Change-Point Detection

Wentao Wang, Yuhang Guo

Présentation du cours APM_52112_EP

13 jan 2026

Overview

- **Tsaknaki et al.** proposed online Bayesian methods to detect change-points, and proved that the regimes are highly related to metaorders and market impacts.
- **Garcin et al.** studied some properties of KL-Divergence and applied them to off-line change-point detection.

Key Problem

- High-frequency order flows exhibit significant directional persistence **(long memory)**.
- Related to execution of large orders split into smaller portions **(metaorders)**.
- **Key Hypothesis:** Order flows are splitted into multiple regimes connected by **change-points**, with stable direction in each regime.
- **Objective:** Detect order flow regimes in order to predict future order flows and market impacts.

Definition and hypothesis

Orderflows (with signs) are cut into groups of N:

$$x_t = \sum_{j=1}^{N} v_{N(t-1+j)}, \quad t = 1, ..., T.$$

Run length:

$$r_t = \begin{cases} 0, & \text{if a CP occurs at time } t, \\ r_{t-1} + 1, & \text{else.} \end{cases}$$

The arrival of a CP is modeled as a Bernoulli process:

$$p(r_t \mid r_{t-1}) = \begin{cases} \frac{1}{h}, & \text{if } r_t = 0, \\ 1 - \frac{1}{h}, & \text{if } r_t = r_{t-1} + 1, \\ 0, & \text{otherwise.} \end{cases}$$

Model 1: BOCPD

The objective is to estimate:

$$p(r_t \mid x_{1:t}) = \frac{p(r_t, x_{1:t})}{\sum_{r_t} p(r_t, x_{1:t})}.$$

Thus we only need to consider the term $p(r_t, x_{1:t})$, which yields:

$$p(r_t, x_{1:t}) = \sum_{r_{t-1}} p(r_t, r_{t-1}, x_{1:t}) \,.$$

With decomposition of each term:

$$p(r_t, r_{t-1}, x_{1:t}) = p(x_t \mid r_t, r_{t-1}, x_{1:t-1})\; p(r_t \mid r_{t-1}, x_{1:t-1})\; p(r_{t-1}, x_{1:t-1}) \,.$$

Two hypothesis:

$$p(r_t \mid r_{t-1}, x_{1:t-1}) = p(r_t \mid r_{t-1}) \,;$$

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

Model 1: BOCPD

Core message-passing recursion: update joint probability by summing over all possible run lengths from previous time.

$$p(r_t, x_{1:t}) = \sum_{r_{t-1}} \underbrace{p(x_t \mid r_{t-1}, x_{1:t-1})}_{\text{UPM}} \underbrace{p(r_t \mid r_{t-1})}_{\text{hazard}} \underbrace{p(r_{t-1}, x_{1:t-1})}_{\text{previous message}}$$

Underlying Predictive Model (**UPM**) hypothesis: rely solely on data within the current regime.

$$p(x_t \mid r_{t-1}, x_{1:t-1}) = p\left(x_t \mid x_{t-1}^{(r_{t-1})}\right).$$

In the case of original BOCPD:

$$x_t \mid \theta_R \sim \mathcal{N}(\theta_R, \sigma^2), \ \textit{with } \theta_R \sim \mathcal{N}(\mu_0, \sigma_0^2)$$

$$\Longrightarrow \boxed{p\left(x_t \mid x_{t-1}^{(r_{t-1})}\right) = \mathcal{N}\left(\mu_{r_{t-1}},\ \sigma^2 + \sigma_{r_{t-1}}^2\right).}$$

Model 2: MBO

However, the i.i.d. assumption within regimes is clearly unrealistic for order flow. We thus introduce the Markovian dependence (structure **AR(1)**):

$$x_t = \theta_R + \rho\big(x_{t-1} - \theta_R\big) + \varepsilon_t,\ \varepsilon_t \sim \mathcal{N}\big(0,\ \sigma^2(1-\rho^2)\big)\ ,\ \rho = \mathit{Cste}$$

As AR(1) remains within the exponential family, conjugacy still holds, yielding a closed-form predictive distribution:

$$\boxed{p\Big(x_{t+1} \mid x_t^{(r_t)}\Big) = \mathcal{N}\Big(\mu_{r_t} + \rho\big(x_t - \mu_{r_t}\big),\ \sigma^2(1-\rho^2) + \sigma_{r_t}^2(1-\rho)^2\Big).}$$

Only the UPM is modified to account for within-regime serial correlation.

Model 3: MBOC

In MBO, the correlation parameter ρ is assumed to be constant, which is too strong. We therefore allow ρ to be time-varying within each regime.

$$x_t = \rho_t\big(x_{t-1} - \theta_R\big) + \theta_R + u_t, \qquad u_t \sim \mathcal{N}(0, \sigma^2).$$

The evolution of ρ_t is modeled using a Score-Driven update:

$$\rho_t = \omega + \alpha\, s_{t-1} + \beta\, \rho_{t-1}.$$

With the scaled score:

$$s_t = \nabla_t = \frac{\partial \log p(u_t)}{\partial \rho_t} = \frac{u_t}{\sigma^2}\big(x_{t-1} - \theta_R\big).$$

Under the condition of a given ρ_t, conjugacy still holds, and the UPM still possesses a closed-form solution:

$$\boxed{p\Big(x_{t+1} \mid x_t^{(r_t)}\Big) = \mathcal{N}\Big(\mu_{r_t} + \rho_t\big(x_t - \mu_{r_t}\big),\ \sigma^2 + \sigma^2_{r_t}\Big).}$$

Model Comparison

- **Goal:** evaluate *one-step-ahead* predictive accuracy for order flow x_t.

Posterior-weighted predictive mean

$$\hat{\mu}_t = \mathbb{E}[x_{t+1} \mid x_{1:t}] = \sum_{r_t} \mathbb{E}[x_{t+1} \mid x_{1:t}, r_t]\; p(r_t \mid x_{1:t})\,.$$

Evaluation metric: Mean Squared Error (one-step-ahead)

$$\mathrm{MSE} = \frac{1}{T}\sum_{t=1}^{T}\left(\hat{\mu}_{t-1} - x_t\right)^2.$$

	TSLA			MSFT		
ρ	0.1	0.2	0.3	0.2	0.3	0.4
ARMA	0.908	0.908	0.908	0.860	0.860	0.860
BOCPD	0.907	0.907	0.907	0.878	0.878	0.878
MBO	0.895	0.896	0.911	0.835	0.834	0.844
MBOC	**0.890**	**0.890**	**0.890**	**0.831**	**0.831**	**0.831**

Empirical Analysis 1

Regime definition

Let $x_{1:T}$ be a time series and $t < s \in \mathbb{N} \cap [1, T]$ such that

$$\arg \max_{i \in \{0,1,\ldots,t\}} p(r_t = i \mid x_{1:t}) = \arg \max_{i \in \{0,1,\ldots,s\}} p(r_s = i \mid x_{1:s}) = 0,$$

and for any $u \in \mathbb{N} \cap (t, s)$,

$$\arg \max_{i \in \{0,1,\ldots,u\}} p(r_u = i \mid x_{1:u}) \neq 0.$$

Then the subsequence $x_{t:s-1}$ is defined as a **regime**.

- **Length distribution:** under a constant hazard, regime lengths should look *geometric*-like; we verify this empirically via the histogram shape.

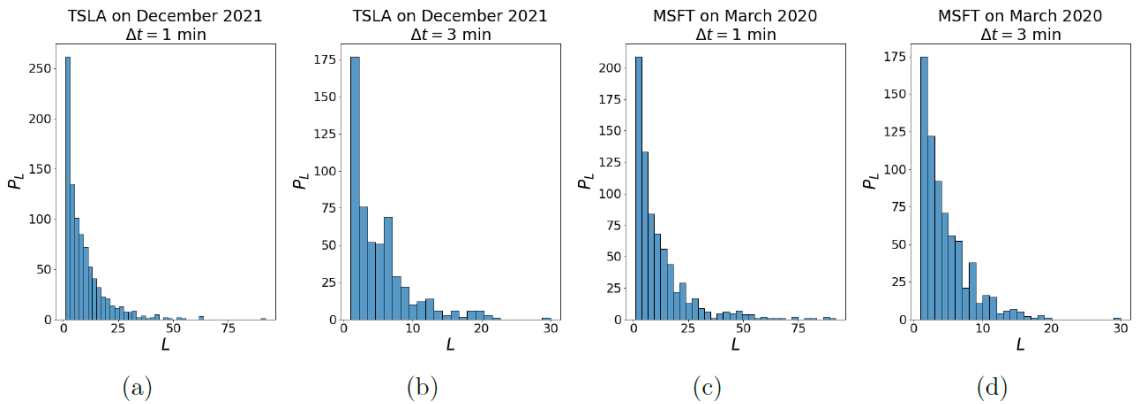


Figure: Regime length histograms (TSLA/MSFT, $\Delta t = 1$min and 3min).

Empirical Analysis 2

- **Test 1: Normality within regimes (Jarque–Bera).**
  - Apply JB test to $x_{t:s-1}$ for each regime.
  - Purpose: check the **Gaussianity** assumption (regime-level distribution closer to normal than the full series).
  - Result: Reject the null the Gaussianity hypothesis of full series (at 1% level). Can't reject the null hypothesis within each regime at 5% level for 3 minute scaler.

- **Test 2: Residual autocorrelation within regimes (Ljung–Box).**
  - Fit the model in each regime and compute residuals (innovations).
  - Apply Ljung–Box test to residuals inside each regime.
  - Purpose: check the assumption (the correlation is driven by the presence of regimes) which means the residual will not have correlation.
  - Result: cannot reject the uncorrelated residual hypothesis at 5% level within regimes for 3 minute scaler.

Price impact during order flow regimes 1

We mesure the relation between the total price change and the net volume exchanged in the same regime:

$$\epsilon_R = \text{sign}\left(\sum_{t_R \le t < s_R} x_t\right).$$

$$Z_R := \left|\frac{\sum_{t_R \le t < s_R} x_t}{\sum_{t_R \le t < s_R} |x_t|}\right| > \Theta, \qquad 0 \le \Theta < 1.$$

$$I^{\Theta}(k) = \mathbb{E}_R[\epsilon_R\,(p_{t_R+k} - p_{t_R-1}) \mid t_R + k < s_R,\ Z_R > \Theta].$$

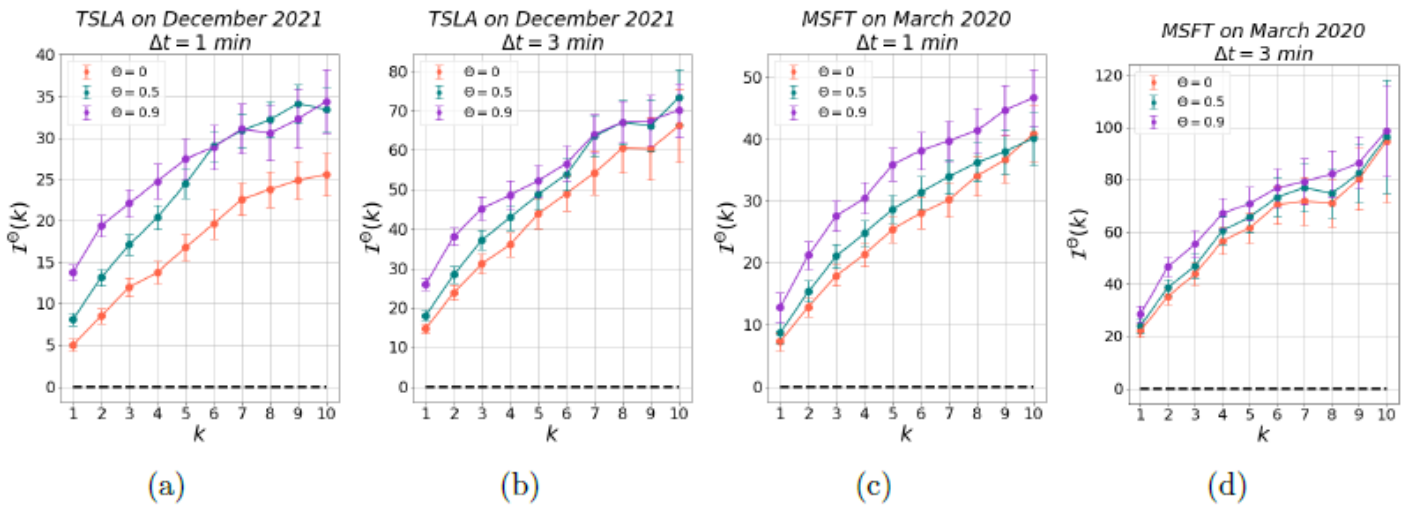


Figure: Market impact

Price impact during order flow regimes 2

We verify that on average the total price impact during a metaorder execution scales with a sublinear power law of metaorder volume:

$$\epsilon_R \Delta p_R = A \left(\epsilon_R \sum_{t_R \leq t < s_R} x_t \right)^{\gamma} + \textit{noise}. \tag{1}$$

	TSLA							
	$\Delta t = 1$ min				$\Delta t = 3$ min			
Θ	A	SE of A	γ	SE of γ	A	SE of A	γ	SE of γ
0	0.121	0.05	0.592	0.041	0.298	0.139	0.52	0.045
0.5	0.159	0.066	0.567	0.041	0.284	0.133	0.528	0.045
0.9	0.172	0.084	0.552	0.049	0.387	0.179	0.504	0.044
	MSFT							
	$\Delta t = 1$ min				$\Delta t = 3$ min			
Θ	A	SE of A	γ	SE of γ	A	SE of A	γ	SE of γ
0	0.498	0.209	0.4	0.038	0.365	0.203	0.458	0.048
0.5	0.469	0.209	0.408	0.04	0.314	0.185	0.47	0.05
0.9	0.334	0.144	0.444	0.038	0.332	0.199	0.47	0.051

Figure: Estimated parameters and their standard errors

Online prediction of order flow 1

- In the PPM regimes are independent, hence in forecasting future value only data from the current regime are useful.
- The parameters could be quite uncertain at the beginning,so it is better to wait for few observations(m steps) into a new regime.
- We want to prove that the correlations based on regimes are larger than the corresponding ones without regimes, especially for large m.

$$I_\epsilon^{(m)}(k) = \mathbb{E}_R\left[\text{sign}\left(\sum_{t=t_R}^{t_R+m-1} x_t\right)\ \text{sign}(x_{t_R+m-1+k})\right], \qquad k = 1, 2, \ldots$$

$$\tilde{I}_\epsilon^{(m)}(k) = \mathbb{E}\left[\text{sign}\left(\sum_{s=t}^{t+m-1} x_s\right)\ \text{sign}(x_{t+m-1+k})\right], \qquad k = 1, 2, \ldots$$

Online prediction of order flow 2

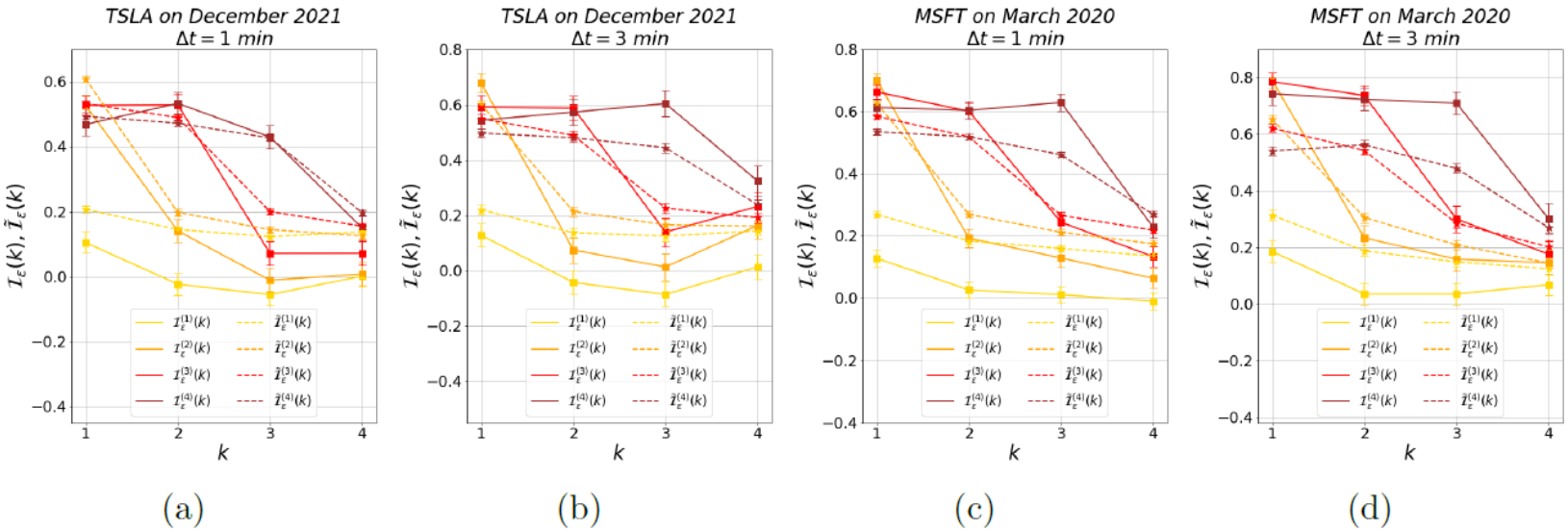


Figure: Online order flow prediction.

This evidence indicates that the knowledge of the order flow regimes improves the short-term predictability of order flow.

Online market impact prediction

$$I_{\Delta p}^{(m)}(k) = \mathbb{E}_R\left[\text{sign}\left(\sum_{t=t_R}^{t_R+m-1} x_t\right)\left(p_{t_R+m-1+k} - p_{t_R+m-1}\right)\right], \qquad k = 1, 2, \ldots$$

$$\tilde{I}_{\Delta p}^{(m)}(k) = \mathbb{E}\left[\text{sign}\left(\sum_{s=t}^{t+m-1} x_s\right)\left(p_{t+m-1+k} - p_{t+m-1}\right)\right], \qquad k = 1, 2, \ldots$$

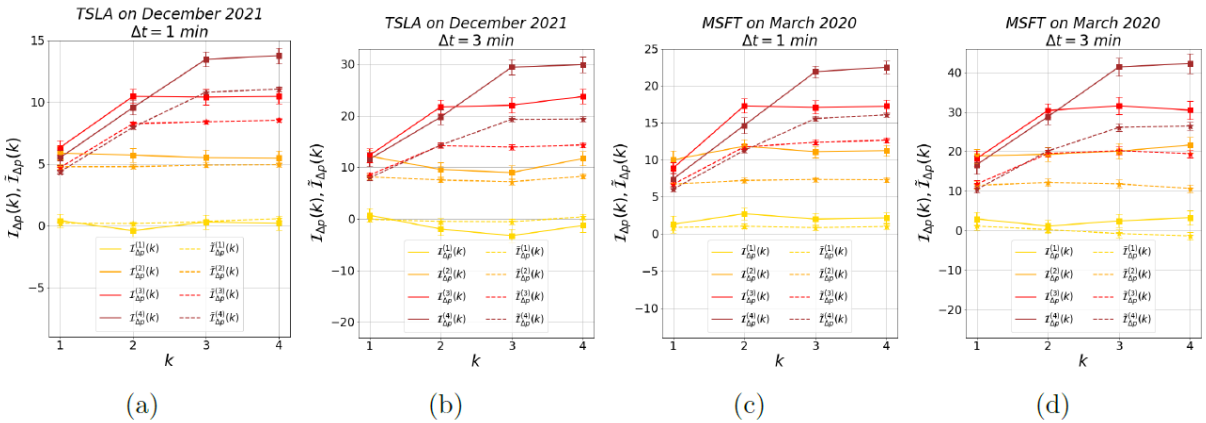


Figure: Online market impact prediction.

- Thus the (online) knowledge that a regime has started provides a significant additional forecasting power to future price change.

Conclusion

This paper constructs online regime identification framework :

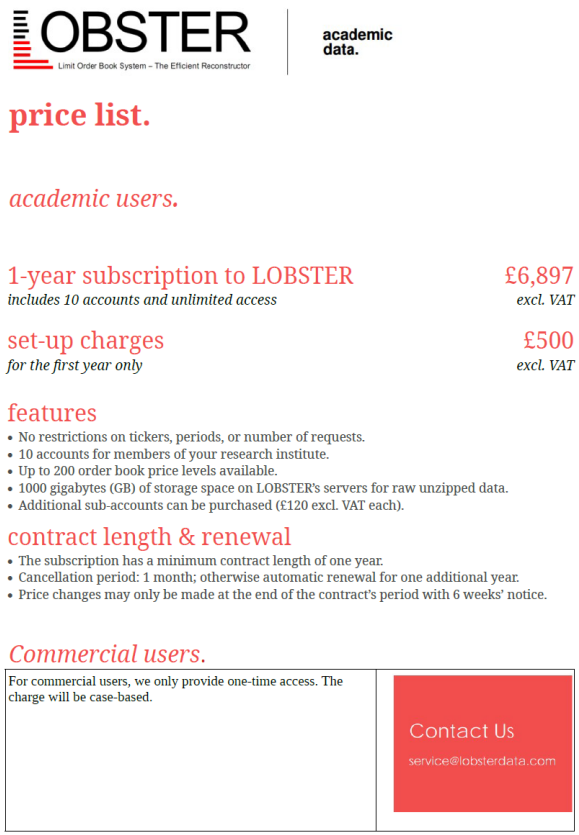
- Starting from **BOCPD** with i.i.d. assumption;
- Passing to AR(1) process which constructs **MBO** algorithm;
- Introducing score-driven mechanism to update correlation in **MBOC**.

Experiments show that:

- Regime-based models outperform regime-free time series models in one-step forecasting of order flow;
- Regime exhibits near-Gaussian properties with weak correlation, shows strong relevance to price fluctuations and market shocks.

$\Longrightarrow$ Bayesian online change-point detection can be extended with AR(1) dependence to model regime-wise serial correlation beyond i.i.d. assumptions.

Tarif de Lobster



LOBSTER
Limit Order Book System – The Efficient Reconstructor

academic data.

price list.

academic users.

1-year subscription to LOBSTER *includes 10 accounts and unlimited access*	£6,897 *excl. VAT*
set-up charges *for the first year only*	£500 *excl. VAT*

features

- No restrictions on tickers, periods, or number of requests.
- 10 accounts for members of your research institute.
- Up to 200 order book price levels available.
- 1000 gigabytes (GB) of storage space on LOBSTER's servers for raw unzipped data.
- Additional sub-accounts can be purchased (£120 excl. VAT each).

contract length & renewal

- The subscription has a minimum contract length of one year.
- Cancellation period: 1 month; otherwise automatic renewal for one additional year.
- Price changes may only be made at the end of the contract's period with 6 weeks' notice.

Commercial users.

For commercial users, we only provide one-time access. The charge will be case-based.	Contact Us service@lobsterdata.com

Figure: Price of LOBSTER

La base eurofidai est suffisante pour reproduire les expériences de cet article

Asymptotic distributions (one- & two-sample empirical KL)

- **Setting:** k bins $\{\Omega_i\}_{i=1}^k$, true pmf $p = (p_1, \ldots, p_k)^\top \in (0,1)^k$, empirical pmfs $\hat{p}_n$, $\hat{q}_m$.
- **One-sample (Thm 1):**

$$2n\, D_{KL}(\hat{p}_n \| p) \xrightarrow{d} \chi^2_{k-1}.$$

Finite-n control (Berry–Esseen type): for $x > 0$,

$$F_{\chi^2_{k-1}}\big(\kappa^{\downarrow}_{n,k}(x)\big) - \mathcal{E}_{n,k} \le \mathbb{P}\big(2nD_{KL}(\hat{p}_n \| p) \le x\big) \le F_{\chi^2_{k-1}}\big(\kappa^{\uparrow}_{n,k}(x)\big) + \mathcal{E}_{n,k},$$

with

$$\mathcal{E}_{n,k} = \Big(42(k-1)^{1/4} + 16\Big) \sum_{i=1}^{k} \frac{(1-p_i)^{3/2}}{\sqrt{np_i}}.$$

- **Two-sample (Thm 2):** if $n, m \to \infty$ and $\frac{n}{n+m} \to \lambda \in (0,1)$,

$$2\frac{nm}{n+m}\, D_{KL}(\hat{p}_n \| \hat{q}_m) \xrightarrow{d} \chi^2_{k-1}.$$

- **Quadratic proxy (Prop 1):**

$$T_{n,m} := \frac{nm}{n+m} \sum_{i=1}^{k} \frac{(\hat{p}_{n,i} - \hat{q}_{m,i})^2}{p_i},$$

and for $x > 0$,

$$\Big|\mathbb{P}(T_{n,m} \le x) - F_{\chi^2_{k-1}}(x)\Big| \le \frac{n^2 + m^2}{\sqrt{nm}(n+m)}\, \mathcal{E}_{n+m,k}.$$

Concentration inequalities (finite-sample tail bounds)

- **Goal:** non-asymptotic control of tails, e.g.

$$\mathbb{P}\big(D_{KL}(\hat{p}_n \| p) \geq x\big) \;\leq\; \text{Bound}(n, k, x),$$

  which gives **conservative thresholds** by solving $\text{Bound}(n, k, x) \leq \alpha$.

- **One-sample bounds:**
  - **Sanov:** $\mathbb{P}(D_{KL}(\hat{p}_n \| p) \geq x) \leq (n+1)^k e^{-nx}$.
  - **Mardia:** refined combinatorial bound (tighter than Sanov in part of the regime).
  - **Agrawal (MGF/Chernoff):** $M_1 \leq M_2 \leq M_3$; simple closed-form

$$\mathbb{P}(D_{KL}(\hat{p}_n \| p) \geq x) \;\leq\; M_3^{k,n}(x) = e^{-nx}\left(\frac{e\,nx}{k-1}\right)^{k-1}.$$

- **Two-sample challenge:** D_{KL} has **no triangle inequality** $\Rightarrow$ cannot directly split into two one-sample terms.
- **Two-sample solution (Thm 3):** use Pinsker / reverse Pinsker to relate KL to $\|p - q\|_1^2$, apply triangle inequality in $\|\cdot\|_1$, then transfer back to KL $\Rightarrow$ two-sample bounds $Mf_1 \leq Mf_2 \leq Mf_3$ (practical use focuses on the simplest Mf_3 and a less conservative variant).

Simulation experiments

Objective: Verification of Approximate Relative Entropy Thresholds + Comparison of Change Point Detection Methods
Results: Asymptotic χ^2 distributions are accurate for medium sample sizes, while Agrawal's concentration bounds are optimal (Sanov/Mardia's are more conservative) for finite samples.
Detection capability:

- t-test: effective only for detecting changes in means;
- AIC: High detection power but prone to false positives;
- Relative Entropy Method: Most robust for detecting general distributional shifts.

Conclusion: Relative entropy outperforms the moment method under complex distribution variations.

Real data experiments

Climate data: Able to detect anomalous climate years (such as the warm winter of 2006–2007).

Financial data:

- VIX: 2008–2009 Financial Crisis;
- N225: 2011 Fukushima Incident;
- OSEAX: 2014–2015 Crude Oil Shock.

Characteristics:

- Captures variations in distribution and dependency structures;
- Short windows are more sensitive to sudden events.

Conclusion: Relative entropy can identify structural inflection points with clear practical significance.

Conclusion

This paper investigates properties and applications of experiential relative entropy (KL divergence) in change point detection.
Theoretically, we derive asymptotic distributions and finite-sample bounds for empirical relative entropy, providing principled testing thresholds.
Methodologically, we use relative entropy for offline change-point detection, directly comparing distributions and clarifying its link to the AIC approach.
Experimental results show that the method is more robust than moment-based approaches under complex distribution shifts and identifies structurally meaningful change-points.

$\Longrightarrow$ KL-divergence can be used for offline change-point detection beyond moment-based methods.