

Name of participant	Ruochong Gu
School	The High School Affiliated to Renmin University of China
Province	Beijing
Country/Region	China
Advisor	Di Ma
Affiliation of the advisor	The High School Affiliated to Renmin University of China
Title	Inverse Gaussian Mixture Recovery and Tail Risk under Regime Shifts: Evidence from U.S. Equities, Artificial Intelligence, and Chinese Small Caps

Inverse Gaussian Mixture Recovery and Tail Risk under Regime Shifts: Evidence from U.S. Equities, Artificial Intelligence, and Chinese Small Caps

Ruochong Gu

September 14, 2026

Abstract

A structural break creates a genuine identification problem for risk measurement: the pre-event sample is statistically precise but plausibly obsolete, whereas the event window is economically relevant but often too short to identify a fat-tailed or multi-modal return distribution. Fitting the event window alone overfits a handful of extreme observations, and pooling it into the longer history dilutes the new regime until its tail consequences vanish. This paper recovers the event-regime density by inverting a mixture identity rather than estimating it in isolation. Because the pooled density of the baseline and event samples is a convex combination of the two sub-population densities, the event density is the residual of that decomposition; restricting the residual to the class of valid Gaussian mixtures makes it a genuine density, gives the recovered components a latent-state interpretation, and supports internally consistent Value at Risk (VaR) and Expected Shortfall (ES) for long and short positions.

The empirical design spans 5 event studies that differ in shock sign, market breadth, and institutional setting: the 2008 GFC and the 2020 COVID-19 crash in U.S. equities, the ChatGPT and DeepSeek episodes for NVIDIA, and China's September 2024 policy-driven rally in small-capitalization equities. The inverse mixture detects substantial two-sided tail amplification in the two S&P 500 crises and in the CSI 1000 episode, but a rotation of exposure rather than a uniform widening in the NVIDIA events: ChatGPT lowers long-side ES while raising short-side ES, and DeepSeek leaves both near pooled levels. The CSI 1000 result is the clearest illustration that a positive event mean does not imply lower downside risk: a policy-induced rightward shift can coexist with a wider reversal tail and elevated two-sided ES. The framework is informative when the event sample is sparse and adapts to structural changes that differ in sign, scale, and economic origin.

Keywords: inverse Gaussian mixture; Gaussian mixture decomposition; expectation–maximization; regime shift; tail risk; Value at Risk; Expected Shortfall.

Contents

1	Introduction	4
1.1	Related literature	4
1.2	Event design and motivation	5
1.3	Contributions and organization	5
2	Mathematical Framework	6
2.1	Mixture representation and latent states	6
2.2	Decomposition of risk and tail thickness	6
2.3	Estimation by expectation–maximization	7
2.4	Inverse recovery of a sparse new regime	8
3	Numerical Schemes for Risk Measures	9
3.1	Loss conventions for long and short positions	9
3.2	VaR Numerical Scheme	9
3.3	Closed-form Expected Shortfall	10
4	Empirical Analysis	10
4.1	Data and empirical design	10
4.2	Inverse-regime parameters and reconstruction performance	11
4.3	Comparative tail-risk evidence	17
4.4	Limitations and dynamic-risk interpretation	21
5	Conclusion	22
	References	23
	Acknowledgement	24

1 Introduction

Financial risk is not invariant to the mechanism that moves prices. A financial crisis, a pandemic, the arrival of a general-purpose technology, an adverse competitive shock, and a policy-led boom all shift the return distribution, but not in the same way. A broad equity index in a crisis is exposed to correlated downside jumps; an individual technology firm responds to firm-specific news and to revisions in expectations; a small-capitalization index can combine a positive policy impulse with unusually large two-sided trading ranges. A single unconditional Gaussian distribution cannot represent this state dependence, and a single unconditional moment cannot summarize the risk it creates.

The difficulty is most severe at the onset of a structural break, when the ordinary and the strained regimes coexist. A long pre-event sample estimates baseline conditions precisely, but it may no longer describe the regime in force. The event window is what a risk manager actually wants to assess, yet it contains too few observations to pin down a fat-tailed or multi-modal distribution. Two standard responses fail for opposite reasons. Estimating a Gaussian mixture on the event window alone overfits a handful of extreme observations; pooling the event window into the baseline history dilutes the new regime, so that when the event is a small share of the combined sample its contribution to the tail is swamped. Markov regime-switching models address the same problem by endowing latent states with transition dynamics, but an externally dated window rarely supplies enough observations to estimate a complete transition structure.

The observation underlying this paper is that the event density need not be guessed from a short window at all; it is pinned down by a population density identity. Let D_O and D_N denote the baseline and event samples, $D_T = D_O \cup D_N$ the combined sample, and $w = |D_N|/|D_T|$ the event weight. In the population, the density of the combined sample is the weighted average of the two sub-population densities, $p_T(r) = (1 - w)p_O(r) + wp_N(r)$. The event density p_N is therefore the residual of this decomposition, not a separate object to be fitted on few observations. Since the baseline and total densities can be estimated with far greater precision than the event density itself, the unknown event density can be recovered. And because both p_O and p_T are well approximated by Gaussian mixtures, we restrict p_N to the same finite-mixture family, which keeps the solution a proper density and gives the recovered components a direct economic reading.

1.1 Related literature

Finite-mixture estimation rests on the EM principle of [Dempster et al. \[1977\]](#) and the systematic treatment in [McLachlan and Peel \[2000\]](#). EM raises the observed-data likelihood monotonically, but [Wu \[1983\]](#) establishes that convergence is to a stationary point rather than necessarily to the global maximum, and that the mixture likelihood is unbounded when a component variance collapses onto a single observation. Such pathologies are not incidental; they are the reason identifiability, label switching, and model dimension are central to finite-mixture practice, as catalogued by [Frühwirth-Schnatter \[2006\]](#). The estimator used here inherits these concerns and addresses them with variance floors, a strict ordering of components, and repeated-start optimization; dimension is selected with the information criterion of [Schwarz \[1978\]](#).

The economic reading of latent states is shared with the Markov regime-switching literature introduced by [Hamilton \[1989\]](#) and surveyed for financial applications by [Ang and Timmermann \[2012\]](#). A hidden Markov model delivers transition probabilities and persistence; the estimator developed here instead recovers an unconditional event density over a pre-dated window. The two tools answer different questions, and we accordingly regard the inverse mixture as a complement to transition-based forecasting rather than a substitute. Externally dated windows follow the evidence that the 2020 crash produced an unprecedented equity-market movement [[Baker et al., 2020](#)]; more broadly, daily returns display fat tails, volatility clustering, and regime-dependent correlation [[Cont, 2001](#)], which is why risk practitioners work with non-Gaussian loss distributions

[McNeil et al., 2015].

Tail risk is measured within the coherent framework of Artzner et al. [1999]. VaR does not satisfy coherence because it ignores the severity of losses beyond a quantile, whereas Acerbi and Tasche [2002] show that Expected Shortfall does, provided the loss distribution is continuous; McNeil et al. [2015] stress that heavy tails make the distinction consequential. A normal-based VaR is the clearest offender: Adams and Thornton [2013] document that a Gaussian VaR understates the frequency and severity of extreme equity losses, so the loss distribution must retain mixture or heavy-tail structure. Finite Gaussian mixtures supply such structure in closed form, delivering VaR and ES that backtest competitively against historical simulation and a single Normal [Contreras, 2014], and the class has been evaluated as a general VaR methodology for fat-tailed returns [Morkūnaitė et al., 2024]. Mixtures complicate the comparison further, because a mixture can move probability toward the center and the extreme at the same time, exactly the setting in which the first moment and the second moment disagree about risk. Our contribution is to connect these risk functionals to an inverse finite-mixture estimator designed for sparse regime changes, and to make the source of an identified change in tail risk explicit.

1.2 Event design and motivation

The five studies span the sign of the shock, the breadth of the market it hits, and the institutional setting. Two are system-wide risk-off episodes in U.S. equities; two are firm-specific information events for a single technology company; one is a policy-driven re-rating of Chinese small-capitalization equities. The design asks whether one mathematically consistent framework can separate a widening of the loss distribution from a directional rotation of it, and whether it remains informative when an improved event mean accompanies intact or larger tails.

The crisis windows are chosen as canonical stress episodes. The GFC window (Sep.–Dec. 2008) and the COVID-19 window (Feb.–Apr. 2020) both realize new-regime volatility near four percent per day against a far calmer baseline (Table 1). For NVIDIA, the ChatGPT episode (30 Nov. 2022–Mar. 2023) is an adoption shock that raised growth expectations and produced a positive event mean, whereas the DeepSeek episode (Jan.–Mar. 2025) is an adverse competitive shock with a negative event mean; the pair isolates the information content of a firm-specific event. China’s CSI 1000 episode (24 Sep.–Nov. 2024) re-rated small-cap equities after a policy package, producing a strongly positive event mean inside a very short window. Taken together, these moments show that neither the first nor the second moment alone is a sufficient risk summary: the same positive mean can accompany a far wider distribution and a larger reversal tail.

The motivational stakes differ across settings. The crises are stress tests of concentrated, correlated exposure, where the dominant risk is a common-jump downside. The NVIDIA events isolate idiosyncratic news, where the identity of the information determines which side of the return distribution is exposed. The CSI 1000 episode examines how a policy shock propagates into a more fragile, retail-weighted market segment in which repricing and rapid liquidity reallocation can widen both tails at once. Across these settings, the framework asks whether the shape of the recovered tail risk carries information about the mechanism of the shock, or whether it merely scales with volatility.

1.3 Contributions and organization

The paper contributes in four ways. It formalizes the recovery of a latent event-regime density as an inverse problem within the Gaussian-mixture family and gives this estimator a well-posed objective. It states the conditions that keep the recovered object a valid density when the event share is small, and it shows why the unconstrained residual of the decomposition is not a usable estimate. It derives long- and short-position VaR and ES from the recovered loss distribution under a coherent sign convention, so that a mixture moving probability toward both the center

and the tail is measured correctly. And it evaluates the framework across 5 event windows that differ in sign, scale, and origin, documenting two-sided tail amplification in the S&P crises and the CSI stimulus, a rotation of long- and short-side exposure in the ChatGPT episode, and near-pooled tails in the DeepSeek episode.

Section 2 develops the mixture representation, the decomposition of risk and tail thickness, EM estimation, and the inverse recovery estimator. Section 3 specifies the loss convention for long and short positions and presents the numerical VaR scheme and the closed-form Expected Shortfall. Section 4 sets out the event design and data, the recovered inverse-regime parameters and reconstruction diagnostics, and the comparative tail-risk evidence, then closes with the analysis's limitations and the practical meaning of dynamic risk updating. Section 5 concludes.

2 Mathematical Framework

2.1 Mixture representation and latent states

Let R_i denote the continuously compounded return on day i , for $i = 1, \dots, N$. A K -component univariate Gaussian mixture has density

$$f_R(r | \theta) = \sum_{k=1}^K \pi_k \phi(r; \mu_k, \sigma_k^2), \quad \pi_k > 0, \quad \sum_{k=1}^K \pi_k = 1, \quad \sigma_k > 0, \quad (1)$$

where $\phi(\cdot; \mu, \sigma^2)$ is the Gaussian density and θ collects all free parameters. The empirical analysis sets $K = 2$. The mixture admits a latent-state reading through an unobserved indicator $Z_i \in \{1, 2\}$ such that

$$\Pr(Z_i = k) = \pi_k, \quad R_i | Z_i = k \sim \mathcal{N}(\mu_k, \sigma_k^2), \quad (2)$$

so that every day belongs to one of two unobserved conditions. Since Z_i is not observed, the state must be integrated out, and the observed-data log-likelihood is

$$\ell(\theta) = \sum_{i=1}^N \log \left\{ \sum_{k=1}^2 \pi_k \phi(R_i; \mu_k, \sigma_k^2) \right\}. \quad (3)$$

The logarithm of a sum prevents separating the two components. Were Z_i observed, the criterion would instead split into the complete-data likelihood

$$\ell_c(\theta) = \sum_{i=1}^N \sum_{k=1}^2 \mathbf{1}\{Z_i = k\} [\log \pi_k + \log \phi(R_i; \mu_k, \sigma_k^2)], \quad (4)$$

where $\mathbf{1}\{\cdot\}$ is the indicator function. The gap between the two expressions is precisely what the estimation procedure in Subsection 2.3 closes. Before turning to it, Subsection 2.2 records what a mixture implies for risk, which is what makes the two-component restriction economically meaningful rather than merely flexible.

2.2 Decomposition of risk and tail thickness

Writing $\bar{\mu} = \mathbb{E}(R) = \sum_k \pi_k \mu_k$, the law of total variance gives

$$\text{Var}(R) = \mathbb{E}[\text{Var}(R | Z)] + \text{Var}[\mathbb{E}(R | Z)] \quad (5)$$

$$= \underbrace{\sum_{k=1}^2 \pi_k \sigma_k^2}_{\text{within-regime variance}} + \underbrace{\sum_{k=1}^2 \pi_k (\mu_k - \bar{\mu})^2}_{\text{between-regime variance}}. \quad (6)$$

The first term is dispersion each state carries on its own; the second is the dispersion induced by states having different conditional means. A single Gaussian collapses both into one scale and loses the distinction, which is one reason a mixture can carry tail risk that a moment-matched Gaussian does not.

The difference appears at the extreme tail. Define $d_k = \mu_k - \bar{\mu}$. The fourth central moment of the mixture is

$$m_4 = \sum_{k=1}^2 \pi_k (3\sigma_k^4 + 6d_k^2\sigma_k^2 + d_k^4), \quad \text{Kurt}(R) = \frac{m_4}{\text{Var}(R)^2}. \quad (7)$$

This combines the kurtosis of each Gaussian with cross-terms that grow in the separation of the means. When the component means coincide, excess kurtosis is strictly positive because $m_4 - 3\text{Var}(R)^2 = 3\text{Var}(\sigma_Z^2) > 0$; with separated means, all four terms in (7) contribute, so kurtosis must be evaluated for a fitted mixture rather than inferred from unequal variances alone. The mixtures estimated in Section 4 are strongly leptokurtic. Together the two results in this subsection justify computing risk measures on the mixture directly instead of on a moment-matched Gaussian, and set the stage for estimating the components in Subsection 2.3.

2.3 Estimation by expectation–maximization

EM treats the unobserved state as missing data [Dempster et al., 1977, McLachlan and Peel, 2000]. At iteration t , the E-step replaces the indicators in the complete-data criterion with their conditional expectations, the responsibilities

$$\gamma_{ik}^{(t)} = \Pr(Z_i = k \mid R_i, \theta^{(t)}) = \frac{\pi_k^{(t)} \phi(R_i; \mu_k^{(t)}, (\sigma_k^2)^{(t)})}{\sum_{j=1}^2 \pi_j^{(t)} \phi(R_i; \mu_j^{(t)}, (\sigma_j^2)^{(t)})}. \quad (8)$$

The M-step maximizes the resulting expected complete-data criterion,

$$Q(\theta \mid \theta^{(t)}) = \sum_{i=1}^N \sum_{k=1}^2 \gamma_{ik}^{(t)} [\log \pi_k + \log \phi(R_i; \mu_k, \sigma_k^2)]. \quad (9)$$

Writing $N_k^{(t)} = \sum_i \gamma_{ik}^{(t)}$, the first-order conditions for the component parameters, together with a Lagrange multiplier for the constraint $\sum_k \pi_k = 1$, give the weighted sample-moment estimates

$$\pi_k^{(t+1)} = \frac{N_k^{(t)}}{N}, \quad \mu_k^{(t+1)} = \frac{1}{N_k^{(t)}} \sum_{i=1}^N \gamma_{ik}^{(t)} R_i, \quad (\sigma_k^2)^{(t+1)} = \frac{1}{N_k^{(t)}} \sum_{i=1}^N \gamma_{ik}^{(t)} \left(R_i - \mu_k^{(t+1)} \right)^2. \quad (10)$$

Each component update is the corresponding weighted moment over the responsibilities, so the step is as tractable as for a single Gaussian with data replaced by their fractional assignments.

EM does not decrease (3): minorizing the log-sum with Jensen’s inequality gives $Q(\theta \mid \theta^{(t)}) \leq \ell(\theta)$, so maximizing the lower bound cannot lower the observed-data likelihood. This monotonicity is not convergence to the global maximum, and the mixture likelihood is unbounded if a component variance collapses onto a single observation. Following Wu [1983], convergence is therefore interpreted as convergence to a stationary point under regularity and identification constraints. In the present setting those constraints rule out a component that exhausts its mass on one observation without representing a persistent economic state.

The likelihood is evaluated in a numerically stable form, and component variances are bounded away from zero, so that no state can absorb an arbitrarily small dispersion. Components are ordered by estimated variance to remove the label-switching symmetry that leaves permutations of a fit indistinguishable in the likelihood. Starting values are spread over the empirical return

distribution, and the reported solution is the one with the largest likelihood across repeated starts. Iteration stops when

$$|\ell^{(t+1)} - \ell^{(t)}| \leq 10^{-7} \left(1 + |\ell^{(t)}|\right). \quad (11)$$

Dimension is chosen with the Bayesian Information Criterion

$$\text{BIC}(K) = -2\hat{\ell}_K + (3K - 1) \log N, \quad (12)$$

where $\hat{\ell}_K$ is the maximized observed-data log-likelihood for a K -component mixture and N is the sample size. The penalty is the number of free parameters $p_K = 3K - 1$, counting K means, K variances, and $K - 1$ weights, the last because $\sum_k \pi_k = 1$. The two terms pull in opposite directions: $-\hat{\ell}_K$ is weakly decreasing in K , since a larger mixture can only improve the fit, while $p_K \log N$ grows linearly in K , so the criterion retains a further component only if it buys a likelihood improvement large enough to outweigh the added penalty. BIC is an asymptotic approximation to the log marginal likelihood, or posterior model probability, which is why its complexity penalty scales with $\log N$: as the sample grows, each additional parameter is charged more, and the criterion is consistent, recovering the true dimension asymptotically. This contrasts with the fixed penalty $2p_K$ of Akaike's criterion, which does not grow with N and therefore tends to retain over-parameterized mixtures [Akaike, 1974]; Frühwirth-Schnatter [2006] discusses the identifiability and model-choice problems that make such a balance necessary for finite mixtures. In the empirical analysis three candidates $K \in \{1, 2, 3\}$ are compared within each combined sample and the selected state dimension is reported. The inverse recovery of Subsection 2.4 keeps the two-component loss distribution throughout, so the latent-state interpretation and the long- and short-position risk measures stay comparable across events.

2.4 Inverse recovery of a sparse new regime

Partition an externally dated event study into baseline observations D_O and new-regime observations D_N , with N_O and N_N elements. Let $D_T = D_O \cup D_N$ and define the event weight

$$w = \frac{N_N}{N_O + N_N}. \quad (13)$$

At the population level, the density of the combined sample satisfies the accounting identity

$$p_T(r) = (1 - w)p_O(r) + wp_N(r), \quad (14)$$

which is exact because D_O and D_N partition D_T . This is the same convex-combination structure as (1), but with the baseline and event samples in the role of the latent groups.

Separate EM fits yield stable approximations $\hat{p}_O$ and $\hat{p}_T$. Reconstructing the event density by direct subtraction, $\{\hat{p}_T - (1 - w)\hat{p}_O\}/w$, is not viable for two reasons. The difference of two estimated densities need not be nonnegative, and when w is small the subtraction multiplies the estimation error in $\hat{p}_T$ and $\hat{p}_O$ by $1/w$, so that any noise in the baseline or pooled fit is amplified by exactly the factor that would have been too small to identify the event directly. We therefore search for the event density inside the valid two-component Gaussian family $p_N(r; \vartheta)$, which constrains the recovered object to be a genuine density. Let p_D denote the direct Gaussian-mixture fit to the short event window, and define the inverse estimator as the constrained minimizer

$$\hat{\vartheta}_N = \arg \min_{\vartheta} \{D_{\text{KL}}[\hat{p}_T \parallel (1 - w)\hat{p}_O + wp_N(\cdot; \vartheta)] + \lambda w D_{\text{KL}}[p_D \parallel p_N(\cdot; \vartheta)]\}, \quad (15)$$

where

$$D_{\text{KL}}(p \parallel q) = \int p(r) \log \left\{ \frac{p(r)}{q(r)} \right\} dr. \quad (16)$$

The first term is the substance of the inverse problem: it requires the candidate event density to reproduce the pooled density once combined with the baseline. The second is a weak anchor on the direct fit, weighted by w so that its pull is commensurate with the event's contribution to the pool rather than with the full pooled sample. Because the entropy of $\hat{p}_T$ is constant in ϑ , minimizing the first term is equivalent to maximizing the expected log of the reconstructed density; in this sense the estimator chooses the event density that accounts for the total distribution, not one that reproduces the sparse event histogram.

Two restrictions keep the problem well posed. Each recovered component must carry at least five percent of the event weight, and its standard deviation must be at least a quarter of the smaller direct-fit standard deviation. These bounds exclude the components that a small w would otherwise encourage: those that are negligible in probability mass or degenerate in scale, which are precisely the two forms a weakly identified inverse problem tends to produce. The anchor $\lambda = 0.25$ is small enough that it does not by itself determine the solution.

Two numerical details follow from the same logic. The mixture weights and scales are parameterized to remain strictly inside the admissible space, so the objective in (15) is evaluated only at proper densities. The KL divergences are computed on a sufficiently broad quadrature domain and renormalized on that domain, which makes the divergence consistent and mutually comparable across candidate parameters. Minimization is initialized at the direct event-window estimate and at several dispersed alternatives. Because the objective is exactly the pooled-density reconstruction, an inverse estimate is evaluated by how accurately its candidate event density reproduces the baseline–event decomposition, not by how closely it traces the sparse event histogram. The direct fit remains a valid description of the short window, but it answers a different question.

3 Numerical Schemes for Risk Measures

3.1 Loss conventions for long and short positions

Risk is measured on one-day loss, so positive values denote adverse outcomes. For a unit long position, $L^{\text{long}} = -R$; for a unit short position, $L^{\text{short}} = R$. Gaussian closure under sign changes implies that both loss laws remain mixtures with weights π_k , scales σ_k , and component means $-\mu_k$ for the long or μ_k for the short. This convention prevents the common error of reporting a lower return quantile as a positive loss without changing signs.

For generic mixture loss L with component means a_k , its density and distribution function are

$$f_L(x) = \sum_{k=1}^2 \pi_k \frac{1}{\sigma_k} \phi\left(\frac{x - a_k}{\sigma_k}\right), \quad (17)$$

$$F_L(x) = \sum_{k=1}^2 \pi_k \Phi\left(\frac{x - a_k}{\sigma_k}\right), \quad (18)$$

where ϕ and Φ without location-scale arguments denote the standard Normal density and CDF.

3.2 VaR Numerical Scheme

For confidence $\alpha \in (0, 1)$, Value at Risk is

$$\text{VaR}_\alpha(L) = \inf\{q : F_L(q) \geq \alpha\}. \quad (19)$$

Because a Gaussian mixture is continuous and has a strictly positive density, the quantile is the unique root of $g(q) = F_L(q) - \alpha$. Since $g'(q) = f_L(q)$, Newton–Raphson gives

$$q_{n+1} = q_n - \frac{F_L(q_n) - \alpha}{f_L(q_n)}. \quad (20)$$

The initial value is the corresponding quantile of the moment-matched Gaussian,

$$\mu_L = \sum_k \pi_k a_k, \quad (21)$$

$$s_L^2 = \sum_k \pi_k (\sigma_k^2 + a_k^2) - \mu_L^2, \quad (22)$$

$$q_0 = \mu_L + s_L \Phi^{-1}(\alpha). \quad (23)$$

Because the denominator in (20) is a density and can be very small in the extreme tail, an unmodified Newton step can leave the region in which the quantile lies. The root is therefore confined to a bracket $[l, u]$ widened until $F_L(l) < \alpha < F_L(u)$; any non-finite or out-of-bracket proposal is replaced by the midpoint $(l + u)/2$, and the bracket is tightened after each accepted step. The scheme thus pairs the local speed of Newton's method with the global reliability of bisection. Iteration terminates when the CDF residual or the step size falls below 10^{-12} in scaled units.

3.3 Closed-form Expected Shortfall

Expected Shortfall is the average loss in the upper $(1 - \alpha)$ tail [Artzner et al., 1999, Acerbi and Tasche, 2002]:

$$\text{ES}_\alpha(L) = \mathbb{E}[L \mid L > q_\alpha] = \frac{1}{1 - \alpha} \int_{q_\alpha}^{\infty} x f_L(x) dx. \quad (24)$$

For component k , put $z_k = (q_\alpha - a_k)/\sigma_k$ and substitute $x = a_k + \sigma_k z$. Using $\int_{z_k}^{\infty} z \phi(z) dz = \phi(z_k)$ gives

$$\int_{q_\alpha}^{\infty} x \frac{1}{\sigma_k} \phi\left(\frac{x - a_k}{\sigma_k}\right) dx = a_k \{1 - \Phi(z_k)\} + \sigma_k \phi(z_k), \quad (25)$$

$$\text{ES}_\alpha(L) = \frac{1}{1 - \alpha} \sum_{k=1}^2 \pi_k [a_k \{1 - \Phi(z_k)\} + \sigma_k \phi(z_k)]. \quad (26)$$

Thus only VaR requires root finding. For the benchmark $L \sim \mathcal{N}(\mu_L, s_L^2)$,

$$\text{VaR}_\alpha^N = \mu_L + s_L z_\alpha, \quad \text{ES}_\alpha^N = \mu_L + s_L \frac{\phi(z_\alpha)}{1 - \alpha}, \quad z_\alpha = \Phi^{-1}(\alpha). \quad (27)$$

4 Empirical Analysis

4.1 Data and empirical design

The spot-price series used to compute returns are adjusted for corporate actions before calculating daily continuously compounded returns. This adjustment prevents mechanical discontinuities associated with splits, dividends, or comparable adjustments from being misidentified as genuine spot-price jumps. Daily continuously compounded returns are

$$R_t = \log(P_t/P_{t-1}). \quad (28)$$

The event definitions in Table 1 use externally dated baseline and new-regime windows. D_T is the union of the stated baseline and new windows rather than every calendar date between disjoint endpoints; this preserves the identity in (14). Estimation uses decimal returns and tables report percentages. The reported density, posterior-state, and tail-risk figures provide complementary views of the same fitted distributions.

The cases differ sharply in both location and dispersion. Both S&P crisis windows have negative means and daily volatility near four percent. ChatGPT has a positive 0.684% mean,

Table 1: Event-study design and realized new-regime moments

Case	Baseline window	New-regime window	N_O	N_N	Mean (%)	Volatility (%)
S&P 500: GFC	Jan. 2004–Aug. 2008	Sep.–Dec. 2008	1,174	85	-0.413	4.022
S&P 500: COVID-19	Jan. 2015–Jan. 2020	Feb.–Apr. 2020	1,279	62	-0.165	3.883
NVIDIA: ChatGPT	Jan. 2020–28 Nov. 2022	30 Nov. 2022–Mar. 2023	732	84	0.684	3.491
NVIDIA: DeepSeek	Jan. 2020–Dec. 2024	Jan.–Mar. 2025	1,257	60	-0.357	4.292
CSI 1000: stimulus	Jan. 2018–23 Sep. 2024	24 Sep.–Nov. 2024	1,633	44	0.740	3.175

whereas DeepSeek has a negative -0.357% mean and the larger NVIDIA event volatility. The CSI 1000 event combines a strongly positive 0.740% mean with 3.175% volatility. These moments already suggest that the first conditional moment cannot summarize the relevant risk: the same positive average return may coexist with a materially wider distribution and a more consequential reversal tail.

4.2 Inverse-regime parameters and reconstruction performance

Table 2: Inverse two-component event-density estimates (daily percentages)

	π_1	μ_1	σ_1	π_2	μ_2	σ_2
S&P: GFC	0.1023	4.515	1.188	0.8977	-1.092	3.782
S&P: COVID-19	0.2049	-4.263	0.941	0.7951	0.200	3.655
NVIDIA: ChatGPT	0.9499	0.402	3.153	0.0501	5.884	5.422
NVIDIA: DeepSeek	0.0501	0.851	1.584	0.9499	-0.360	4.236
CSI 1000: stimulus	0.5465	0.799	1.879	0.4535	0.327	4.034

Table 2 reports the components of the recovered event-regime density. They are latent economic states defined by their conditional location and scale, not observed labels attached to individual trading days, and the table describes no transition dynamics among them. In the GFC and COVID-19 estimates the component locations are well separated, assigning probability to materially different market conditions ranging from severe downside to rebound or high-dispersion days; the crisis distribution is therefore not adequately described by a single high-variance Normal. The CSI 1000 estimate combines a positive-location state with a more diffuse one, consistent with a policy-led repricing in which optimistic information shifts the center of the distribution while uncertainty, crowded positioning, and sharp reversals preserve a substantial tail. For NVIDIA, a broad dominant state is accompanied by a smaller state that absorbs event-specific changes in location or dispersion.

Table 3: BIC comparison for each combined sample; lower is better

Case	$K = 1$	$K = 2$	$K = 3$	Selected K
S&P: GFC	-7263.9	-7931.3	-7951.7	3
S&P: COVID-19	-8098.2	-8813.1	-8876.7	3
NVIDIA: ChatGPT	-3132.3	-3151.8	-3141.1	2
NVIDIA: DeepSeek	-5139.6	-5227.0	-5217.2	2
CSI 1000: stimulus	-9137.9	-9340.8	-9338.6	2

Table 3 rejects a single Gaussian in every case. Three components are selected for both S&P crises, where very long, calm histories are combined with extreme event observations, and two for ChatGPT, DeepSeek, and the CSI 1000 index. The two-component inverse target is

nevertheless retained throughout, so that the economic interpretation and the dimension of the risk measures stay constant across events. The S&P result indicates that a separation into calm, intermediate, and crisis states may fit broad-index data better than a baseline paired with a single event state.

Table 4: KL reconstruction comparison; the direct event-window GMM is retained as a benchmark, while density figures display the baseline and pooled GMMs.

Case	Direct GMM	Inverse GMM	Reduction (%)
S&P: GFC	0.024915	0.022405	10.1
S&P: COVID-19	0.032638	0.022481	31.1
NVIDIA: ChatGPT	0.000905	0.000145	84.0
NVIDIA: DeepSeek	0.000772	0.000273	64.6
CSI 1000: stimulus	0.000872	0.000494	43.3

The inverse estimator’s reconstruction KL is compared with the direct benchmark in Table 4. The comparison is not a visual goodness-of-fit contest against the event histogram: the Direct GMM is expressly optimized for that histogram, whereas the Inverse GMM is optimized for the pooled-density target. Both reported KL entries measure the same reconstruction object, $D_{\text{KL}}(\hat{p}_T || \hat{p}_R)$, where $\hat{p}_R$ combines the fixed baseline fit with an event-regime estimate. The distinction is substantive. The direct fit describes the observations available in the short window; the inverse fit explains how the event regime must differ from the baseline in order to account for the observed total distribution.

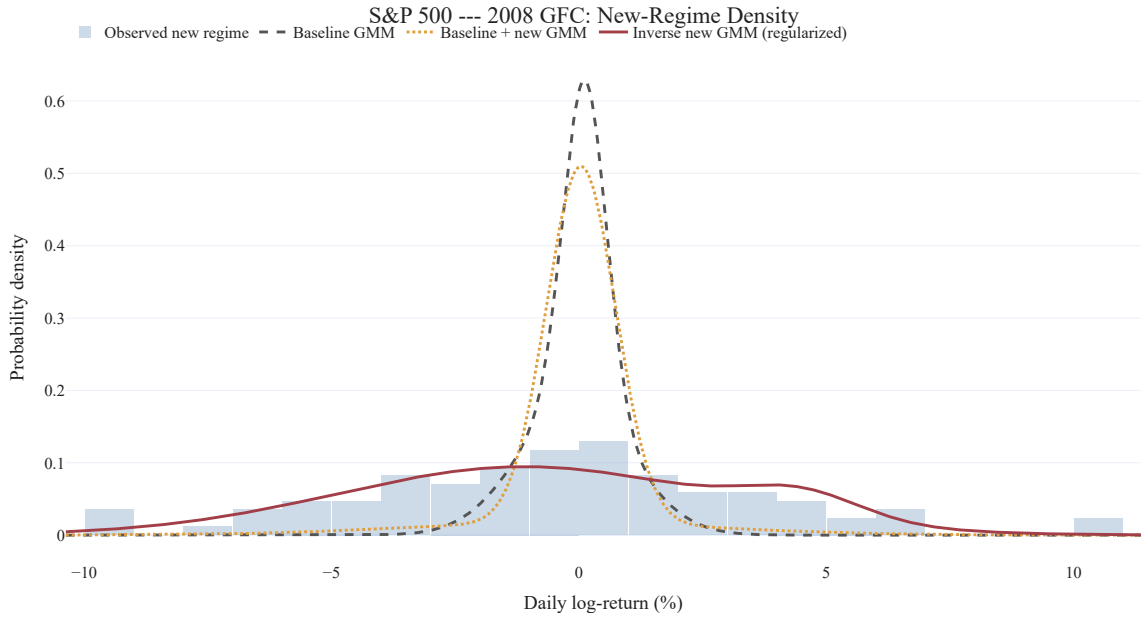


Figure 1: S&P 500 GFC event study: observed new-regime returns and fitted densities. The Baseline GMM is fitted to the baseline window, the Baseline + new GMM to the combined window, and the regularized Inverse GMM recovers the new-regime density from their reconstruction.

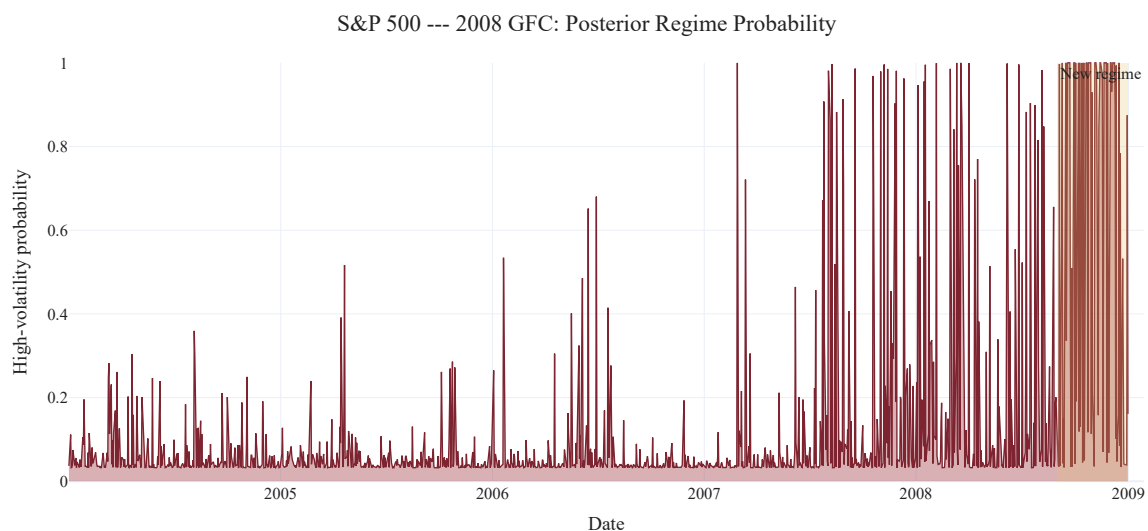


Figure 2: S&P 500 GFC event study: posterior probability of the high-volatility component. The shaded interval is the new regime.

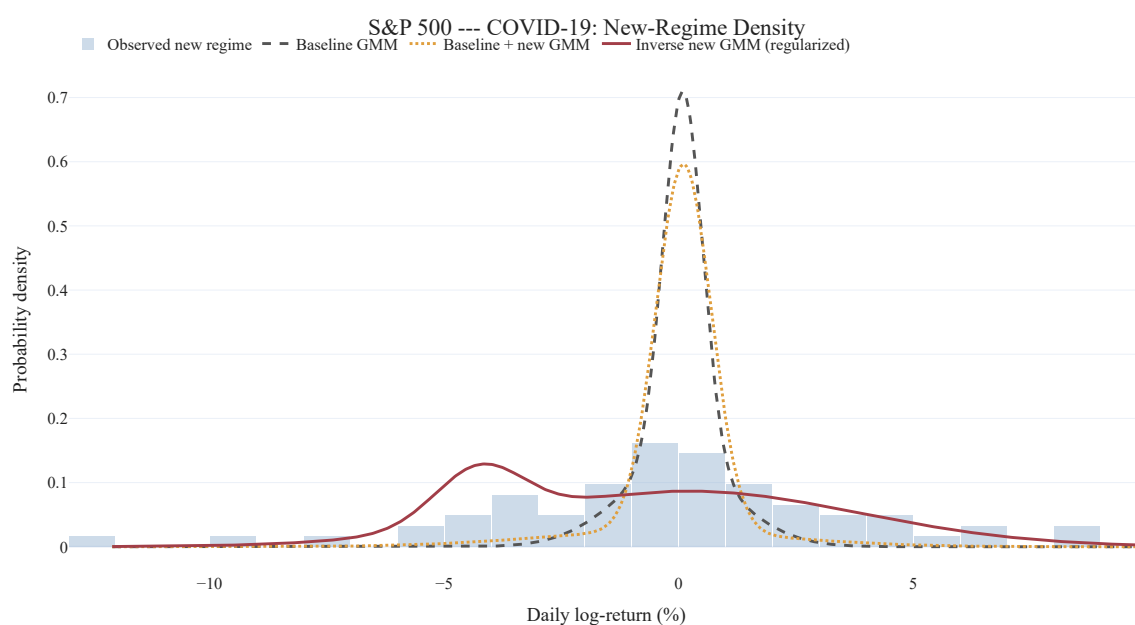


Figure 3: S&P 500 COVID-19 event study: observed new-regime returns and fitted densities. The Baseline GMM is fitted to the baseline window, the Baseline + new GMM to the combined window, and the regularized Inverse GMM recovers the new-regime density from their reconstruction.

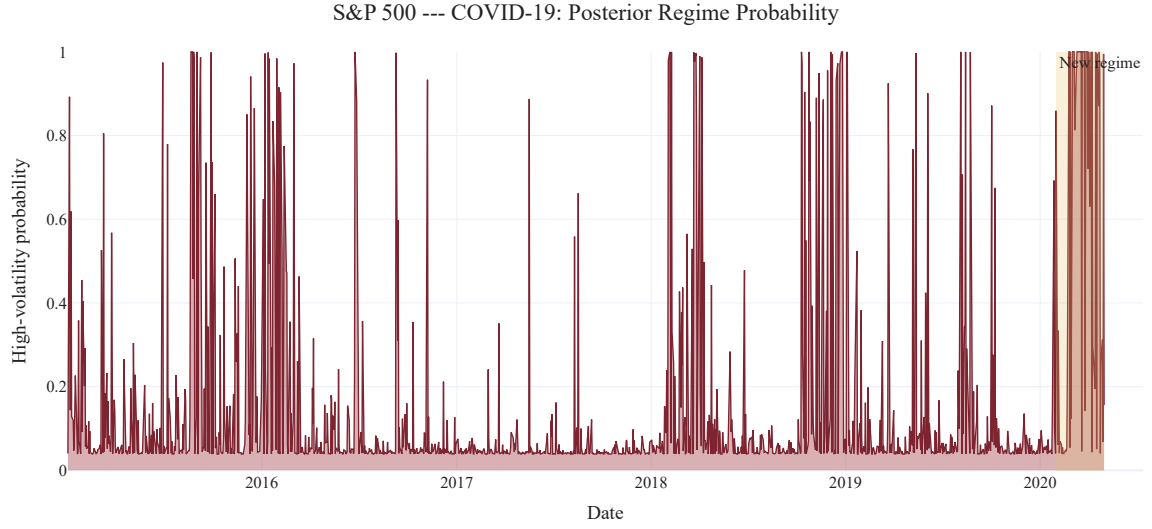


Figure 4: S&P 500 COVID-19 event study: posterior probability of the high-volatility component. The shaded interval is the new regime.

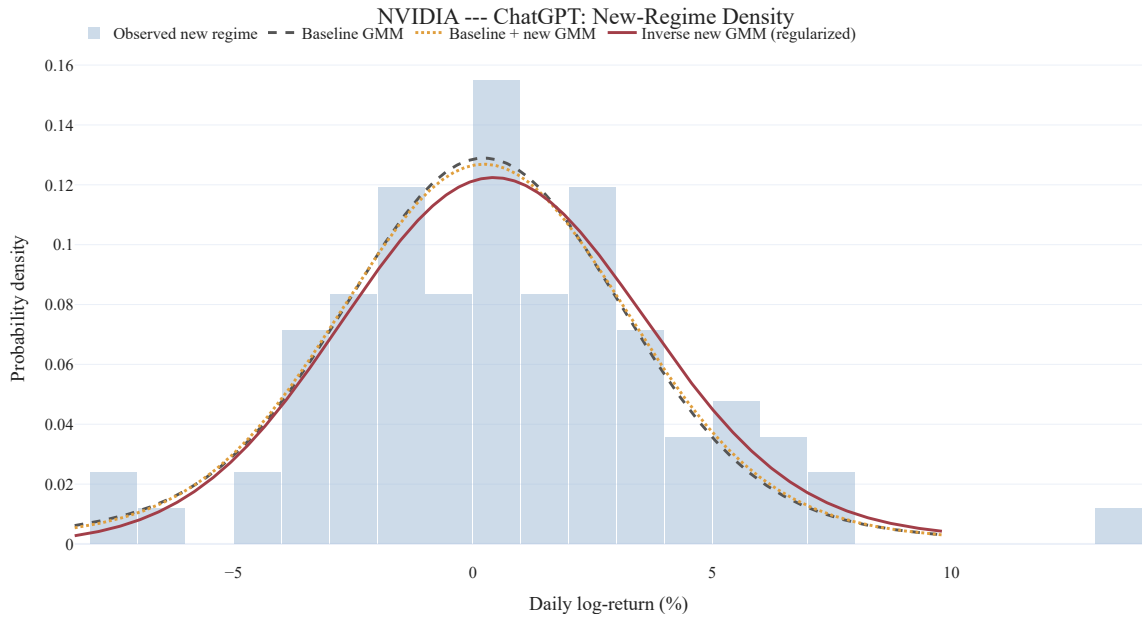


Figure 5: NVIDIA ChatGPT event study: observed new-regime returns and fitted densities. The Baseline GMM is fitted to the baseline window, the Baseline + new GMM to the combined window, and the regularized Inverse GMM recovers the new-regime density from their reconstruction.

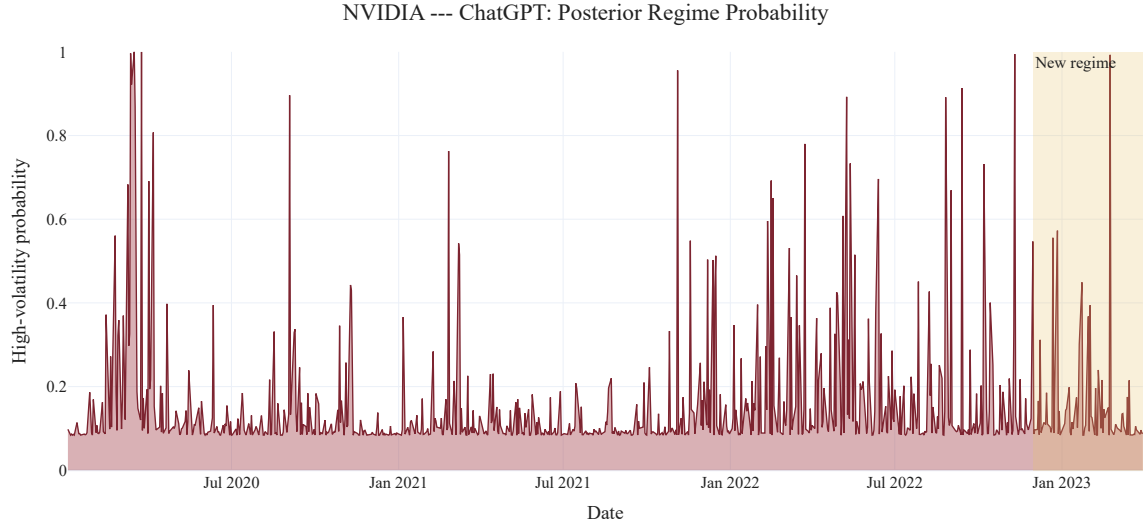


Figure 6: NVIDIA ChatGPT event study: posterior probability of the high-volatility component. The shaded interval is the new regime.

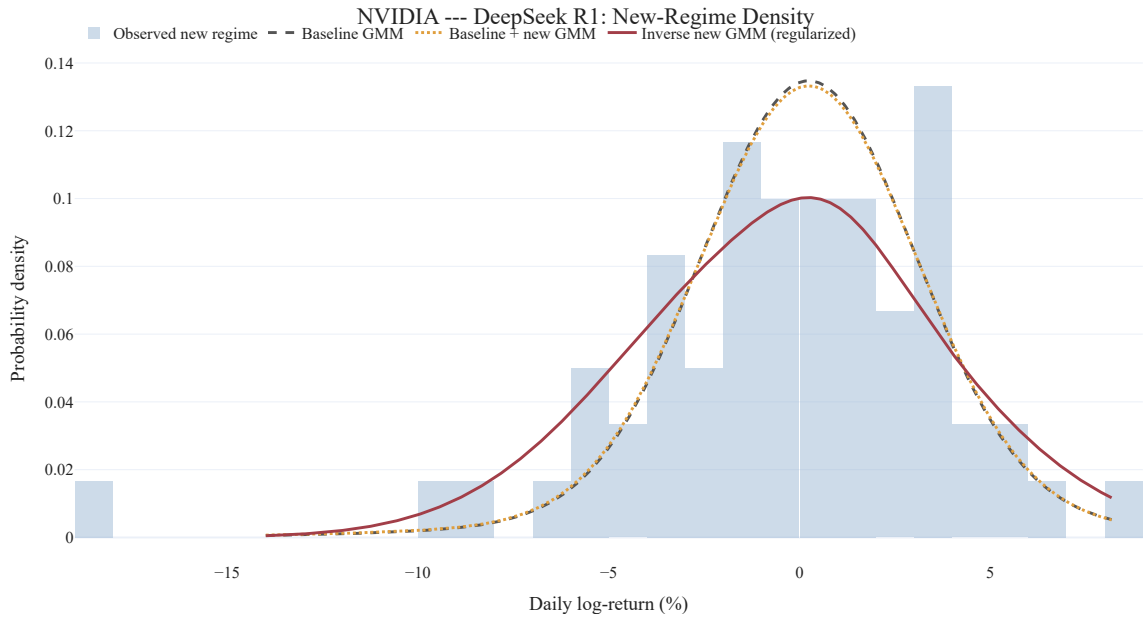


Figure 7: NVIDIA DeepSeek R1 event study: observed new-regime returns and fitted densities. The Baseline GMM is fitted to the baseline window, the Baseline + new GMM to the combined window, and the regularized Inverse GMM recovers the new-regime density from their reconstruction.

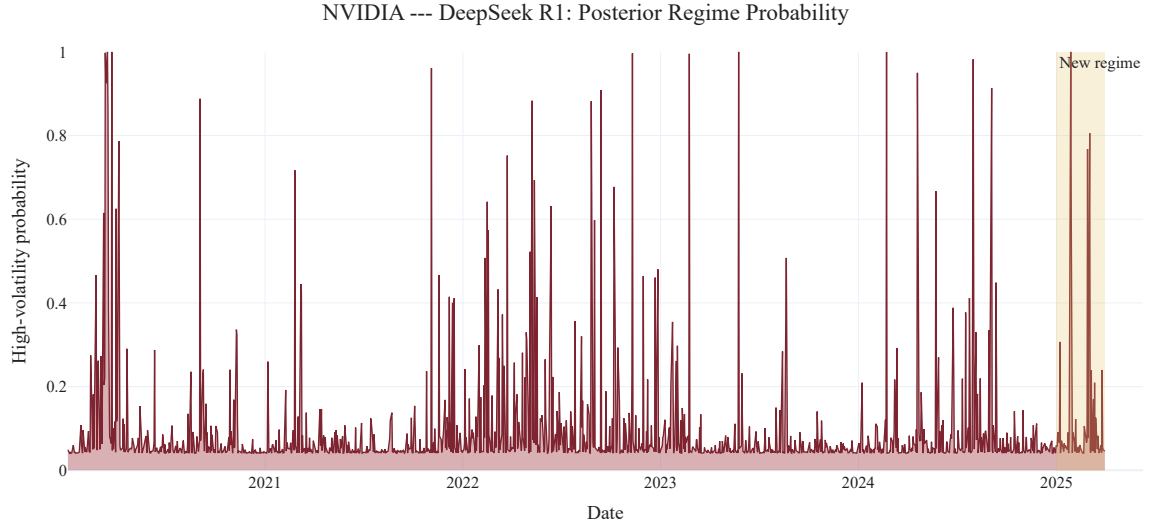


Figure 8: NVIDIA DeepSeek R1 event study: posterior probability of the high-volatility component. The shaded interval is the new regime.

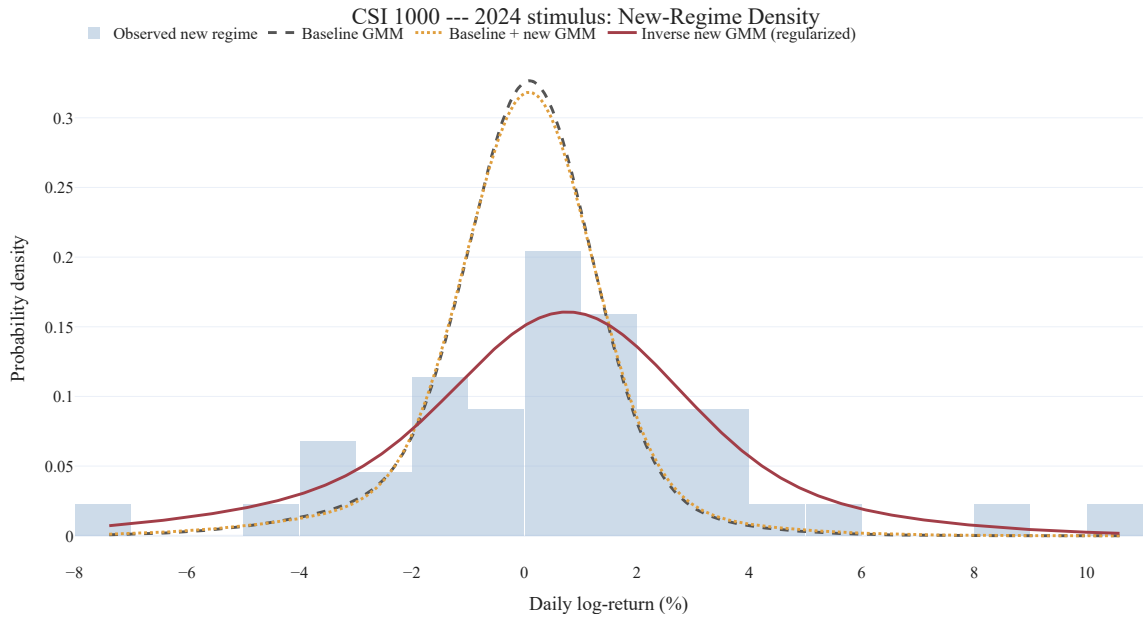


Figure 9: CSI 1000 index stimulus event study: observed new-regime returns and fitted densities. The Baseline GMM is fitted to the baseline window, the Baseline + new GMM to the combined window, and the regularized Inverse GMM recovers the new-regime density from their reconstruction.

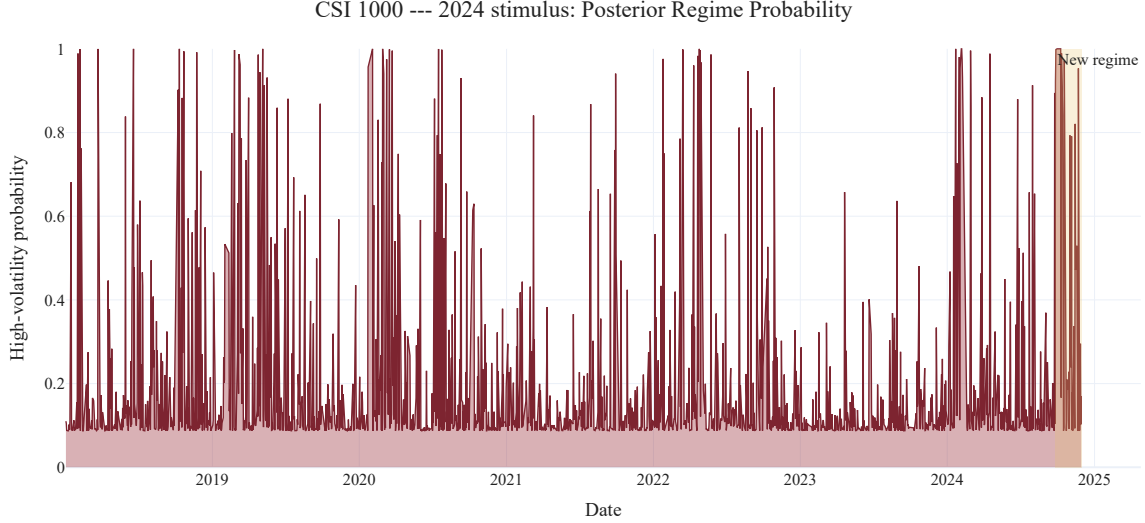


Figure 10: CSI 1000 index stimulus event study: posterior probability of the high-volatility component. The shaded interval is the new regime.

4.3 Comparative tail-risk evidence

Table 5 and the figures that follow compare four candidates for the event-window loss distribution. *Total* is the two-component Gaussian mixture fitted to the pooled baseline-and-event sample, which averages pre-event and event dynamics. *Inverse* is the event-regime mixture recovered by the inverse estimator of Section 2.4, which isolates the event contribution while preserving the pooled-density identity. *Direct* is the two-component mixture fitted to the event-window returns alone and is retained as a benchmark for what fitting only the short window can deliver. *New Normal* — shown as “Normal” in the table and as the final bar of each figure — is the single Gaussian distribution fitted to the event-window returns, with VaR and ES evaluated in closed form by (27) from the sample mean and standard deviation of the event-window returns. The name records exactly what it is: the Normal distribution of the new regime. It supplies the conventional thin-tailed counterfactual, the tail risk implied by treating the short event window as Gaussian; the mixture estimates, which preserve the event window’s asymmetry and excess kurtosis, should be read against it.

Table 5: One-day 99% ES as a percentage of position value

Case	Long position					Short position		
	Total	Inverse	Direct	New Normal	Inv./Total	Total	Inverse	Inv./Total
S&P: GFC	6.416	11.033	11.638	11.131	1.72	5.643	8.867	1.57
S&P: COVID-19	5.502	9.254	12.608	10.514	1.68	4.709	9.654	2.05
NVIDIA: ChatGPT	11.515	8.030	8.067	8.619	0.70	11.436	13.532	1.18
NVIDIA: DeepSeek	11.928	11.576	18.659	11.796	0.97	11.940	10.855	0.91
CSI 1000: stimulus	6.052	9.292	9.724	7.723	1.54	5.229	9.948	1.90

The broad-market crises display the clearest amplification. GFC inverse long ES is 11.033%, versus 6.416% in the combined model, and inverse short ES is 8.867%. COVID-19 produces a similar pattern: long and short inverse ES are 9.254% and 9.654%, respectively. The recovered event states therefore assign materially more probability to adverse losses than does the pooled distribution. Their ES values remain below the direct short-window estimates, which is consistent with using the longer baseline to distinguish persistent state change from sampling variation in a sparse crisis window.

NVIDIA reveals why cross-asset inference must be directional. ChatGPT-period returns average 0.684% daily. Its inverse long ES falls to 8.030%, or 0.70 times total-sample ES, while inverse short ES rises modestly to 13.532%. This asymmetry is consistent with favorable adoption news improving the right side of the return distribution while increasing the risk of adverse repricing for short positions. DeepSeek returns average -0.357%, but the longer baseline absorbs much of the event’s contribution to the pooled density: inverse 99% long and short ES are 11.576% and 10.855%, respectively, equal to 0.97 and 0.91 of total-sample ES. The two NVIDIA episodes thus illustrate distributional rotation rather than a uniform volatility response; the sign and information content of the shock matter for which tail is exposed.

The CSI 1000 provides the clearest illustration that a positive mean does not imply lower tail risk. The inverse long and short ES rise to 9.292% and 9.948%, factors of 1.54 and 1.90 over total-sample risk. The economic mechanism is two-sided. The policy announcement can shift the conditional mean to the right through repricing and improved expectations, while simultaneously increasing dispersion through crowded trades, rapid liquidity reallocation, and uncertainty about the durability and transmission of the policy impulse. For a long position, this creates a material left-tail reversal risk after a sharp rally; for a short position, continuation of the policy-driven rally creates a large right-tail loss. In mixture terms, the mean shift is accompanied by greater within-state dispersion and a larger separation between latent states, so the between-state component of variance and the extreme loss probabilities can both increase.

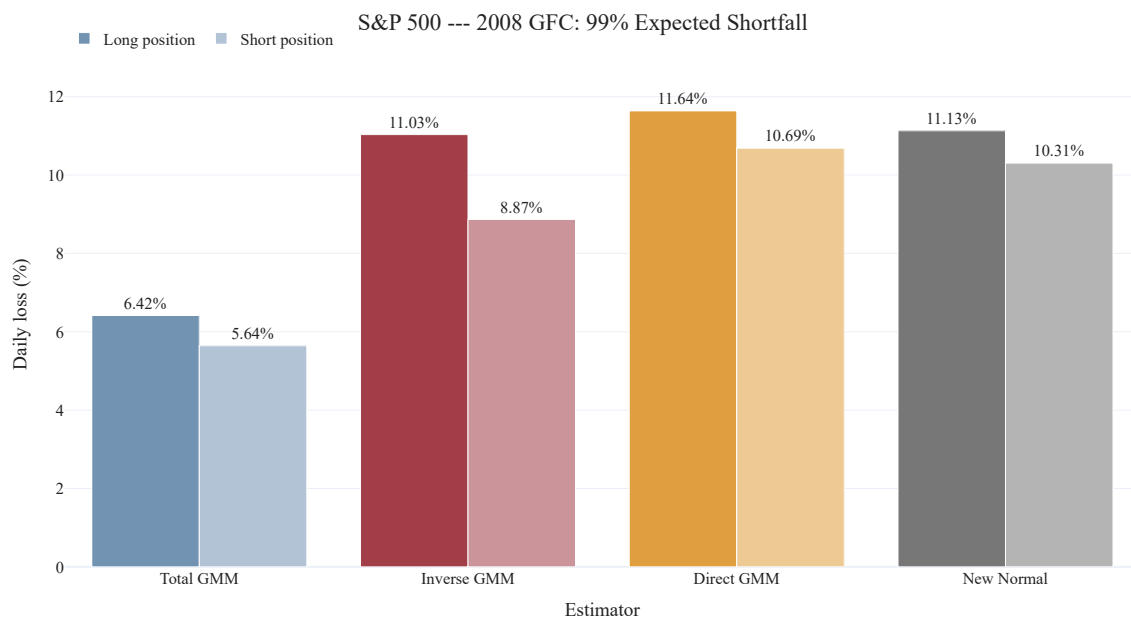


Figure 11: S&P 500 GFC event study: comparison of 99% Expected Shortfall for long and short positions under the pooled (Total), inverse, direct, and Gaussian (New Normal) event-window estimates.

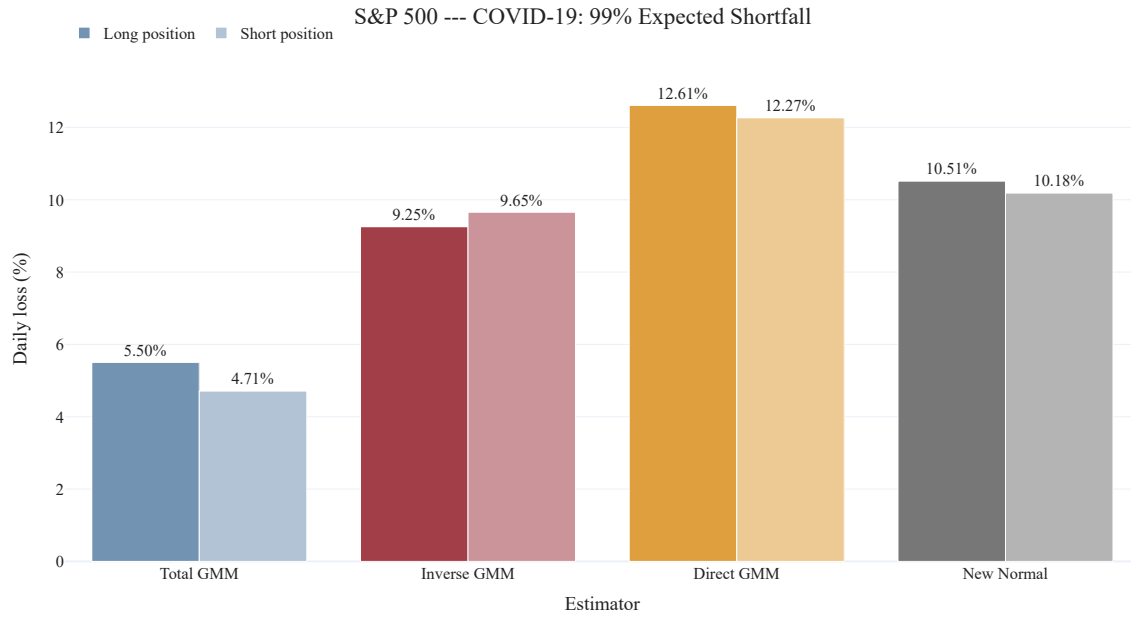


Figure 12: S&P 500 COVID-19 event study: comparison of 99% Expected Shortfall for long and short positions under the pooled (Total), inverse, direct, and Gaussian (New Normal) event-window estimates.

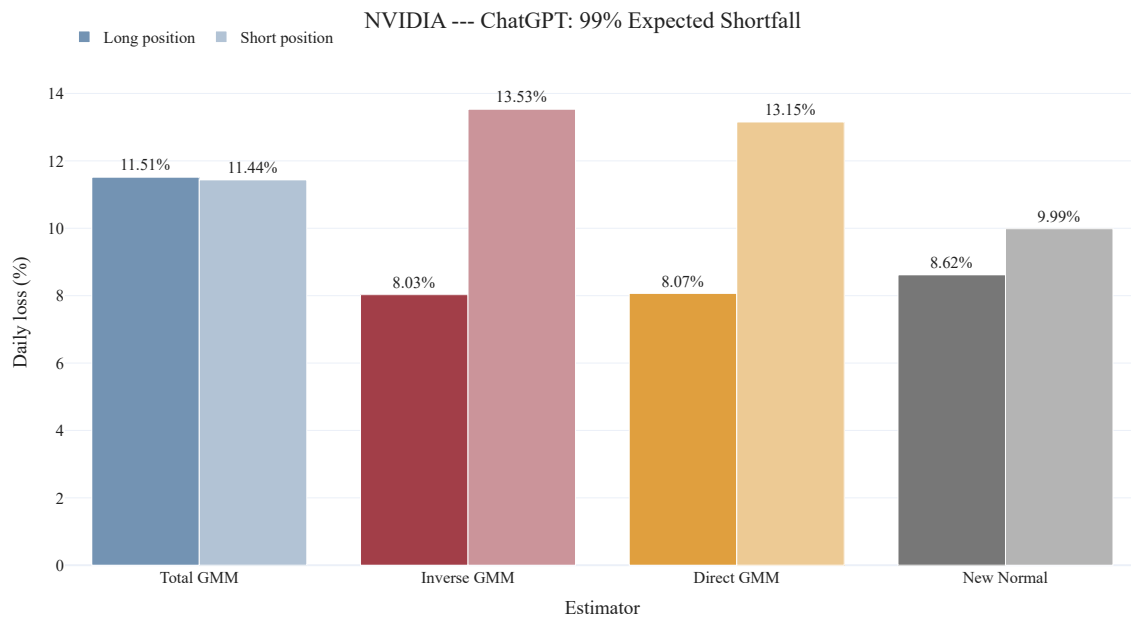


Figure 13: NVIDIA ChatGPT event study: comparison of 99% Expected Shortfall for long and short positions under the pooled (Total), inverse, direct, and Gaussian (New Normal) event-window estimates.

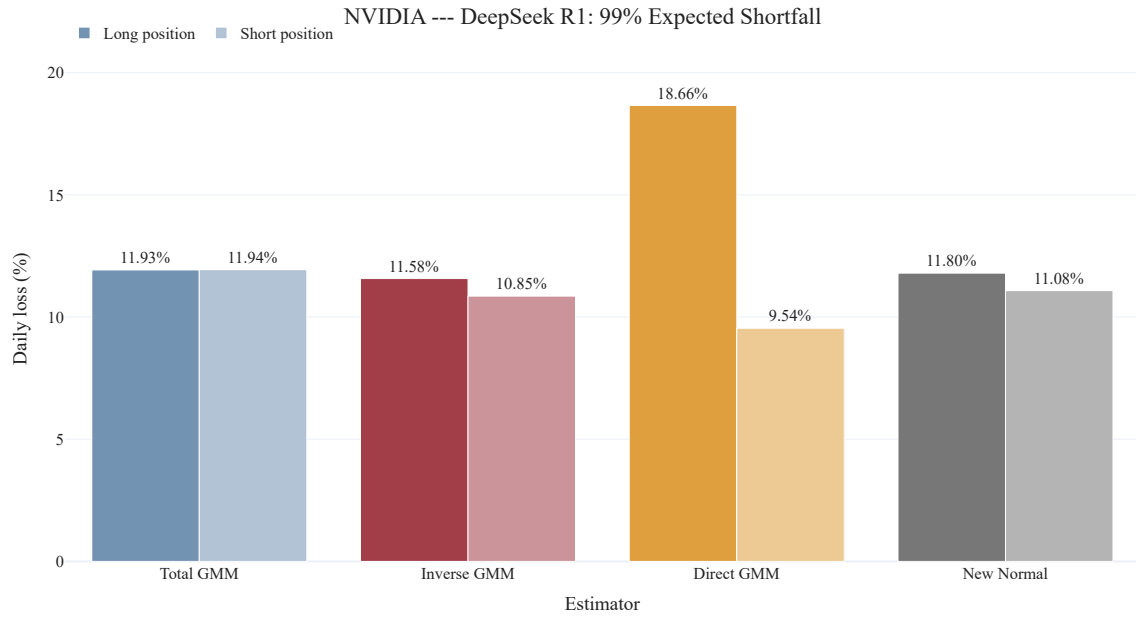


Figure 14: NVIDIA DeepSeek R1 event study: comparison of 99% Expected Shortfall for long and short positions under the pooled (Total), inverse, direct, and Gaussian (New Normal) event-window estimates.

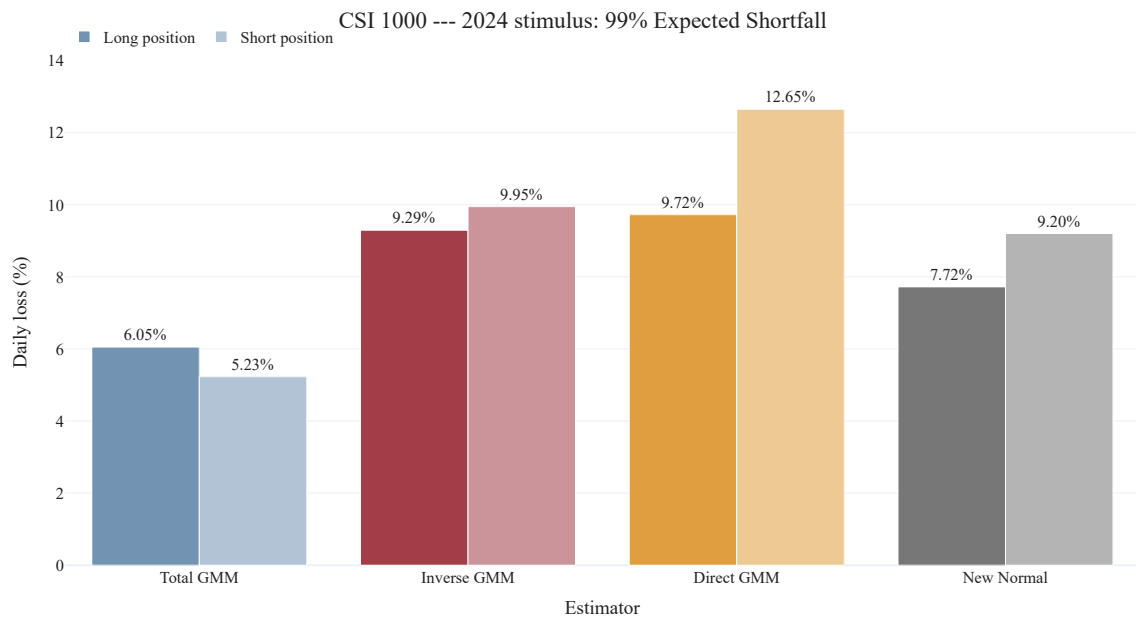


Figure 15: CSI 1000 index stimulus event study: comparison of 99% Expected Shortfall for long and short positions under the pooled (Total), inverse, direct, and Gaussian (New Normal) event-window estimates.

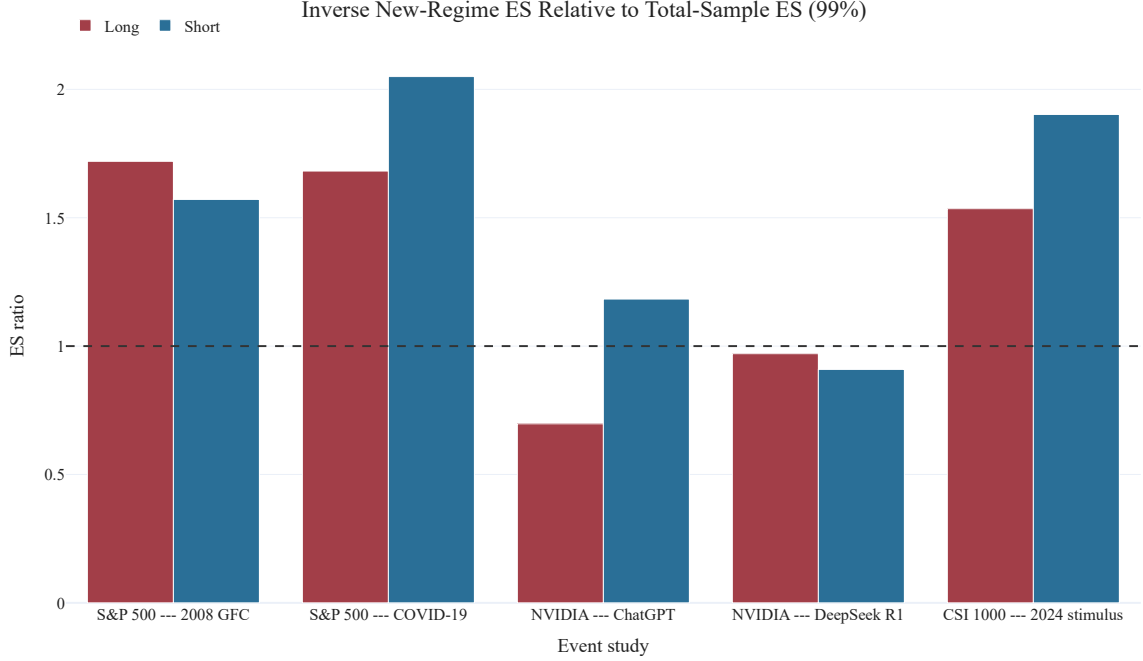


Figure 16: Cross-event inverse-to-total 99% ES ratios.

4.4 Limitations and dynamic-risk interpretation

The inverse estimator is a static, ex post object rather than a filter. It recovers an unconditional density over a pre-dated window, and the responsibilities plotted in the posterior figures are retrospective classifications, not filtered state probabilities: nothing in the model carries a state forward from one day to the next, so it supplies no transition probabilities and no persistence. Rolling deployment is therefore a repeated computation — recompute p_O , p_T , and w when a new observation or a newly declared regime arrives and solve the inverse problem again — rather than an online update. The natural extension is a sequential hidden-Markov mixture that propagates its states, but that requires enough within-window data to identify a transition matrix, which is precisely what the sparse event windows do not supply [Hamilton, 1989, Frühwirth-Schnatter, 2006].

The binding limits are properties of the model class, not of the estimator’s implementation. Each component is Gaussian, so the recovered density is asymptotically thin-tailed: beyond roughly three or four standard deviations of the most volatile component, the mixture tail decays like $\exp(-x^2/2\sigma^2)$ and cannot reproduce the power-law behaviour that Cont [2001] and Adams and Thornton [2013] emphasise for equity returns. The inverse estimator therefore reweights tail mass already present in the baseline and pooled Gaussian fits; it cannot manufacture tails heavier than its components allow. This is the standing caveat behind Gaussian-based risk measurement, in which a Normal VaR understates the frequency and severity of extreme losses [Adams and Thornton, 2013]. Fixing $K = 2$ compounds the restriction: the BIC results show that the S&P crises prefer three states, and the component count is held fixed across events so that the long- and short-position risk measures remain comparable, at the cost of some in-sample fit. A mixture with heavy-tailed components, such as Student- t , is the natural relaxation, but it abandons the closed-form risk functionals in (26) used here.

Within these limits, “dynamic risk management” should be read narrowly. It denotes a disciplined, event-time updating of an estimated loss distribution: when a new regime is declared or a new observation arrives, the baseline and pooled mixtures are recomputed, the inverse problem is re-solved, and the VaR and ES are read off a distribution that reflects the prevailing

state. It is not a claim that the present iid mixture models persistence or forecasts the next state. The framework is best understood as a diagnostic that identifies where and in which direction a shock moved the loss distribution, rather than as a predictive filter.

5 Conclusion

This paper extends conventional mixture risk measurement to the sparse-regime inverse problem. Rather than estimate a fragile event density in isolation or dilute it in a long pooled sample, the method recovers a valid two-component GMM that reconciles the baseline and total-sample densities. The resulting distribution has a direct interpretation in terms of latent economic states and supports internally consistent calculation of long- and short-position VaR and ES.

Across five event studies, the S&P 500 crises and China’s stimulus boom exhibit large increases in both long- and short-position tail risk, although their economic mechanisms differ. NVIDIA shows that regime detection is not synonymous with symmetric amplification: ChatGPT lowers long-loss ES but raises short-loss ES modestly, whereas the longer-baseline DeepSeek estimates leave long ES near its pooled level and modestly reduce short ES. This heterogeneity is the central cross-asset result.

The inverse GMM is therefore a useful tool for identifying and quantifying event-specific distributional change when risk managers must update from few observations. Its value is not that it imposes a uniformly more conservative estimate, but that it preserves mathematical consistency while making the source of tail-risk change explicit: volatility amplification, a directional rotation, or a change in the relative weight of latent states. A natural next step is a sequential inverse hidden-Markov mixture with BIC-selected state dimension and out-of-sample VaR/ES backtesting.

References

- Carlo Acerbi and Dirk Tasche. On the coherence of expected shortfall. *Journal of Banking & Finance*, 26(7):1487–1503, 2002.
- Mike Adams and Barry Thornton. Black swans and VaR. *Journal of Finance and Accountancy*, 14:1–17, 2013.
- Hirotsugu Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.
- Andrew Ang and Allan Timmermann. Regime changes and financial markets. *Annual Review of Financial Economics*, 4:313–337, 2012.
- Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, 1999.
- Scott R. Baker, Nicholas Bloom, Steven J. Davis, Kyle Kost, Marco Sammon, and Tasaneeya Viratyosin. The unprecedented stock market reaction to COVID-19. *Review of Asset Pricing Studies*, 10(4):742–758, 2020.
- Rama Cont. Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, 2001.
- J. R. Contreras. Market risk measures using finite Gaussian mixtures. *Journal of Applied Finance and Banking*, 4(6):29–45, 2014.
- Arthur P. Dempster, Nan M. Laird, and Donald B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, 39(1):1–38, 1977.
- Sylvia Frühwirth-Schnatter. *Finite Mixture and Markov Switching Models*. Springer, New York, 2006.
- James D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–384, 1989.
- Geoffrey J. McLachlan and David Peel. *Finite Mixture Models*. Wiley, New York, 2000.
- Alexander J. McNeil, Ruediger Frey, and Paul Embrechts. *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, revised edition, 2015.
- Indrė Morkūnaitė, Dmitrij Celov, and Remigijus Leipus. Evaluation of value-at-risk (VaR) using the Gaussian Mixture Models. *Research in Statistics*, 2(1):1–14, 2024. doi: 10.1080/27684520.2024.2346075.
- Gideon Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.
- C. F. Jeff Wu. On the convergence properties of the EM algorithm. *The Annals of Statistics*, 11(1):95–103, 1983.

Acknowledgement

In presenting this paper, I would like to express my sincere gratitude to everyone who supported me throughout this research. I thank my advisor, Di MA, for serving as my school advisor during this project. Although the subject-specific development of the work was outside her academic field, she encouraged me to pursue the project seriously, reminded me to keep a clear research schedule, and offered general comments on the organization and presentation of my writing. I am also grateful to my math teacher Guodong JIANG and computer science teacher Robert Morrison, for their comments and encouragement, and to my family and classmates for their patience and support during the long process of reading, coding, checking, and revising.

The motivation for this project came from my own interest in the way financial risk changes when the market enters a new regime. Both of my parents work in frontline financial roles, so discussions about markets, price movements, and the practical consequences of risk have been part of my everyday environment for a long time. Those conversations did not provide the model or the conclusions of this paper, but they encouraged me to connect abstract statistical ideas with questions that matter in real financial decision-making. A crisis, a technology shock, and a policy intervention can all alter the return distribution, but they need not affect the two tails in the same way. I became particularly interested in the practical difficulty of measuring a new regime when its event window is short: a long historical sample is stable but may no longer describe current conditions, while a short sample is relevant but can make a flexible model overreact to a few extreme observations. Pooling the two periods creates the opposite problem by diluting the new regime. I therefore asked whether the new distribution could be recovered indirectly from the relationship between the baseline distribution and the total distribution, rather than estimated from the short event window alone.

To develop this idea, I first studied probability distributions, Gaussian mixture models, expectation-maximization, Kullback-Leibler divergence, Value at Risk, and Expected Shortfall. The central observation was the density identity $p_T = (1 - w)p_O + wp_N$, which states that the total distribution is a weighted combination of the baseline and new-regime distributions. From this identity, I formulated the recovery of a sparse event regime as an inverse estimation problem. I then designed the study to examine whether a single mathematically consistent framework could distinguish genuine tail amplification from a directional rotation of risk across broad-market crises, firm-specific technology events, and a policy-driven movement in Chinese small-capitalization equities.

I was responsible for the main theoretical work in this project. I formulated the inverse density-recovery problem, derived the baseline-total-new-regime decomposition, and analyzed why direct subtraction can produce a negative or unstable density when the event share is small. I also established the restrictions needed to prevent the inverse solution from collapsing into an implausibly narrow component. For the risk-measurement part, I defined the loss conventions for long and short positions, derived the closed-form Expected Shortfall for a Gaussian mixture, and implemented a numerical procedure for the corresponding VaR quantile.

I also completed the computational and empirical work myself. I wrote the Python implementation of the EM estimator and BIC comparisons, the regularized inverse-KL optimizer, the long- and short-position risk measures. I prepared the corporate action adjusted spot return series from the publicly available price data, defined the five event studies, fitted the baseline, pooled, direct, and inverse models, and generated the tables and figures used in the paper. I tested the code on synthetic mixtures, checked the monotonicity and convergence of EM, verified that the VaR solution satisfied the target CDF equation, compared the analytic Expected Shortfall with numerical integration, and examined whether inverse recovery improved reconstruction relative to a direct fit of the short event window. I then interpreted the empirical results and wrote the first complete version of the manuscript.

Several difficulties arose during the project. The first was that the algebraic residual suggested by the density identity was not automatically a valid probability density; it could

become negative, and the division by a small event weight amplified estimation error. I addressed this by returning to the definition of a density and replacing direct subtraction with a constrained search over proper Gaussian mixtures. A second difficulty was numerical weak identification. In some trials, a very narrow component with little probability mass could improve the objective without representing a meaningful event regime. By varying the initial values and examining the fitted weights and scales, I identified this failure mode and introduced minimum component weights, lower bounds on component scales, and multiple dispersed starting points. I also had to be careful with the opposite loss conventions of long and short positions. I resolved this by transforming the component means explicitly and checking the results against the mixture CDF and numerical integration. These difficulties taught me that a plausible-looking numerical output is not enough; every result must be connected to a definition, an assumption, and an independent check.

This project has been an opportunity for me to learn not only how to build a statistical model, but also how to conduct research independently: to begin with a question, return to definitions when a method fails, test an implementation from more than one direction, and state clearly what the evidence can and cannot show. I am deeply grateful to all those who supported me during this process and helped me complete this work.