

# Audit Report: Workflow 1308532de7a9446d813e57129826aa71

## Overview

- Source JSON: reports\workflow-report-1308532de7a9446d813e57129826aa71.json
- Run ID: 1308532de7a9446d813e57129826aa71
- Papers input: papers\Use\_of\_Cumulative\_Sums\_of\_Squares\_for\_Re.pdf
- Started at: 2026-02-15T17:40:01.994514+00:00
- Finished at: 2026-02-15T17:53:00.314525+00:00
- Duration: 0:12:58.320011

## Effective Configuration

- Chat model: gpt-5
- 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: n/a
- enrich: n/a
- econ\_data: fetched
- agentic: completed
- index: skipped (reason: db\_unreachable)
- evaluate: n/a
- report\_store: skipped (reason: db\_unreachable)

## Agentic Summary

- Status: completed
- Main question: What is the key contribution?
- Report question set: structured
- Structured questions generated: 83
- Confidence mean/median: 0.27249828745446014 / 0.268446744281151
- Confidence labels: low=77, medium=6, high=0

## Final Answer

- Primary: Introduces the ICSS procedure for variance change-point detection—built on a statistic  $D_k$  that is a monotone transform of the two-sample variance F statistic  $F_{T-k,k} = \frac{(C_T - C_k)/(T-k)}{C_k/k}$  with  $C_k = \sum_{t=1}^k a_t^2$ —and locates breaks by maximizing  $\max_k |D_k|$ ; under homogeneity, the  $D_k$  process converges to a Brownian bridge, enabling use of fixed critical values (e.g., 5%  $\approx 1.358$ ) for inference [page 3 words 0–349; page 12 words 0–349].

- Supporting:
  - Provides asymptotic moments under no breaks (e.g.,  $E[C_k] = k\sigma^2$ ,  $E[C_k C_T] = k(T+2)\sigma^4$ ) and shows  $E[D_k] \rightarrow 0$  as  $T$  grows, justifying the null behavior used for thresholding [page 11 words 275–507].
  - Benchmarks ICSS against likelihood-ratio and Bayesian posterior-odds approaches via extensive simulations (various  $T$ , change locations, and variance ratios), showing strong detection except when changes are very early and small; performance improves with larger  $T$  and variance ratios [page 7 words 0–349; page 7 words 275–556; page 9 words 550–899].
  - Establishes ICSS as a computationally light alternative with stable, lower CPU cost versus LR (roughly  $O(T^3)$ ), especially beneficial for long series and multiple changes [page 9 words 825–904; page 11 words 275–507].
  - Extends application to AR(1) settings via prewhitened residuals, yielding results close to the i.i.d. case [page 9 words 550–899; page 10 words 0–84].
  - Outlines complementary testing frameworks: sequential LR tests for the number of change points with extreme-value (Gumbel) standardization, and Bayesian posterior odds with per-time change probability, to compare modeling strategies [page 8 words 0–329].

## Sub-Answers

### Sub-answer 1

- Question: ResponseTextConfig(format=ResponseFormatText(type='text'), verbosity='medium')
- Answer:

Could you clarify what you'd like from these excerpts (e.g., statistic definitions, testing procedure, simulation setup/results, or computational complexity)? Key points available here include:

- ICSS statistic and F-test link
  - For a fixed  $k$ , the usual two-sample variance F statistic is  $F_{T-k,k} = \frac{(C_T - C_k)/(T-k)}{C_k/k}$ , where  $C_k = \sum_{t=1}^k a_t^2$ ;  $D_k$  is a monotone function of  $F_{T-k,k}$  and depends on  $k$  only via  $k/T$ . Unlike the F-test (known  $k$ ), ICSS searches  $\max_k |D_k|$  to locate a change point [page 3 words 0–349].
  - Asymptotically (Normal noise), the  $D_k$  process is a Brownian bridge; an example uses the 5% critical value 1.358 with detection at  $k^* = 342$  when  $\max_k |D_k|$  exceeds this boundary [page 3 words 0–349].
- Asymptotics and moments under homogeneity
  - Under  $N_T = 0$ :  $E[C_k] = k\sigma^2$ ,  $E[C_k C_T] = k(T+2)\sigma^4$ , and  $E[D_k] \rightarrow 0$  as  $T$  grows [page 11 words 275–507].
  - Appendix sketch: with  $t_i = a_i^2 - \sigma^2$ , a Donsker-type argument yields convergence to a Brownian bridge  $W^0$  [page 12 words 0–349].
- Likelihood ratio (LR) testing for number of change points
  - Strategy: sequentially test  $H_0 : N_T = n-1$  vs  $H_a : N_T = n$  using  $LR_{n-1,n}$  for  $n = 1, 2, \dots$ ; critical values require simulation because the problem is nonregular

[page 8 words 0–329].

- Standardize  $LR_{n-1,n}$  to reference the extreme value limit with cdf  $F(x) = \exp(-e^{-x})$ ; formulas for standardization are given in Table 3 (referenced) [page 8 words 0–329].
- Bayesian posterior odds (multiple change points)
  - Posterior odds compare  $p(N_T = n \mid a)$  across  $n$ , using  $p(a, K \mid N_T)$  and a per-time prior change probability  $X$ ; details and Table 5 summarize two-change cases [page 8 words 0–329].
- Simulation design (comparisons: ICSS, LR, Bayesian)
  - 1,000 replicates per design;  $T \in \{100, 200, 500\}$ ; one-change locations  $K/T \in \{.25, .5, .75\}$ ; variance ratios  $A \in \{2, 3\}$ . Two-change locations  $(K_1/T, K_2/T) \in \{(.33, .66), (.20, .80), (.58, .80)\}$  with permutations of  $(\sigma_1^2, \sigma_2^2, \sigma_3^2)$  constrained by  $A_1 = \sigma_2^2/\sigma_1^2$ ,  $A_2 = \sigma_3^2/\sigma_2^2$  and symmetry reductions. Data are iid  $N(0, 1)$ ; Fortran implementation; RNG via Bratley–Fox–Schrage; Sparcstation 1+ [page 7 words 0–349; page 7 words 275–556].
- Main empirical findings (one change)
  - With  $T = 100$  and small variance ratio ( $A = 2$ ), detection is hard—especially if the change is near the beginning—with ICSS. With  $T = 200$  or  $A = 3$ , correct identification exceeds 80% when the change is near the middle. ICSS improves with larger  $T$  and  $A$ . The Bayesian approach performs best for  $A = 2$  and is comparable to others for  $A = 3$  [page 9 words 550–899].
- AR(1) residual application
  - Generate  $Y_t = \phi Y_{t-1} + a_t$ ; apply ICSS to residuals  $\varepsilon_t = Y_t - \hat{\phi} Y_{t-1}$ . The percentage with  $N_T = 1$  is close to the iid results in Table 4 (see Table 8) [page 9 words 550–899; page 10 words 0–84].
- Computational burden
  - ICSS was designed to avoid the heavy computation of posterior odds and the need to evaluate functions at all possible change locations. LR uses Baufays–Rasson’s algorithm for maximizing the likelihood but remains heavy for long series and multiple changes [page 9 words 825–904].
  - CPU-time scaling: LR is roughly  $O(T^3)$  and of similar order for  $N_T = 2$  and  $N_T = 3$ ; ICSS’s dependence on  $T$  is the same across values of  $N_T$  (lower and more stable cost) [page 11 words 275–507].

## 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 (0.25738287842443736)

- Retrieval method: local
- Evidence type: text
- Assumption flag: False
- Answer:

To develop and assess a method—based on cumulative sums of squares (ICSS)—for retrospectively detecting changes in variance in time series, including determining the number and locations of variance changepoints, and to compare its performance with likelihood ratio and Bayesian approaches.

- Quote snippet:

The performance of different procedures used to determine the number of variance changes in a series

- Citation anchors:
  - page=7, words=0-120, section=4. SIMULATION EXPERIMENT, note=States the study evaluates procedures to determine the number of variance changes and compares ICSS with likelihood ratio and Bayesian methods.
  - page=4, words=0-40, section=Inclán and Tiao: Retrospective Detection of Changes in Variance, note=Section header emphasizes focus on retrospective detection of variance changes.
  - page=6, words=550-640, section=ICSS algorithm application, note=Describes the ICSS procedure as a systematic search for points of change.

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

- Category: A) Research question / contribution
- Confidence: low (0.27419868133310066)
- Retrieval method: local
- Evidence type: direct textual evidence
- Assumption flag: False
- Answer:

It introduces the ICSS algorithm—a centered cumulative-sum-of-squares, iterative method—to retrospectively detect multiple variance change points efficiently; it develops properties of the centered CSS statistic and shows via simulations that ICSS matches Bayesian/likelihood-ratio performance for moderate samples without their heavy computational burden.

- 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, note=States proposal of ICSS, study of centered CSS properties, and comparable performance without heavy computation.
  - page=2, words=275-624, section=Introduction, note=Describes using centered cumulative sums of squares and an iterative algorithm to find multiple change points.

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

- Category: A) Research question / contribution
- Confidence: low (0.25484008361822535)
- Retrieval method: local
- Evidence type: textual
- Answer:

H0: no change in variance (NT = 0) versus H1: one change in variance (NT = 1).

- Quote snippet:  
testing one change against no change in variance
- Citation anchors:
  - page=3, words=300-430, section=Section 2.1, note=Defines NT=0 (no change) and NT=1 (one change) and states testing one change against no change in variance.
  - page=7, words=310-400, section=Section 4.1, note=Defines LR\_mn for testing H0: NT = m against Ha: NT = n (number of variance changes).

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

- Category: A) Research question / contribution
- Confidence: low (0.26037546805984413)
- Retrieval method: local
- Evidence type: explicit statement
- Data source: Section 4. SIMULATION EXPERIMENT description
- Assumption flag: False
- Answer:

The main outcomes are the number of correct identifications of variance changes and the computational time (CPU time).

- Quote snippet:  
the main way is by the number of 'correct identifications.' Another aspect to consider is the computational requirements of each approach, which can be measured in terms of the CPU time.
- Citation anchors:
  - page=7, words=0-349, section=4. SIMULATION EXPERIMENT, note=Defines primary performance measures: correct identifications and CPU time.

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

- Category: A) Research question / contribution
- Confidence: low (0.2180384234455463)
- Retrieval method: local
- Evidence type: methods/simulation design description
- Data source: Inclán and Tiao (1994), Journal of the American Statistical Association
- Assumption flag: False
- Answer:

Series length ( $T$ ); location(s) of changepoint(s) ( $K$  or  $K_1, K_2$ ); variance ratio(s) ( $A$  for one change;  $A_1 = \tau_1/\tau_0$  and  $A_2 = \tau_2/\tau_1$  for two changes); configuration/permutation of segment variances; and the number of changepoints ( $NT$ ).

- 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  $3$ ). For Part Two, we used the same three series lengths ... three different sets of locations of the change points ... and six possible configurations of variances ... keep the corresponding ratios  $A_1 = T_1 / T_0$  and  $A_2 = T_2 / T_1$ .

- Citation anchors:
  - page=page 7, words=0-349, section=Simulation Experiment (Part One, Part Two), note=Enumerates manipulated factors:  $T, K$  (or  $K_1, K_2$ ),  $A$  (or  $A_1, A_2$ ), and variance configurations.

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

- Category: A) Research question / contribution
- Confidence: low (0.29809896205083153)
- Retrieval method: local
- Evidence type: direct\_text
- Data source: Box and Jenkins (1976)
- Assumption flag: False
- Answer:

Applied example: IBM stock closing prices (first differences of log prices) over May 17, 1961 to November 2, 1962; country not specified in the provided text.

- 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=360-430, section=3.1 Example: IBM Stock Prices, note=States the data series and its time span.

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

- Category: A) Research question / contribution
- Confidence: low (0.23553237245887174)
- Retrieval method: local
- Evidence type: direct textual evidence
- Data source: Inclán & Tiao (1994), Journal of the American Statistical Association
- Assumption flag: False
- Answer:

The Iterated Cumulative Sums of Squares (ICSS) algorithm—an iterative use of the centered cumulative sum of squares ( $D_k$ ) to systematically locate multiple variance-change points.

- 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/Abstract, note=States proposal of ICSS to detect variance changes.
  - page=2, words=275-624, section=Introduction, note=Describes using cumulative sums of squares and iterative algorithm.
  - page=5, words=825-1174, section=ICSS Algorithm description, note=Algorithm labeled and illustrated.

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

- Category: A) Research question / contribution
- Confidence: low (0.242333650293353)
- Retrieval method: local
- Evidence type: textual
- Assumption flag: True
- Assumption notes: Interpreted 'mechanisms' as alternative explanations for observed features/regime shifts: heavy-tailed vs variance-shift modeling, and variance vs AR-parameter changes.
- Answer:

Two alternative mechanisms are discussed: (1) modeling heavy tails in returns (e.g., Pareto/heavy-tailed distributions) versus explaining them by variance shifts (nonstationary variance with step changes), and (2) changes in variance versus changes in the autoregressive parameter as drivers of regime shifts.

- Quote snippet:

A Bayesian analysis that allows both the variance and the autoregressive parameter to change at the same points ... indicates that there are both variance changes and parameter changes.

- Citation anchors:
  - page=2, words=275-624, section=Introduction, note=Contrasts heavy-tailed distributions with a normal model having variance step changes.
  - page=6, words=550-899, section=Section 3.1 Example, note=States that both variance and autoregressive parameter changes can occur.

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

- Category: A) Research question / contribution
- Confidence: low (0.2661350049764556)
- Retrieval method: local
- Evidence type: textual
- Assumption flag: False
- Answer:

The paper does not present policy implications; it focuses on developing and evaluating the ICSS statistical method for retrospectively detecting variance changes and comparing computational costs.

- 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=560-650, section=2. CENTERED CUMULATIVE SUMS OF SQUARES, note=States the paper's main methodological objective (variance change detection).
  - page=6, words=550-640, section=ICSS algorithm application, note=Describes the procedure's systematic search for change points.
  - page=7, words=0-120, section=4. SIMULATION EXPERIMENT, note=Frames performance comparison of procedures rather than policy discussion.
  - page=9, words=720-870, section=4.6 CPU Time Requirements, note=Emphasizes computational burden and efficiency, not policy.

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

- Category: A) Research question / contribution
- Confidence: low (0.2910150337102933)
- Retrieval method: local
- Evidence type: text
- Assumption flag: False
- Assumption notes: The provided text discusses statistical detection accuracy and computation; it does not introduce any welfare framework.
- Answer:

None. The results are statistical—focused on detecting variance changes, their locations, and computational performance—not on welfare or economic surplus.

- Quote snippet:

The performance of different procedures used to determine the number of variance changes in a series can be measured in several ways; the main way is by the number of "correct identifications."

- Citation anchors:
  - page=7, words=0-349, section=4. SIMULATION EXPERIMENT, note=Defines performance as correct identifications and CPU time; no welfare concepts are introduced.
  - page=9, words=550-899, section=4.2–4.6 Results, note=Discusses detection success rates and computational burden; still no welfare interpretation.

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

- Category: A) Research question / contribution
- Confidence: low (0.2484759032229551)
- Retrieval method: local



- Evidence type: Direct textual evidence from results and method discussion
- Assumption flag: False
- Answer:

They note several limitations: (1) heavy computational burden for likelihood/Bayesian methods when searching multiple change points (posterior-odds runs can be infeasible for long series); (2) a masking effect with multiple changepoints, where  $D_k$  tends to highlight only one; (3) ICSS struggles with small variance shifts or early changes, and detecting two monotone-increasing changes often requires large samples; and (4) the location estimate  $k^*$  is biased toward the middle of the series, improving only with larger  $T$  or variance ratios.

- Quote snippet:

it is hard to detect a small variance change (variance ratio  $A = 2$ )—particularly when it appears at the beginning of the series and we are using the ICSS algorithm.

- Citation anchors:
  - page=2, words=550-899, section=Background/Section 2, note=Heavy computational burden when looking for several points of change simultaneously
  - page=4, words=0-349, section=Discussion of  $E[D_k]$  with two changepoints, note=Masking effect when more than one changepoint
  - page=9, words=550-899, section=Results (4.2, 4.3) and CPU time (4.6), note=Hard to detect small/early variance change with ICSS; posterior odds infeasible for  $T=500$  (1,358 CPU hours)
  - page=9, words=275-624, section=4.4 Sampling Distribution of  $k^*$ , note= $k^*$  distribution skewed and biased toward the middle; concentrates with larger  $T$  or variance ratio

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

- Category: A) Research question / contribution
- Confidence: low (0.2866564534027284)
- Retrieval method: local
- Evidence type: paraphrase\_of\_context
- Data source: Inclán & Tiao (1994), Journal of the American Statistical Association
- Assumption flag: False
- Answer:

The paper's novelty is an iterated cumulative sums of squares (ICSS) algorithm based on a centered cumulative sum of squares statistic to systematically and retrospectively detect multiple variance changepoints, achieving performance comparable to Bayesian and likelihood-ratio methods for moderate samples while avoiding their heavy computational burden and mitigating masking when multiple changes exist.

- Quote snippet:

the ICSS algorithm offers results comparable to those obtained by a Bayesian approach or by likelihood ratio tests, without the heavy computational burden required by these approaches.

- Citation anchors:
  - page=2, words=0-200, section=Introduction, note=Introduces ICSS and claims comparable performance without heavy computational burden.

- page=2, words=275-410, section=Introduction, note=States use of cumulative sums of squares and iterative search for multiple changepoints.
- page=2, words=550-750, section=2. Centered Cumulative Sums of Squares, note=Defines the centered and normalized cumulative sum of squares statistic  $D_k$  used by ICSS.
- page=6, words=0-140, section=3.1 Example: IBM Stock Prices / ICSS illustration, note=Shows ICSS eliminates masking when more than one variance change exists.
- page=9, words=825-900, section=4.6 CPU Time Requirements, note=Notes ICSS avoids evaluating functions at all possible changepoint locations, reducing computational burden.

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

- Category: B) Identification strategy / causal design
- Confidence: low (0.27741585874335584)
- Retrieval method: local
- Evidence type: direct description from text
- Assumption flag: False
- Answer:

Detect variance change points by computing the centered cumulative sum of squares  $D_k$ , flagging  $k^*$  where  $|D_k|$  crosses asymptotic boundaries, then iteratively splitting the series at  $k^*$  and reapplying on subsegments with adjacent-point checks to fine-tune until convergence (ICSS).

- Quote snippet:

A solution is an iterative scheme based on successive application of  $D_k$  to pieces of the series, dividing consecutively after a possible changepoint is found.

- Citation anchors:
  - page=5, words=825-1174, section=ICSS algorithm steps, note=Iterative scheme applying  $D_k$  to pieces, dividing after a possible changepoint is found.
  - page=6, words=550-899, section=Fine-tuning and convergence, note=Systematic search from extremes toward middle; fine-tune by checking each point given adjacent ones until convergence.
  - page=3, words=275-624, section= $D_k$  boundaries, note= $D_k$  exhibits excursions beyond specified boundaries derived from asymptotic distribution under constant variance.
  - page=2, words=0-349, section=Introduction, note=Procedure based on an iterated cumulative sums of squares (ICSS) algorithm.

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

- Category: B) Identification strategy / causal design
- Confidence: low (0.26840387625640166)
- Retrieval method: local
- Evidence type: text
- Data source: Inclan & Tiao (1994), Journal of the American Statistical Association
- Assumption flag: False
- Answer:

## Experimental (simulation experiment)

- Quote snippet:

The simulation experiment has two separate parts, the first for one changepoint and the second for two changepoints... For each part, we used 1,000 replicates per design point

- Citation anchors:
  - page=7, words=0-349, section=4. SIMULATION EXPERIMENT, note=Introduces and describes the simulation experiment design.
  - page=7, words=275-556, section=4.1 and setup details, note=Details controlled simulation with replicates and generated data.

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

- Category: B) Identification strategy / causal design
- Confidence: low (0.3252492401531617)
- Retrieval method: local
- Evidence type: textual
- Data source: Inclán & Tiao (1994), Journal of the American Statistical Association
- Assumption flag: False
- Answer:

Not applicable. The paper is methodological and does not use exogenous variation for identification; it detects variance changes via the ICSS cumulative sums of squares approach.

- Quote snippet:

We study the detection of multiple changes of variance in a sequence of independent observations. We propose a procedure to detect variance changes based on an ICSS algorithm.

- Citation anchors:
  - page=2, words=0-349, section=Introduction, note=States the study focuses on detecting multiple variance changes using ICSS, not on causal identification.
  - page=2, words=550-899, section=Centered cumulative sums of squares, note=Defines the centered cumulative sum of squares used for detection.
  - page=3, words=275-624, section=Methodological details, note=Discusses testing for a change in variance and likelihood ratio—again methodological, not exogenous variation.

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

- Category: B) Identification strategy / causal design
- Confidence: low (0.12010050431450066)
- Retrieval method: local
- Evidence type: `inferred_from_context`
- Data source: Inclán & Tiao (1994), Use of Cumulative Sums of Squares for Retrospective Detection of Changes of Variance, JASA 89(427): 913-923.
- Assumption flag: False
- Answer:

Not applicable—the article is a methodological study (ICSS for detecting variance change points) and does not define a treatment or its timing.

- 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=States study focus on detecting variance change points; no treatment is described.
  - page=11, words=0-349, section=Summary/Appendix, note=Recommends ICSS for analyzing series with multiple change points; still no treatment definition.

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

- Category: B) Identification strategy / causal design
- Confidence: low (0.1971543586339832)
- Retrieval method: local
- Evidence type: text
- Data source: <http://links.jstor.org/sici?sici=0162-1459%28199409%2989%3A427%3C913%3AUOCSOS%3E2.0.CO%3B2-H>
- Answer:

There is no traditional control group; performance is compared by applying alternative procedures (likelihood ratio tests and a Bayesian approach) alongside ICSS to the same simulated series for paired comparisons.

- Quote snippet:

The same series is subject to each of the procedures considered to facilitate paired comparisons between procedures.

- Citation anchors:
  - page=7, words=40-140, section=4. SIMULATION EXPERIMENT, note=States the same series is subjected to each procedure for paired comparisons.
  - page=7, words=0-60, section=4. SIMULATION EXPERIMENT, note=Introduces simulation comparing ICSS with likelihood ratio and a Bayesian approach.
  - page=2, words=0-120, section=Abstract/Intro, note=Mentions simulation results comparing ICSS to other approaches.

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

- Category: B) Identification strategy / causal design
- Confidence: low (0.2852559821489955)
- Retrieval method: local
- Evidence type: text\_and\_formula
- Data source: Inclán and Tiao (1994), Journal of the American Statistical Association
- Assumption flag: True

- Assumption notes: No regression is estimated; the method assumes a zero-mean Normal series with piecewise-constant variance and uses the centered cumulative sum of squares  $D_k$  for detection.
- Answer:

Baseline model:  $a_t$  are independent  $\text{Normal}(0, \sigma_t^2)$  with piecewise-constant variance. Estimating equation/statistic:  $D_k = C_k/C_T - k/T$ , where  $C_k = \sum_{i=1}^k a_i^2$ ; inference is based on  $\max_k |D_k|$ .

- Quote snippet:  
Let  $D_k = C_k/C_n - k/n$ , where  $C_k = \sum_{i=1}^k a_i^2$ .
- Citation anchors:
  - page=5, words=60-130, section=Theorem 1, note=Definition of  $D_k$  and  $C_k$
  - page=3, words=120-200, section=Main text, note=Assumption:  $\{a_t\}$  Normally distributed with mean 0 and variances  $\sigma_t^2$
  - page=11, words=430-500, section=Main text, note=Under variance homogeneity,  $E[D_k] \rightarrow 0$
  - page=2, words=860-910, section=Intro/overview, note= $D_k$  oscillates around 0 for homogeneous variance

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

- Category: B) Identification strategy / causal design
- Confidence: low (0.2763758758257676)
- Retrieval method: local
- Evidence type: textual absence within provided context
- Data source: Inclán and Tiao (1994), Journal of the American Statistical Association
- Assumption flag: False
- Answer:

None (not applicable). The context discusses variance-change detection in time series and does not include unit, time, two-way, or higher-dimensional fixed effects.

- Quote snippet:  
This behavior leads to the search for a variance change point via  $\max_k |D_k|$ .
- Citation anchors:
  - page=3, words=550-882, section=Discussion of  $D_k$  for variance change, note=Focuses on  $D_k$  and detecting variance change points; no mention of fixed effects.
  - page=6, words=550-899, section=IBM example and ICSS algorithm, note=Applies ICSS to time series variance changes; no fixed-effects framework referenced.

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

- Category: B) Identification strategy / causal design
- Confidence: low (0.3315274087668142)
- Retrieval method: local
- Evidence type: textual
- Table/Figure: Table 1 and Table 2 (SE formulas described in text)

- Assumption flag: False
- Answer:

Standard errors are Monte Carlo (simulation-based), not robust or clustered. For quantile estimates they use  $SE \approx \sqrt{p(1-p)} / (\sqrt{n} \cdot f)$ , with  $n=10,000$  and  $f$  the estimated density at the  $p$ th quantile; for proportions they use  $SE = \sqrt{\hat{p}(1-\hat{p})/10,000}$ . No clustering level applies.

- Quote snippet:

The standard errors were obtained as  $\sqrt{p(1-p)/n}$ , where  $n = 10,000$  is the number of replicates and  $f$  is the density at the  $p$ th quantile... The standard errors in this table are obtained as  $SE = \sqrt{\hat{p}(1-\hat{p})/10,000}$

- Citation anchors:
  - page=5, words=275-450, section=Discussion of Table 1 quantiles, note=SE formula for quantile estimates from simulations
  - page=5, words=450-624, section=Discussion of Table 2 proportions, note=SE formula for proportion of series meeting criterion

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

- Category: B) Identification strategy / causal design
- Confidence: low (0.2725651917028813)
- Retrieval method: local
- Evidence type: text
- Assumption flag: False
- Answer:

Not specified in the provided context; the document concerns variance-change detection (ICSS, likelihood ratio) rather than causal identification, so none of parallel trends, exclusion restriction, or ignorability is stated.

- Quote snippet:

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

- Citation anchors:
  - page=7, words=0-60, section=4. SIMULATION EXPERIMENT, note=Context frames the paper around detecting variance changes, not causal inference assumptions.
  - page=4, words=275-350, section=2.3 Asymptotic Behavior of  $D_k$  Under Homogeneous Variance, note=Asymptotic discussion of  $D_k$  (Brownian bridge) indicates focus on statistical testing for variance changes.

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

- Category: B) Identification strategy / causal design
- Confidence: low (0.2664735476513488)
- Retrieval method: local
- Evidence type: theoretical + simulation
- Table/Figure: Table 2

- Data source: 10,000 Monte Carlo replicates of independent  $N(0,1)$  series; asymptotic theory (Theorem 1)
- Assumption flag: True
- Assumption notes: Interprets the 'identifying assumption' as (i) the null distribution of  $D_k$  being Brownian-bridge and (ii)  $\max|D_k|$  locating the true variance-change point.
- Answer:

They provide both theory and simulations: Theorem 1 shows the centered cumulative sum of squares converges to a Brownian bridge under homoscedastic Normal noise, supporting use of Brownian-bridge critical values; and Monte Carlo evidence (10,000  $N(0,1)$  series) shows the empirical probability that  $\max|D_k|$  is below the 1.358 critical value is  $\approx 95\%$ . They also report that the mode of  $k^*$  (where  $\max|D_k|$  occurs) is at the true variance-change point and concentrates there as  $T$  or the variance ratio increases.

- Quote snippet:

What makes the ICSS algorithm work well is that the mode of  $k^*$  is exactly at the point where the change in variance occurs.

- Citation anchors:
  - page=5, words=140-310, section=Theorem 1, note=Asymptotic null:  $D_k$  converges to a Brownian bridge ( $W_0$ ).
  - page=5, words=0-120, section=Table 2, note=Monte Carlo: percentage with  $\max|D_k| < 1.358$  near 95% (10,000  $N(0,1)$  replicates).
  - page=9, words=360-500, section=4.4, note=Distribution of  $k^*$ : mode exactly at the true change point; concentration increases with  $T$  and variance ratio.

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

- Category: B) Identification strategy / causal design
- Confidence: medium (0.38895839421501494)
- Retrieval method: local
- Evidence type: textual description
- Data source: Inclán and Tiao (1994), JASA 89(427): 913–923
- Assumption flag: True
- Assumption notes: Within the provided excerpts, there is no mention of event-study or pre-trend tests; content centers on ICSS variance change detection, simulations, and an IBM stock example.
- Answer:

No. The article does not report event-study or pre-trend tests; it focuses on detecting variance changes using the ICSS algorithm and compares it with likelihood ratio and Bayesian 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=0-349, section=Introduction/scope, note=States the paper's focus on ICSS for variance change detection.

- page=2, words=275-624, section=Method overview, note=Describes cumulative sums of squares approach and paper organization; no mention of event-study or pre-trends.
- page=7, words=0-349, section=Simulation experiment, note=Details simulations comparing ICSS to likelihood ratio and Bayesian approaches; no event-study or pre-trend tests.
- page=6, words=275-624, section=Empirical example, note=IBM stock price example analyzed via ICSS; no event-study framework discussed.

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

- Category: B) Identification strategy / causal design
- Confidence: low (0.1772272201535017)
- Retrieval method: local
- Evidence type: text
- Data source: Provided context excerpts from Inclán & Tiao on ICSS and variance change detection
- Assumption flag: True
- Assumption notes: No IV terminology or regression framework appears in the provided excerpts; inferred that instruments and first-stage definitions are not applicable to this method-focused variance-change study.
- Answer:

Not applicable. The provided context describes the ICSS algorithm for detecting variance changes in time series and does not use instrumental variables (IV). Therefore, there are no instruments or first-stage regressions to define in this study.

- Quote snippet:  
Iterated Cumulative Sums of Squares (ICSS) Algorithm
- Citation anchors:
  - page=5, words=900-1120, section=3. Detections of Multiple Changes: The Iterated Cumulative Sums of Squares Algorithm, note=Introduces and details the ICSS algorithm; no mention of IV or instruments.
  - page=6, words=0-120, section=ICSS discussion and convergence, note=Describes algorithmic convergence and usage; context is variance-change diagnostics, not IV.
  - page=6, words=550-880, section=Procedure refinement and example, note=Focuses on systematic search and diagnostics for variance changes; no IV framework.

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

- Category: B) Identification strategy / causal design
- Confidence: low (0.3120605828645241)
- Retrieval method: local
- Evidence type: textual
- Data source: Inclán and Tiao (1994), Journal of the American Statistical Association
- Assumption flag: False
- Answer:

Not available in the provided context. The text does not report any first-stage regression results



(e.g., IV first-stage F-statistic, partial  $R^2$ , or other relevance evidence). The only F-statistic discussed pertains to testing equality of variances in variance-change detection, not instrument relevance.

- 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...  $F_{\{T-k, k\}} = ((CT - C_k)/(T - k))/(C_k/k)$ .

- Citation anchors:
  - page=3, words=0-349, section=2.1 Relationship of  $D_k$  to the F Statistic and the Likelihood Ratio, note=F-stat here is for equality of variances, not IV first-stage strength.

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

- Category: B) Identification strategy / causal design
- Confidence: low (0.3424343411773066)
- Retrieval method: local
- Evidence type: text
- Table/Figure: Figure 4
- Data source: Inclán and Tiao (1994), Journal of the American Statistical Association
- Assumption flag: False
- Answer:

Not applicable. The study uses ICSS/change-point detection for variance shifts, not a regression discontinuity design; thus there is no running variable, no cutoff, and no bandwidth choice.

- Quote snippet:

we will be looking for  $\max_k |D_k|$  to determine the location of the change point.

- Citation anchors:
  - page=3, words=0-349, section=Section 2.1, note=Method looks for  $\max |D_k|$  to locate a change point, not an RDD setup.
  - page=3, words=275-624, section=Sections 2.1–2.2, note=Focus on testing variance changes and using  $D_k$ /CUSUM of squares.
  - page=6, words=275-624, section=Section 3.1, note=ICSS algorithm performs a systematic search for points of change.
  - page=9, words=275-624, section=Section 4.4, note=Discussion of  $k^*$  where  $\max |D_k|$  is attained; confirms change-point framework.

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

- Category: B) Identification strategy / causal design
- Confidence: medium (0.41074100604643277)
- Retrieval method: local
- Evidence type: text
- Table/Figure: Table 1
- Data source: Journal of the American Statistical Association, September 1994 (Inclán & Tiao)
- Assumption flag: False

- Assumption notes: The provided context covers variance change-point detection (ICSS), not difference-in-differences.
- Answer:

Not applicable. The study is not a DiD design; it detects variance change points using the ICSS (Iterative Cumulative Sum of Squares) algorithm, estimating changepoints via  $\max_k |D_k|$  with Brownian-bridge critical values (Table 1). There is no staggered adoption or DiD estimator.

- Quote snippet:

This behavior leads to the search for a variance change point via  $\max_k |D_k|$ . Let  $k^*$  be the value of  $k$  at which  $\max_k |D_k|$  is attained.

- Citation anchors:
  - page=3, words=550-882, section=Method: change-point search, note=Defines search via  $\max_k |D_k|$  to locate variance change points.
  - page=5, words=0-349, section=Asymptotics and boundaries, note= $D_k \rightarrow$  Brownian bridge; Table 1 gives quantiles for decision thresholds.
  - page=5, words=1100-1198, section=Algorithm steps, note=Uses boundary  $D^*$  from Table 1 to declare changepoints.
  - page=6, words=0-349, section=ICSS algorithm usage, note=Describes the ICSS algorithm and its purpose in detecting variance changes.

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

- Category: C) Data, sample, and measurement
- Confidence: low (0.26690114886370353)
- Retrieval method: local
- Evidence type: direct\_quote\_and\_summary
- Data source: Box and Jenkins (1976); author-generated simulated series and AR(1) residuals
- Assumption flag: False
- Answer:

The paper uses: (1) a real dataset of IBM stock closing prices (analyzed as first differences of log prices) from May 17, 1961 to November 2, 1962, as reported by Box and Jenkins (1976); and (2) simulated time series (and AR(1) residuals) generated by the authors for the simulation experiments.

- Quote snippet:

The series analyzed is the first difference of the logarithm of the IBM stock closing prices ... as reported by Box and Jenkins (1976).

- Citation anchors:
  - page=6, words=275-624, section=3.1 Example: IBM Stock Prices, note=Describes the IBM stock prices dataset and cites Box and Jenkins (1976).
  - page=7, words=0-349, section=4. SIMULATION EXPERIMENT, note=States that simulation experiments were conducted with generated series.
  - page=9, words=550-899, section=4.5 The ICSS Algorithm Applied to Residuals, note=Details application to AR(1) residuals from simulated series.

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

- Category: C) Data, sample, and measurement
- Confidence: low (0.24640099422268022)
- Retrieval method: local
- Evidence type: text
- Data source: IBM stock closing prices (Box and Jenkins, 1976)
- Assumption flag: True
- Assumption notes: Unit inferred from the empirical example focused on IBM, a single firm; other parts use simulated series.
- Answer:

firm

- 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=Empirical example analyzes IBM stock closing prices (a single company's series).

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

- Category: C) Data, sample, and measurement
- Confidence: low (0.268446744281151)
- Retrieval method: local
- Evidence type: article\_text
- Data source: Box and Jenkins (1976) IBM stock closing prices
- Assumption flag: False
- Assumption notes: Geographic coverage is not specified in the provided context.
- Answer:

Sample period: May 17, 1961 to November 2, 1962. Geographic coverage: Not specified in the provided context (IBM stock closing prices dataset).

- 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=States the IBM stock series dates: May 17, 1961 to November 2, 1962.

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

- Category: C) Data, sample, and measurement
- Confidence: low (0.23034433773356772)
- Retrieval method: local

- Evidence type: textual description (methods)
- Assumption flag: True
- Assumption notes: Interprets 'sample' as the simulated series used in the study (not an observational dataset).
- Answer:

Simulated time-series only. Inclusion criteria: (1) innovations assumed Normal with mean 0 and piecewise-constant variance; (2) series lengths  $T$  in  $\{100, 200, 500\}$ ; (3) one-change cases use  $K1$  at  $\{0.25T, 0.50T, 0.75T\}$  with variance ratio  $A$  in  $\{2, 3\}$ ; (4) two-change cases use  $(K1, K2)$  in  $\{(.33T, .66T), (.20T, .80T), (.58T, .80T)\}$  with base variance set to 1 and varying ratios  $A1 = \sigma_1^2/\sigma_0^2$  and  $A2 = \sigma_2^2/\sigma_1^2$ ; symmetric variance permutations excluded; (5) the same generated series are subjected to ICSS, likelihood ratio, and Bayesian procedures for paired comparisons; (6) additional runs apply ICSS to residuals from AR(1) models fitted to the same simulated series.

- Quote snippet:

The same series is subject to each of the procedures considered to facilitate paired comparisons between procedures.

- Citation anchors:
  - page=7, words=0-349, section=4. SIMULATION EXPERIMENT, note=Defines  $T$ , change-point locations, variance ratios; same series used across procedures; variance permutations and base variance constraint.
  - page=3, words=0-349, section=2.1 Relationship of  $D_k$  to the F Statistic and the Likelihood Ratio, note=States the Normal, mean-0 assumption for  $\{a_t\}$ .
  - page=9, words=550-899, section=4.5 The ICSS Algorithm Applied to Residuals, note=ICSS applied to residuals from AR(1) fitted to the same series.

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

- Category: C) Data, sample, and measurement
- Confidence: low (0.29216650868207844)
- Retrieval method: local
- Evidence type: text
- Table/Figure: Figure 1
- Assumption flag: True
- Assumption notes: Interpreted 'main analysis' as the primary empirical series analyzed around Figure 1/Section 3, where indices run  $k=1, \dots, 700$ .
- Answer:

700

- Quote snippet:

$D_k$  plot  $k=1, \dots, 700$

- Citation anchors:
  - page=3, words=200-230, section=Figure 1 caption/context, note=Shows  $D_k$  plot indexed  $k=1, \dots, 700$ , indicating series length 700.
  - page=6, words=820-870, section=ICSS algorithm discussion, note=Refers to  $D_k(a[398:700])$ , confirming  $T$  ends at 700.

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

- Category: C) Data, sample, and measurement
- Confidence: low (0.2937680977760846)
- Retrieval method: local
- Evidence type: Direct quotes and contextual summary from article text
- Table/Figure: Table 1 (Empirical and Asymptotic Quantiles of  $\max |D_k|$ )
- Data source: Inclán and Tiao (1994), Journal of the American Statistical Association
- Answer:

The key outcome is the variance-change statistic  $D_k$ , defined as a centered and normalized cumulative sum of squares of the series. Evidence for a change is assessed by  $\max_k |D_k|$  exceeding asymptotic Brownian-bridge boundaries (e.g., Table 1), with the change-point estimated at  $k^*$  where the maximum occurs. Transformations used include: (1) the intrinsic centering/normalization in  $D_k$ ; (2) applying ICSS to transformed data such as the first difference of log prices (IBM example); and (3) applying ICSS to residuals from an AR(1) fit to remove autocorrelation.

- Quote snippet:  
the centered (and normalized) cumulative sum of squares.
- Citation anchors:
  - page=2, words=825-916, section=Definition/overview, note=Defines  $D_k$  as the centered and normalized cumulative sum of squares; behavior around 0 for homogenous variance.
  - page=3, words=550-882, section=Detection rule and asymptotics, note=Decision via  $\max |D_k|$  vs boundary; Brownian-bridge behavior; Table 1 quantiles.
  - page=6, words=275-624, section=Example (IBM), note=Uses first difference of the logarithm of IBM stock prices.
  - page=9, words=550-899, section=Residuals application, note=Applies ICSS to residuals from an AR(1) model.

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

- Category: C) Data, sample, and measurement
- Confidence: low (0.2466897861561292)
- Retrieval method: local
- Evidence type: textual\_evidence
- Data source: Methodological simulations with independent  $N(0,1)$  series; no treatment/exposure variables
- Assumption flag: False
- Answer:

Not applicable. This is a methodological/statistical paper; no treatment or exposure is measured. The study analyzes sequences of independent observations (often simulated  $N(0,1)$  series) to detect variance changes. Constructed variables/statistics include the centered cumulative sum of squares  $D_k = C_k/C_n - k/n$  (with  $C_k$  the cumulative sum of squares), its linkage to the F statistic, and likelihood ratio statistics standardized to an extreme value distribution within the ICSS framework.

- Quote snippet:

Let  $D_k = C_k/C_n - k/n$ , where  $C_k = \sum_{i=1}^k C_i$ .

- Citation anchors:
  - page=2, words=0-349, section=Introduction, note=States focus on detecting multiple changes of variance in a sequence of independent observations; no treatment/exposure construct.
  - page=5, words=0-349, section=Theorem 1 and definitions, note=Defines  $D_k = C_k/C_n - k/n$  with  $C_k$  as cumulative sum of squares; also notes simulations with independent  $N(0,1)$  observations.
  - page=3, words=0-349, section=Relationship of  $D_k$  to the F Statistic, note=Explains  $D_k$  as a function of the usual F statistic, reinforcing that constructed statistics are used.
  - page=8, words=0-329, section=Likelihood ratio standardization, note=Describes standardizing LR statistics to an extreme value distribution; another constructed diagnostic.

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

- Category: C) Data, sample, and measurement
- Confidence: low (0.2842639867015846)
- Retrieval method: local
- Evidence type: textual
- Assumption flag: False
- Answer:

No. The method analyzes sequences of observations (or residuals) to detect variance shifts and does not include covariates or control variables. None are always included.

- 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=States the focus on variance changes in a sequence of independent observations, not on covariates.
  - page=2, words=550-899, section=2. Centered Cumulative Sums of Squares, note=Defines the setup using uncorrelated mean-zero variables and cumulative sums of squares; no covariates.
  - page=3, words=275-624, section=Methods/Likelihood formulation, note=Assumes  $\{a_t\}$  Normally distributed with mean 0 and time-varying variances; focuses on variance change testing.
  - page=10, words=0-84, section=Results (tables), note=Mentions application to AR(1) residuals, reinforcing that inputs are residual series, not covariates.

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

- Category: C) Data, sample, and measurement
- Confidence: low (0.24359517343422313)
- Retrieval method: local
- Evidence type: not specified in context

- Data source: Inclán and Tiao (1994), Journal of the American Statistical Association (provided context)
- Assumption flag: False
- Answer:

Not discussed in the provided context; there is no description of handling missing data (e.g., dropping, imputation, or weighting).

- Quote snippet:

It is important to look at the plots, because a big outlier would create a significant peak in the Dk plot... advisable to complement the search with a procedure for outlier detection.

- Citation anchors:
  - page=6, words=0-349, section=ICSS algorithm discussion / 3.1 Example: IBM Stock Prices, note=Context discusses outliers and ICSS procedure but gives no guidance on missing data handling.

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

- Category: C) Data, sample, and measurement
- Confidence: low (0.28048630291205834)
- Retrieval method: local
- Evidence type: Inference from absence in provided methodological description
- Data source: Inclán and Tiao (1994), Journal of the American Statistical Association: Use of Cumulative Sums of Squares for Retrospective Detection of Changes of Variance
- Assumption flag: True
- Assumption notes: Weights are not mentioned in the provided excerpts; conclusion is based on the absence of any reference to survey/propensity weighting and the focus on unweighted cumulative sums of squares.
- Answer:

No. The methods and simulations described rely on unweighted cumulative sums of squares (the ICSS algorithm) and related likelihood/Bayesian comparisons; survey or propensity weights are not mentioned or used.

- 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, note=Describes ICSS approach; no mention of any weighting scheme.
  - page=2, words=825-916, section=Introduction/Method, note=Defines centered/normalized cumulative sum of squares Dk; no weights referenced.
  - page=7, words=0-349, section=Simulation experiment setup, note=Details simulation design and procedures compared; no use of survey or propensity weights.

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

- Category: C) Data, sample, and measurement
- Confidence: low (0.20254398734717888)
- Retrieval method: local
- Evidence type: direct\_text
- Table/Figure: Table 1 (note)
- Data source: IBM stock prices (Box and Jenkins, 1976); simulated  $N(0,1)$  series
- Assumption flag: False
- Answer:

No. The analyses use either a single time series (IBM stock prices from Box and Jenkins, 1976) or simulated independent  $N(0,1)$  series. There is no cross-source linkage or merging performed.

- Quote snippet:

The series analyzed is the first difference of the logarithm of the IBM stock closing prices

- Citation anchors:
  - page=6, words=275-624, section=3.1 Example: IBM Stock Prices, note=Describes use of a single real-world series (IBM stock prices) from a cited source.
  - page=3, words=550-882, section=Table 1 note, note=States results are from simulations of independent  $N(0,1)$  series.

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

- Category: C) Data, sample, and measurement
- Confidence: medium (0.3529576886460369)
- Retrieval method: local
- Evidence type: Tables and accompanying text in the provided excerpts
- Table/Figure: Table 1; Table 2; Table 7; Table 8
- Data source: Inclán and Tiao (1994), Journal of the American Statistical Association
- Assumption flag: True
- Assumption notes: Interpreted the 'main variables' as  $k^*$ ,  $\max|Dk|$ , and ICSS detection proportions ( $NT=1$ ). If a different set of variables was intended, the reported summaries may differ.
- Answer:

They report: (1) for  $k^*$ : mode, mean, and standard deviation; (2) for  $\max|Dk|$ : empirical and asymptotic quantiles with standard errors; and (3) detection rates as percentages of series meeting criteria (e.g.,  $\max|Dk| < 1.358$ ) with standard errors, plus percentages of  $NT = 1$  from ICSS runs.

- Quote snippet:

Mode of  $k^*$  Mean of  $k^*$  SD of  $k^*$

- Citation anchors:
  - page=3, words=550-882, section=Table 1, note=Empirical and asymptotic quantiles of  $\max|Dk|$  with SEs are presented.
  - page=5, words=0-349, section=Table 2, note=Percentages of series with  $\max|Dk| < 1.358$  and their standard errors.



- page=10, words=0-84, section=Table 7 and Table 8, note=Summaries for  $k^*$  (Mode, Mean, SD) and percentage of NT = 1.

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

- Category: C) Data, sample, and measurement
- Confidence: low (0.289297076290918)
- Retrieval method: local
- Evidence type: textual\_figure\_reference
- Table/Figure: Figure 1; Figure 2; Figure 3; Figure 4
- Assumption flag: False
- Answer:

Yes. The paper includes descriptive figures (e.g., Figure 1) that show baseline patterns under homogeneous variance and illustrate expected  $C_k$  and  $D_k$  behaviors, with additional figures for change-point scenarios.

- 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=275-624, section=Figure 1 discussion, note=Mentions Figure 1 and example series used to illustrate baseline and changed-variance patterns.
  - page=3, words=550-882, section=Main text on baseline behavior, note=Describes baseline pattern for homogeneous variance (straight line for  $C_k$ , characteristic  $D_k$  behavior).
  - page=4, words=550-869, section=Figure 3 caption/discussion, note=References Figure 3 showing expected  $D_k$  patterns with specified change points.
  - page=6, words=275-624, section=Figure 4 caption/discussion, note=References Figure 4 sequence of  $D_k$  functions, further illustrating patterns.

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

- Category: D) Results, magnitudes, heterogeneity, robustness
- Confidence: low (0.276356738767391)
- Retrieval method: local
- Evidence type: text
- Answer:

Not reported in the provided context; no main-effect estimate (sign or magnitude) is given.

- Quote snippet:  
Note that although  $L_{Ro,}$  and  $\max_k |D_k|$  are related, they are not equivalent for finite  $T$ .
- Citation anchors:
  - page=3, words=0-349, section=2.1, note=Context discusses  $D_k$ , F-statistic relation, and critical value, but no main-effect coefficient is presented.
  - page=4, words=275-624, section=2.3, note=Explains properties of  $\max_k |D_k|$  and likelihood ratio but does not report any main-effect estimate.

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

- Category: D) Results, magnitudes, heterogeneity, robustness
- Confidence: low (0.22501087583053234)
- Retrieval method: local
- Evidence type: text
- Assumption flag: True
- Assumption notes: Interpreted 'preferred specification' as the preferred method/approach for variance-change detection in this context; the article recommends ICSS for long series with multiple changepoints.
- Answer:

Use the ICSS algorithm for detecting variance changepoints—especially in long series with multiple changes—because it has  $O(T)$  computational cost and, in simulations, performs as well or better than likelihood ratio tests as  $T$  grows, while Bayesian posterior odds are computationally burdensome.

- 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=Explicit recommendation to use ICSS for long series with multiple change points
  - page=11, words=275-507, section=4.6 CPU Time Requirements, note=CPU time order: LR roughly  $O(T^2)$  for  $NT \geq 2$ ; ICSS dependence consistent across  $NT$
  - page=9, words=0-349, section=Results of Part Two (Two Changepoints), note=ICSS often better than likelihood ratio tests with  $\geq 200$  observations
  - page=9, words=550-899, section=4.6 CPU Time Requirements (lead-in), note=Heavy computation for LR/Bayesian; ICSS avoids intensive function evaluations

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

- Category: D) Results, magnitudes, heterogeneity, robustness
- Confidence: low (0.30583567056319455)
- Retrieval method: local
- Evidence type: textual
- Assumption flag: False
- Answer:

Not specified. The context describes a statistical method (ICSS) for detecting variance change points and reports thresholds/quantiles, but it does not quantify an economic effect in percent change, elasticity, or dollars.

- Quote snippet:  
This behavior leads to the search for a variance change point via  $\max_k |D_k|$ .
- Citation anchors:
  - page=3, words=560-700, section=2. CENTERED CUMULATIVE SUMS OF SQUARES,

note=Describes searching for variance change points via  $\max|Dk|$ ; no economic magnitude reported.

- page=6, words=275-420, section=3.1 Example: IBM Stock Prices, note=Example uses stock returns and reports change points only, not economic effect sizes.
- page=5, words=0-120, section=Asymptotic behavior/quantiles, note=Presents asymptotic results and percentages for  $\max|Dk|$ ; still methodological, not economic magnitudes.

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

- Category: D) Results, magnitudes, heterogeneity, robustness
- Confidence: low (0.33810902789115976)
- Retrieval method: local
- Evidence type: Simulation and empirical corroboration
- Table/Figure: Table 2; Table 4; Figure 4
- Data source: Simulated  $N(0,1)$  series; IBM stock prices (1961–1962)
- Assumption flag: False
- Answer:

Key checks: (1) Monte Carlo simulations varying series length, changepoint location, and variance ratios for both one and two changepoints; results show correct identification improves with longer  $T$  and larger variance ratios, exceeding 80% for  $T \geq 200$  or  $A=3$  when the change is central, while  $A=2$  with  $T=100$  and early change is harder. (2) Applying ICSS to  $AR(1)$  residuals yields detection rates very close to the baseline simulation, indicating robustness to modeling residuals. (3) Real-data corroboration: on IBM returns, ICSS changepoints match maximum-likelihood estimates and Bayesian posterior modes at the same points. (4) Type I error check: empirical 95th quantile of  $\max|Dk|$  is below the asymptotic 1.358, implying conservative size under the null. (5) Localization property: the mode of  $k^*$  is at the true changepoint and concentrates with larger  $T$  or variance ratio. Overall, the main results survive these robustness checks.

- Quote snippet:

the percentage of series for which  $NT = 1$  is very close to the figures presented in Table 4.

- Citation anchors:
  - page=7, words=0-349, section=Simulation design ( $T$ ,  $K/T$ ,  $A$ ; one and two changepoints), note=Design spans  $T=100, 200, 500$ ;  $K/T=.25, .50, .75$ ;  $A=2, 3$ ; and multiple two-changepoint configurations.
  - page=9, words=550-899, section=Simulation results summary, note=Performance patterns: harder at  $T=100$ ,  $A=2$  (early change);  $>80\%$  when  $T \geq 200$  or  $A=3$  in middle; improvement with  $T$  and  $A$ .
  - page=9, words=550-899, section=ICSS applied to  $AR(1)$  residuals, note=Residuals check shows  $NT=1$  rates very close to Table 4.
  - page=6, words=550-899, section=IBM example; cross-method corroboration, note=ICSS points agree with ML; Bayesian posterior modes at same points.
  - page=5, words=275-624, section=Type I error/quantiles, note=Empirical 95th quantile  $<$  asymptotic 1.358, implying smaller empirical type I error.
  - page=9, words=275-624, section=Distribution of  $k^*$ , note=Mode at true change; concentration increases with  $T$  and variance ratio.

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

- Category: D) Results, magnitudes, heterogeneity, robustness
- Confidence: low (0.20050309178168896)
- Retrieval method: local
- Evidence type: simulation (Table 2) and asymptotic theory
- Table/Figure: Table 2
- Data source: 10,000 simulated IID  $N(0,1)$  series per  $T$  under no change
- Assumption flag: False
- Answer:

Placebo (null) tests simulate IID  $N(0,1)$  series with no variance changes and compute  $\max_k |D_k|$  against the asymptotic boundary 1.358. Across  $T=100,200,300,400,500$  (10,000 reps each), 97.13%, 96.51%, 96.31%, 96.07%, and 95.53% of series stay below 1.358, showing false positives are near the intended ~5% and consistent with Brownian-bridge asymptotics.

- Quote snippet:  
NOTE: 10,000 replicates, series of  $T$  independent  $N(0, 1)$  observations.
- Citation anchors:
  - page=5, words=0-349, section=Table 2: Null simulation results, note=Percent of series with  $\max |D_k| < 1.358$  for  $T=100-500$ ; 10,000 replicates of IID  $N(0,1)$
  - page=5, words=0-349, section=Theorem 1, note=Asymptotic result:  $D_k$  converges to a Brownian bridge, basis for boundary 1.358
  - page=3, words=0-349, section=Figure discussion / critical value, note=Asymptotic critical value  $D_{.35} = 1.358$  and boundary interpretation for  $|D_k|$

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

- Category: D) Results, magnitudes, heterogeneity, robustness
- Confidence: low (0.28755518983303757)
- Retrieval method: local
- Evidence type: text
- Data source: Inclán and Tiao (1994), JASA, provided excerpts
- Assumption flag: False
- Answer:

Not specified in the provided context; no falsification (unaffected) outcomes are discussed or tested.

- 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=300-460, section=4. SIMULATION EXPERIMENT, note=Context focuses on simulation to detect variance changes; no mention of falsification or unaffected outcomes.

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

- Category: D) Results, magnitudes, heterogeneity, robustness
- Confidence: low (0.32051980795754204)
- Retrieval method: local
- Evidence type: text
- Assumption flag: False
- Answer:

None. The provided context reports no heterogeneity results 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

- Citation anchors:
  - page=7, words=275-556, section=4. Simulation experiment setup and focus, note=Describes simulations focused on detecting variance changepoints, not subgroup heterogeneity.
  - page=3, words=550-882, section=Method description and Table 1 context, note=Focus on variance change detection via  $D_k$  and empirical/asymptotic quantiles.
  - page=9, words=275-624, section=Posterior odds and sampling distribution of  $k^*$ , note=Discusses number of variance changes and distribution of the estimated changepoint.

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

- Category: D) Results, magnitudes, heterogeneity, robustness
- Confidence: low (0.2055530935249787)
- Retrieval method: local
- Evidence type: textual\_evidence\_and\_simulation\_results
- Data source: Inclán and Tiao (1994), JASA – provided excerpts
- Assumption flag: False
- Answer:

They perform three mechanisms to test for variance changes: (1) the ICSS  $\max|D_k|$  threshold test applied iteratively to segments (if  $\max|D_k| \leq D^*$ , no change; if  $> D^*$ , declare a changepoint and continue), (2) sequential likelihood-ratio tests  $LR_{\{n-1, n\}}$  to assess the number of changepoints, standardized to an extreme-value reference distribution, and (3) Bayesian posterior-odds  $K_{\{n, n-1\}}$  to select the number of changepoints up to a specified maximum. Implications:  $\max|D_k|$  is biased toward mid-series but its mode occurs at the true changepoint and concentrates there as  $T$  or the variance ratio increases; small changes ( $A=2$ ) in short series ( $T=100$ ), especially early in the series, are hard to detect with ICSS, while for  $T \geq 200$  or  $A=3$ , correct identification exceeds ~80% when the change is near the middle; the Bayesian approach performs best for  $A=2$ , and ICSS improves with larger  $T$  and  $A$ . They also note F-test power features for variance differences.

- Quote snippet:

If  $M(t1:T) < D^*$ , there is no evidence of variance changes... If  $M(t1:t2) > D^*$ , then we have a new point of change

- Citation anchors:
  - page=5, words=0-349, section=ICSS algorithm steps, note=Decision rule for ICSS: compare  $\max|Dk|$  to  $D^*$ ; if below, stop; if above, declare a change and iterate.
  - page=8, words=0-329, section=4.1 Likelihood ratios and posterior odds, note= $LR_{\{n-1,n\}}$  tests, standardization to extreme value distribution; posterior odds framework.
  - page=7, words=275-556, section=4.1 Likelihood ratios and posterior odds, note=Definition/use of  $LR_{\{n-1,n\}}$  for testing numbers of changepoints.
  - page=9, words=275-624, section=Distribution of  $k^*$  and implications, note= $k^*$  is biased toward middle; mode at true changepoint; concentration increases with  $T$  or variance ratio.
  - page=9, words=550-899, section=4.2 Results (one changepoint), note=Performance implications: hard to detect  $A=2$  with  $T=100$ ;  $\geq 80\%$  correct for  $T \geq 200$  or  $A=3$  when change is in the middle; Bayesian best for  $A=2$ ; ICSS improves with  $T$  and  $A$ .
  - page=4, words=0-349, section=Discussion of tests, note=F-test power characteristics for variance differences and interaction with degrees of freedom.

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

- Category: D) Results, magnitudes, heterogeneity, robustness
- Confidence: medium (0.4155334148582157)
- Retrieval method: local
- Evidence type: simulation results and asymptotic theory from the provided text
- Table/Figure: Tables 2, 4, 6–7 (as referenced in text)
- Assumption flag: True
- Assumption notes: No bandwidths or control-variable choices are discussed; sensitivity is interpreted in terms of sample size, variance ratio, change-point location, and configuration as reported.
- Answer:

Sensitivity is driven by sample length ( $T$ ), variance ratio ( $A$ ), and where/how variance changes occur—bandwidths/controls are not discussed. Specifically: (1) With  $T=100$  and small  $A=2$ , detection is hard, especially for early changes; with  $T \geq 200$  or  $A=3$  and a mid-series change, correct identifications exceed  $\sim 80\%$ , and performance improves with larger  $T$  and  $A$ . (2) With two change-points, monotone variance increases are most difficult;  $T \approx 500$  is often needed to get  $NT=2$  more than half the time; best results occur when changes are equally spaced and the largest variance is in the middle. (3) Finite-sample critical values are conservative: the 95th quantile of  $\max|Dk|$  is below the asymptotic 1.358, lowering type I error. (4) The location estimator  $k^*$  is skewed toward the middle but has its mode at the true change and concentrates there as  $T$  or  $A$  increase.

- Quote snippet:
 

Once we have 200 observations or the variance ratio is larger ( $A = 3$ ), the correct identifications occur more than 80% of the time
- Citation anchors:

- page=9, words=550-899, section=Simulation results (one changepoint), note=Detection depends on T, A, and position; >80% with  $T \geq 200$  or  $A=3$  when change is in the middle; harder with  $T=100$  and  $A=2$ , especially early changes.
- page=9, words=0-349, section=Simulation results (two changepoints), note=Monotone increases are hardest; need large T (e.g., 500) to get  $NT=2$  more than half the time; best when changes are equally spaced and large variance is in the middle.
- page=9, words=275-624, section=Sampling distribution of  $k^*$ , note= $k^*$  is biased toward the center but its mode is at the true change; concentration improves with larger T or A.
- page=5, words=275-624, section=Finite-sample quantiles, note=95th quantile of  $\max|Dk|$  is below asymptotic 1.358, implying smaller empirical type I error (conservative critical value).

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

- Category: D) Results, magnitudes, heterogeneity, robustness
- Confidence: low (0.204945056307743)
- Retrieval method: local
- Evidence type: Paraphrase with direct citations
- Table/Figure: Tables 4–8
- Assumption flag: True
- Assumption notes: The explicit conclusion bullets were not provided; key takeaways are summarized from the provided conclusion and results excerpts.
- Answer:
- ICSS is recommended for analyzing long series with multiple variance change points, based on results and CPU-time evidence.
- Computationally efficient: iterative search averages  $O(T)$  operations and avoids exhaustive evaluation; convergence typically in few iterations.
- One change point: with  $T=100$  and small change ( $A=2$ ), detection is hard—especially early in the series; with  $T \geq 200$  or larger change ( $A=3$ ) near the middle, correct identification exceeds ~80%. Bayesian is best for  $A=2$  and comparable for  $A=3$ .
- Two change points: with  $T \geq 200$ , ICSS generally outperforms likelihood ratio tests; best when change points are evenly spaced (.33T, .67T) and when the largest variance is in the middle ( $A_1, A_2$ )=(4,2). Monotone-increasing variances require larger T (e.g., 500) to identify both changes >50% of the time.
- Posterior-odds approach is computationally prohibitive (e.g., ~1,358 CPU hours for  $T=500$ ), limiting practicality.
- Location statistic  $k^*$ : distribution is skewed and biased toward the middle, but the mode aligns with the true change; concentration around the true point improves with larger T and variance ratios.
- Applying ICSS to AR(1) residuals yields similar identification rates to the i.i.d. case.

- 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=CPU Time Requirements / Conclusion, note=Recommends ICSS for long series with multiple change points;  $O(T)$  operations.
  - page=5, words=825-1174, section=ICSS Algorithm and Convergence, note=Iterative scheme; convergence in few iterations.
  - page=9, words=550-899, section=4.2 Results Part One; 4.6 CPU Time (setup), note=One change point detection difficulty at  $T=100$ ,  $A=2$ ; improves with  $T \geq 200$  or  $A=3$ ; Bayesian best at  $A=2$ .
  - page=9, words=0-349, section=4.3 Results Part Two, note=ICSS better than LR with  $T \geq 200$ ; best at (.33,.67) and when large variance in the middle; posterior odds computational burden (1,358 CPU hours).
  - page=9, words=275-624, section=4.4 Sampling Distribution of  $k^*$ , note= $k^*$  mode at true change; distribution biased to middle; concentration improves with  $T$  and variance ratio.
  - page=9, words=550-899, section=4.5 ICSS on Residuals, note=Similar identification rates on AR(1) residuals.

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

- Category: E) Citations and related literature
- Confidence: low (0.27717837898716885)
- Retrieval method: local
- Evidence type: textual citations from the article
- Data source: Inclán & Tiao (1994), JASA 89(427):913–923 (provided context)
- Assumption flag: False
- Answer:

Key prior works are: (1) Brown, Durbin, and Evans (1975) because the ICSS procedure is built on their centered cumulative sum of squares; (2) Hsu, Miller, and Wichern (1974) for initiating the variance-change literature with a step-change variance model for stock returns; (3) Hsu (1977, 1979, 1982) and Booth & Smith (1982) for single-change detection methods (frequentist and Bayesian) that highlight the computational burden of multiple changes and serve as baselines; (4) Baufays & Rasson (1985) for MLE methods handling several change points and for providing the algorithm used to compute LR benchmarks; (5) Wichern, Miller, & Hsu (1976) and Abraham & Wei (1984) for variance-change models in autocorrelated (AR) series, relevant to applying ICSS to residuals; (6) Worsley (1986) as a related MLE framework for change-point estimation (mean changes) that informs single-change inference.

- Quote snippet:  
It is based on a centered version of the cumulative sum of squares presented by Brown, Durbin, and Evans (1975).
- Citation anchors:
  - page=2, words=275-624, section=Introduction, note=ICSS is based on the centered CUSUM of squares of Brown, Durbin, and Evans (1975); foundational change-of-variance literature and single-change Bayesian/frequentist works cited.



- page=2, words=0-349, section=Introduction, note=Hsu, Miller, and Wichern (1974) noted as starting the variance-change literature with step changes in variance for stock returns.
- page=2, words=550-899, section=2. Centered Cumulative Sums of Squares, note=Worsley (1986) MLE for change in mean; Wichern, Miller, & Hsu (1976) and Abraham & Wei (1984) on AR(1) with variance change; Baufays & Rasson (1985) MLE for multiple change points.
- page=9, words=550-899, section=4.6 CPU Time Requirements, note=Baufays & Rasson (1985) algorithm used to obtain the ML in LR tests; computational motivation for ICSS.

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

- Category: E) Citations and related literature
- Confidence: low (0.25288654187197895)
- Retrieval method: local
- Evidence type: explicit statements in the text
- Data source: Inclán & Tiao (1994) article text (JSTOR excerpt)
- Assumption flag: False
- Answer:

The paper builds most directly on Brown, Durbin, and Evans (1975) for the centered cumulative sum of squares, and extends the variance change-point literature initiated by Hsu, Miller, and Wichern (1974) and developed by Hsu (1977, 1979, 1982) and Booth & Smith (1982).

- Quote snippet:  
It is based on a centered version of the cumulative sum of squares presented by Brown, Durbin, and Evans (1975).
- Citation anchors:
  - page=2, words=275-624, section=Introduction/Overview, note=States the method is based on a centered version of CUSUM of squares by Brown, Durbin, and Evans (1975); lists prior variance change-point works including Hsu (1977, 1979, 1982) and Booth & Smith (1982).
  - page=2, words=0-349, section=Introduction, note=Notes the statistical literature on changes of variance started with Hsu, Miller, and Wichern (1974).

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

- Category: E) Citations and related literature
- Confidence: low (0.30012011333351307)
- Retrieval method: local
- Evidence type: textual evidence from methods/results description and cited references
- Data source: JASA 1994 article text (provided context)
- Assumption flag: False
- Answer:

The results compare ICSS against (1) likelihood ratio tests implemented using the Baufays and

Rasson (1985) algorithm, and (2) a Bayesian posterior-odds approach specified with an inverted-gamma prior as in Inclán (1991).

- Quote snippet:  
compare the performance of the ICSS algorithm with the likelihood ratio and a Bayesian approach.
- Citation anchors:
  - page=7, words=0-349, section=4. SIMULATION EXPERIMENT, note=States comparison of ICSS with likelihood ratio and a Bayesian approach.
  - page=9, words=550-899, section=4.6 CPU Time Requirements, note=LR tests computed using the Baufays and Rasson (1985) algorithm.
  - page=9, words=275-624, section=4.4 Sampling Distribution of  $k^*$  / 4.2 Results context, note=Bayesian posterior odds uses inverted-gamma prior (see Inclán 1991).

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

- Category: E) Citations and related literature
- Confidence: low (0.2478824092067886)
- Retrieval method: local
- Evidence type: textual
- Table/Figure: Tables 6–8 (simulation results and AR(1) residuals)
- Data source: Simulated time series ( $T=100, 200, 500$ ) with specified changepoints and variance ratios; AR(1) residuals derived from these simulations
- Assumption flag: False
- Answer:

Only simulated series are cited. The study uses artificially generated time series with specified lengths ( $T = 100, 200, 500$ ), changepoint locations, and variance ratios ( $A = 2, 3$ ) for simulation experiments; it also applies the ICSS algorithm to AR(1) residuals derived from these simulated series. No external empirical datasets are cited in the provided excerpt; an illustrative series is mentioned but not identified.

- Quote snippet:  
For Part One, we used three series lengths ( $T = 100, 200$ , and  $500$ )... and two values for the variance ratio ( $A = 2$  and  $3$ )... Table 8... Residuals from AR(1).
- Citation anchors:
  - page=7, words=0-349, section=4. SIMULATION EXPERIMENT, note=Details of simulated series: lengths, changepoint locations, variance ratios; paired comparisons.
  - page=9, words=0-349, section=4.4 Sampling Distribution of  $k^*$ , note=States there are 1,000 replicates per configuration, confirming simulation-based data.
  - page=10, words=0-84, section=Tables 7–8, note=Mentions applying ICSS to AR(1) residuals from the simulated series.
  - page=6, words=550-899, section=Application example, note=Refers to an analyzed series illustrating ICSS, but the dataset is not identified in the excerpt.

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

- Category: E) Citations and related literature
- Confidence: low (0.3465047268726521)
- Retrieval method: local
- Evidence type: explicit references list
- Data source: Inclán & Tiao (1994), Journal of the American Statistical Association (provided context)
- Assumption flag: False
- Answer:

The article cites methodological/econometric works on change-point detection, CUSUM/CUSUM-of-squares, likelihood-ratio and Bayesian methods, time-series models with changing variance/parameters, and related asymptotics/simulation, including: Hinkley (1971); Brown, Durbin, and Evans (1975); Smith (1975, 1980); Booth and Smith (1982, Journal of Econometrics); Menzefricke (1981); Hsu (1977, 1979, 1982); Haccou and Meelis (1988); Haccou, Meelis, and van de Geer (1988); Worsley (1986); Wichern, Miller, and Hsu (1976); Abraham and Wei (1984); Baufays and Rasson (1985); Tyssedal and Tjostheim (1988); Tsay (1988); Chang, Tiao, and Chen (1988); Cobb (1978); Commenges, Seal, and Pinatel (1986); Billingsley (1968); Box and Jenkins (1976); Bratley, Fox, and Schrage (1987); Broemeling and Tsurumi (1987). No DiD, IV, or RDD methods are cited.

- Quote snippet:  
Booth & Smith (1982), 'A Bayesian Approach to Retrospective Identification of Change-Points'; Brown, Durbin & Evans (1975), 'Techniques for Testing the Constancy of Regression Relationships Over Time'
- Citation anchors:
  - page=12, words=0-349, section=References and Appendix, note=Multiple methodological references listed (Booth & Smith 1982; Brown, Durbin & Evans 1975; Billingsley 1968; Box & Jenkins 1976; etc.)
  - page=12, words=275-624, section=References and Appendix, note=Additional change-point and likelihood-ratio references (Haccou & Meelis 1988; Worsley 1986; Hsu 1977, 1979, 1982).
  - page=12, words=550-858, section=References and Appendix, note=Further references on change-point tests and asymptotics (Menzefricke 1981; Smith 1975, 1980; Worsley 1986).

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

- Category: E) Citations and related literature
- Confidence: low (0.23712234626341297)
- Retrieval method: local
- Evidence type: textual
- Data source: Inclán & Tiao (1994), JASA (JSTOR excerpt)
- Assumption flag: True
- Assumption notes: Interpreted 'positions itself against' as contrasting with prior benchmark methods and frameworks rather than refuting them.

- Answer:

Yes. The paper contrasts its ICSS method with established Bayesian and likelihood ratio change-point procedures and related maximum likelihood approaches, citing classic works such as Smith (1975, 1980), Booth & Smith (1982), Hsu (1977, 1979, 1982), Worsley (1986), and Baufays & Rasson (1985). It also builds on the classic CUSUM-of-squares work of Brown, Durbin & Evans (1975) and references the foundational variance-change formulation of Hsu, Miller & Wichern (1974).

- Quote snippet:

the ICSS algorithm offers results comparable to those obtained by a Bayesian approach or by likelihood ratio tests, without the heavy computational burden

- Citation anchors:

- page=2, words=0-349, section=Introduction, note=Positions ICSS against Bayesian and likelihood ratio tests due to computational burden.
- page=2, words=275-624, section=Introduction, note=States approach is based on centered CUSUM of squares by Brown, Durbin, and Evans (1975); lists classic change-point references (Hinkley 1971; Smith 1975, 1980; Booth & Smith 1982; Hsu 1977, 1979, 1982).
- page=2, words=550-899, section=Section 2 setup, note=Mentions Worsley (1986) maximum-likelihood tests and AR variance-change works (Wichern et al. 1976; Abraham & Wei 1984; Baufays & Rasson 1985).

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

- Category: E) Citations and related literature
- Confidence: low (0.24801479150047653)
- Retrieval method: local
- Evidence type: explicit textual mention
- Assumption flag: False
- Answer:

Yes. The paper references appendices (Appendix A with derivations and Appendix B with a proof) that underpin core claims, and it notes Fortran code for the ICSS algorithm is available upon request. No data repositories are cited.

- 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=825-968, section=Main text, note=Mentions availability of Fortran program for the ICSS algorithm.
- page=11, words=0-349, section=Appendix A, note=Appendix A provides derivations (Approximate Expected Value of  $D_k$ ).
- page=12, words=0-349, section=Appendix B, note=Appendix B provides the 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 (0.23077763319438027)
- Retrieval method: local
- Evidence type: textual
- Assumption flag: False
- Answer:

The cited literature mainly treats a single, unknown variance change and tests changepoints one at a time due to heavy computational burden, leaving a gap for an efficient, systematic method to detect multiple variance changes in moderate-length series.

- Quote snippet:

single change of variance at an unknown point... focusing on the detection of points of change one at a time because of the heavy computational burden involved in looking for several points of change simultaneously.
- Citation anchors:
  - page=2, words=275-624, section=Introduction, note=Prior work focuses on single changepoint; multiple points are computationally burdensome.
  - page=2, words=0-349, section=Introduction, note=Authors position ICSS as comparable to Bayesian/LR without heavy computational burden.
  - page=9, words=550-899, section=Simulation: CPU Time Requirements, note=Motivation cites heavy computational burden of posterior odds and likelihood-based searches.

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

- Category: F) Replication and transparency
- Confidence: low (0.19380720232206966)
- Retrieval method: local
- Evidence type: textual
- Data source: Article text (Inclán & Tiao, 1994) in provided context
- Assumption flag: False
- Answer:

Yes. The Fortran program (ICSS algorithm) is available from the authors upon request via email to [inclan@guvax.georgetown.edu](mailto: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](mailto:inclan@guvax.georgetown.edu).
- Citation anchors:
  - page=6, words=825-968, note=States the Fortran program is available upon request via email to [inclan@guvax.georgetown.edu](mailto:inclan@guvax.georgetown.edu).

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

- Category: F) Replication and transparency

- Confidence: low (0.22306541673272023)
- Retrieval method: local
- Evidence type: textual
- Data source: Inclán & Tiao (1994), Journal of the American Statistical Association
- Assumption flag: False
- Answer:

No. The provided excerpts do not mention any pre-analysis plan or registered trial; they describe a methodological study with simulations, not a registered study.

- 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=Describes simulation setup; no registration or pre-analysis plan referenced.

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

- Category: F) Replication and transparency
- Confidence: low (0.18228175986289236)
- Retrieval method: local
- Evidence type: Terms of Use notice in article front matter
- Data source: JSTOR Terms and Conditions of Use notice (page 1)
- Assumption flag: False
- Answer:

Yes. The article's front matter discloses restricted-use terms via JSTOR's Terms and Conditions (personal, non-commercial use only; no downloading entire issues or multiple copies without prior permission). No NDAs are mentioned.

- Quote snippet:  
  
you may use content in the JSTOR archive only for your personal, non-commercial use.
- Citation anchors:
  - page=1, words=0-230, section=JSTOR Terms and Conditions notice, note=States personal, non-commercial use only and prohibits downloading entire issues or multiple copies without prior permission.

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

- Category: F) Replication and transparency
- Confidence: low (0.2456742970693667)
- Retrieval method: local
- Evidence type: textual evidence
- Data source: IBM stock prices as reported by Box and Jenkins (1976)
- Assumption flag: False
- Answer:

Yes—key construction steps are documented (e.g., transforming IBM prices to first differences of logs and detailed simulation setups). Explicit data-cleaning is limited to advising outlier checks.

- Quote snippet:

The series analyzed is the first difference of the logarithm of the IBM stock closing prices...

- Citation anchors:
  - page=6, words=275-624, section=3.1 Example: IBM Stock Prices, note=Documents construction of the empirical series as first differences of log IBM prices.
  - page=7, words=0-349, section=4. SIMULATION EXPERIMENT, note=Details parameters for constructing simulated data (lengths, changepoint locations, variance ratios).
  - page=6, words=0-349, section=ICSS Algorithm discussion, note=Advises complementing variance-change search with outlier detection, a cleaning consideration.

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

- Category: F) Replication and transparency
- Confidence: low (0.309728123445472)
- Retrieval method: local
- Evidence type: text
- Assumption flag: False
- Answer:

Partially omitted

- 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=200-340, section=4.3 Results of Part Two, note=Posterior odds not obtained for  $T=500$ ; reduced replicates for  $T=200$ .
  - page=9, words=700-780, section=4.5 The ICSS Algorithm Applied to Residuals, note=Indicates further work pending, suggesting incomplete reporting.

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

- Category: G) External validity and generalization
- Confidence: low (0.21142405836654996)
- Retrieval method: local
- Evidence type: Direct text statements on scope/assumptions and simulation design
- Assumption flag: False
- Answer:

Sequences of independent (or prewhitened/uncorrelated) observations with piecewise-constant variance—especially financial time series—with moderate-to-large sample sizes (about 200+ observations). Theory and simulations assume i.i.d. Normal innovations; autocorrelated series can be handled by applying ICSS to AR-model residuals.

- Quote snippet:

We study the detection of multiple changes of variance in a sequence of independent observations. There are series, particularly in the area of finance,

- Citation anchors:
  - page=2, words=0-349, section=Introduction, note=Scope: independent observations; finance context; moderate size (200+) applicability.
  - page=2, words=550-899, section=Centered Cumulative Sums of Squares, note=Defines method for uncorrelated mean-zero series with time-varying variances.
  - page=5, words=0-349, section=Asymptotics (Brownian bridge), note=Assumes i.i.d. Normal innovations for asymptotic distribution.
  - page=9, words=550-899, section=ICSS Applied to Residuals, note=Application to AR(1) residuals indicates use with autocorrelated series after prewhitening.

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

- Category: G) External validity and generalization
- Confidence: low (0.20618870169803688)
- Retrieval method: local
- Evidence type: Simulation results and theoretical properties
- Data source: Inclán and Tiao (1994), Journal of the American Statistical Association
- Assumption flag: False
- Answer:

Settings least likely to generalize are short series with small variance changes occurring near the ends, and series with two changepoints where the variance increases monotonically. The  $\max|D_k|$  criterion is biased toward the middle, further reducing reliability for edge changes.

- Quote snippet:

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=550-899, section=4.2 Results of Part One: Series With One Change-point, note=Hard to detect small change ( $A=2$ ) in short series, especially at the beginning.
  - page=9, words=0-349, section=4.3 Results of Part Two: Series With Two Changepoints, note=Most difficult when variances increase monotonically; needs large  $T$  (~500).
  - page=9, words=275-624, section=4.4 Sampling Distribution of  $k^*$ , note= $k^*$  biased toward the middle of the series.
  - page=4, words=275-624, section=2.3 Asymptotic Behavior of  $D_k$  Under Homogeneous Variance, note= $\max|D_k|$  puts more weight near the middle, inducing bias toward  $T/2$ .

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

- Category: G) External validity and generalization
- Confidence: low (0.2617001163362451)
- Retrieval method: local
- Evidence type: explicit text



- Table/Figure: Table 1; Figure 2
- Assumption flag: False
- Answer:

Yes. They define statistical boundaries for  $D_k$  from its asymptotic distribution and use a critical value  $D^*$  to decide changes; they also limit iterations and set a convergence rule (within two observations).

- Quote snippet:

These boundaries can be obtained from the asymptotic distribution of  $D_k$  assuming constant variance.

- Citation anchors:
  - page=3, words=275-624, section=Main text on  $D_k$  plots and boundaries, note=Boundaries from asymptotic distribution of  $D_k$  assuming constant variance
  - page=5, words=0-349, section=ICSS Algorithm steps, note=Decision rule using critical value  $D^*$ ; stopping condition
  - page=5, words=825-1174, section=ICSS implementation details, note=Controls to avoid indefinite cycling; iteration limit mentioned
  - page=6, words=550-899, section=Convergence criteria, note=Convergence when change points move by no more than two observations
  - page=4, words=550-869, section=Asymptotic boundaries illustration, note=Reference to asymptotic boundaries (e.g., 1.358) shown in plots

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

- Category: G) External validity and generalization
- Confidence: medium (0.37517108644342595)
- Retrieval method: local
- Evidence type: text
- Assumption flag: True
- Assumption notes: Interprets different time periods or markets as time series that differ in sample length, spacing and configuration of variance shifts, variance ratios, outliers, and autocorrelation; conclusions follow the stated simulation and algorithm behavior.
- Answer:

Results would vary with the series' characteristics. Detection performance depends on sample size, spacing and configuration of variance changes, variance ratios, and data idiosyncrasies. Longer samples improve identification and localization; equal spacing between change points tends to perform best; monotone variance increases are hardest (often requiring many more observations). Performance improves as variance ratios increase. Multiple change points can mask each other, but the iterative ICSS procedure mitigates this. Outliers can create false peaks and should be checked. Applying ICSS to autoregressive residuals yields results similar to independent data. Hence, different time periods or markets—if they differ in length, volatility regime patterns, outlier prevalence, or autocorrelation—can produce different numbers and locations of detected variance changes and different accuracy levels.

- Quote snippet:

The most difficult situation is when the variances change in a monotone way; that is, the

variance increases at the first change point and increases again at the second change point.

- Citation anchors:
  - page=9, words=0-349, section=Simulation results (two changepoints), note=Hardest when variances change monotonically; need larger  $T$  (e.g., 500); best when spacing is (.33,.67) and large variance in the middle; ICSS often better with  $T \geq 200$ .
  - page=9, words=275-624, section=Sampling distribution of  $k^*$ , note= $k^*$  biased toward middle; concentration around true changepoint increases with sample size and variance ratio.
  - page=4, words=0-349, section=Behavior of  $E[Dk]$  with multiple changepoints, note=Piecewise linear  $E[Dk]$ ; masking effect when  $>1$  changepoint; dependence on configuration of variances.
  - page=6, words=0-349, section=Practical considerations, note=Outliers can create significant  $Dk$  peaks; recommend outlier detection; ICSS on AR residuals performs similarly to independent data.
  - page=6, words=275-624, section=Example and convergence, note=Applied example on IBM returns; illustrates practical variance-change detection in a market series.

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

- Category: H) Measurement validity
- Confidence: low (0.2784010202574155)
- Retrieval method: local
- Evidence type: textual
- Assumption flag: True
- Assumption notes: Interpreted 'key variables' as variance changes/changepoints; underlying series values (e.g., IBM price log-differences) are directly observed, but variance changes are inferred.
- Answer:

Proxies/inferred. The key quantities—variance changes and changepoints—are not observed directly; they are inferred from the data using statistics like  $\max|Dk|$  (and LR), while the raw series (e.g., IBM price log-differences) are directly observed.

- Quote snippet:

This behavior leads to the search for a variance change point via  $\max_k |Dk|$  ... If this maximum absolute value exceeds a predetermined boundary, then we may conclude that there is a change point

- Citation anchors:
  - page=3, words=550-882, section=Detection via  $Dk$ , note=Variance change points concluded from the statistic  $\max|Dk|$  crossing a boundary.
  - page=6, words=275-624, section=IBM example data, note=Raw data described as first differences of log IBM closing prices (directly observed then transformed).

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

- Category: H) Measurement validity

- Confidence: low (0.2677151175224959)
- Retrieval method: local
- Evidence type: textual\_evidence
- Table/Figure: Table 2
- Assumption flag: True
- Assumption notes: Inference that autocorrelation in raw series may distort detection is based on authors' application of ICSS to AR(1) residuals rather than an explicit statement of risk.
- Answer:

Acknowledged/likely risks include: (1) miscalibrated thresholds from asymptotic approximations, yielding conservative type I error and potential under-detection; (2) biased location estimates of the change point ( $k^*$  skewed toward the series middle); (3) low power for small samples or small variance ratios, especially when the change occurs early, and in monotone-increasing variance scenarios, which may require large  $T$  to detect multiple changes; (4) dependence in data can affect detection, prompting use on AR(1) residuals rather than raw series.

- Quote snippet:

This distribution is skewed;  $k^*$  is biased towards the middle of the series.

- Citation anchors:
  - page=5, words=500-610, section=Asymptotics and empirical quantiles, note=Empirical 95th quantile below asymptotic 1.358 implying conservative type I error
  - page=9, words=210-270, section=4.4 Sampling Distribution of  $k^*$ , note= $k^*$  biased toward the middle
  - page=9, words=550-760, section=Results and detection difficulty, note=Hard to detect small variance change, especially early in series
  - page=9, words=0-120, section=Two changepoints results, note=Most difficult when variances increase monotonically; need large  $T$
  - page=8, words=120-200, section=Likelihood ratio testing, note=Critical values must be obtained by simulation; non-regular problem
  - page=9, words=760-899, section=AR(1) residuals application, note=Applied ICSS to AR(1) residuals, indicating concern about dependence

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

- Category: H) Measurement validity
- Confidence: low (0.2598324384323857)
- Retrieval method: local
- Evidence type: asymptotic theory; simulation benchmarks; algorithmic convergence checks; LR standardization
- Table/Figure: Table 1 (Empirical vs asymptotic quantiles of  $\max|D_k|$ ); Table 2 (percent below 1.358); Table 3 (LR standardization)
- Data source: 10,000 replicates of  $T$  independent  $N(0,1)$  series (noted in Tables 1–2)
- Answer:

Yes. The paper validates key measures by (1) deriving asymptotic reference values for  $\max|D_k|$  from a Brownian bridge and (2) benchmarking them with large-scale simulations (Tables 1–2, with SEs). It also standardizes likelihood-ratio tests to an extreme-value distribution (Table 3) and uses algorithmic stopping and convergence checks for detected changepoints.

- Quote snippet:

Table 1 compares selected asymptotic quantiles of  $\max, |D|, I$  with corresponding empirical quantiles for various series lengths obtained by simulation.

- Citation anchors:
  - page=5, words=0-349, section=Brownian bridge asymptotics; empirical vs asymptotic quantiles (Table 1), note=Asymptotic distribution of  $D_k$  and comparison to simulations
  - page=3, words=550-882, section=Table 1 details and simulation note, note=Empirical and asymptotic quantiles of  $\max|D_k|$ ; 10,000 replicates; Brownian-bridge reference
  - page=5, words=275-624, section=Table 2 percentages and SE formula, note=Percentage of series with  $\max|D_k| < 1.358$ ; standard errors from simulations
  - page=6, words=550-899, section=Algorithm convergence checks (Step 3), note=Convergence when number of points stabilizes and locations move by  $\leq 2$
  - page=8, words=0-329, section=LR test standardization (Table 3) and extreme value reference, note=LR statistics standardized for reference to extreme value distribution; simulation for critical values

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

- Category: H) Measurement validity
- Confidence: low (0.21229758667580595)
- Retrieval method: local
- Evidence type: no direct mention in provided context
- Data source: Inclán & Tiao (1994) Use of Cumulative Sums of Squares for Retrospective Detection of Changes of Variance, JASA
- Assumption flag: True
- Assumption notes: Conclusion is limited to the provided excerpts; other parts of the full article might address different topics.
- Answer:

No. In the provided excerpts, the authors focus on detecting variance change points using the ICSS algorithm and related simulations; there is no discussion of construct validity for core outcomes.

- 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 ICSS algorithm.

- Citation anchors:
  - page=2, words=0-349, section=Introduction/Abstract, note=States study scope on variance change points and ICSS; no mention of construct validity or core outcomes.
  - page=2, words=275-624, section=Introduction/Organization, note=Outlines  $D_k$  function, relation to F statistic, ICSS algorithm, and simulations; no construct validity discussion.
  - page=7, words=0-349, section=Simulation experiment setup, note=Describes simulation design and performance comparisons; no core outcomes or validity constructs.

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

- Category: I) Policy counterfactuals and welfare

- Confidence: low (0.2852302375069697)
- Retrieval method: local
- Evidence type: textual (scope/absence)
- Assumption flag: False
- Answer:

None. The provided text focuses on statistical methods (ICSS, likelihood ratio, Bayesian) for detecting variance changes, simulations, and an IBM stock price example; it does not consider or imply any policy counterfactuals.

- 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=560-620, section=Centered Cumulative Sums of Squares, note=States the methodological focus on detecting variance changes retrospectively.
  - page=7, words=20-120, section=4. SIMULATION EXPERIMENT, note=Describes simulation to compare procedures, not policy evaluation.
  - page=6, words=300-380, section=3.1 Example: IBM Stock Prices, note=Application example is stock prices, not a policy setting.

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

- Category: I) Policy counterfactuals and welfare
- Confidence: low (0.24546938630512277)
- Retrieval method: local
- Evidence type: absence\_in\_context
- Data source: Inclán and Tiao (1994), Journal of the American Statistical Association excerpt
- Assumption flag: False
- Answer:

None. The provided context is a methodological/statistical discussion (ICSS algorithm, likelihood ratios, sampling distributions, and CPU-time costs). It does not address welfare tradeoffs or socioeconomic distributional impacts.

- Quote snippet:

the main way is by the number of "correct identifications." Another aspect to consider is the computational requirements of each approach, which can be measured in terms of the CPU time.

- Citation anchors:
  - page=7, words=0-349, section=4. SIMULATION EXPERIMENT, note=Focuses on performance metrics and computational requirements, not welfare or distributional impacts.
  - page=2, words=550-899, section=2. CENTERED CUMULATIVE SUMS OF SQUARES, note=Mentions measuring cost in CPU time; methodological emphasis.
  - page=9, words=550-899, section=4.2 Results of Part One, note=Discusses detection performance and computational burden; no welfare analysis.

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

- Category: I) Policy counterfactuals and welfare
- Confidence: low (0.28150735669307536)
- Retrieval method: local
- Evidence type: textual (CPU time comparison; simulation summaries)
- Data source: Provided article text (Sections 4 and 4.6)
- Assumption flag: True
- Assumption notes: Interprets 'cost-benefit' as an economic evaluation and 'incidence analyses' as occurrence rates of outcomes; neither is presented—only computational cost (CPU time) and simulation frequencies.
- Answer:

No. The paper reports computational cost comparisons (CPU time) and simulation frequencies but does not provide cost-benefit or incidence analyses.

- Quote snippet:

Another aspect to consider is the computational requirements of each approach, which can be measured in terms of the CPU time.

- Citation anchors:
  - page=2, words=550-899, section=Introduction/Background, note=Cost framed explicitly as CPU time for comparing approaches.
  - page=7, words=0-349, section=4. SIMULATION EXPERIMENT, note=States computational requirements measured in CPU time.
  - page=9, words=550-899, section=4.6 CPU Time Requirements, note=Details on CPU time burden and comparisons.
  - page=11, words=0-349, section=CPU time regression summary, note=Regression models summarizing CPU time; recommendation based on computational cost.

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

- Category: I) Policy counterfactuals and welfare
- Confidence: low (0.21524732197355106)
- Retrieval method: local
- Evidence type: textual
- Data source: Inclán & Tiao (1994), Journal of the American Statistical Association, 89(427), 913–923.
- Assumption flag: True
- Assumption notes: Interprets “policy recommendations” as methodological/practical guidance for applying variance-change detection procedures; no public policy is discussed.
- Answer:

No explicit public policy is given; the article offers methodological recommendations: (1) Use the ICSS algorithm for detecting variance changes, especially when computational burden is a concern. (2) Control ICSS iterations to avoid cycling (cap iterations; declare convergence when change-point locations move by <2 observations). (3) After detecting change points, estimate segment variances to build a homogeneous “corrected series” and compute residual diagnostics (e.g., Q statistic) on that corrected series. (4) For likelihood-ratio testing, proceed sequentially with  $LR_{\{n-1,n\}}$ , obtain

critical values by simulation, and standardize to an extreme value distribution. (5) For Bayesian inference, compute posterior-odds  $K_{\{n,n-1\}}$  up to a specified maximum number of changes  $N_{\max}$  and choose NT maximizing the conditional probability. (6) For time series with dynamics, apply ICSS to residuals from fitted AR models. (7) Practical power guidance: small samples ( $T \approx 100$ ) and small variance ratios ( $A=2$ ), especially with early changes, are hard to detect; performance improves notably with longer series ( $T \geq 200$ ) or larger variance ratios ( $A=3$ ).

- Quote snippet:

This "corrected series" should then be used to obtain other residual diagnostics, like the Q statistic, which would be distorted if

- Citation anchors:
  - page=5, words=825-1174, section=ICSS algorithm and implementation, note=Recommends iteration controls to avoid cycling; mentions practical cap (~20 iterations).
  - page=6, words=550-899, section=ICSS convergence and post-detection use, note=Convergence rule ( $< 2$  observations movement) and recommendation to create a corrected series for subsequent diagnostics.
  - page=8, words=0-329, section=Likelihood ratio strategy, note=Sequential  $LR_{\{n-1,n\}}$ ; critical values by simulation; extreme value standardization.
  - page=9, words=275-624, section=Bayesian posterior odds procedure, note=Systematic use of posterior odds with specified  $N_{\max}$ ; choose NT maximizing conditional probability.
  - page=9, words=550-899, section=Performance and computational guidance, note=ICSS for computational efficiency; detection difficulty at  $T=100$  and  $A=2$ ; improved performance with  $T \geq 200$  or  $A=3$ ; apply ICSS to AR residuals.

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

- Category: J) Data quality and integrity
- Confidence: low (0.304980734681012)
- Retrieval method: local
- Evidence type: text
- Data source: Simulation experiment description and results text
- Assumption flag: True
- Assumption notes: Interprets 'missingness/attrition' as data loss or dropouts; none reported. Only computational omissions in posterior-odds analyses are noted.
- Answer:

No data missingness or attrition is reported. All simulated series were complete of length  $T$ . The only omission was analytical: posterior odds were not computed for  $T=500$ , and for  $T=200$  only 100 (vs 1,000) replicates were run.

- Quote snippet:

The posterior odds ratio was not obtained for series with  $T = 500$ ... the runs with  $T = 200$  for the log posterior odds have only 100 replicates instead of the 1,000...

- Citation anchors:
  - page=7, words=275-556, section=Simulation experiment methods, note=States each replicate is a full series of length  $T$  of independent  $N(0,1)$  variables.

- page=9, words=0-349, section=Results/notes on computational feasibility, note=Explains posterior odds not obtained for T=500 and only 100 replicates for T=200.

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

- Category: J) Data quality and integrity
- Confidence: low (0.3496926880881717)
- Retrieval method: local
- Evidence type: textual
- Assumption flag: False
- Answer:

Outliers are not handled by winsorization or trimming. Instead, users should visually inspect Dk plots—outliers appear as sudden jumps without a change in slope—and complement the variance-change search with a dedicated outlier-detection procedure (e.g., Chang, Tiao, and Chen 1988).

- Quote snippet:  
a big outlier would create a significant peak in the Dk plot... it is advisable to complement the search for variance changes with a procedure for outlier detection
- Citation anchors:
  - page=6, words=0-349, section=Section 3 (ICSS algorithm guidance), note=Outliers create significant peaks; use outlier detection alongside ICSS.

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

- Category: J) Data quality and integrity
- Confidence: low (0.2451173391958245)
- Retrieval method: local
- Evidence type: Algorithm description and simulation study reported in the text
- Table/Figure: Table 1 and Table 2
- Data source: Inclán & Tiao (1994), Journal of the American Statistical Association
- Assumption flag: False
- Answer:

Yes. The ICSS procedure includes internal validation checks of detected changepoints (re-checking each candidate given adjacent ones and repeating until convergence with a defined tolerance), iteration controls to prevent cycling, and external validation via simulation (empirical quantile comparisons, standard errors, and a simulation study assessing correct identifications and CPU time).

- Quote snippet:  
Check each possible changepoint by calculating  $Dk(a[cpj-1+1:cpj+1])$ ... Repeat Step 3 until the number of changepoints does not change... we consider that if each changepoint is within two observations... then the algorithm has converged.
- Citation anchors:



- page=5, words=550-899, section=ICSS Algorithm Step 3 (validation and convergence), note=Re-check each candidate changepoint; repeat until number/locations stabilize within tolerance.
- page=5, words=825-1174, section=ICSS implementation safeguards, note=Controls over number of iterations to avoid cycling; practical convergence behavior.
- page=5, words=0-349, section=Asymptotic vs empirical quantiles, note=Table 1 compares asymptotic and empirical quantiles for max  $|D_k|$  via simulation.
- page=5, words=275-624, section=Empirical validation and Type I error, note=Percentages with standard errors; empirical 95th quantile lower than asymptotic value.
- page=7, words=0-349, section=Simulation experiment design, note=Simulation to compare ICSS with alternatives; uses correctness and CPU time as metrics.

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

- Category: J) Data quality and integrity
- Confidence: low (0.2814155547201351)
- Retrieval method: local
- Evidence type: Textual evidence from methods/results and table notes
- Table/Figure: Table 4 note; Section 4 design notes
- Data source: Journal article text (JASA 1994)
- Assumption flag: False
- Answer:

No. The paper transparently describes its simulation design, notes when certain results were not computed due to CPU-time limits, and explains the elimination of redundant variance permutations by symmetry—none of which indicate reporting bias or selective sample inclusion.

- 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=Section 4 (Simulation Experiment), results discussion, note=Explicitly states omission of posterior-odds results for  $T=500$  and reduced replicates for  $T=200$  due to computational cost.
  - page=7, words=275-556, section=Section 4 (Simulation Experiment), design details, note=Explains elimination of some variance-permutation cases based on symmetry, not selective inclusion.
  - page=10, words=0-84, section=Table note, note=Clarifies differing replicate counts across conditions; transparent reporting rather than selective.

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

- Category: K) Model fit and diagnostics
- Confidence: low (0.31894037148092247)
- Retrieval method: local
- Evidence type: text\_and\_tables
- Table/Figure: Table 4; Table 5; Table 6
- Data source: Inclán & Tiao (1994), Journal of the American Statistical Association

- Assumption flag: False
- Answer:

They report residual diagnostics using the Q statistic (to be applied to the variance-corrected series), frequency distributions of the number of detected variance changes (NT) for ICSS, LR tests, and Bayesian posterior odds, and summaries of the sampling distribution of the estimated change-point location  $k^*$  (mode, mean, and SD), noting its skewness and bias toward the series middle.

- Quote snippet:  
other residual diagnostics, like the Q statistic, which would be distorted if calculated
- Citation anchors:
  - page=6, words=550-899, section=Diagnostics (residuals/Q statistic), note=Mentions using residual diagnostics like the Q statistic on corrected series.
  - page=7, words=275-556, section=4.1 Assessing the Evidence, note=Introduces Table 4: frequency distribution of NT for one change.
  - page=8, words=0-329, section=Two changepoints results, note=Introduces Table 5: frequency distribution of NT for two changes.
  - page=9, words=0-349, section=4.4 Sampling Distribution of  $k^*$ , note=Table 6 provides Mode of  $k^*$ , Mean of  $k^*$ , SD of  $k^*$ ; discussion of skew/bias.
  - page=9, words=275-624, section=4.4 Sampling Distribution of  $k^*$ , note=Explains  $k^*$  distribution is skewed and biased toward the middle.

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

- Category: K) Model fit and diagnostics
- Confidence: low (0.3410795504216201)
- Retrieval method: local
- Evidence type: text
- Assumption flag: True
- Assumption notes: Conclusion based on absence of any mention of testing logs vs. levels or nonlinearities in the provided excerpts; only a log-difference transformation is described for the example.
- Answer:

No. In the provided context, there is no evidence that functional-form choices (logs vs. levels or nonlinearities) were tested. The example simply uses first-differenced log prices and proceeds with variance-change detection.

- Quote snippet:  
The series analyzed is the first difference of the logarithm of the IBM stock closing prices
- Citation anchors:
  - page=6, words=275-624, section=3.1 Example: IBM Stock Prices, note=Uses first difference of log prices; no discussion of testing alternative functional forms.
  - page=7, words=275-556, section=4. Simulation Experiment, note=Simulations use independent  $N(0,1)$  series; focus is on variance changes, not functional-form tests.

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

- Category: K) Model fit and diagnostics
- Confidence: low (0.32987787498865473)
- Retrieval method: local
- Evidence type: textual
- Table/Figure: Table 8
- Data source: Inclán, C., & Tiao, G. C. (1994). Use of Cumulative Sums of Squares for Retrospective Detection of Changes of Variance. *JASA*, 89(427), 913–923.
- Assumption flag: False
- Answer:

Yes. The paper discusses residual diagnostics (e.g., the Q statistic) and applies the ICSS algorithm to AR(1) residuals, reporting results in Table 8.

- Quote snippet:

This "corrected series" should then be used to obtain other residual diagnostics, like the Q statistic

- Citation anchors:
  - page=6, words=825-968, note=Mentions using residual diagnostics like the Q statistic after variance correction.
  - page=9, words=550-899, note=States the ICSS algorithm was applied to AR(1) residuals and summarizes results.
  - page=10, words=0-84, note=Table 8 title confirms residual-based results are reported.

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

- Category: K) Model fit and diagnostics
- Confidence: medium (0.3882154722563691)
- Retrieval method: local
- Evidence type: Simulation comparisons, methodological notes, and algorithmic diagnostics
- Table/Figure: Table 1; Tables 4–8
- Data source: Journal of the American Statistical Association (1994), Inclán and Tiao ICSS study
- Assumption flag: False
- Answer:

Results are sensitive to the chosen test/estimator, critical values, and data-generation specifications. Method choice: For small variance changes ( $A=2$ ), the Bayesian posterior-odds approach performs best, while ICSS improves notably with larger samples or larger variance ratios and often outperforms likelihood-ratio (LR) tests; LR is frequently weaker in multi-change settings. Threshold choice: Using empirical quantiles for  $\max|D_k|$  yields lower critical values than the asymptotic 1.358, reducing type I error relative to the asymptotic boundary, so conclusions depend on whether asymptotic or simulation-based cutoffs are used. Model structure and iteration: Accounting for adjacent changepoints can remove previously 'significant' candidates (e.g., point 342 becomes non-significant after neighboring points are included), indicating sensitivity to the iterative specification. Estimation context: Applying ICSS to AR(1) residuals using OLS produces detection rates very

close to the i.i.d. case, suggesting limited sensitivity to estimating the AR parameter. Estimator of location  $k^*$ : its sampling distribution is skewed (biased toward the center), but the mode is at the true change; concentration improves with larger  $T$  or larger variance ratios. Performance also depends on changepoint locations and variance configurations (hardest under monotone increases; easier with equal segment lengths or when the large variance is in the middle).

- Quote snippet:

The Bayesian procedure is the best for  $A = 2$  and gives comparable results to the other approaches with  $A = 3$ .

- Citation anchors:

- page=9, words=640-880, section=4.2, note=Bayesian best for  $A=2$ ; ICSS improves with longer series/larger ratios; LR comparison
- page=5, words=470-620, section=Quantiles/Type I error, note=Empirical 95th quantile lower than asymptotic 1.358; smaller empirical type I error
- page=6, words=560-720, section=Algorithm Step 3, note=Point 342 becomes non-significant after considering adjacent changepoints
- page=9, words=820-899, section=Residuals, note=ICSS on AR(1) residuals: percentages close to Table 4 (i.i.d. case)
- page=9, words=275-540, section=4.4, note= $k^*$  distribution skewed; biased to middle; mode at true change; concentration increases with  $T$  or  $A$
- page=9, words=0-220, section=4.3, note=Hardest when variances change monotonically; ICSS often better than LR with  $T \geq 200$ ; best at equal segment lengths or large variance in middle
- page=3, words=190-349, section=2.1, note=Distinction between  $F$  (known  $k$ ) and searching  $\max|Dk|$  (unknown  $k$ )