, note=Mentions using the corrected cpl series to obtain residual diagnostics like the Q statistic and the distortion that arises if calculated on the variance-changing series.

**K04: How sensitive are results to alternative specifications or estimators?**

- Category: K) Model fit and diagnostics
- Confidence: medium (0.38826670153547127)
- Retrieval method: local
- Evidence type: Comparative performance across estimators (ICSS, likelihood ratio tests, Bayesian) and sensitivity to sample size and change magnitude
- Data source: Simulation study in Tsay (1994) JASA
- Table/Figure: Table 4
- Answer:

Results are quite sensitive to the estimator/specification used. In simulations, ICSS generally outperforms likelihood ratio tests for longer series and larger variance changes, while detecting small changes in short series (e.g.,  $T=100$ ,  $A=2$ ) is hard, especially when the change is early. Bayesian posterior odds are computationally heavy and not feasible for large  $T$ . CPU time grows with the number of changepoints, indicating sensitivity to the chosen method and problem size; conclusions therefore depend on sample length, change magnitudes, and the estimator used.

- Quote snippet:
 

We conclude that for series of 100 observations, it is hard to detect a small variance change (variance ratio  $A = 2$ )— particularly when it appears at the beginning of the series.
- Citation anchors:
  - page=9, words=0-349, section=4.1 Assessing the Evidence with Respect to the Number of Change Points Using Likelihood Ratio Tests and the Posterior Odds Ratio, note=Hard to detect a small variance change with 100 observations ( $A=2$ ); comparison among ICSS, LR tests, and Bayesian approach.
  - page=9, words=275-624, section=4.1 Assessing the Evidence with Respect to the Number of Change Points Using Likelihood Ratio Tests and the Posterior Odds Ratio, note=Continuation of results comparing estimators; discussion of series with one changepoint and method performance.
  - page=11, words=275-507, section=4.6 CPU Time Requirements, note=Discussion of computational burden and how CPU time depends on number of changepoints and series length, highlighting

method sensitivity.