DSSR

Discussion Paper No. 156

Fitting Dynamic Conditional Densities to Cross-Sectional Returns with Persistent Tail Intensities

Yasumasa Matsuda and Keigo Sato

October, 2026

Data Science and Service Research
Discussion Paper

Center for Data Science and Service Research
Graduate School of Economic and Management
Tohoku University
27-1 Kawauchi, Aobaku
Sendai 980-8576, JAPAN

Fitting Dynamic Conditional Densities to Cross-Sectional Returns with Persistent Tail Intensities

Yasumasa Matsuda and Keigo Sato
Graduate School of Economics and Management
Tohoku University, Japan

Correspondence: Yasumasa Matsuda, Graduate School of Economics and Management, Tohoku University, Japan. Email: yasumasa.matsuda.a4@tohoku.ac.jp

Abstract

We propose dynamic conditional densities with spline mixtures (DCD-SM) to detect whether the tail intensity of financial returns depends on past information after accounting for volatility and tail-order dynamics. The target is a weighted mixture of the conditional return densities of individual assets in a cross section. Individual distributions may differ, and returns within a date may be dependent. A normalized spline-mixture correction uses score-driven updates to change left and right tail intensities while preserving the reference tail order. We use Wald tests for persistence and assess empirical size, power, and forecast stability in simulations. With fitted references that have dynamic tail order, Wald tests at the nominal five-percent level detect dynamics that affect both tail intensities in the same direction in all 19 US and Japanese equity cross sections, for both raw returns and residuals after principal-component removal with training-fixed loadings. Evidence for dynamics in the imbalance between the two tail intensities is more limited. In the residuals, the correction improves out-of-sample one-percent lower-tail quantile loss relative to these references in 16 of 19 panels, with three significant positive differences at the five-percent level. Forecast gains are most evident in log scores and one-percent quantile losses and are less uniform for five-percent quantile loss.

Keywords: ARCH, B-spline mixtures, Cross-sectional returns, Score-driven models, Student-t distribution, Tail intensity.

JEL classification: C22, C58, C53.

1 Introduction

Volatility persistence is a central feature of financial returns. ARCH and GARCH models describe this persistence through conditional variance dynamics (Engle, 1982; Bollerslev, 1986). We ask whether tail characteristics remain predictable after location and scale dynamics have been taken into account. We distinguish tail order, which determines the rate of tail decay, from tail intensity, which scales the tail density for a given order. Our main question is whether dynamic tail order is sufficient, or whether tail intensity has additional persistence. We study this question for conditional return mixtures formed from equity cross sections. With equal weights, the cumulative distribution function of the mixture gives the probability that a stock selected at random from the cross section has a return below a given threshold, conditional on past information.

Hansen (1994) extends ARCH modeling to conditional distributional parameters beyond the mean and variance. Conditional skewness and higher-moment persistence are studied by Harvey and Siddique (1999) and Jondeau and Rockinger (2003). Related score-driven models directly describe degrees-of-freedom dynamics (Blazsek and Monteros, 2017), time-varying peaks-over-threshold densities (Massacci, 2017), and densities with asymmetric tail indices (Thiele, 2020). D'Innocenzo et al. (2024) model score-driven tail shape and tail scale in a generalized Pareto distribution and document persistence in both. Their tail-scale parameter can also reflect changes in the scale of the full return distribution. Catania (2021) develops dynamic mixtures of return densities with score-driven updates of both component parameters and mixture weights. More recently, van Brummelen et al. (2026) use a spline-based error density to construct score-driven location and scale updates. In those specifications, the spline density parameters are fixed while location or scale changes over time.

To capture tail-intensity persistence beyond dynamic location, scale, and tail order, we propose dynamic conditional densities with spline mixtures (DCD-SM). We transform returns to $[0, 1]$ using a given cumulative distribution function. We call the distribution used for this transformation the reference distribution. We represent the conditional mixture density of the transformed returns by a spline mixture. Its weights evolve through two score-driven updates (Creal et al., 2013). One changes both tail intensities together, and the other changes their right–left balance. The spline construction preserves the reference tail order for any fixed number of components. We use references with dynamic location and scale, and references that also have dynamic tail order. In contrast to Catania (2021), DCD-SM uses a spline correction on the reference PIT scale to test additional tail-intensity dynamics while preserving the tail order of a fitted reference.

The analyst defines the target as a weighted mixture of constituent conditional return densities. The nonnegative weights sum to one and are chosen before the current returns are observed. Constituent densities may differ, and returns within a date may be dependent. The target is the conditional density of a constituent selected according to these weights. It generally differs from the distribution of a portfolio return. Kelly and Jiang (2014) use the cross section of stock returns to estimate time-varying tail risk. Kokoszka et al. (2019) forecast monthly cross-sectional return densities for the S&P 500 and Dow Jones using functional time series methods. Zhang et al. (2022) develop Wasserstein autoregressive models for density time series and apply them to the same cross sections. In their empirical applications, estimated density curves serve as inputs to the functional time series models. In a related spline-mixture formulation, Matsuda and Iwafuchi (2025) develop density-valued ARMA models through logit transformations of the mixture weights. DCD-SM fits the conditional mixture density directly to the observed returns using score-driven updates. The state update has no additional noise term, so parameter values and initial conditions determine the state from past returns. The weighted log score can therefore be evaluated recursively without integrating over unobserved states or first estimating a density curve for each date. We estimate the chosen mixture density using a weighted cross-sectional log score. The conditional expectation of this weighted log score depends only on the chosen mixture density, even when the constituent conditional densities differ and returns are dependent within a date.

With the reference distribution treated as known and the spline dimension fixed, we establish consistency and asymptotic normality of the DCD-SM estimator under stability and identification conditions. Conditional Wald tests use a date-level sandwich covariance that allows dependence within a date. Simulations re-estimate all stages to assess the feasible procedure's finite-sample size, power, and forecast

stability under independent and common-factor sampling. We separately evaluate strict out-of-sample (OOS) density and tail-quantile forecasts, since in-sample detection and predictive gains concern different questions.

The empirical analysis covers 19 US and Japanese equity cross sections, including the S&P 500 and Nikkei 225 constituent panels and their sectors. Conditional on a fitted reference with dynamic tail order, nominal five-percent Wald tests detect dynamics that affect both tail intensities in the same direction in all 19 raw-return panels and all 19 panels after removal of a first principal component with training-fixed means and loadings. Evidence for dynamics in the right–left imbalance is less widespread. For the principal-component residuals, the additional correction improves one-percent lower-tail quantile loss relative to that reference in 16 of 19 panels, with three significant positive differences at the five-percent level. Forecast gains are most evident in log scores and one-percent quantile losses, while five-percent quantile-loss gains are less uniform.

Our contribution is a framework for directly estimating the dynamics of a chosen conditional mixture from cross-sectional returns whose conditional densities may differ and whose dependence need not be specified. Within this framework, we test additional tail-intensity persistence beyond dynamic location, scale, and tail order. We use cross-sectional returns to study density dynamics that can be difficult to detect from a single return series.

Derivations, proofs, and additional simulation and data-construction details are provided in the online supplementary material.

2 Dynamic Conditional Densities with Spline Mixtures

We model dynamic conditional densities with spline mixtures (DCD-SM) relative to a given reference distribution. Observations are mapped to the probability-integral-transform (PIT) scale of the reference, where the density correction is represented by a spline mixture with score-driven dynamics (Creal et al., 2013).

For cross-sectional observations, the target is a weighted mixture of constituent conditional densities, with constituent weights determined before the current observations arrive. The constituent distributions may differ, and observations within a date may be conditionally dependent.

2.1 Spline-mixture density and embedding

Let $B_1, \ldots, B_J$ be nonnegative B-spline basis functions on $[0, 1]$ forming a partition of unity (de Boor, 2001),

$$\sum_{j=1}^{J} B_j(u) = 1,$$

and define

$$m_j = \int_0^1 B_j(u)\, du, \qquad \phi_j(u) = \frac{B_j(u)}{m_j}, \qquad j = 1, \ldots, J.$$

Then each ϕ_j integrates to one and $\sum_j m_j = 1$. For a probability vector $p = (p_1, \ldots, p_J)'$ with $p_j > 0$ for every j, define the spline-mixture density with a parameter $q \in \mathbb{R}^{J-1}$ by

$$h_{\mathrm{SM}}(u \mid q) = \sum_{j=1}^{J} p_j(q)\phi_j(u). \tag{1}$$

The parameter q is specified using a Helmert contrast matrix $Q \in \mathbb{R}^{(J-1)\times J}$ satisfying

$$QQ' = I_{J-1}, \qquad Q\mathbf{1}_J = 0.$$

The simplex weights are embedded in Euclidean coordinates by

$$q = Q \log(p/m), \tag{2}$$

where the logarithm and division are componentwise. The inverse mapping is the weighted softmax

$$p_j(q) = \frac{m_j \exp\{(Q'q)_j\}}{\sum_{k=1}^{J} m_k \exp\{(Q'q)_k\}}, \qquad j = 1, \ldots, J. \tag{3}$$

Consequently,

$$q = 0 \quad \Longleftrightarrow \quad p = m \quad \Longleftrightarrow \quad h_{\mathrm{SM}}(u \mid q) = 1.$$

Thus the origin in the logit coordinates corresponds exactly to the uniform density on the PIT scale. We set the static spline anchor at $q_0 = 0$. We do not fit a separate static spline correction.

In the empirical analysis we use clamped cubic B-splines (de Boor, 2001, Chapter IX, pp. 99–100) with breakpoints

$$(0, .05, .10, .20, .30, .40, .50, .60, .70, .80, .90, .95, 1),$$

which gives $J = 15$ normalized spline-density components.

2.2 Model description

Let $r_{t,i}$ denote the return on stock i at date t. The observed cross section is

$$r_{t,1}, \ldots, r_{t,N_t}.$$

Let $\mathcal{F}_{t-1}$ denote the information available before this cross section is observed. The constituent set and its size $N_t \geq 1$ are taken to be $\mathcal{F}_{t-1}$-measurable. Let $g_{t,i}$ denote the conditional density of $r_{t,i}$ given $\mathcal{F}_{t-1}$. These densities may differ across i, and observations within a date may be conditionally dependent.

The analyst specifies the distributional target by choosing nonnegative, $\mathcal{F}_{t-1}$-measurable weights $w_{t,i}$ satisfying $\sum_{i=1}^{N_t} w_{t,i} = 1$. We take the target density to be

$$\bar{g}_t(r) := \sum_{i=1}^{N_t} w_{t,i} g_{t,i}(r).$$

This is the conditional density of the return on a constituent selected at random with probabilities $w_{t,i}$, independently of the current returns given $\mathcal{F}_{t-1}$. To fit this target, we evaluate each observed return

under the same modeled mixture density. The individual returns need not have this density and may be dependent. In general, the target differs from the density of an index or portfolio return. When $N_t = 1$, it reduces to the usual single-series conditional density. If all constituents have the same conditional density g_t, then $\bar{g}_t = g_t$ for any such choice of weights. This special case does not require conditional independence. More generally, the weights define the mixture of interest to the analyst. They may reflect, for example, market capitalization or nonnegative portfolio holdings known before date t, and are held fixed with respect to candidate model parameters. The empirical study in Section 5 uses equal weights, $w_{t,i} = 1/N_t$, so every constituent has the same selection probability.

Let F_t be a given, $\mathcal{F}_{t-1}$-measurable reference CDF, continuously differentiable and strictly increasing with positive density f_t on the return support. Models M1–M3 in Section 2.3 provide the specific reference CDFs used below. Define

$$u_{t,i} = F_t(r_{t,i}), \qquad i = 1, \ldots, N_t.$$

Write $\phi_{t,i}(u)$ for the conditional density of $u_{t,i}$ given $\mathcal{F}_{t-1}$.

DCD-SM represents this target on the PIT scale using the spline mixture defined in Section 2.1:

$$\bar{\phi}_t(u) := \sum_{i=1}^{N_t} w_{t,i}\phi_{t,i}(u) = h_{\mathrm{SM}}(u \mid q_t), \tag{4}$$

where q_t is $\mathcal{F}_{t-1}$-measurable. Let

$$H_{\mathrm{SM}}(u \mid q) = \int_0^u h_{\mathrm{SM}}(v \mid q)\, dv.$$

The resulting model CDF for the weighted return mixture is

$$G_t(r \mid \mathcal{F}_{t-1}) = H_{\mathrm{SM}}\{F_t(r) \mid q_t\}. \tag{5}$$

Its density is

$$\bar{g}_t(r) = g_t(r \mid \mathcal{F}_{t-1}) = f_t(r)\, h_{\mathrm{SM}}\{F_t(r) \mid q_t\}. \tag{6}$$

Setting $q_t = 0$ leaves the reference mixture distribution unchanged exactly. This identity anchor is what makes the nested comparisons below transparent.

The decomposition into reference and correction is not generally identified if both are unrestricted. Indeed, writing $H_t(u) = H_{\mathrm{SM}}(u \mid q_t)$, any smooth increasing bijection A of $[0, 1]$ gives the same return CDF through $\widetilde{F}_t = A \circ F_t$ and $\widetilde{H}_t = H_t \circ A^{-1}$. Such a transformed correction need not belong to the fixed finite-dimensional spline family, but joint identification of a parameterized reference and correction would still require separate restrictions. We do not assert that identification here. We define and fit the correction with the reference held fixed. Given this reference, linear independence of the spline basis and the invertible Helmert embedding identify q_t from the corrected density. Thus the correction and its interpretation are relative to the chosen reference, rather than an intrinsic decomposition of the return law.

General score-driven dynamics. The distributional specification in (4) is completed by a recursion for the predictable state q_t. The basic DCD-SM specification can be summarized as

$$\begin{aligned} u_{t,i} \mid \mathcal{F}_{t-1} &\sim \phi_{t,i}(\cdot), \\ \sum_{i=1}^{N_t} w_{t,i}\phi_{t,i}(u) &= h_{\mathrm{SM}}(u \mid q_t), \\ q_t &= \sum_{\ell=1}^{L} K_\ell \, s_{q,t-\ell}(q_{t-\ell}). \end{aligned} \tag{7}$$

Here $K_\ell \in \mathbb{R}^{(J-1)\times(J-1)}$, and $s_{q,t}(q)$ is the derivative with respect to q of the date-t weighted log likelihood defined below. Only scores from dates before t enter this update. Equation (7) expresses q_t as a linear combination of past density scores, following the generalized autoregressive score (GAS) principle of Creal et al. (2013). The GAS framework recovers Gaussian ARCH/GARCH dynamics under suitable scaling and parameter restrictions. DCD-SM extends this score-driven construction from a conditional variance parameter to a vector of spline-mixture coordinates.

We first give the explicit score function, then introduce its Fisher-projected version and a parsimonious HAR specification for tail-intensity dynamics.

Analytic spline score. For a candidate state q, define the weighted cross-sectional log-score criterion

$$\ell_{q,t}(q) = \sum_{i=1}^{N_t} w_{t,i} \log h_{\mathrm{SM}}(u_{t,i} \mid q). \tag{8}$$

The state score is $s_{q,t}(q) = \partial \ell_{q,t}(q)/\partial q$. It has an explicit expression in terms of the component responsibilities

$$\tau_j(u, q) = \frac{p_j(q)\phi_j(u)}{h_{\mathrm{SM}}(u \mid q)}. \tag{9}$$

Write

$$\tau(u, q) = \{\tau_1(u, q), \ldots, \tau_J(u, q)\}', \qquad \bar{\tau}_t(q) = \sum_{i=1}^{N_t} w_{t,i}\tau(u_{t,i}, q).$$

Differentiating analytically gives

$$s_{q,t}(q) = Q\{\bar{\tau}_t(q) - p(q)\}. \tag{10}$$

We write $s_{q,t} = s_{q,t}(q_t)$ when the score is evaluated at the current state. It compares the weighted component responsibilities with the predicted mixture weights in Helmert coordinates. No numerical differentiation is required to evaluate this state score.

We use the weighted log score to fit the mixture target chosen by the analyst. Its algebraic form is that of a normalized working-independence composite log likelihood (Varin et al., 2011), but it requires neither identical constituent densities nor within-date independence. Section 3 establishes the score identities under the weighted-average-density specification in (4).

Fisher orthogonalization of the spline score. To make the location–scale coordinates explicit, write the given reference CDF as

$$F_t(r) = F_t^\circ\left(\frac{r - m_t}{\sigma_t}\right),$$

where F_t° is the standardized reference CDF with density f_t°. The PIT observations and their modeled mixture density are then

$$u_{t,i} := F_t^\circ\left(\frac{r_{t,i} - m_t}{\sigma_t}\right), \qquad u_{t,i} \mid \mathcal{F}_{t-1} \sim \phi_{t,i}(\cdot),$$
$$\sum_{i=1}^{N_t} w_{t,i}\phi_{t,i}(u) = h_{\mathrm{SM}}(u \mid q_t).$$

The reference density on the return scale is

$$f_t(r) = \frac{1}{\sigma_t} f_t^\circ\left(\frac{r - m_t}{\sigma_t}\right).$$

Let

$$ms = (m, \log\sigma^2)$$

denote the location–scale block. Its weighted reference score is

$$s_{ms,t} = \sum_{i=1}^{N_t} w_{t,i} \frac{\partial}{\partial ms} \log\left\{\frac{1}{\sigma} f_t^\circ\left(\frac{r_{t,i} - m}{\sigma}\right)\right\}\Big|_{ms=(m_t,\log\sigma_t^2)}. \tag{11}$$

The remaining reference coordinates are held fixed in this state derivative. Thus $s_{ms,t}$ depends on the reference distribution at date t, but not on the spline state q_t.

We project the spline score $s_{q,t}(q_t)$ off these location–scale score directions using fixed reference values, with location–scale block ms_0 and spline state $q_0 = 0$. Their choice is specified below. Let I_0 denote the Fisher information matrix of the return density in (6) with respect to (ms, q), evaluated at these fixed reference values:

$$I_0 = \begin{pmatrix} I_{ms,ms,0} & I_{ms,q,0} \\ I_{q,ms,0} & I_{q,q,0} \end{pmatrix}.$$

All Fisher blocks are evaluated once at these fixed reference values and held fixed over time. Assume that $I_{ms,ms,0}$ is nonsingular and define

$$\Pi_{q,ms,0} = I_{q,ms,0} I_{ms,ms,0}^{-1}. \tag{12}$$

The projected spline score is

$$s_{q,t}^{\perp} = s_{q,t}(q_t) - \Pi_{q,ms,0} s_{ms,t}. \tag{13}$$

In the Fisher metric of the fixed reference, this projection removes the location–scale score directions from the score used to update q_t.

Parsimonious HAR specification. Each unrestricted coefficient matrix K_ℓ contains $(J-1)^2$ parameters. To obtain a parsimonious specification, we restrict the spline state to two coordinates and use the lag

structure of the heterogeneous-autoregressive (HAR) model of Corsi (2009), with horizons of 1, 5, and 22 trading days. Define two vectors in Helmert coordinates by

$$e_{\text{even}} = Q\begin{pmatrix}1 & 0 & \cdots & 0 & 1\end{pmatrix}', \qquad e_{\text{odd}} = Q\begin{pmatrix}-1 & 0 & \cdots & 0 & 1\end{pmatrix}'. \tag{14}$$

The first and last spline components control the left and right tails under $u = F_t(r)$. The even vector changes both tail intensities together, and the odd vector changes their right–left balance. The sign convention for e_{odd} is right minus left and matches the empirical implementation. Starting from the orthogonalized spline score $s^{\perp}_{q,t}$, define the two scalar scores

$$s_{\text{even},t} = e'_{\text{even}} s^{\perp}_{q,t}, \qquad s_{\text{odd},t} = e'_{\text{odd}} s^{\perp}_{q,t}. \tag{15}$$

For a scalar score sequence s_t, define

$$\bar{s}^{(d)}_t = \frac{1}{d}\sum_{\ell=0}^{d-1} s_{t-\ell}, \qquad d = 5, 22. \tag{16}$$

The windows of 5 and 22 trading days correspond to approximately one week and one month, respectively. The two amplitudes follow HAR$(1, 5, 22)$ updates:

$$\begin{aligned} a_{\text{even},t+1} &= \kappa_{\text{even},1} s_{\text{even},t} + \kappa_{\text{even},5}\bar{s}^{(5)}_{\text{even},t} + \kappa_{\text{even},22}\bar{s}^{(22)}_{\text{even},t}, \\ a_{\text{odd},t+1} &= \kappa_{\text{odd},1} s_{\text{odd},t} + \kappa_{\text{odd},5}\bar{s}^{(5)}_{\text{odd},t} + \kappa_{\text{odd},22}\bar{s}^{(22)}_{\text{odd},t}. \end{aligned} \tag{17}$$

The spline state is reconstructed as

$$q_{t+1} = a_{\text{even},t+1} e_{\text{even}} + a_{\text{odd},t+1} e_{\text{odd}}. \tag{18}$$

This is a projected-score version of (7) with $L = 22$. Each lag matrix is a linear combination of $e_{\text{even}}e'_{\text{even}}$ and $e_{\text{odd}}e'_{\text{odd}}$, with the lag coefficients determined by the HAR structure. There are therefore six DCD-SM dynamic coefficients, three for the even update and three for the odd update. The state is restricted to span$\{e_{\text{even}}, e_{\text{odd}}\}$. Its component in the orthogonal complement is fixed at zero. There is no autoregressive state term, and the longest HAR window contains scores from the preceding 22 dates. The odd component allows the corrected density to be asymmetric even when the reference density is symmetric.

2.3 Empirical model example: M1–M5

Our aim is to detect persistence in tail behavior beyond location and scale dynamics. We distinguish tail order, which determines the rate of tail decay, from tail intensity, which determines the magnitude of the tail density for a given order. M1–M3 are benchmark models based on existing methods. M1 is static, M2 adds location and scale dynamics, and M3 adds tail-order dynamics. We introduce M4 and M5 as DCD-SM extensions of M2 and M3, respectively. Their spline corrections preserve the reference tail order and add dynamics in the left and right tail intensities. M5 therefore tests whether tail-intensity dynamics remain after accounting for dynamic tail order. The common parametric anchor is a symmetric Student-t distribution with location m, scale $\sigma > 0$, and degrees of freedom $\eta > 2$. We

use the variance-standardized parameterization associated with Hansen (1994):

$$f_t(r \mid m, \sigma, \eta) = \frac{\Gamma\{(\eta+1)/2\}}{\sigma\sqrt{\pi(\eta-2)}\,\Gamma(\eta/2)} \left\{1 + \frac{(r-m)^2}{\sigma^2(\eta-2)}\right\}^{-(\eta+1)/2}. \tag{19}$$

Thus m and σ^2 are the mean and variance of this reference density. Write its CDF as $F_t(r \mid m, \sigma, \eta)$ and denote the static estimates of location, variance, and tail order by $(m_0, \sigma_0^2, \eta_0)$. In a cross section, these are parameters of the reference mixture density, rather than common parameters imposed on each security. M1–M3 retain symmetric references. M4 and M5 add both even and odd spline corrections, so their final densities need not be symmetric. The five specifications are defined sequentially as follows.

M1: static symmetric Student-t. Student-t distributions have long been used to model stock returns (Praetz, 1972; Blattberg and Gonedes, 1974). The reference CDF is

$$F_t(r) = F_t(r \mid m_0, \sigma_0, \eta_0),$$

with $q_t = 0$. The parameters $(m_0, \sigma_0^2, \eta_0)$ are estimated by maximizing the static return log-score criterion subject to $2.05 < \eta_0 < 100$ and are then held fixed. This fitted reference supplies the fixed anchor for the Fisher orthogonalization above, with $ms_0 = (m_0, \log\sigma_0^2)$ and $q_0 = 0$.

M2: dynamic location and scale. Starting from M1, keep η_0 fixed and define

$$v_t = \log\sigma_t^2, \qquad z_{t,i} = \frac{r_{t,i} - m_t}{\sigma_t}.$$

Differentiating the log of (19) gives the location and log-variance scores

$$\begin{aligned} s_{m,t} &= \sum_{i=1}^{N_t} w_{t,i} \frac{(\eta_0+1)z_{t,i}}{\sigma_t(\eta_0 - 2 + z_{t,i}^2)}, \\ s_{v,t} &= \frac{1}{2}\sum_{i=1}^{N_t} w_{t,i} \left\{\frac{(\eta_0+1)z_{t,i}^2}{\eta_0 - 2 + z_{t,i}^2} - 1\right\}. \end{aligned} \tag{20}$$

The HAR$(1, 5, 22)$ updates are

$$m_{t+1} = m_0 + \kappa_{m,1}s_{m,t} + \kappa_{m,5}\bar{s}_{m,t}^{(5)} + \kappa_{m,22}\bar{s}_{m,t}^{(22)},$$

$$v_{t+1} = \log\sigma_0^2 + \kappa_{v,1}s_{v,t} + \kappa_{v,5}\bar{s}_{v,t}^{(5)} + \kappa_{v,22}\bar{s}_{v,t}^{(22)}.$$

Thus

$$F_t(r) = F_t(r \mid m_t, \sigma_t, \eta_0),$$

and $q_t = 0$. Setting all six M2 HAR coefficients to zero gives M1 exactly.

Under a Gaussian reference, the score-driven framework includes ARCH models (Creal et al., 2013). M2 belongs to the existing class of Student-t score-driven models studied by Harvey and Sucarrat (2014). Here we use weighted cross-sectional scores and HAR$(1, 5, 22)$ lags. These scores account for heavy tails and limit the effect of extreme returns on the updates.

M3: dynamic Student-t tail order. M3 builds on existing Student-t models with score-driven degrees of freedom. Lucas and Zhang (2016) update variance and degrees of freedom through score-driven exponentially weighted moving averages, using the inverse diagonal of the Fisher information for scaling. Blazsek and Monteros (2017) use a score-driven autoregression for degrees of freedom. Palumbo (2026) models symmetric and asymmetric tail-order dynamics and develops Lagrange multiplier tests for their presence.

Our M3 specification uses weighted cross-sectional scores to update the tail order of the chosen conditional mixture density. We project the tail-order score off location and log variance using Fisher coefficients evaluated once at M1, and use HAR$(1, 5, 22)$ averages of the projected scores in the update. The projection removes these score directions in the Fisher metric of M1. We hold the M2 estimates of m_t and σ_t fixed for each date. Tail-order dynamics are specified in a scalar unconstrained latent coordinate $\xi_{\eta,t}$. Let

$$\Lambda(x) = \frac{1}{1 + \exp(-x)},$$

and map the latent state to the degrees of freedom by

$$\eta_t = \eta_L + (\eta_U - \eta_L)\Lambda(\xi_{\eta,t}). \tag{21}$$

The implementation uses

$$\eta_L = 2.05, \qquad \eta_U = 100.$$

We set $\eta_L = 2.05$ to ensure finite variance and maintain a numerical margin above the boundary $\eta = 2$. Hence $2.05 < \eta_t < 100$ automatically. The M1 tail-order estimate determines the latent anchor through

$$\xi_{\eta,0} = \log\left(\frac{\eta_0 - \eta_L}{\eta_U - \eta_0}\right). \tag{22}$$

The latent coordinate follows an anchored score-HAR$(1, 5, 22)$ recursion,

$$\xi_{\eta,t+1} = \xi_{\eta,0} + \kappa_{\eta,1} s^{\perp}_{\xi_\eta,t} + \kappa_{\eta,5} \bar{s}^{\perp,(5)}_{\xi_\eta,t} + \kappa_{\eta,22} \bar{s}^{\perp,(22)}_{\xi_\eta,t}. \tag{23}$$

The score entering this filter is with respect to the latent coordinate. If $s_{\eta,t}$ denotes the weighted cross-sectional score with respect to the degrees of freedom, the chain rule gives

$$s_{\xi_\eta,t} = \frac{\partial \eta_t}{\partial \xi_{\eta,t}} s_{\eta,t} = (\eta_U - \eta_L)\Lambda(\xi_{\eta,t})\{1 - \Lambda(\xi_{\eta,t})\} s_{\eta,t}. \tag{24}$$

For M3, the same fixed-anchor projection is applied to the latent tail-order score after the chain-rule transformation in (24). Define

$$\dot{\eta}_0 = \left.\frac{\partial \eta}{\partial \xi_\eta}\right|_{\xi_{\eta,0}} = \frac{(\eta_0 - \eta_L)(\eta_U - \eta_0)}{\eta_U - \eta_L}.$$

The Fisher information is evaluated once at M1 and held fixed. Both scores in the following projection

are evaluated at the current M3 state (m_t, σ_t, η_t):

$$\begin{aligned} s^{\perp}_{\xi_\eta,t} &= s_{\xi_\eta,t} - I_{\xi_\eta,ms,0} I^{-1}_{ms,ms,0} s_{ms,t} \\ &= s_{\xi_\eta,t} - \frac{6\dot{\eta}_0}{\eta_0(\eta_0-2)(\eta_0+1)} s_{v,t}. \end{aligned} \tag{25}$$

Here $s_{v,t}$ is the log-variance component of $s_{ms,t}$. The location coefficient is zero by symmetry. Section S1 gives the closed-form Fisher blocks and their derivation. The tail-order state in (21) then defines

$$F_t(r) = F_{\mathrm{t}}(r \mid m_t, \sigma_t, \eta_t),$$

again with $q_t = 0$. Setting the three coefficients in (23) to zero fixes $\xi_{\eta,t} = \xi_{\eta,0}$ and therefore $\eta_t = \eta_0$, giving M2 exactly.

M4: M2 reference plus DCD-SM. We hold the M2 estimates of m_t and σ_t fixed for each date and form the transformed observations as

$$u_{t,i} = F_{\mathrm{t}}(r_{t,i} \mid m_t, \sigma_t, \eta_0).$$

For M4, (13) projects the spline score off location and log variance. At the symmetric M1 anchor, the symmetric knot sequence makes the even signal orthogonal to location and the odd signal orthogonal to log variance. Thus (15) becomes

$$\begin{aligned} s_{\mathrm{even},t} &= e'_{\mathrm{even}} s_{q,t} - c_{\mathrm{even},0} s_{v,t}, \\ s_{\mathrm{odd},t} &= e'_{\mathrm{odd}} s_{q,t} - c_{\mathrm{odd},0} s_{m,t}, \end{aligned} \tag{26}$$

where

$$\begin{aligned} c_{\mathrm{even},0} &= \frac{2(\eta_0+3)}{\eta_0} e'_{\mathrm{even}} I_{q,v,0}, \\ c_{\mathrm{odd},0} &= \frac{\sigma_0^2(\eta_0-2)(\eta_0+3)}{\eta_0(\eta_0+1)} e'_{\mathrm{odd}} I_{q,m,0}. \end{aligned} \tag{27}$$

The vectors $I_{q,m,0}$ and $I_{q,v,0}$ are the columns of $I_{q,ms,0}$. Their expectations involve the spline basis and are evaluated by numerical quadrature once at M1, as detailed in Section S1. The resulting even and odd scores are propagated by (17)–(18).

M5: M3 reference plus DCD-SM. We hold the M3 estimates of m_t, σ_t, and η_t fixed for each date and use $F_{\mathrm{t}}(\cdot \mid m_t, \sigma_t, \eta_t)$ as the reference CDF. The same DCD-SM recursion is then fitted. The spline score is projected off location and log variance only. The dynamic tail-order direction associated with $\xi_{\eta,t}$ is deliberately retained. M5 therefore combines dynamic Student-t tail order with an additional even/odd tail-intensity correction. The even and odd scores use (26) with the same M1 coefficients as M4, evaluating the PIT observations and reference scores using the M3 estimates for each date.

Within this sequential specification, setting the newly introduced dynamic coefficients to zero gives the exact zero-restriction relations

$$M1 \subset M2 \subset M3 \subset M5, \qquad M2 \subset M4.$$

M3 and M4 are two different ways of adding shape dynamics beyond the location and scale dynamics estimated by M2. M5 asks whether tail-intensity dynamics remain useful conditional on the dynamic Student-t reference. Its spline score retains the latent tail-order direction, so M5 does not test a Fisher-orthogonal complement to M3.

2.4 Estimation, inference, and forecasting

All models are fitted directly to the return observations, using either raw returns or PC1 residuals in the empirical analysis, without separate RM/RV standardization. M1 is estimated first. M2 is then estimated from score-HAR updates of location and log variance. For M3, we hold the M2 estimates of m_t and σ_t fixed for each date and estimate the three HAR coefficients for $\xi_{\eta,t}$. For M4, we retain the same M2 estimates and estimate the six DCD-SM coefficients. For M5, we hold the M3 estimates of m_t, σ_t, and η_t fixed for each date and estimate the six DCD-SM coefficients. Section 3 establishes consistency and asymptotic normality of the DCD-SM coefficient estimators with the reference distribution and the Fisher projection matrices treated as known.

For M4 and M5, let

$$\kappa = (\kappa_{\text{even},1}, \kappa_{\text{even},5}, \kappa_{\text{even},22}, \kappa_{\text{odd},1}, \kappa_{\text{odd},5}, \kappa_{\text{odd},22})'.$$

With the reference distribution at each date and the M1 Fisher projection matrices held fixed, the objective is the time average of the weighted cross-sectional return log score

$$L_T(\kappa) = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N_t} w_{t,i} \log \bar{g}_t(r_{t,i} \mid \kappa), \tag{28}$$

where $\bar{g}_t$ is given by (6). The criterion uses the same constituent weights that define the target mixture. Each return is evaluated under the target mixture density $\bar{g}_t$. For conditionally iid returns within each date and equal weights, the date-t contribution is the joint log likelihood divided by N_t. In general, the individual returns need not have this density and may be dependent. The objective fits the target mixture density without specifying the joint distribution of the observed cross section. It has the same algebraic form as a normalized working-independence composite log likelihood (Lindsay, 1988; Varin et al., 2011).

The reference distribution at each date and the Fisher projection matrices are held fixed during estimation and covariance calculation. For every candidate κ, the state, responsibilities, projected score, even and odd scores, and next-period state are recomputed forward. We estimate κ by directly maximizing (28), without an EM step. Inference uses Wald tests with conditional sandwich covariance estimates.

Strict out-of-sample forecasts are generated recursively. All fitted coefficients and Fisher projection matrices are held fixed at their training estimates, and the scores from the last 22 training dates initialize the HAR histories. The density forecast for date t uses only information and scores dated $t-1$ or earlier. After the date-t return or cross section is observed, its score is used only for subsequent forecasts. In a cross section, forecast densities and their quantiles describe the target mixture distribution, rather than each constituent's conditional distribution. The same procedure applies when $N_t = 1$, although the numerical comparisons in this paper concentrate on multiple-observation cross sections.

2.5 Tail intensities as $J \to \infty$

We study how the even and odd amplitudes relate to tail intensity as the number of spline components J increases. Let $B_{j,J}$ be a clamped B-spline basis of fixed degree, with masses $m_{j,J} = \int_0^1 B_{j,J}(u)\,du$, and write h_J for its correction density. The vectors in (14) now depend on J, and the state is

$$q_{t,J} = a_{\text{even},t}\, e_{\text{even},J} + a_{\text{odd},t}\, e_{\text{odd},J}.$$

We hold the amplitude sequences fixed as J increases. Define

$$\alpha_{L,t} = a_{\text{even},t} - a_{\text{odd},t}, \qquad \alpha_{R,t} = a_{\text{even},t} + a_{\text{odd},t},$$

and write the normalizing factor as

$$Z_{t,J} = 1 + m_{1,J}(e^{\alpha_{L,t}} - 1) + m_{J,J}(e^{\alpha_{R,t}} - 1). \tag{29}$$

For each J, the clamped basis gives

$$h_J(0 \mid q_{t,J}) = \frac{e^{\alpha_{L,t}}}{Z_{t,J}}, \qquad h_J(1 \mid q_{t,J}) = \frac{e^{\alpha_{R,t}}}{Z_{t,J}}. \tag{30}$$

Proposition 2.1. *Suppose that the support widths of $B_{1,J}$ and $B_{J,J}$ tend to zero as $J \to \infty$, so that $m_{1,J} \to 0$ and $m_{J,J} \to 0$, and that*

$$\sup_t \big(|a_{\text{even},t}| + |a_{\text{odd},t}|\big) < \infty.$$

Then $\sup_t |Z_{t,J} - 1| \to 0$ and

$$\begin{aligned} h_J(0 \mid q_{t,J}) &= e^{a_{\text{even},t} - a_{\text{odd},t}} + o(1), \\ h_J(1 \mid q_{t,J}) &= e^{a_{\text{even},t} + a_{\text{odd},t}} + o(1), \end{aligned} \tag{31}$$

uniformly in t. For a reference density $f_t > 0$ on $\mathbb{R}$, the corresponding return density $\bar{g}_{t,J}(r) = f_t(r) h_J\{F_t(r) \mid q_{t,J}\}$ satisfies

$$\begin{aligned} \lim_{J\to\infty} \lim_{r\to-\infty} \frac{\bar{g}_{t,J}(r)}{f_t(r)} &= e^{a_{\text{even},t} - a_{\text{odd},t}}, \\ \lim_{J\to\infty} \lim_{r\to+\infty} \frac{\bar{g}_{t,J}(r)}{f_t(r)} &= e^{a_{\text{even},t} + a_{\text{odd},t}}. \end{aligned} \tag{32}$$

The tail limit is taken for each fixed J before letting J increase.

This gives a semiparametric interpretation of the correction. The Student-t reference determines tail order, and the spline correction adjusts the left and right tail intensities. The even amplitude controls their common change in log intensity, and the odd amplitude controls the difference between the right and left log intensities. Under the same conditions, the integrated absolute difference satisfies

$$\sup_t \int_0^1 |h_J(u \mid q_{t,J}) - 1|\,du = O(m_{1,J} + m_{J,J}) \to 0.$$

The tail intensity ratios can therefore stay different from one while this integral tends to zero. Section S2 proves both results.

3 Asymptotic theory

The analyst chooses nonnegative weights $w_{t,i}$ summing to one, with zero weights allowed, to define the conditional mixture of interest. We study the dynamics of this chosen mixture while allowing constituent densities to differ and returns within a date to be dependent. We treat the reference distribution and the Fisher projection matrices as known and study estimation of $h_{\mathrm{SM}}(\cdot \mid q_t)$ with a fixed spline dimension J. We establish consistency and asymptotic normality of the estimators of its dynamic coefficients as the number of dates T increases. Detailed assumptions and proofs appear in Section S3 of the supplement.

3.1 Estimation criterion and regularity conditions

Let $E = (e_{\mathrm{even}}, e_{\mathrm{odd}})$ and $z_t = (a_{\mathrm{even},t}, a_{\mathrm{odd},t})'$, so that $q_t = Ez_t$. For a candidate vector θ of the six HAR coefficients, define

$$s_t(q) = \sum_i w_{t,i}\psi(u_{t,i}, q), \qquad \psi(u, q) = Q\{\tau(u, q) - p(q)\}, \qquad u_{t,i} = F_t(r_{t,i}).$$

The projected signal driving the state recursion is

$$d_t(\theta) = E'\{s_t(q_t(\theta)) + b_t\}, \qquad b_t = -I_{q,ms,0}I^{-1}_{ms,ms,0}s_{ms,t}.$$

The projection coefficient in b_t is evaluated once at the fixed reference values $(ms, q) = (ms_0, 0)$, as in (12). The reference score $s_{ms,t}$ in (11) is computed from the date-t returns using the known reference distribution. With $C_w(\theta) = \mathrm{diag}(\kappa_{\mathrm{even},w}, \kappa_{\mathrm{odd},w})$, the score update equation is

$$z_{t+1}(\theta) = \sum_{w\in\{1,5,22\}} C_w(\theta)\frac{1}{w}\sum_{j=0}^{w-1} d_{t-j}(\theta). \tag{33}$$

This is a stochastic recurrence in the two amplitudes and their 21 lags. Stability depends on the derivative of the score feedback as well as on the HAR coefficients.

The local criterion and estimator are

$$\ell_t(\theta) = \sum_i w_{t,i}\log h_{\mathrm{SM}}\{u_{t,i} \mid q_t(\theta)\}, \qquad \widehat{\theta}_T \in \arg\max_{\theta\in\mathcal{N}} \frac{1}{T}\sum_{t=1}^{T}\ell_t(\theta), \tag{34}$$

where $\mathcal{N}$ is a compact convex neighborhood of the true parameter. The fixed reference log density is constant in θ, so this criterion has the same maximizer as the weighted cross-sectional return log score. Let θ_0 denote the true coefficient vector. Define S_t and the negative expected Hessian $\mathcal{H}$ by

$$S_t := \frac{\partial\ell_t(\theta_0)}{\partial\theta}, \qquad \mathcal{H} := -E\left[\frac{\partial^2\ell_t(\theta_0)}{\partial\theta\,\partial\theta'}\right].$$

The assumptions are summarized below. Their full versions appear in the supplement.

1. The reference CDF, constituent set, N_t, and weights $w_{t,i}$ are given before the date-t returns are observed and do not depend on θ. The weighted average conditional PIT density equals $h_{\mathrm{SM}}\{\cdot \mid q_t(\theta_0)\}$.

2. The true parameter is interior to $\mathcal{N}$. For $\theta \in \mathcal{N}$, $q_t(\theta)$ is continuous in θ and jointly stationary and ergodic. The effect of initialization decays exponentially, uniformly on $\mathcal{N}$. There is a finite constant C such that $\|b_t\| \le C$ for all t almost surely.

3. If $q_t(\theta) = q_t(\theta_0)$ almost surely for every t, then $\theta = \theta_0$.

4. The first and second derivatives of $q_t(\theta)$ with respect to θ are continuous in θ. Together with $q_t(\theta)$, they are jointly stationary and ergodic. Their initialization effects decay exponentially, uniformly on $\mathcal{N}$. For some $\delta > 0$, assume

$$E\left[\sup_{\theta\in\mathcal{N}}\left\|\frac{\partial q_t(\theta)}{\partial\theta'}\right\|_F^{2+\delta}\right] < \infty,$$
$$E\left[\sup_{\theta\in\mathcal{N}}\left\|\frac{\partial^2 q_{t,j}(\theta)}{\partial\theta\,\partial\theta'}\right\|_F^{1+\delta}\right] < \infty, \qquad j = 1,\dots,J-1.$$

   Here $q_{t,j}$ is the jth coordinate of q_t, and $\|\cdot\|_F$ denotes the Frobenius norm.

5. The negative expected Hessian $\mathcal{H}$ is positive definite. There is a finite nonrandom matrix Ω such that

$$\frac{1}{T}\sum_{t=1}^{T}\mathrm{Var}(S_t \mid \mathcal{F}_{t-1}) \xrightarrow{p} \Omega. \tag{35}$$

These conditions concern the score update equation and aggregate conditional moments. They do not impose stationarity or independence on the full observation process. Local stability and identification are assumptions, rather than consequences of optimizer convergence.

3.2 Consistency and asymptotic normality

Theorem 3.1. *Under Assumptions 1–3, the local estimator in (34) converges almost surely to* θ_0.

Let $L(\theta) := E[\ell_t(\theta)]$. Under Assumption 1,

$$\begin{aligned} E[\ell_t(\theta) \mid \mathcal{F}_{t-1}] &= \sum_i w_{t,i}\int_0^1 \phi_{t,i}(u)\log h_{\mathrm{SM}}(u \mid q_t(\theta))\,du \\ &= \int_0^1 h_{\mathrm{SM}}(u \mid q_t(\theta_0))\log h_{\mathrm{SM}}(u \mid q_t(\theta))\,du. \end{aligned}$$

This equality relies on using the same weights in the criterion and in the target mixture. The expected log-score gap is therefore

$$L(\theta_0) - L(\theta) = E\left[\mathrm{KL}\left\{h_{\mathrm{SM}}(\cdot \mid q_t(\theta_0)) \,\|\, h_{\mathrm{SM}}(\cdot \mid q_t(\theta))\right\}\right] \ge 0.$$

Linear independence of the spline basis and the invertible logit embedding identify the density state for

the given reference. Dynamic identification then identifies the HAR coefficients. The full argument is given in the supplement.

Let $J_t = \partial q_t(\theta_0)/\partial\theta'$. The chain rule gives

$$S_t = J_t' s_t(q_t(\theta_0)). \tag{36}$$

Similarly, Assumption 1 gives

$$\begin{aligned} E[s_t(q_t(\theta_0)) \mid \mathcal{F}_{t-1}] &= \sum_i w_{t,i} \int_0^1 \psi(u, q_t(\theta_0))\phi_{t,i}(u)\, du \\ &= \int_0^1 \psi(u, q_t(\theta_0)) h_{\mathrm{SM}}(u \mid q_t(\theta_0))\, du \\ &= 0. \end{aligned}$$

Hence $E(S_t \mid \mathcal{F}_{t-1}) = 0$. Thus S_t is a martingale difference even when individual constituent scores have nonzero conditional means.

Under Assumption 1, the negative expected Hessian has the following representation:

$$\begin{aligned} \mathcal{I}_q(q) &= \int_0^1 \psi(u,q)\psi(u,q)' h_{\mathrm{SM}}(u \mid q)\, du, \\ \mathcal{H} &= E[J_t' \mathcal{I}_q(q_t(\theta_0)) J_t], \qquad \Sigma_t = \mathrm{Var}\{s_t(q_t(\theta_0)) \mid \mathcal{F}_{t-1}\}, \\ \mathrm{Var}(S_t \mid \mathcal{F}_{t-1}) &= J_t' \Sigma_t J_t. \end{aligned} \tag{37}$$

For an ordinary correctly specified log likelihood, the negative expected Hessian equals the score variance. Here the weighted log score is used without specifying the joint distribution within a date. Therefore Ω cannot in general be recovered from $\mathcal{H}$ alone, and the limiting covariance takes the sandwich form used in composite likelihood inference (Varin et al., 2011).

Theorem 3.2. *Under Assumptions 1–5, the local estimator satisfies*

$$\sqrt{T}(\widehat{\theta}_T - \theta_0) \Longrightarrow N(0, \mathcal{H}^{-1}\Omega\mathcal{H}^{-1}). \tag{38}$$

The conclusion also holds when the states and their derivatives are computed from initial values satisfying the forgetting conditions.

With equal weights and a constant N, conditional iid sampling gives $\Omega = \mathcal{H}/N$, so the covariance in (38) reduces to $\mathcal{H}^{-1}/N$. This simplification is not generally available with heterogeneous or dependent returns. The proof combines a martingale central limit theorem for S_t, a uniform law of large numbers for the recursive Hessian, and consistency. These results concern the DCD-SM dynamic coefficients with an M2 or M3 reference treated as known.

3.3 Sandwich covariance and conditional Wald tests

Using the score update equation with fixed initial values, let $\widehat{S}_t = \partial_\theta \widehat{\ell}_t(\widehat{\theta}_T)$ and define

$$H_T = -\sum_t \partial_\theta^2 \widehat{\ell}_t(\widehat{\theta}_T), \qquad B_T = \sum_t (\widehat{S}_t - \bar{S})(\widehat{S}_t - \bar{S})', \qquad \widehat{V}_T = H_T^{-1} B_T H_T^{-\prime}. \tag{39}$$

Under the assumptions of Theorem 3.2, $T\widehat{V}_T \to \mathcal{H}^{-1}\Omega\mathcal{H}^{-1}$ in probability. For a full-row-rank restriction matrix R, with $R\theta_0 = r$ and positive-definite restricted limiting covariance,

$$W_T = (R\widehat{\theta}_T - r)'(R\widehat{V}_T R')^{-1}(R\widehat{\theta}_T - r) \Longrightarrow \chi^2_{\mathrm{rank}(R)}.$$

The even and odd block restrictions each have rank three. Their joint restriction has rank six.

Lag-zero outer products are sufficient because the parameter score is a martingale difference under the stated assumptions. A fixed-lag HAC estimator has the same limit. The main implementation uses lag zero. This parameter covariance is separate from the HAC(22) variance used for OOS forecast-score differences.

Under the stated assumptions, we can therefore estimate and test the dynamics of the chosen mixture without specifying each constituent density or the joint distribution within a date. Sections 4 and 5 study the numerical performance of the procedure with estimated reference distributions.

4 Simulation evidence

All designs in this section use $w_{t,i} = 1/N$, the equal-weight special case of the weighted-mixture framework in Section 2.2. Section 3 treats the reference distribution and Fisher projection matrices as known. Here we fit M1–M5 and compute the Fisher projection coefficients from the M1 estimates in each replication to assess the finite-sample size and power of the conditional Wald tests. The experiment uses a symmetric Student-t reference. The main comparison is between conditionally independent draws and draws with a common factor explaining 10 percent of individual return variance on average. We use $N \in \{20, 50, 100\}$ and 100 Monte Carlo replications for each design. We evaluate empirical size under the null of no shape persistence and power against alternatives with persistent tail dynamics. As an additional assessment, we examine the forecast consequences of fitting shape dynamics through out-of-sample (OOS) accuracy.

4.1 Generating dynamics and sequential estimation

The generating baseline is the static symmetric Student-t M1, with location zero, scale one, and degrees of freedom $\eta_0 = 2.6750886053$. Returns are expressed in percent. Since location and scale have no generating dynamics, M2 coincides with M1 and is omitted as a separate generating model. M2 is still fitted in estimation. These values determine generation and are not supplied to estimation. We generate M3 by adding tail-order dynamics through (40), with no spline correction ($q_t = 0$). M4 keeps tail order fixed and generates the even/odd amplitudes through (41). M5 combines the tail-order update in (40) with the amplitude updates in (41). The conditional generating CDF is

$$G_t(y) = H_{\mathrm{SM}}\{F_t(y) \mid q_t\}, \qquad q_t = e_{\mathrm{even}}a_{\mathrm{even},t} + e_{\mathrm{odd}}a_{\mathrm{odd},t}.$$

Here F_t is a symmetric Student-t CDF with the fixed generating location and scale and either static or dynamic tail order, according to the design. We use the spline basis and the vectors e_{even} and e_{odd} defined in Section 2.

We use the latent tail coordinate $\xi_{\eta,t}$ in (21). Let $d_{\eta,t}$ be its date-average score, projected off location and log variance at the fixed baseline Fisher geometry, and let $d_{b,t}$ denote the corresponding projected

score for $b \in \{\text{even}, \text{odd}\}$. The generating updates are

$$\xi_{\eta,t+1} = \xi_{\eta,0} + \sum_{w\in\{1,5,22\}} \kappa_{\eta,w}(N)\bar{d}^{(w)}_{\eta,t}, \tag{40}$$

$$a_{b,t+1} = \sum_{w\in\{1,5,22\}} \kappa_{b,w}(N)\bar{d}^{(w)}_{b,t}. \tag{41}$$

Each date's signals are calculated from the actual N generated returns. For generation, the Fisher projection coefficients are computed once at $q = 0$ and the true static Student-t reference values, and held fixed. Only location and log variance are projected out. The tail-order direction is retained in the spline signal, and no score whitening is applied.

For active components, the adopted S&P 500 HAR vectors are multiplied by

$$\kappa_b(N) = \sqrt{N/500}\,\kappa_b^{\text{SP}}, \qquad b \in \{\eta, \text{even}, \text{odd}\}.$$

The baseline HAR coefficient vectors κ_b^{SP} are reported in Section S4.1 of the supplement. We use the same coefficients for independent draws and the common-factor designs. Under independence, the $N^{-1/2}$ fluctuation scale of an average score motivates this adjustment. The same multiplier is used in the common-factor design, without an additional dependence-specific variance adjustment. Consequently, comparisons across N concern this scaled design, not a fixed-coefficient experiment isolating only sample size. Comparisons between dependence specifications at a given N use the same coefficients.

We consider five designs. Case 0 uses the static M1=M2 baseline. Case 1 generates M3 with persistent η. Case 2 generates M4 with only the even amplitude persistent. Case 3 generates M5 with persistent η and even amplitude. Case 4 adds persistence in the M5 odd amplitude. Inactive coefficient vectors are zero. We generate 500 burn-in dates, 1,886 estimation dates, and 642 OOS dates. For each candidate coefficient vector, the score update equations are run through the burn-in observations. These dates are excluded from the estimation criterion and parameter-score outer products.

Estimation begins with a static symmetric-t M1 fit of location, scale, and tail order. For fitting, the Fisher projection coefficients are computed once at $q = 0$ and the fitted M1 reference values, and held fixed. M2 estimates six HAR coefficients for location and log variance, holding the M1 estimates fixed. M3 holds the M2 estimates of m_t and σ_t fixed for each date and estimates three HAR coefficients for $\xi_{\eta,t}$. M4 uses the same M2 estimates and estimates six even/odd DCD-SM coefficients. M5 holds the M3 estimates of m_t, σ_t, and η_t fixed for each date and estimates six DCD-SM coefficients. Although the generating location and scale are static, M2 allows dynamics in both coordinates. For estimation, returns are converted from percent to decimal units. Forecast scores are converted back to the original percentage units.

4.2 Dependence with identical conditional marginals

Returns are generated in two steps. Given a uniform variable $V_{t,i}$, first draw a PIT value from the spline-mixture density and then apply the inverse reference CDF to return to the return scale:

$$u_{t,i} = H_{\text{SM}}^{-1}(V_{t,i} \mid q_t), \qquad Y_{t,i} = F_t^{-1}(u_{t,i}).$$

Thus $u_{t,i}$ has conditional density $h_{\mathrm{SM}}(\cdot \mid q_t)$, and $Y_{t,i}$ has conditional CDF $G_t = H_{\mathrm{SM}} \circ F_t$. Equivalently, $Y_{t,i} = G_t^{-1}(V_{t,i})$. For independent draws, the $V_{t,i}$ are conditionally independent uniforms. For the common-factor design, draw $a_i \sim U[-0.1, 1.1]$ independently at the start of each replication and hold the loadings fixed over time and across the five designs. Let C_t and $x_{t,i}$ be independent standard normal innovations, independent over dates. Introduce γ_t to adjust the strength of the common component, as specified below, and define

$$\beta_{t,i} = \frac{\gamma_t a_i}{\sqrt{1 + (\gamma_t a_i)^2}}, \tag{42}$$

$$Z_{t,i} = \beta_{t,i} C_t + \sqrt{1 - \beta_{t,i}^2}\, x_{t,i}, \tag{43}$$

$$u_{t,i} = H_{\mathrm{SM}}^{-1}\{\Phi(Z_{t,i}) \mid q_t\}, \qquad Y_{t,i} = F_t^{-1}(u_{t,i}). \tag{44}$$

Conditional on past information and the fixed loadings, every $Z_{t,i}$ is standard normal. Thus the $V_{t,i} = \Phi(Z_{t,i})$ are identically uniform but dependent across i through the common factor C_t. Applying the same two transformations gives each return the conditional CDF G_t. This construction relaxes independence without introducing heterogeneous conditional marginals.

Before drawing the current innovations, γ_t is calibrated to satisfy

$$\frac{1}{N} \sum_{i=1}^{N} \frac{\mathrm{Var}\{E(Y_{t,i} \mid C_t, \mathcal{F}_{t-1}) \mid \mathcal{F}_{t-1}\}}{\mathrm{Var}(Y_{t,i} \mid \mathcal{F}_{t-1})} = 0.10. \tag{45}$$

The conditioning information includes the fixed loadings and the current predictable density state. The share concerns the transformed returns and averages individual explained variance fractions. It is neither $\beta_{t,i}^2$, a common pairwise correlation, nor a PCA eigenvalue share. Individual shares differ with the loadings. In particular, the calculation integrates over individual noise and common-factor realizations.

4.3 Size and power

Estimation maximizes the time average of daily cross-sectional mean log scores, as in (28). Parameter inference uses the lag-zero sandwich covariance in (39). We test the three coefficients of each block at the nominal five-percent level. We jointly estimate the even and odd blocks and conduct a separate Wald test of the three HAR coefficients in each block. Covariance calculations hold the M2 estimates fixed for M3 and M4, and the M3 estimates fixed for M5. The Fisher projection coefficients are also held fixed.

Table 1: Empirical size of nominal five-percent Wald tests

Dependence	N	M3 η	M4 even	M4 odd	M5 even	M5 odd
Independent	20	3	6	8	6	8
Common 10%	20	4	6	7	5	7
Independent	50	6	8	5	7	5
Common 10%	50	6	12	3	5	3
Independent	100	3	5	8	2	8
Common 10%	100	3	5	4	2	5

All states are static (case 0). Entries are rejection percentages over 100 replications. All displayed tests are numerically valid. M1–M5 and the M1 Fisher geometry are estimated. The lag-zero covariance holds M2 estimates fixed for M3 and M4, and M3 estimates fixed for M5.

Table 1 reports case 0, where every displayed test is a size diagnostic. Under independence, rejection

frequencies range from 2 to 8 percent. In the 10-percent factor design, they range from 2 to 12 percent. M4 even rejects in 6, 12, and 5 percent of replications for $N = 20, 50, 100$, respectively. Dependence does not produce uniformly severe overrejection in this design, but the results do not justify nominal size control in every cell. Each cell contains only 100 replications. At a true rejection probability of 0.05, its Monte Carlo standard error is about 2.2 percentage points.

Table 2: Rejection frequencies under persistent shape dynamics

Case	Dependence	N	M3 η	M4 even	M4 odd	M5 even	M5 odd
1: η only	Independent	20	21	6	5	2	6
	Common 10%	20	17	10	10	5	10
	Independent	50	48	13	4	9	4
	Common 10%	50	30	12	7	8	7
	Independent	100	71	14	5	3	5
	Common 10%	100	46	15	7	4	7
2: Even only	Independent	20	36	97	10	93	10
	Common 10%	20	26	92	8	90	8
	Independent	50	37	100	7	100	7
	Common 10%	50	16	100	8	99	9
	Independent	100	65	100	6	100	6
	Common 10%	100	13	100	6	100	6
3: η and even	Independent	20	22	97	8	97	8
	Common 10%	20	26	97	8	92	7
	Independent	50	65	100	6	100	5
	Common 10%	50	43	99	6	99	6
	Independent	100	93	100	5	100	5
	Common 10%	100	58	100	3	100	3
4: η, even, and odd	Independent	20	19	94	23	91	23
	Common 10%	20	20	91	26	88	28
	Independent	50	59	100	57	100	57
	Common 10%	50	46	98	34	97	34
	Independent	100	97	100	73	100	73
	Common 10%	100	69	100	68	100	68

Percentages over 100 replications at nominal size 5%. Columns for active components describe detection; rejections in other columns need not identify a separate true shape component. True coefficients depend on N through the stated calibration. M1–M5 are estimated. Wald calculations hold M2 estimates fixed for M3 and M4, and M3 estimates fixed for M5.

Table 2 shows substantial detection of persistent even dynamics. In case 2, M4 even rejects in 92, 100, and 100 percent of the 10-percent-factor replications. Detection of tail order alone is weaker: M3 rejects in 17, 30, and 46 percent in case 1. In case 4, M5 odd rejects in 28, 34, and 68 percent. These are rejection frequencies at nominal five-percent critical values, not size-adjusted power estimates. The relative strengths of these components therefore matter for finite-sample detection.

Rejection need not uniquely identify the generating component. In case 2, M3 rejects despite static generating tail order, while in case 1 the M4 even block can reject without a generating even update. A candidate reference can leave another type of shape variation for its fitted block to approximate. Such cells do not automatically diagnose the size of a correctly specified component null. These results suggest that tail-order and tail-intensity dynamics can be difficult to distinguish in finite samples.

4.4 Forecast accuracy

The estimated coefficients are held fixed during the OOS period. We evaluate the log score and one-percent quantile loss, averaged across returns within each date, with M2 as the common benchmark. M2 is estimated separately in each replication.

For each date t, the model forecasts one mixture density $\widehat{g}_t$ for the cross section. The daily log score

is

$$\text{Log score}_t = \frac{1}{N}\sum_{i=1}^{N} \log \widehat{g}_t(Y_{t,i}).$$

Higher values indicate better forecasts.

For a quantile level $\alpha \in (0,1)$ and the common predictive return quantile $Q_t(\alpha)$, the quantile loss is

$$L_{t,i}(\alpha) = \big[\alpha - \mathbf{1}\{Y_{t,i} < Q_t(\alpha)\}\big]\big[Y_{t,i} - Q_t(\alpha)\big].$$

Smaller values indicate better forecasts. This section uses $\alpha = 0.01$, and the empirical analysis in Section 5.5 uses $\alpha \in \{0.01, 0.05\}$.

Within each replication, percentage improvements are

$$\text{Loss improvement} = 100(\bar{L}_{\text{M2}} - \bar{L}_m)/\bar{L}_{\text{M2}},$$
$$\text{Log-score improvement} = 100\{\exp(\overline{\log g_m - \log g_{\text{M2}}}) - 1\}.$$

We report the median percentage improvement over 100 replications.

Table 3: Median OOS density and tail improvements relative to M2

Criterion			Log score			1% quantile loss		
Case	Dependence	N	M3	M4	M5	M3	M4	M5
0: Static	Independent	20	-0.001	-0.008	-0.011	-0.001	-0.079	-0.078
	Common 10%	20	-0.003	-0.011	-0.015	+0.001	-0.092	-0.092
	Independent	50	-0.002	-0.004	-0.006	+0.001	-0.033	-0.035
	Common 10%	50	-0.002	-0.006	-0.007	+0.002	-0.053	-0.049
	Independent	100	-0.000	-0.001	-0.002	-0.000	-0.015	-0.016
	Common 10%	100	-0.001	-0.004	-0.005	-0.000	-0.034	-0.032
1: η only	Independent	20	-0.001	-0.007	-0.007	-0.001	-0.014	-0.012
	Common 10%	20	-0.002	-0.009	-0.011	-0.004	-0.064	-0.071
	Independent	50	+0.001	-0.004	-0.003	+0.000	-0.038	-0.033
	Common 10%	50	+0.001	-0.005	-0.005	-0.003	-0.045	-0.042
	Independent	100	+0.002	-0.001	+0.001	+0.001	-0.011	-0.011
	Common 10%	100	+0.002	-0.003	-0.003	-0.001	-0.041	-0.038
2: Even only	Independent	20	+0.000	+0.033	+0.030	+0.032	+0.296	+0.286
	Common 10%	20	+0.000	+0.035	+0.028	+0.034	+0.356	+0.396
	Independent	50	+0.002	+0.034	+0.030	+0.023	+0.312	+0.309
	Common 10%	50	+0.000	+0.033	+0.028	+0.024	+0.365	+0.367
	Independent	100	+0.001	+0.027	+0.025	+0.020	+0.301	+0.298
	Common 10%	100	+0.000	+0.031	+0.030	+0.017	+0.319	+0.313
3: η and even	Independent	20	+0.001	+0.033	+0.031	+0.019	+0.312	+0.333
	Common 10%	20	+0.000	+0.027	+0.026	+0.019	+0.111	+0.120
	Independent	50	+0.002	+0.026	+0.026	+0.014	+0.255	+0.254
	Common 10%	50	+0.001	+0.031	+0.029	+0.014	+0.298	+0.310
	Independent	100	+0.004	+0.028	+0.030	+0.013	+0.255	+0.256
	Common 10%	100	+0.003	+0.038	+0.040	+0.012	+0.358	+0.368
4: η, even, and odd	Independent	20	-0.001	+0.037	+0.036	+0.010	+0.293	+0.272
	Common 10%	20	+0.001	+0.031	+0.027	+0.009	+0.378	+0.382
	Independent	50	+0.003	+0.033	+0.033	+0.014	+0.329	+0.335
	Common 10%	50	+0.001	+0.043	+0.045	+0.012	+0.337	+0.347
	Independent	100	+0.003	+0.036	+0.035	+0.013	+0.298	+0.294
	Common 10%	100	+0.003	+0.048	+0.055	+0.005	+0.353	+0.360

Median percentage improvements over 100 replications. Within each replication, log-score improvements are $100\{\exp(\overline{\Delta \log g}) - 1\}$ and quantile-loss improvements are percentage reductions relative to M2. All models, including the M2 benchmark, are estimated.

Table 3 reports log-score and one-percent quantile-loss improvements for M3–M5 in cases 0–4.

Median gains for M4 and M5 are positive in cases 2–4. The static design gives essentially zero or slightly negative gains, as expected when the generating distribution is static. Tail-order-only forecast gains are small, and adding tail-intensity dynamics in that design does not systematically improve on M2. Table S1 of the supplement additionally compares M5 with M3.

In case 4 with $N = 100$ and the 10-percent factor, M5 reduces the one-percent quantile loss by 0.360 percent at the median. These results support adding tail-intensity dynamics, without implying a uniform advantage of M5 over M4. Detection of persistence and improvement in forecast scores are distinct questions. The parameter tests assess shape dynamics, whereas the OOS comparisons assess the predictive consequences of the fitted dynamics.

5 Empirical evidence

This section studies conditional density dynamics in 19 equity cross sections: the S&P 500 and Nikkei 225 constituent panels, 11 US sectors, and six Japanese sectors. We use the symmetric-t M1–M5 specification and compare raw returns with residuals obtained by removing a first principal component estimated once from the training sample and then held fixed separately within each panel. The broad-market panels contain constituent returns, not single index returns. "Raw" and "PC1 residual" use the same observed constituent–dates, restricted to constituents with estimable training PC1 loadings. We use $w_{t,i} = 1/N_t$ throughout, so the target is the equally weighted conditional mixture $\bar{g}_t = N_t^{-1} \sum_i g_{t,i}$, a special case of the weighted mixture defined in Section 2.2. Constituent distributions may differ, and returns within a date may be dependent.

5.1 Data and sample construction

The common core sample begins on July 5, 2016, except for the S&P 500 Real Estate sector, which begins on September 1, 2016. Estimation ends on December 29, 2023. Strict one-step-ahead OOS evaluation runs through July 24, 2026, beginning on January 2, 2024 for the US panels and January 4, 2024 for the Japanese panels. There are 642 US and 624 Japanese OOS trading dates. Table S2 in the supplement reports panel sizes and the number of training dates.

Daily prices are obtained from Yahoo Finance. We use adjusted closing prices when available and otherwise reported closing prices, and calculate returns as daily log differences. Raw returns are not normalized by realized means or variances.

Membership is reconstructed point in time from the community-maintained `fja05680/sp500` S&P 500 histories and the official Nikkei constituent list and component-change history. Sector panels preserve these histories. Source and classification details are given in Section S5.1 of the supplement. Price data are unavailable for some historical constituents, including delisted stocks. On each date, we use only constituents with available returns. Section 3 assumes that inclusion and N_t are determined before the current returns are observed. Membership reconstruction alone does not establish this condition for missing observations.

Median cross-sectional sizes range from 18 to 70 among sectors, compared with 455 and 218 in the broad-market panels. Some sector samples contain fewer than 20 observations on individual dates.

5.2 Symmetric reference, estimation, and inference

M1 estimates a static symmetric standardized Student-t reference, with location, scale, and tail-order parameter η free. M2 fixes the fitted M1 tail order and estimates six HAR$(1, 5, 22)$ coefficients for location and log variance. M3 holds the M2 estimates of m_t and σ_t fixed for each date and estimates three HAR coefficients for latent tail order ξ_η. No parametric skewness dynamics are estimated.

M4 uses the same M2 estimates and estimates six even/odd DCD-SM coefficients. M5 holds the M3 estimates of m_t, σ_t, and η_t fixed for each date and estimates six DCD-SM coefficients. The Fisher projection coefficients are computed once at $q = 0$ and the fitted M1 reference values, and held fixed. M4 computes its PIT observations and reference scores using the M2 estimates for each date, and M5 uses the M3 estimates. The spline scores are orthogonalized with respect to the location and log-variance scores, while they are not orthogonalized with respect to the η score. Although the references are symmetric, the odd spline correction allows the final densities to be asymmetric.

Every estimation stage uses the same equal-date criterion,

$$L_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{N_t}\sum_{i=1}^{N_t}\log \bar{g}_t(r_{t,i} \mid \theta). \tag{46}$$

Each date receives equal weight as the number of covered constituents changes over time. HAR windows contain exactly 1, 5, and 22 previous dates, with unobserved presample scores initialized to zero.

In-sample inference uses fitted-alternative conditional sandwich-Wald tests with lag zero and the covariance in (39). The fitted reference distribution and Fisher projection coefficients are held fixed in each covariance calculation. For residual panels, the fitted PCA means and loadings are also held fixed. Their estimation uncertainty is not added to the parameter covariance. In the tables, "η" denotes the latent tail-order block ξ_η, whose HAR coefficients are defined in (23). Each of location, scale, η, even, and odd has a three-coefficient χ^2_3 Wald test. The M2 location test sets the three location HAR coefficients to zero, retaining fitted log-variance dynamics. The scale test sets the three log-variance HAR coefficients to zero, retaining fitted location dynamics. The M4/M5 joint null has six coefficients. Detection counts use the nominal five-percent level without adjustment for multiple tests.

5.3 Persistence in the raw cross sections

Table 4 gives broad-market p-values for raw returns and PC1 residuals. In the S&P 500 raw panel, the M3 η test has $p = 5.06 \times 10^{-5}$, while M4 and M5 even have $p = 1.38 \times 10^{-9}$ and 3.46×10^{-19}. For the Nikkei 225, the corresponding p-values are 3.83×10^{-4}, 2.81×10^{-4}, and 4.20×10^{-4}. Odd persistence is detected in the S&P 500 panel, with $p = 0.0201$ for M4 and $p = 0.0456$ for M5, but not in the Nikkei 225 raw panel.

Across the 19 raw panels, M2 detects location dynamics in four and scale dynamics in all 19. Among the 17 sectors, the corresponding counts are three and 17. M3 detects η dynamics in 15 panels, including 13 sectors, whereas M4 and M5 detect even dynamics in all 19 panels. Odd detections are less frequent: five panels for M4 and three for M5, including four and two sectors. The evidence is consequently strongest for changes in both tail intensities together. The broad-market and sector samples overlap. These counts are descriptive summaries, not independent replications of the same test.

The η and even results are related descriptions of changing tails, not separate economic mechanisms

Table 4: Broad-market conditional sandwich-Wald p-values

Model	Location	Scale	η	Even	Odd	Joint
Raw; S&P 500						
M2	0.0095	6.91×10^{-49}	–	–	–	–
M3	–	–	5.06×10^{-5}	–	–	–
M4	–	–	–	1.38×10^{-9}	0.0201	1.34×10^{-14}
M5	–	–	–	3.46×10^{-19}	0.0456	2.99×10^{-21}
Raw; Nikkei 225						
M2	0.6742	1.33×10^{-20}	–	–	–	–
M3	–	–	3.83×10^{-4}	–	–	–
M4	–	–	–	2.81×10^{-4}	0.4576	3.74×10^{-5}
M5	–	–	–	4.20×10^{-4}	0.3712	1.07×10^{-4}
PC1 residual; S&P 500						
M2	0.3554	3.73×10^{-81}	–	–	–	–
M3	–	–	8.56×10^{-12}	–	–	–
M4	–	–	–	4.55×10^{-7}	2.37×10^{-4}	1.69×10^{-8}
M5	–	–	–	6.33×10^{-6}	4.64×10^{-4}	4.54×10^{-7}
PC1 residual; Nikkei 225						
M2	0.0011	2.54×10^{-100}	–	–	–	–
M3	–	–	7.20×10^{-14}	–	–	–
M4	–	–	–	3.51×10^{-106}	2.12×10^{-5}	5.50×10^{-117}
M5	–	–	–	8.59×10^{-104}	5.04×10^{-5}	3.76×10^{-113}

The reference is symmetric Student-t. Each M2 test sets the three HAR coefficients for the indicated component to zero, retaining fitted dynamics of the other component. M3 has one three-coefficient η block. M4 and M5 have even and odd three-coefficient blocks and a six-coefficient joint null. All parameter covariances use lag zero and hold M1 estimates fixed for M2, M2 estimates fixed for M3 and M4, and M3 estimates fixed for M5.

Table 5: Persistence rejections at the five-percent level

Target	Universe	M2 location	M2 scale	M3 η	M4 even	M4 odd	M5 even	M5 odd
Raw	All panels	4/19	19/19	15/19	19/19	5/19	19/19	3/19
Raw	Sectors	3/17	17/17	13/17	17/17	4/17	17/17	2/17
PC1 residual	All panels	3/19	19/19	17/19	19/19	5/19	19/19	5/19
PC1 residual	Sectors	2/17	17/17	15/17	17/17	3/17	17/17	3/17

Raw and PC1 residual use the same constituent–dates. Counts are descriptive and unadjusted for multiplicity. All displayed tests are numerically valid. Numerical validity of a Wald statistic does not establish exact finite-sample size.

established by the tests. The M5 even results support adding tail-intensity dynamics to the fitted M3 reference.

5.4 Removing a within-panel principal component

Our main adjustment removes one common linear component separately in each of the 19 panels. We estimate a covariance PCA once using all training dates through December 29, 2023, and hold its means and leading loading vector fixed throughout training and OOS evaluation. Each constituent is demeaned by its observed training mean, without volatility standardization. For PCA only, missing training returns are replaced by that mean, or equivalently by zero after demeaning. Constituents without nonconstant observed training returns are excluded. No imputed return enters the density estimation or forecast evaluation. This training-sample preprocessing is retrospective within the training period. It uses no OOS information and is not reestimated during OOS evaluation.

Let $\widehat{\mu}$ and $\widehat{b}$ denote the fixed means and loadings, and let O_t identify eligible constituents observed at date t. For their return vector y_t, define

$$\widehat{F}_t = \frac{\widehat{b}'_{O_t}(y_t - \widehat{\mu}_{O_t})}{\widehat{b}'_{O_t}\widehat{b}_{O_t}}, \qquad e_t = y_t - \widehat{\mu}_{O_t} - \widehat{b}_{O_t}\widehat{F}_t.$$

The subscript O_t selects entries of the fixed vectors. The residual removes both the training mean and the projected PC1 component. We fit all five models to raw returns and residuals on the same constituent–dates. The two targets give 38 panels and 190 fitted models. There is no initial rolling-window exclusion.

The broad-market mean pairwise correlations decline from 0.356 to 0.009 for the S&P 500 and from 0.361 to 0.002 for the Nikkei 225. The leading eigenvalues account for 36.5 and 37.5 percent of their respective training covariance traces. These PCA shares are not the conditional explained-return variance fractions calibrated in Section 4.2. Construction details are given in Section S5.2 of the supplement.

M2 detects location dynamics in three residual panels, including two sectors, and scale dynamics in all 19 residual panels. Tail-intensity dynamics remain after this removal. Both M4 and M5 detect even persistence in all 19 residual panels, as they do in the raw panels. M3 η is detected in 17 of 19 residual panels. Odd persistence is detected in five residual panels for M4 and in five for M5. Both models detect odd persistence in the S&P 500 and Nikkei 225 residual panels. For the US Communication Services residuals, the fitted M1 latent tail-order parameter reaches its lower search bound, giving $\eta = 2.0506$. The M3, M4, and M5 persistence tests for this panel are numerically valid.

5.5 Strict out-of-sample forecasting

Having documented shape dynamics through the proposed conditional tests, we next examine their consequences for OOS forecast accuracy. These comparisons address the predictive gains from fitting the dynamics, separately from testing whether the shape-dynamics coefficients are zero. We compare M4 with M2 and M5 with M3 separately for raw returns and PC1 residuals. These comparisons assess the incremental contribution of tail-intensity dynamics relative to their respective references. The M5–M3 comparison matches the simulation comparison in Table S1 of the supplement.

All fitted coefficients, Fisher geometry, and PCA preprocessing are held fixed at the training estimates. After observing returns on date $t-1$, we update the dynamic parameters to construct the density forecast

for date t. We evaluate the log score and one- and five-percent quantile losses, using the definitions in Section 4.4. Higher log scores and lower losses indicate better forecasts. We average the daily log scores and quantile losses over the OOS period.

Table 6: Panels with positive OOS forecast improvements

Target	Comparison	Log score	1% loss	5% loss
Raw	M4–M2	11 (2)	14 (0)	8 (0)
Raw	M5–M3	14 (4)	15 (2)	13 (0)
Raw	M5–M4	18 (1)	14 (0)	17 (1)
PC1 residual	M4–M2	13 (4)	15 (3)	10 (0)
PC1 residual	M5–M3	14 (4)	16 (3)	11 (0)
PC1 residual	M5–M4	14 (5)	13 (1)	17 (4)

Each entry is the number of positive point-estimate improvements out of 19 panels, followed by the number with a two-sided HAC(22) forecast-difference p-value below 0.05 in parentheses. The HAC correction here concerns forecast differences, not the lag-zero parameter Wald tests.

Table 7: Broad-market OOS improvements within each target

Target	Market	Comparison	Log score (%)	1% loss (%)
Raw	S&P 500	M4–M2	+0.292	-0.169
Raw	S&P 500	M5–M3	+0.220	+0.837
Raw	Nikkei 225	M4–M2	+0.263	+6.981
Raw	Nikkei 225	M5–M3	+0.300	+7.031
PC1 residual	S&P 500	M4–M2	-0.005	-0.065
PC1 residual	S&P 500	M5–M3	-0.018	+0.009
PC1 residual	Nikkei 225	M4–M2	+0.047	+1.527
PC1 residual	Nikkei 225	M5–M3	+0.101	+1.810

Log-score improvements are $100\{\exp(\overline{\Delta \log g}) - 1\}$, using the proposed model minus the benchmark. Loss improvements are percentage reductions relative to the indicated benchmark. Positive values improve forecasting.

Table 6 counts positive panel-level improvements and, in parentheses, positive differences with two-sided forecast-comparison $p < 0.05$. These forecast-difference standard errors retain HAC(22) (Newey and West, 1987). They are separate from the lag-zero parameter Wald covariances.

For raw returns, M4 improves continuous log score relative to M2 in 11 of 19 panels, while M5 improves on M3 in 14. Their one-percent quantile-loss improvement counts are 14 and 15, and their five-percent counts are eight and thirteen.

For the training-fixed PC1 residuals, M4 and M5 improve log score relative to their respective references in 13 and 14 panels, with four significant positive differences each. One-percent quantile loss improves in 15 and 16 residual panels, with three significant positive differences each. Five-percent quantile loss improves in ten and eleven panels, without significant positive differences.

Broad-market magnitudes in Table 7 illustrate these differences. In raw Nikkei 225 returns, M4 and M5 improve log score by 0.263 and 0.300 percent relative to M2 and M3 and reduce one-percent quantile loss by approximately 6.98 and 7.03 percent. For S&P 500 PC1 residuals, strong parameter-test detections coexist with small negative log-score differences for both corrections. The M5 one-percent quantile-loss reduction is only about 0.009 percent. For Nikkei 225 residuals, the corresponding M4 and M5 loss reductions are 1.53 and 1.81 percent. In-sample persistence detection does not ensure a material improvement in each OOS score.

Removing a common component also removes any predictable variation it contains and changes the forecast target. We therefore do not require the incremental gains to increase after residualization. The

relevant finding is that additional tail-intensity dynamics and some forecast gains remain after the main contemporaneous linear component has been removed.

5.6 Empirical interpretation

The empirical evidence supports persistent tail-intensity dynamics in equity cross-sectional mixture densities. The even component describes changes in left and right tail intensities in the same direction relative to the fitted reference. Asymmetry in a return distribution does not by itself imply persistent changes in its asymmetry. In both M4 and M5, persistence in the imbalance between left and right tail intensities is detected in only a minority of panels, whereas same-direction tail-intensity dynamics are detected in all 19 raw panels and all 19 PC1-residual panels. In M5, the even component remains significant after fitting location, scale, and tail-order dynamics. These results support adding tail-intensity dynamics to the reference model, including after the main within-panel linear common component has been removed.

The OOS results show that fitting these dynamics can also improve density and lower-tail quantile forecasts. Relative to M2 and M3, respectively, M4 and M5 improve log score and one-percent quantile loss in a majority of both raw and PC1-residual panels. The size of these gains varies across panels, and statistically significant gains occur less frequently than positive point estimates. Five-percent quantile-loss improvements are less consistent. The predictive contribution of tail-intensity dynamics is therefore most evident in the log score and the lower one-percent quantile loss, with gains remaining in many panels after PC1 removal.

Taken together, the persistence tests and forecast comparisons suggest that the detected tail-intensity dynamics have predictive value. Persistence is detected in all panels, and including these dynamics improves log score and one-percent quantile loss in most panels, both before and after PC1 removal. The detection of persistent tail variation is thus accompanied by a tendency toward better OOS forecasts. The size of the forecast gains depends on the panel and the criterion, as illustrated by the small gains for the S&P 500 residuals despite strong persistence detections.

6 Conclusion

This paper has developed dynamic conditional densities with spline mixtures (DCD-SM) for a conditional mixture density chosen by the analyst. The model is fitted directly to cross-sectional returns using a weighted log score. Individual conditional densities may differ, and dependence within each date need not be specified. The spline correction allows left and right tail intensities to change while preserving the reference tail order. With the reference distribution and Fisher projection matrices known and the spline dimension fixed, we establish consistency and asymptotic normality of the dynamic coefficient estimators.

Simulations evaluate the finite-sample size and power of the Wald tests and the accuracy of out-of-sample forecasts under independent and dependent cross-sectional sampling. In US and Japanese equity cross sections, we detect persistent changes that affect left and right tail intensities in the same direction, even after fitting dynamic location, scale, and tail order and removing the leading principal component. Adding tail-intensity dynamics improves log scores and one-percent lower-tail quantile losses in most panels, although the magnitude and statistical significance of these gains vary.

The same spline-mixture representation can also be used in density-valued VAR models fitted to previously estimated densities. Those models treat the estimated densities as observations and place stochastic innovations in their dynamics. DCD-SM instead fits the chosen conditional mixture density directly to returns and updates its parameters from past observations through a score-driven recursion, without separate innovations in the density dynamics. Introducing such innovations is a natural extension. This would make the density parameters latent and require a model linking them to cross-sectional returns, together with a filtering method, potentially using particle filtering.

Funding

This work was supported by the Japan Society for the Promotion of Science KAKENHI [grant number 26H01952].

Acknowledgements

The first author used OpenAI GPT-6.1 Sol to assist with manuscript revision and code development. The first author directed these tasks and retained responsibility for checking the generated text and code and for the final manuscript and results.

Data availability

The data sources and sample construction are described in Section 5.1 and Section S5.1 of the supplement. The simulation code, empirical estimation code, table-generation code, and aggregated numerical results are available in the companion repository at https://github.com/matsuda318/DCD-SM. The repository records sample construction and input hashes; the simulations use synthetic returns. Empirical input files and intermediate daily paths are retained locally and are not included in the public repository.

Conflict of interest

The authors declare no conflicts of interest.

References

Blattberg, R. C. and Gonedes, N. J. (1974). A comparison of the stable and Student distributions as statistical models for stock prices. *The Journal of Business*, 47(2):244–280.

Blazsek, S. and Monteros, L. A. (2017). Dynamic conditional score models of degrees of freedom: filtering with score-driven heavy tails. *Applied Economics*, 49(53):5426–5440.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Catania, L. (2021). Dynamic adaptive mixture models with an application to volatility and risk. *Journal of Financial Econometrics*, 19(4):531–564.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196.

Creal, D., Koopman, S. J., and Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28(5):777–795.

de Boor, C. (2001). *A Practical Guide to Splines*. Springer, revised edition.

D'Innocenzo, E., Lucas, A., Schwaab, B., and Zhang, X. (2024). Modeling extreme events: Time-varying extreme tail shape. *Journal of Business & Economic Statistics*, 42(3):903–917.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007.

Hansen, B. E. (1994). Autoregressive conditional density estimation. *International Economic Review*, 35(3):705–730.

Harvey, A. C. and Sucarrat, G. (2014). EGARCH models with fat tails, skewness and leverage. *Computational Statistics & Data Analysis*, 76:320–338.

Harvey, C. R. and Siddique, A. (1999). Autoregressive conditional skewness. *Journal of Financial and Quantitative Analysis*, 34(4):465–487.

Jondeau, E. and Rockinger, M. (2003). Conditional volatility, skewness, and kurtosis: existence, persistence, and comovements. *Journal of Economic Dynamics and Control*, 27(10):1699–1737.

Kelly, B. and Jiang, H. (2014). Tail risk and asset prices. *Review of Financial Studies*, 27(10):2841–2871.

Kokoszka, P., Miao, H., Petersen, A., and Shang, H. L. (2019). Forecasting of density functions with an application to cross-sectional and intraday returns. *International Journal of Forecasting*, 35(4):1304–1317.

Lindsay, B. G. (1988). Composite likelihood methods. In *Statistical Inference from Stochastic Processes*, volume 80 of *Contemporary Mathematics*, pages 221–239. American Mathematical Society.

Lucas, A. and Zhang, X. (2016). Score-driven exponentially weighted moving averages and Value-at-Risk forecasting. *International Journal of Forecasting*, 32(2):293–302.

Massacci, D. (2017). Tail risk dynamics in stock returns: Links to the macroeconomy and global markets connectedness. *Management Science*, 63(9):3072–3089.

Matsuda, Y. and Iwafuchi, R. (2025). Density-valued ARMA models by spline mixtures. *Journal of Time Series Analysis*.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

Palumbo, D. (2026). Testing and modelling time varying (a)symmetric tails. *Quantitative Finance*.

Praetz, P. D. (1972). The distribution of share price changes. *The Journal of Business*, 45(1):49–55.

Thiele, S. (2020). Modeling the conditional distribution of financial returns with asymmetric tails. *Journal of Applied Econometrics*, 35(1):46–60.

van Brummelen, J., Gorgi, P., and Koopman, S. J. (2026). Score-driven time-varying parameter models with spline-based densities. *Statistics and Computing*, 36(2):94.

Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, 21(1):5–42.

Zhang, C., Kokoszka, P., and Petersen, A. (2022). Wasserstein autoregressive models for density time series. *Journal of Time Series Analysis*, 43(1):30–52.

Supplementary material

This supplement gives the Fisher projection formulas, tail-intensity proofs, and proofs of consistency and asymptotic normality. It also reports the simulation coefficients, the M5–M3 forecast comparison, and empirical data and PCA construction details. References prefixed by "S" concern this supplement. References to the main paper are identified explicitly.

S1 Fisher projections at the symmetric Student-t anchor

Let (m_0, σ_0, η_0) be the fixed M1 reference parameters, with $\sigma_0 > 0$ and $\eta_0 > 2$, and let $v_0 = \log \sigma_0^2$. At $q_0 = 0$, the correction density is one, so the corrected return density equals the variance-standardized Student-t reference in (19). Write E_0 for expectation under this reference. The Fisher blocks below are expectations of single-observation score products. They define the fixed projection applied to the weighted date-average scores in the main paper. The formulas are evaluated separately at each raw-return or residual M1 estimate.

S1.1 Reference Fisher blocks and the M3 projection

For $z = (r - m_0)/\sigma_0$, the location and log-variance scores at the anchor are

$$\begin{aligned} \rho_m(z) &= \frac{(\eta_0 + 1)z}{\sigma_0(\eta_0 - 2 + z^2)}, \\ \rho_v(z) &= -\frac{1}{2} + \frac{(\eta_0 + 1)z^2}{2(\eta_0 - 2 + z^2)}. \end{aligned} \tag{S1.1}$$

The location score is odd in z, whereas the log-variance and tail-order scores are even. Hence $I_{m,v,0} = I_{\eta,m,0} = 0$. The remaining entries give

$$I_{ms,ms,0} = \begin{pmatrix} \dfrac{\eta_0(\eta_0 + 1)}{\sigma_0^2(\eta_0 - 2)(\eta_0 + 3)} & 0 \\ 0 & \dfrac{\eta_0}{2(\eta_0 + 3)} \end{pmatrix}. \tag{S1.2}$$

With $\dot{\eta}_0$ as defined in the main paper, the latent tail-order cross-information is

$$I_{\xi_\eta,ms,0} = \left(0,\ \frac{3\dot{\eta}_0}{(\eta_0 - 2)(\eta_0 + 1)(\eta_0 + 3)}\right). \tag{S1.3}$$

Multiplying (S1.3) by the inverse of (S1.2) gives (25). These blocks and the resulting M3 projection coefficient can therefore be computed without numerical integration.

To derive these expressions, let Z have the standardized anchor distribution and set

$$X = \frac{Z^2}{\eta_0 - 2 + Z^2}.$$

Then $X \sim \mathrm{Beta}(1/2, \eta_0/2)$, with

$$E_0 X = \frac{1}{\eta_0 + 1}, \qquad \mathrm{Var}_0(X) = \frac{2\eta_0}{(\eta_0 + 1)^2(\eta_0 + 3)}.$$

Also,

$$E_0\{X(1-X)\} = \frac{\eta_0}{(\eta_0+1)(\eta_0+3)}.$$

Substitution into $E_0\rho_m(Z)^2$ and $E_0\rho_v(Z)^2$ gives (S1.2). For the cross-information, differentiate the log-variance score with respect to η while holding r, m_0, σ_0 fixed. The information identity yields

$$\begin{aligned} I_{\eta,v,0} &= -E_0\left[\left.\frac{\partial\rho_v(Z;\eta)}{\partial\eta}\right|_{\eta=\eta_0}\right] \\ &= -\frac{1}{2(\eta_0+1)} + \frac{\eta_0}{2(\eta_0-2)(\eta_0+3)} \\ &= \frac{3}{(\eta_0-2)(\eta_0+1)(\eta_0+3)}. \end{aligned}$$

The chain rule $\rho_{\xi_\eta} = \dot\eta_0\rho_\eta$ then gives (S1.3).

S1.2 Spline cross-information and the M4–M5 projections

Let $B(u) = (B_1(u), \ldots, B_J(u))'$ and let $m_j = \int_0^1 B_j(u)\,du$ denote the spline masses, distinct from the scalar location m_0. At $q_0 = 0$, the mixture weights equal the mass vector, and the responsibilities equal $B(u)$. The single-observation spline score is thus

$$\psi_q(u,0) = Q\{B(u) - (m_1, \ldots, m_J)'\}.$$

Let $z_0(u)$ be the quantile function of the mean-zero, unit-variance Student-t distribution with η_0 degrees of freedom. Since the reference PIT is uniform, the spline cross-information is

$$I_{q,ms,0} = \int_0^1 \psi_q(u,0)\begin{pmatrix}\rho_m\{z_0(u)\} & \rho_v\{z_0(u)\}\end{pmatrix}du. \tag{S1.4}$$

This is an explicit one-dimensional integral. For general spline knots, its entries are evaluated numerically because they involve the Student-t quantile function together with the spline basis.

For a symmetric knot sequence, $B_j(1-u) = B_{J+1-j}(u)$ and $m_j = m_{J+1-j}$. The even and odd scores at the anchor simplify to

$$\begin{aligned} e'_{\text{even}}\psi_q(u,0) &= B_1(u) + B_J(u) - m_1 - m_J, \\ e'_{\text{odd}}\psi_q(u,0) &= B_J(u) - B_1(u). \end{aligned} \tag{S1.5}$$

The first is even about $u = 1/2$ and the second is odd. Also, $z_0(1-u) = -z_0(u)$, so

$$e'_{\text{even}}I_{q,m,0} = 0, \qquad e'_{\text{odd}}I_{q,v,0} = 0.$$

The two remaining cross-information entries are

$$e'_{\text{even}}I_{q,v,0} = \int_0^1 \{B_1(u) + B_J(u) - m_1 - m_J\}\rho_v\{z_0(u)\}\,du, \tag{S1.6}$$

$$e'_{\text{odd}}I_{q,m,0} = \int_0^1 \{B_J(u) - B_1(u)\}\rho_m\{z_0(u)\}\,du. \tag{S1.7}$$

Combining these entries with (S1.2) gives the coefficients in (27) and the scalar updates in (26). M4

and M5 use the same projection coefficients, computed once at $q = 0$ and the fixed M1 reference values. Their PIT observations and reference scores are computed at each date using the M2 and M3 reference distributions, respectively. The symmetry cancellations concern the fixed anchor and do not require the corrected density at every date to be symmetric.

In the saved experiments, both the reference Fisher blocks and the spline cross-information are computed by Gauss–Legendre quadrature with 128 nodes per PIT subinterval. The reference calculation subdivides the probability interval near the tails, and the spline calculation uses the knot spans. The closed forms above identify the same reference quantities and provide analytic checks.

S2 Proof of the tail-intensity limit

This section proves Proposition 2.1. We first derive the formulas for fixed J, then let the supports of the first and last B-splines shrink as J increases. The amplitude sequences are held fixed in this calculation. The estimation theory in Section 3 uses a fixed J.

S2.1 Formulas for fixed J

Let $B_{1,J}, \dots, B_{J,J}$ be a clamped B-spline basis of fixed degree on $[0, 1]$, and let

$$m_{j,J} = \int_0^1 B_{j,J}(u)\, du.$$

Write e_j for the jth coordinate vector in $\mathbb{R}^J$. With the Helmert matrix Q_J, define

$$e_{\mathrm{even},J} = Q_J(e_1 + e_J), \qquad e_{\mathrm{odd},J} = Q_J(e_J - e_1).$$

For given amplitudes, write

$$q_{t,J} = a_{\mathrm{even},t} e_{\mathrm{even},J} + a_{\mathrm{odd},t} e_{\mathrm{odd},J}, \tag{S2.1}$$

and set

$$\alpha_{L,t} = a_{\mathrm{even},t} - a_{\mathrm{odd},t}, \qquad \alpha_{R,t} = a_{\mathrm{even},t} + a_{\mathrm{odd},t}.$$

Lemma S2.1. *For each J, the mixture weights are*

$$\begin{aligned} p_{1,t,J} &= \frac{m_{1,J} e^{\alpha_{L,t}}}{Z_{t,J}}, \quad p_{J,t,J} = \frac{m_{J,J} e^{\alpha_{R,t}}}{Z_{t,J}}, \\ p_{j,t,J} &= \frac{m_{j,J}}{Z_{t,J}}, \qquad\qquad j = 2, \dots, J-1, \end{aligned} \tag{S2.2}$$

where

$$Z_{t,J} = 1 + m_{1,J}(e^{\alpha_{L,t}} - 1) + m_{J,J}(e^{\alpha_{R,t}} - 1). \tag{S2.3}$$

The correction density is

$$h_J(u \mid q_{t,J}) = \frac{1 + (e^{\alpha_{L,t}} - 1)B_{1,J}(u) + (e^{\alpha_{R,t}} - 1)B_{J,J}(u)}{Z_{t,J}}, \tag{S2.4}$$

and its values at $u = 0$ *and* $u = 1$ *are*

$$h_J(0 \mid q_{t,J}) = \frac{e^{\alpha_{L,t}}}{Z_{t,J}}, \qquad h_J(1 \mid q_{t,J}) = \frac{e^{\alpha_{R,t}}}{Z_{t,J}}. \tag{S2.5}$$

Hence

$$a_{\text{odd},t} = \frac{1}{2}\{\log h_J(1 \mid q_{t,J}) - \log h_J(0 \mid q_{t,J})\}, \tag{S2.6}$$

$$a_{\text{even},t} = \frac{1}{2}\{\log h_J(0 \mid q_{t,J}) + \log h_J(1 \mid q_{t,J})\} + \log Z_{t,J}. \tag{S2.7}$$

Proof. Define the vector

$$c_{t,J} = (\alpha_{L,t}, 0, \ldots, 0, \alpha_{R,t})'.$$

Then $q_{t,J} = Q_J c_{t,J}$. Since

$$Q_J' Q_J = I_J - \frac{1}{J}\mathbf{1}_J\mathbf{1}_J',$$

the entries of $Q_J' q_{t,J}$ differ from those of $c_{t,J}$ by a common constant. This constant cancels between the numerator and denominator of the softmax formula in (3). The unnormalized weights are therefore $m_{1,J}e^{\alpha_{L,t}}$ and $m_{J,J}e^{\alpha_{R,t}}$ for the first and last components, and $m_{j,J}$ for the interior components. Their sum is $Z_{t,J}$, which gives (S2.2). Thus every weight includes the same normalizing factor.

Substituting these weights into the mixture density and using $\sum_j B_{j,J}(u) = 1$ gives (S2.4). The clamped basis satisfies $B_{1,J}(0) = B_{J,J}(1) = 1$. At $u = 0$, all basis functions except $B_{1,J}$ are zero. At $u = 1$, all except $B_{J,J}$ are zero. This gives (S2.5). Subtracting and adding the logarithms of these two values gives (S2.6) and (S2.7). □

S2.2 Limit as J increases

Proof of Proposition 2.1. Choose a finite A such that

$$|a_{\text{even},t}| + |a_{\text{odd},t}| \le A \quad \text{for every } t.$$

Then $|\alpha_{L,t}|, |\alpha_{R,t}| \le A$. With $C = e^A + 1$,

$$|Z_{t,J} - 1| \le C(m_{1,J} + m_{J,J}).$$

The right side tends to zero and does not depend on t. Also, $Z_{t,J} \ge e^{-A}$ because it is a weighted average of $e^{\alpha_{L,t}}$, $e^{\alpha_{R,t}}$, and one. It follows that

$$\sup_t |Z_{t,J} - 1| \to 0, \qquad \sup_t |\log Z_{t,J}| \to 0. \tag{S2.8}$$

We have

$$\left|\frac{e^{\alpha_{L,t}}}{Z_{t,J}} - e^{\alpha_{L,t}}\right| \le e^{2A}|Z_{t,J} - 1|,$$

with the same bound for $\alpha_{R,t}$. Hence

$$\begin{aligned} h_J(0 \mid q_{t,J}) &= e^{a_{\text{even},t} - a_{\text{odd},t}} + o(1), \\ h_J(1 \mid q_{t,J}) &= e^{a_{\text{even},t} + a_{\text{odd},t}} + o(1), \end{aligned} \tag{S2.9}$$

uniformly in t. For each fixed J, the reference CDF tends to zero and one in the left and right tails. Since h_J is continuous at $u = 0$ and $u = 1$, the corrected return density satisfies

$$\begin{aligned} \lim_{r \to -\infty} \frac{\bar{g}_{t,J}(r)}{f_t(r)} &= h_J(0 \mid q_{t,J}), \\ \lim_{r \to +\infty} \frac{\bar{g}_{t,J}(r)}{f_t(r)} &= h_J(1 \mid q_{t,J}). \end{aligned} \tag{S2.10}$$

Taking $J \to \infty$ gives (32). For a Student-t reference with degrees of freedom η_t, the density decays as $|r|^{-(\eta_t+1)}$ in both tails. Multiplication by a positive constant changes the tail coefficient and preserves this exponent. □

S2.3 Integrated difference from the reference

The values $h_J(0 \mid q_{t,J})$ and $h_J(1 \mid q_{t,J})$ need not tend to one, but the integrated absolute difference between h_J and the uniform density does tend to zero. To see this, subtract one from (S2.4):

$$h_J(u \mid q_{t,J}) - 1 = \frac{(e^{\alpha_{L,t}} - 1)\{B_{1,J}(u) - m_{1,J}\} + (e^{\alpha_{R,t}} - 1)\{B_{J,J}(u) - m_{J,J}\}}{Z_{t,J}}.$$

Because the basis functions are nonnegative,

$$\int_0^1 |B_{1,J}(u) - m_{1,J}|\, du \le 2m_{1,J},$$

and the same bound holds with $B_{J,J}$ and $m_{J,J}$. Therefore,

$$\begin{aligned} \int_0^1 |h_J(u \mid q_{t,J}) - 1|\, du &\le \frac{2}{Z_{t,J}} \left\{ |e^{\alpha_{L,t}} - 1| m_{1,J} + |e^{\alpha_{R,t}} - 1| m_{J,J} \right\} \\ &\le 2e^A C(m_{1,J} + m_{J,J}). \end{aligned}$$

The constants A and C are those used in the proof above. Thus

$$\sup_t \int_0^1 |h_J(u \mid q_{t,J}) - 1|\, du = O(m_{1,J} + m_{J,J}) \to 0.$$

Thus the values of h_J at $u = 0$ and $u = 1$ can have limits different from one even though h_J converges to one in integrated absolute difference.

S3 Detailed assumptions and proofs

Throughout, $\|\cdot\|$ denotes the Euclidean norm for vectors and the corresponding induced operator norm for matrices. For a third-order tensor $\mathcal{T} = (T_{rkl})$, $\|\mathcal{T}\|_F$ denotes the Frobenius norm,

$$\|\mathcal{T}\|_F := \left(\sum_{r,k,l} T_{rkl}^2\right)^{1/2}.$$

The reference distribution and Fisher projection matrices are known, and the spline dimension J is fixed. We prove consistency and asymptotic normality of the six DCD-SM dynamic coefficient estimators in M4 and M5. The weights $w_{t,i}$ are nonnegative and sum to one. Individual conditional densities may differ, and returns within a date may be dependent. Bounds on individual spline scores and their state derivatives also bound their weighted averages.

S3.1 Consistency

We establish consistency for the score update equation used in M4 and M5. We assume that, near the true parameter, it has a unique stationary causal solution and that the effect of initialization decays uniformly at an exponential rate.

Use the weights $p(q)$, density $h(u \mid q)$, and responsibilities $\tau(u, q)$ from Sections 2.1 and 2.2 of the main paper, with $QQ' = I_{J-1}$ and $Q\mathbf{1}_J = 0$. For $u_{t,i} = F_t(r_{t,i})$, the weighted spline score is $s_t(q) = \sum_i w_{t,i}\psi(u_{t,i}, q)$, where

$$\psi(u, q) = Q\{\tau(u, q) - p(q)\}. \tag{S3.1}$$

Lemma S3.1. *For fixed J,*

$$\sup_{u,q} \|\psi(u, q)\| \le \sqrt{2}, \qquad \sup_{u,q} \left\| \frac{\partial \psi(u, q)}{\partial q'} \right\| \le 1.$$

Let

$$\mathcal{T}_\psi(u, q) := \left[\frac{\partial^2 \psi_r(u, q)}{\partial q_k \, \partial q_l} \right]_{r,k,l=1}^{J-1}$$

denote the third-order tensor of second derivatives of the vector score. Then there exists a finite constant C_J, depending only on the fixed spline dimension J, such that

$$\sup_{u,q} \|\mathcal{T}_\psi(u, q)\|_F \le C_J < \infty. \tag{S3.2}$$

Therefore, a first-order Taylor expansion in q gives, for any $q, \widetilde{q} \in \mathbb{R}^{J-1}$,

$$\|\psi(u, q) - \psi(u, \widetilde{q})\| \le \|q - \widetilde{q}\|, \tag{S3.3}$$

and

$$\left\| \frac{\partial \psi(u, q)}{\partial q'} - \frac{\partial \psi(u, \widetilde{q})}{\partial q'} \right\| \le C_J \|q - \widetilde{q}\|. \tag{S3.4}$$

Proof. Both $\tau(u, q)$ and $p(q)$ are probability vectors, so $\|\tau(u, q) - p(q)\| \le \sqrt{2}$. Since $\|Q\| = 1$, the

score bound follows. Writing $V(a) = \mathrm{diag}(a) - aa'$, differentiation gives

$$\frac{\partial \psi(u,q)}{\partial q'} = Q\{V(\tau(u,q)) - V(p(q))\}Q'.$$

For a probability vector a, the diagonal entry of $V(a)$ and the sum of the absolute off-diagonal entries in the same row both equal $a_i(1-a_i)$. Gershgorin's theorem gives $\|V(a)\| \le \max_i 2a_i(1-a_i) \le 1/2$, so the score Jacobian has norm at most one.

For a weighted softmax vector $a_j(z) = c_j e^{z_j}/\sum_s c_s e^{z_s}$ with $c_j \ge 0$ and $\sum_j c_j > 0$,

$$\frac{\partial a_j(z)}{\partial z_l} = a_j(z)\{\delta_{jl} - a_l(z)\}. \tag{S3.5}$$

Consequently,

$$\frac{\partial V_{jk}\{a(z)\}}{\partial z_l} = a_j(\delta_{jl} - a_l)(\delta_{jk} - a_k) - a_j a_k(\delta_{kl} - a_l). \tag{S3.6}$$

The right side is uniformly bounded. Both $p(q)$ and $\tau(u,q)$ are weighted softmax vectors with index $Q'q$. Applying the chain rule to the score Jacobian therefore gives a uniform bound on every entry of $\mathcal{T}_\psi(u,q)$. Since J and Q are fixed, its Frobenius norm is bounded by a finite constant C_J. The first-order Taylor formula gives (S3.3) and (S3.4). □

DCD-SM as a stochastic recurrence equation

Let $E = (e_{\mathrm{even}}, e_{\mathrm{odd}})$ contain the vectors in (14). Define

$$z_t = (a_{\mathrm{even},t}, a_{\mathrm{odd},t})', \qquad q_t = Ez_t. \tag{S3.7}$$

For M4 and M5, write the fixed-anchor Fisher projection as

$$s_t^{\perp}(q) = s_t(q) + b_t, \qquad b_t := -I_{q,ms,0}I_{ms,ms,0}^{-1}s_{ms,t}, \qquad ms = (m, \log\sigma^2).$$

The projection coefficient $I_{q,ms,0}I_{ms,ms,0}^{-1}$ is computed once at $(ms, q) = (ms_0, 0)$ and held fixed. The reference score $s_{ms,t}$ is computed from the date-t returns using the known reference distribution, as in (11). Define

$$r_t(z) := E's_t^{\perp}(Ez) = E'\{s_t(Ez) + b_t\} \in \mathbb{R}^2. \tag{S3.8}$$

With the observed returns and reference distribution held fixed, its state Jacobian is

$$G_t(z) := \frac{\partial r_t(z)}{\partial z'} = E'H_t(Ez)E, \qquad H_t(q) := \frac{\partial s_t(q)}{\partial q'}. \tag{S3.9}$$

Lemma S3.1 implies that $G_t(z)$ is globally bounded and globally Lipschitz in z.

The parameter vector contains the six HAR coefficients,

$$\theta = \kappa = (\kappa_{\mathrm{even},1}, \kappa_{\mathrm{even},5}, \kappa_{\mathrm{even},22}, \kappa_{\mathrm{odd},1}, \kappa_{\mathrm{odd},5}, \kappa_{\mathrm{odd},22})', \qquad d_\theta = 6. \tag{S3.10}$$

For $h \in \{1, 5, 22\}$, define

$$C_h(\theta) := \mathrm{diag}\left(\kappa_{\mathrm{even},h}, \kappa_{\mathrm{odd},h}\right). \tag{S3.11}$$

Expanding the HAR$(1,5,22)$ moving averages in (18) gives

$$z_{t+1} = \sum_{j=0}^{21} K_j(\theta) r_{t-j}(z_{t-j}), \tag{S3.12}$$

where

$$K_j(\theta) = \begin{cases} C_1(\theta) + \frac{1}{5}C_5(\theta) + \frac{1}{22}C_{22}(\theta), & j = 0, \\ \frac{1}{5}C_5(\theta) + \frac{1}{22}C_{22}(\theta), & j = 1, \dots, 4, \\ \frac{1}{22}C_{22}(\theta), & j = 5, \dots, 21. \end{cases}$$

Now define the companion state

$$X_t := \begin{pmatrix} z_t' & z_{t-1}' & \cdots & z_{t-21}' \end{pmatrix}' \in \mathbb{R}^{44}. \tag{S3.13}$$

Then (S3.12) is the stochastic recurrence equation

$$X_{t+1} = \Phi_t(X_t; \theta). \tag{S3.14}$$

For $x = (z_t', z_{t-1}', \dots, z_{t-21}')'$, let

$$A_t(x;\theta) := \frac{\partial \Phi_t(x;\theta)}{\partial x'}.$$

Its companion form is

$$A_t(x;\theta) = \begin{pmatrix} K_0(\theta)G_t(z_t) & K_1(\theta)G_{t-1}(z_{t-1}) & \cdots & K_{21}(\theta)G_{t-21}(z_{t-21}) \\ I_2 & 0 & \cdots & 0 \\ 0 & I_2 & \ddots & \vdots \\ \vdots & & \ddots & 0 \end{pmatrix}. \tag{S3.15}$$

Thus local stability of the score update equation is governed by products of the feedback matrices $K_j(\theta)G_{t-j}$ rather than by the HAR lag matrices alone.

Assumptions

Assumption 1 (Conditional weighted-average-density specification). The reference distribution F_t, the constituent set and its finite size $N_t \geq 1$, and the weights $w_{t,i} \geq 0$ with $\sum_i w_{t,i} = 1$ are given before the date-t returns are observed. They are $\mathcal{F}_{t-1}$-measurable and do not depend on θ. The Fisher projection matrices are known and held fixed. Conditional on $\mathcal{F}_{t-1}$, each $u_{t,i}$ has density $\phi_{t,i}$ on $[0,1]$, and

$$\bar{\phi}_t(u) := \sum_{i=1}^{N_t} w_{t,i}\phi_{t,i}(u) = h\{u \mid q_t(\theta_0)\} \quad \text{Lebesgue-a.e., a.s.} \tag{S3.16}$$

The individual densities need not be equal, and conditional independence across i is not required. No stationarity or ergodicity assumption is imposed here on the full cross-sectional observation process.

For later use, Assumption 1 implies that any jointly measurable random function $a_t(u)$, measurable

with respect to $\mathcal{F}_{t-1}$ for each u and satisfying $E[\int_0^1 |a_t(u)|\bar{\phi}_t(u)\,du] < \infty$, obeys

$$E\left[\sum_{i=1}^{N_t} w_{t,i} a_t(u_{t,i}) \,\middle|\, \mathcal{F}_{t-1}\right] = \sum_{i=1}^{N_t} w_{t,i} \int_0^1 a_t(u)\phi_{t,i}(u)\,du = \int_0^1 a_t(u) h\{u \mid q_t(\theta_0)\}\,du. \tag{S3.17}$$

The identity also applies componentwise to vector- and matrix-valued functions.

Assumption 2 (Stationarity of q_t and local stability). The true parameter θ_0 belongs to the interior of a compact parameter neighborhood $\mathcal{N}$. For every $\theta \in \mathcal{N}$, the recursion (S3.14) admits a unique causal solution $X_t(\theta)$, measurable with respect to $\mathcal{F}_{t-1}$. The associated spline-state process $q_t(\theta)$ is almost surely continuous in θ on $\mathcal{N}$, and

$$\{q_t(\cdot)\}_{t\in\mathbb{Z}} \quad \text{in } C(\mathcal{N}; \mathbb{R}^{J-1})$$

is strictly stationary and ergodic. This is a joint condition on the candidate state family, including $q_t(\theta_0)$, rather than a separate condition for each parameter value. Since E has full column rank, $z_t(\theta) = (E'E)^{-1}E'q_t(\theta)$. The companion state $X_t(\theta)$ stacks finitely many lags of $z_t(\theta)$. Thus $X_t(\cdot)$ inherits continuity, strict stationarity, and ergodicity. There is a finite constant C_b such that $\|b_t\| \le C_b$ almost surely for all t. Moreover, for any fixed initialization in a common local basin of attraction, there exist a finite random variable C and a constant $\rho \in (0,1)$ such that

$$\sup_{\theta\in\mathcal{N}} \|\widehat{X}_t(\theta) - X_t(\theta)\| \le C\rho^t \qquad \text{a.s. for all sufficiently large } t. \tag{S3.18}$$

Assumption 3 (Local dynamic identification). For the given reference distribution and Fisher projection matrices, and for $\theta \in \mathcal{N}$,

$$q_t(\theta) = q_t(\theta_0) \quad \text{a.s. for all } t$$

implies $\theta = \theta_0$.

Uniform convergence of the local criterion

For $\theta \in \mathcal{N}$, define

$$\ell_t(\theta) = \sum_{i=1}^{N_t} w_{t,i} \log h\{u_{t,i} \mid q_t(\theta)\}, \qquad L_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} \ell_t(\theta),$$

where $q_t(\theta)$ is obtained from the first two entries of $X_t(\theta)$ using (S3.7). Let

$$L(\theta) = E[\ell_t(\theta)].$$

Because the reference distribution at each date does not depend on θ, maximizing $L_T(\theta)$ is equivalent to maximizing the weighted cross-sectional return log-score criterion over $\mathcal{N}$. Define the local estimator

by

$$\widehat{\theta}_T \in \arg\max_{\theta \in \mathcal{N}} L_T(\theta). \tag{S3.19}$$

The criterion is uniformly bounded. Indeed, by Lemma S3.1, $s_t(q)$ is uniformly bounded, and b_t is uniformly bounded by Assumption 2. Since the matrices $K_j(\theta)$ are bounded on the compact neighborhood $\mathcal{N}$, (S3.12) implies

$$\sup_{t,\theta \in \mathcal{N}} \|z_t(\theta)\| \le M_z < \infty,$$

and hence

$$\sup_{t,\theta \in \mathcal{N}} \|q_t(\theta)\| \le M_q < \infty.$$

If $v = Q'q$, then

$$h(u \mid q) = \frac{\sum_{j=1}^{J} B_j(u) e^{v_j}}{\sum_{j=1}^{J} m_j e^{v_j}}.$$

Using the partition-of-unity identities $\sum_j B_j(u) = 1$ and $\sum_j m_j = 1$ gives

$$e^{-2M_q} \le h\{u \mid q_t(\theta)\} \le e^{2M_q} \tag{S3.20}$$

uniformly in u, t, and $\theta \in \mathcal{N}$.

Lemma S3.2. *Under Assumptions 1–2,*

$$\sup_{\theta \in \mathcal{N}} |L_T(\theta) - L(\theta)| \longrightarrow 0 \qquad a.s. \tag{S3.21}$$

Proof. For each fixed $\theta \in \mathcal{N}$, decompose the criterion as

$$\ell_t(\theta) = d_t(\theta) + m_t(\theta), \qquad m_t(\theta) := E\{\ell_t(\theta) \mid \mathcal{F}_{t-1}\}.$$

The conditional-average identity gives

$$m_t(\theta) = \int_0^1 h\{u \mid q_t(\theta_0)\} \log h\{u \mid q_t(\theta)\}\, du. \tag{S3.22}$$

Thus $m_t(\theta)$ is a measurable function of the jointly stationary and ergodic family $q_t(\theta)$ in Assumption 2. Its expectation $E[m_t(\theta)] = L(\theta)$ does not depend on t. The bound (S3.20) implies $|\ell_t(\theta)|, |m_t(\theta)| \le 2M_q$, so $d_t(\theta)$ is a bounded martingale difference. In particular,

$$\sum_{t=1}^{\infty} \frac{E[d_t(\theta)^2]}{t^2} < \infty, \qquad \frac{1}{T} \sum_{t=1}^{T} d_t(\theta) \longrightarrow 0 \quad \text{a.s.}$$

The second statement follows from martingale convergence and Kronecker's lemma. Applying the ergodic theorem to $m_t(\theta)$ now gives

$$L_T(\theta) \to L(\theta) \qquad \text{a.s.}$$

for every fixed θ.

To obtain uniform convergence without assuming stationarity of $\ell_t(\theta)$, define

$$\omega_t(\delta) := \sup_{\substack{\theta,\widetilde{\theta}\in\mathcal{N} \\ \|\theta-\widetilde{\theta}\|\le\delta}} \|q_t(\theta) - q_t(\widetilde{\theta})\|.$$

Assumption 2 implies that $\omega_t(\delta) \to 0$ almost surely as $\delta \downarrow 0$, while $\omega_t(\delta) \le 2M_q$. The bounded score in Lemma S3.1 gives the bound

$$|\ell_t(\theta) - \ell_t(\widetilde{\theta})| \le \sqrt{2}\,\omega_t(\delta) \quad \text{if } \|\theta - \widetilde{\theta}\| \le \delta.$$

The same bound holds for $|m_t(\theta) - m_t(\widetilde{\theta})|$, since $\omega_t(\delta)$ is predictable. Dominated convergence gives

$$E[\omega_t(\delta)] \to 0.$$

For fixed δ, $\omega_t(\delta)$ is stationary and ergodic by Assumption 2, so

$$\frac{1}{T}\sum_{t=1}^{T} \omega_t(\delta) \to E[\omega_t(\delta)] \qquad \text{a.s.}$$

A finite δ-net of the compact set $\mathcal{N}$, followed by $\delta \downarrow 0$, proves (S3.21). Only the conditional expectation $m_t(\theta)$ and $\omega_t(\delta)$ require an ergodic theorem. The observation-dependent remainder is handled by a martingale law of large numbers. □

Consistency theorem

Theorem S3.3. *Suppose that Assumptions 1–3 hold. Then the local estimator in* (S3.19) *satisfies*

$$\widehat{\theta}_T \longrightarrow \theta_0 \qquad a.s.$$

The same conclusion holds for the criterion evaluated from any fixed initialization covered by Assumption 2.

Proof. Lemma S3.2 gives uniform convergence of the criterion on $\mathcal{N}$. It remains to identify its population maximizer.

Since $q_t(\theta)$ is $\mathcal{F}_{t-1}$-measurable, (S3.17) applied to the log density ratio gives

$$\begin{aligned} L(\theta_0) - L(\theta) &= E\left[\sum_{i=1}^{N_t} w_{t,i} \log \frac{h\{u_{t,i} \mid q_t(\theta_0)\}}{h\{u_{t,i} \mid q_t(\theta)\}}\right] \\ &= E\left[\int_0^1 h\{u \mid q_t(\theta_0)\} \log \frac{h\{u \mid q_t(\theta_0)\}}{h\{u \mid q_t(\theta)\}}\, du\right] \\ &= E\left[\mathrm{KL}\left\{h(\cdot \mid q_t(\theta_0)) \,\|\, h(\cdot \mid q_t(\theta))\right\}\right] \ge 0. \end{aligned} \tag{S3.23}$$

If equality holds, the nonnegative KL divergence is zero almost surely, so

$$h(\cdot \mid q_t(\theta)) = h(\cdot \mid q_t(\theta_0)) \quad \text{Lebesgue-a.e., a.s.}$$

Linear independence of the normalized B-spline basis implies

$$p\{q_t(\theta)\} = p\{q_t(\theta_0)\} \quad \text{a.s.}$$

The weighted-softmax mapping is one-to-one because

$$Q \log\{p(q)/m\} = QQ'q - \log\left(\sum_{j=1}^{J} m_j e^{(Q'q)_j}\right) Q\mathbf{1}_J = q.$$

Hence

$$q_t(\theta) = q_t(\theta_0) \quad \text{a.s. for all } t,$$

and Assumption 3 yields $\theta = \theta_0$. Thus $L(\theta)$ has the unique maximizer θ_0 on $\mathcal{N}$. Combining identification with (S3.21) gives

$$\widehat{\theta}_T \to \theta_0 \qquad \text{a.s.}$$

by the usual argmax argument.

Finally, local-uniform forgetting in (S3.18), together with the bounded score, implies that computing the score update equation from fixed initial values in the same local basin changes the average criterion by $o(1)$ uniformly on $\mathcal{N}$. The initialized estimator therefore has the same consistency limit. □

S3.2 Asymptotic Normality

We now derive the asymptotic distribution of the local estimator. The derivatives of $X_t(\theta)$ are obtained by differentiating the score update equation. Their existence, continuity, and moments are stated directly below.

Let

$$S := \begin{pmatrix} I_2 & 0 & \cdots & 0 \end{pmatrix} \in \mathbb{R}^{2\times 44},$$

so that $z_t(\theta) = SX_t(\theta)$ and

$$q_t(\theta) = ESX_t(\theta).$$

Take the compact neighborhood $\mathcal{N}$ in Assumption 2 to be convex, shrinking it if necessary. The true parameter θ_0 remains an interior point.

Derivative stochastic recursions

Define the first parameter derivative matrix

$$D_t(\theta) := \frac{\partial X_t(\theta)}{\partial \theta'} \in \mathbb{R}^{44\times d_\theta}. \tag{S3.24}$$

Differentiating the SRE (S3.14) by the ordinary chain rule gives

$$D_{t+1}(\theta) = A_t\{X_t(\theta);\theta\}D_t(\theta) + C_t(\theta), \tag{S3.25}$$

where

$$C_t(\theta) := \left.\frac{\partial \Phi_t(x;\theta)}{\partial \theta'}\right|_{x=X_t(\theta)} \in \mathbb{R}^{44\times d_\theta}.$$

In the empirical HAR specification the matrices $K_j(\theta)$ are linear in the six HAR coefficients. Because $r_t(z)$ is uniformly bounded, $C_t(\theta)$ is locally uniformly bounded.

For the second derivative, define the third-order tensor

$$R_t(\theta) := \left[\frac{\partial^2 X_{t,r}(\theta)}{\partial\theta_j \, \partial\theta_k} \right]_{\substack{r=1,\dots,44 \\ j,k=1,\dots,d_\theta}}. \tag{S3.26}$$

Write $A_{t,ra}$ and $D_{t,aj}$ for the entries of $A_t\{X_t(\theta);\theta\}$ and $D_t(\theta)$, respectively. A second application of the chain rule gives, for $r = 1,\dots,44$ and $j,k = 1,\dots,d_\theta$,

$$\begin{aligned} R_{t+1,rjk} = {} & \sum_{a=1}^{44} A_{t,ra} R_{t,ajk} \\ & + \sum_{a=1}^{44}\sum_{b=1}^{44} \frac{\partial^2 \Phi_{t,r}}{\partial x_a \, \partial x_b} D_{t,aj} D_{t,bk} \\ & + \sum_{a=1}^{44} \frac{\partial^2 \Phi_{t,r}}{\partial x_a \, \partial\theta_k} D_{t,aj} + \sum_{a=1}^{44} \frac{\partial^2 \Phi_{t,r}}{\partial x_a \, \partial\theta_j} D_{t,ak} \\ & + \frac{\partial^2 \Phi_{t,r}}{\partial\theta_j \, \partial\theta_k}. \end{aligned} \tag{S3.27}$$

All derivatives of Φ_t in (S3.27) are evaluated at $\{X_t(\theta),\theta\}$. For the empirical HAR parameterization,

$$\frac{\partial^2 \Phi_{t,r}}{\partial\theta_j \, \partial\theta_k} = 0.$$

Moreover, Lemma S3.1 implies that the second derivatives of Φ_t with respect to the state and the mixed state–parameter derivatives are locally uniformly bounded for $X_t(\theta), \theta \in \mathcal{N}$. Hence the nonhomogeneous part of (S3.27) is bounded in norm by a constant multiple of

$$1 + \|D_t(\theta)\| + \|D_t(\theta)\|^2.$$

Assumption 4 (Derivative regularity and moments). On $\mathcal{N}$, $X_t(\theta)$ is twice continuously differentiable in θ almost surely. The derivative processes $D_t(\theta)$ and $R_t(\theta)$ defined above are causal and almost surely continuous in θ. The random-function process $\{(X_t(\cdot), D_t(\cdot), R_t(\cdot))\}$ is jointly strictly stationary and ergodic in the corresponding product of continuous-function spaces. For some $\delta > 0$,

$$E\left[\sup_{\theta\in\mathcal{N}} \|D_t(\theta)\|^{2+\delta}\right] < \infty, \qquad E\left[\sup_{\theta\in\mathcal{N}} \|R_t(\theta)\|_F^{1+\delta}\right] < \infty. \tag{S3.28}$$

For any fixed initialization covered by Assumption 2, the corresponding first- and second derivatives computed from these initial values satisfy

$$\sup_{\theta\in\mathcal{N}} \left\{ \|\widehat{D}_t(\theta) - D_t(\theta)\| + \|\widehat{R}_t(\theta) - R_t(\theta)\|_F \right\} \le C_D \rho_D^t \quad \text{a.s.} \tag{S3.29}$$

for some finite random variable C_D and some $\rho_D \in (0,1)$.

Define

$$J_t(\theta) := \frac{\partial q_t(\theta)}{\partial \theta'} = ES\frac{\partial X_t(\theta)}{\partial \theta'} \in \mathbb{R}^{(J-1)\times d_\theta}, \tag{S3.30}$$

and write $J_t = J_t(\theta_0)$.

Let

$$\mathcal{I}_q(q) := \int_0^1 \psi(u,q)\psi(u,q)'h(u \mid q)\, du.$$

Assumption 5 (Local information and variance stabilization). The negative expected Hessian

$$\mathcal{H} := -E[\nabla_\theta^2 \ell_t(\theta_0)] = E\left[J_t' \mathcal{I}_q\{q_t(\theta_0)\} J_t\right] \tag{S3.31}$$

is positive definite. There is a finite nonrandom matrix Ω such that

$$\frac{1}{T}\sum_{t=1}^{T} J_t'\Sigma_t J_t \longrightarrow \Omega \qquad \text{in probability,} \tag{S3.32}$$

where

$$\Sigma_t := \mathrm{Var}\{s_t(q_t(\theta_0)) \mid \mathcal{F}_{t-1}\}.$$

This condition is used for asymptotic normality, not consistency. It controls the aggregate effect of within-date heterogeneity and dependence without requiring stationarity or ergodicity of the full observation process.

Martingale score and Hessian

The parameter score at the truth is

$$S_t := \nabla_\theta \ell_t(\theta_0) = J_t' s_t\{q_t(\theta_0)\}. \tag{S3.33}$$

Its conditional covariance is

$$E[S_t S_t' \mid \mathcal{F}_{t-1}] = J_t'\Sigma_t J_t, \tag{S3.34}$$

whose time average converges to Ω by Assumption 5. Assumption 4 and the bounded DCD-SM score imply $\sup_t E\|S_t\|^{2+\delta} < \infty$.

Lemma S3.4. *Under Assumptions 1, 2, 4, and 5,*

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} S_t \Longrightarrow N(0,\Omega). \tag{S3.35}$$

Proof. The solution $X_t(\theta_0)$ and its derivative D_t are measurable with respect to $\mathcal{F}_{t-1}$. Hence J_t is $\mathcal{F}_{t-1}$-measurable. Individual scores may have nonzero conditional means. Applying (S3.17) to $\psi\{u, q_t(\theta_0)\}$ instead gives

$$\begin{aligned} E\left[s_t\{q_t(\theta_0)\} \mid \mathcal{F}_{t-1}\right] &= \int_0^1 \psi\{u, q_t(\theta_0)\}\bar{\phi}_t(u)\, du \\ &= \int_0^1 \psi\{u, q_t(\theta_0)\} h\{u \mid q_t(\theta_0)\}\, du = 0. \end{aligned}$$

The final equality is the density-score identity for the weighted-average-density model. Neither identical conditional marginals nor conditional independence across i is used in this argument. Therefore

$$E(S_t \mid \mathcal{F}_{t-1}) = 0,$$

and $\{S_t, \mathcal{F}_t\}$ is a martingale difference sequence. Its conditional covariance is

$$E(S_t S_t' \mid \mathcal{F}_{t-1}) = J_t' \Sigma_t J_t.$$

By the variance-stabilization condition in Assumption 5,

$$\frac{1}{T}\sum_{t=1}^{T} E(S_t S_t' \mid \mathcal{F}_{t-1}) \longrightarrow \Omega \qquad \text{in probability.}$$

The uniform $(2+\delta)$ moment bound implies the conditional Lindeberg condition in probability: for every $\epsilon > 0$, the expectation of

$$\frac{1}{T}\sum_{t=1}^{T} E\left[\|S_t\|^2 \mathbf{1}\{\|S_t\| > \epsilon\sqrt{T}\} \,\middle|\, \mathcal{F}_{t-1}\right]$$

is $O(T^{-\delta/2})$. The martingale central limit theorem gives (S3.35). □

For $\theta \in \mathcal{N}$, define

$$H_t\{q_t(\theta)\} := \frac{\partial s_t\{q_t(\theta)\}}{\partial q'} = \sum_{i=1}^{N_t} w_{t,i} \frac{\partial \psi\{u_{t,i}, q_t(\theta)\}}{\partial q'}.$$

Let

$$R_{q,t}(\theta) := \left[\frac{\partial^2 q_{t,r}(\theta)}{\partial\theta_j\, \partial\theta_k}\right]_{\substack{r=1,\ldots,J-1 \\ j,k=1,\ldots,d_\theta}}$$

denote the third-order tensor of second parameter derivatives of the embedded spline state. Since $q_t(\theta) = ESX_t(\theta)$,

$$R_{q,t,rjk}(\theta) = \sum_{a=1}^{44} (ES)_{ra} R_{t,ajk}(\theta).$$

Write $J_{t,\cdot j}(\theta)$ for the jth column of $J_t(\theta)$. The (j,k) element of the criterion Hessian is

$$\begin{aligned}\frac{\partial^2 \ell_t(\theta)}{\partial\theta_j\, \partial\theta_k} &= J_{t,\cdot j}(\theta)' H_t\{q_t(\theta)\} J_{t,\cdot k}(\theta) \\ &\quad + \sum_{r=1}^{J-1} s_{t,r}\{q_t(\theta)\} R_{q,t,rjk}(\theta).\end{aligned} \tag{S3.36}$$

At the truth, the conditional expectation of the second term is zero because $R_{q,t}(\theta_0)$ is predictable and the weighted cross-sectional score has conditional mean zero. Applying (S3.17) to the score Jacobian

and then using the density information identity gives

$$\begin{aligned} E\left[H_t\{q_t(\theta_0)\} \mid \mathcal{F}_{t-1}\right] &= \int_0^1 \frac{\partial \psi\{u, q_t(\theta_0)\}}{\partial q'} \bar{\phi}_t(u)\, du \\ &= \int_0^1 \frac{\partial \psi\{u, q_t(\theta_0)\}}{\partial q'} h\{u \mid q_t(\theta_0)\}\, du \\ &= -\mathcal{I}_q\{q_t(\theta_0)\}. \end{aligned}$$

This is an identity for the expected average Hessian, not an information equality for the actual weighted cross-sectional score covariance. Consequently,

$$-E\{\nabla_\theta^2 \ell_t(\theta_0)\} = \mathcal{H}. \tag{S3.37}$$

Lemma S3.5. *Under Assumptions 1, 2, and 4,*

$$\sup_{\theta \in \mathcal{N}} \left\| \frac{1}{T} \sum_{t=1}^{T} \nabla_\theta^2 \ell_t(\theta) - E\{\nabla_\theta^2 \ell_t(\theta)\} \right\| \longrightarrow 0 \qquad a.s. \tag{S3.38}$$

The map $\theta \mapsto E\{\nabla_\theta^2 \ell_t(\theta)\}$ is continuous at θ_0.

Proof. Lemma S3.1 gives

$$\sup_{t,q} \|H_t(q)\| \leq 1$$

and a uniform bound on the third-order tensor of first derivatives of $H_t(q)$ with respect to q. Therefore the first term of (S3.36) is bounded by a constant multiple of $\|D_t(\theta)\|^2$, while the second is bounded by a constant multiple of $\|R_t(\theta)\|_F$. The moment bounds in Assumption 4 apply to both terms uniformly over $\theta \in \mathcal{N}$.

To separate the observations from $q_t(\theta)$ and its derivatives, let $\Xi_t(u, \theta)$ be the one-observation Hessian contribution, with entries

$$\Xi_{t,jk}(u, \theta) = J_{t,\cdot j}(\theta)' \frac{\partial \psi\{u, q_t(\theta)\}}{\partial q'} J_{t,\cdot k}(\theta) + \sum_{r=1}^{J-1} \psi_r\{u, q_t(\theta)\} R_{q,t,rjk}(\theta).$$

Then

$$\begin{aligned} \nabla_\theta^2 \ell_t(\theta) &= \sum_i w_{t,i} \Xi_t(u_{t,i}, \theta), \\ A_t^H(\theta) &:= E\{\nabla_\theta^2 \ell_t(\theta) \mid \mathcal{F}_{t-1}\} \\ &= \int_0^1 \Xi_t(u, \theta) h\{u \mid q_t(\theta_0)\}\, du. \end{aligned}$$

By Assumption 4, $A_t^H(\theta)$ is stationary and ergodic for each fixed θ. Moreover,

$$\sup_{u,\theta} \|\Xi_t(u, \theta)\| \leq K_t := C \sup_{\theta \in \mathcal{N}} \{\|D_t(\theta)\|^2 + \|R_t(\theta)\|_F\},$$

where K_t is stationary and ergodic and has a finite pth moment for some $1 < p \leq 2$, by (S3.28). Consequently the centered Hessian is a martingale difference with uniformly bounded pth moments.

Its time average converges to zero almost surely by the martingale strong law, since $\sum_t t^{-p} < \infty$. The ergodic theorem applied to $A_t^H(\theta)$ yields pointwise convergence of the average Hessian to its expectation, which is independent of t.

For uniformity, use the predictable modulus

$$\omega_t^H(\varepsilon) := \sup_{\substack{\theta,\widetilde{\theta}\in\mathcal{N} \\ \|\theta-\widetilde{\theta}\|\le\varepsilon}} \sup_{u\in[0,1]} \|\Xi_t(u,\theta) - \Xi_t(u,\widetilde{\theta})\|.$$

The score bounds and Assumption 4 imply almost-sure convergence of this modulus to zero as $\varepsilon \downarrow 0$, uniformly in u. It is a stationary and ergodic function of $q_t(\theta)$ and its derivative families and is bounded by $2K_t$. It bounds both the observed Hessian modulus and that of its conditional expectation. Dominated convergence, the ergodic theorem for this modulus, and the finite-net argument in Lemma S3.2 prove (S3.38). The bound $2K_t$ gives continuity of the expectation by dominated convergence. □

Asymptotic-normality theorem

Theorem S3.6. *Suppose that Assumptions 1–5 hold. Then the locally stable estimator in* (S3.19) *satisfies*

$$\sqrt{T}(\widehat{\theta}_T - \theta_0) \Longrightarrow N\left(0,\, \mathcal{H}^{-1}\Omega\mathcal{H}^{-1}\right). \tag{S3.39}$$

If Ω *is positive definite, the limiting covariance matrix is positive definite.*

Proof. By Theorem S3.3, $\widehat{\theta}_T \to \theta_0$ almost surely. Since θ_0 is an interior point of the convex stable neighborhood $\mathcal{N}$, the first-order condition holds with probability tending to one. The integral Taylor expansion gives

$$\begin{aligned} 0 &= \nabla_\theta L_T(\theta_0) \\ &\quad + \left[\int_0^1 \nabla_\theta^2 L_T\{\theta_0 + x(\widehat{\theta}_T - \theta_0)\}\, dx\right](\widehat{\theta}_T - \theta_0). \end{aligned}$$

Lemma S3.5, (S3.37), and consistency imply that the matrix in brackets converges in probability to $-\mathcal{H}$. Lemma S3.4 gives

$$\sqrt{T}\,\nabla_\theta L_T(\theta_0) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} S_t \Longrightarrow N(0,\Omega).$$

Assumption 5 and Slutsky's theorem therefore yield (S3.39). □

Initialization and sandwich inference

Write $\widehat{\ell}_t(\theta)$ for the criterion evaluated from initialized solutions of the score update equation, with the same known reference distribution and Fisher projection matrices as before. Define

$$S_t(\theta) = \nabla_\theta \ell_t(\theta), \qquad \widehat{S}_t(\theta) = \nabla_\theta \widehat{\ell}_t(\theta),$$

so that $S_t(\theta_0) = S_t$ in (S3.33). The score bounds, the Hessian formula (S3.36), and the forgetting conditions in Assumptions 2 and 4 imply, for all sufficiently large t,

$$\begin{aligned}&\sup_{\theta\in\mathcal{N}}\big\{\|\widehat{S}_t(\theta) - S_t(\theta)\| \\ &\qquad + \|\nabla^2_\theta \widehat{\ell}_t(\theta) - \nabla^2_\theta \ell_t(\theta)\|\big\} \le C_* \rho_*^t M_t, \qquad M_t = 1 + \sup_{\theta\in\mathcal{N}}\{\|D_t(\theta)\|^2 + \|R_t(\theta)\|_F\},\end{aligned} \tag{S3.40}$$

where $C_* < \infty$ almost surely and $\rho_* \in (0,1)$ is deterministic. Here the bounded and Lipschitz spline score controls the score difference. The bounded score Jacobian and its Lipschitz bound control the Hessian difference. The remaining factors are bounded by M_t.

Assumption 4 makes M_t stationary with finite expectation. In particular,

$$E\left[\sum_{t=1}^{\infty} \rho_*^t M_t\right] = \frac{\rho_*}{1-\rho_*} E[M_1] < \infty,$$

so the weighted series is finite almost surely. The finitely many initial terms are also finite, by continuity on the compact parameter neighborhood. Thus both derivative differences in (S3.40) are summable uniformly in θ, almost surely. Their partial sums are therefore $O_p(1)$. After the $T^{-1/2}$ score and T^{-1} Hessian normalizations, these differences vanish. Consistency of the initialized estimator and the same Taylor argument then give the limit in (S3.39). In the remainder, $\widehat{\theta}_T$ denotes the estimator computed from these initial values.

Remark S3.7. *Under Assumption 1, $\{S_t\}$ is a martingale difference sequence. Its covariance limit is the conditional-variance limit (S3.32). The score process itself need not be stationary. The uniform $(2+\delta)$ moment bound and a martingale strong law imply*

$$\frac{1}{T}\sum_{t=1}^{T}\{S_t S_t' - E[S_t S_t' \mid \mathcal{F}_{t-1}]\} \longrightarrow 0 \quad a.s.$$

For each fixed lag $k \ge 1$, $S_t S_{t-k}'$ is also a martingale difference with uniformly bounded moments of some order greater than one, so its sample average converges to zero almost surely. Thus the outer-product estimator and a fixed-lag HAC estimator, including HAC(22) (Newey and West, 1987), converge in probability to Ω. This equivalence concerns parameter scores under Assumption 1. The empirical parameter tests use lag zero; the HAC(22) calculations for OOS forecast-score differences concern a different sequence and are not justified by this martingale argument. The continuity and moment conditions in Assumption 4 justify evaluating these estimators at $\widehat{\theta}_T$. The forgetting conditions handle initialization. In the special case where the score and its conditional covariance form a jointly stationary and ergodic process, the limit simplifies to $\Omega = E[S_t S_t']$. The sandwich form $\mathcal{H}^{-1}\Omega\mathcal{H}^{-1}$ remains necessary because Assumption 1 allows both heterogeneity and dependence within a cross section, so in general $\Omega \ne \mathcal{H}$. More explicitly, with $\psi_{t,i} = \psi\{u_{t,i}, q_t(\theta_0)\}$,

$$\Sigma_t = \sum_{i=1}^{N_t}\sum_{j=1}^{N_t} w_{t,i} w_{t,j} \operatorname{Cov}(\psi_{t,i}, \psi_{t,j} \mid \mathcal{F}_{t-1}).$$

Under identical conditionally independent marginals,

$$\Sigma_t = (\sum_i w_{t,i}^2)\mathcal{I}_q\{q_t(\theta_0)\}.$$

For equal weights, this reduces to $\mathcal{I}_q\{q_t(\theta_0)\}/N_t$. Under heterogeneous marginals, conditional independence alone does not give this simplification, because individual score covariances may differ and conditional score means may be nonzero. The variance estimator must use the outer products of the date-level parameter scores, using lag zero in the main implementation, rather than treating constituent scores as independent replications. Wald tests also require a nonsingular asymptotic covariance for the restrictions being tested.

Corollary S3.8 (Sandwich covariance and Wald inference)**.**
Suppose that Assumptions 1–5 hold. At the estimator computed from fixed initial values, let $\widehat{S}_t = \widehat{S}_t(\widehat{\theta}_T)$ and $\bar{S}_T = T^{-1}\sum_t \widehat{S}_t$. Define the time-average matrices

$$\widehat{\mathcal{H}}_T = -\frac{1}{T}\sum_{t=1}^{T} \nabla_\theta^2 \widehat{\ell}_t(\widehat{\theta}_T), \tag{S3.41}$$

$$\widehat{\Omega}_T = \frac{1}{T}\sum_{t=1}^{T} (\widehat{S}_t - \bar{S}_T)(\widehat{S}_t - \bar{S}_T)'. \tag{S3.42}$$

Then $\widehat{\mathcal{H}}_T \to \mathcal{H}$ and $\widehat{\Omega}_T \to \Omega$ in probability. Consequently,

$$\widehat{V}_T = \frac{1}{T}\widehat{\mathcal{H}}_T^{-1}\widehat{\Omega}_T\widehat{\mathcal{H}}_T^{-\prime} = H_T^{-1}B_T H_T^{-\prime}, \qquad H_T = T\widehat{\mathcal{H}}_T, \quad B_T = T\widehat{\Omega}_T, \tag{S3.43}$$

satisfies

$$T\widehat{V}_T \longrightarrow \mathcal{V} := \mathcal{H}^{-1}\Omega\mathcal{H}^{-1} \quad \text{in probability}.$$

This agrees with (39).

For a fixed full-row-rank restriction matrix R with $R\theta_0 = r$ and $R\mathcal{V}R'$ positive definite,

$$W_T = (R\widehat{\theta}_T - r)'(R\widehat{V}_T R')^{-1}(R\widehat{\theta}_T - r) \Longrightarrow \chi^2_{\mathrm{rank}(R)}. \tag{S3.44}$$

The even and odd block tests each have three restrictions. Their joint test has six. Positive definiteness is required for the covariance of the tested restrictions, without requiring Ω to be positive definite on every parameter direction.

Proof. The Hessian uniform law, consistency, and (S3.40) give $\widehat{\mathcal{H}}_T \to \mathcal{H}$ in probability. For the meat matrix at θ_0, the martingale strong law in the preceding remark and Assumption 5 yield

$$\frac{1}{T}\sum_t S_t S_t' \longrightarrow \Omega \quad \text{in probability}, \qquad \frac{1}{T}\sum_t S_t \longrightarrow 0 \quad \text{a.s.}$$

To justify substitution of the estimator, consider the predictable one-observation parameter-score map $J_t(\theta)'\psi\{u, q_t(\theta)\}$. Its modulus of continuity is uniform in u, by the score bounds and Assumption 4. The corresponding modulus for outer products of weighted cross-sectional scores is bounded by a constant times $\sup_{\theta\in\mathcal{N}} \|D_t(\theta)\|^2$. This is an integrable, stationary bound. Dominated convergence, the ergodic

theorem for the predictable modulus, and consistency therefore make the time-average change in score outer products $o_p(1)$. The same argument applies to the score mean. This step does not require stationarity of the observations.

Initialization also has a summable effect on the outer products. Indeed, the bounded and Lipschitz score and the forgetting conditions give

$$\sup_{\theta\in\mathcal{N}} \|\widehat{S}_t(\theta)\widehat{S}_t(\theta)' - S_t(\theta)S_t(\theta)'\| \le C_{**}\rho_*^t\left(1 + \sup_{\theta\in\mathcal{N}} \|D_t(\theta)\|^2\right)$$

for all sufficiently large t, with $C_{**} < \infty$ almost surely. The summability argument above applies again. Hence $\bar{S}_T = o_p(1)$, and subtracting $\bar{S}_T\bar{S}_T'$ does not affect the meat limit. This proves $\widehat{\Omega}_T \to \Omega$. Positive definiteness of $\mathcal{H}$ and continuity of inversion give the sandwich limit. Theorem S3.6 for the estimator computed from fixed initial values and Slutsky's theorem then give (S3.44) under the stated restriction null. □

S4 Simulation calibration and additional forecasts

S4.1 Generating coefficients

The S&P 500 HAR coefficient vectors used in generation, in horizon order $(1, 5, 22)$, are

$$\begin{aligned}
\kappa_\eta^{\mathrm{SP}} &= (-5.745736, -4.815923, 30.255739)',\\
\kappa_{\mathrm{even}}^{\mathrm{SP}} &= (-0.074906, 10.692708, 259.195621)',\\
\kappa_{\mathrm{odd}}^{\mathrm{SP}} &= (1.976005, -4.297596, 63.792928)'.
\end{aligned}$$

Each active vector is multiplied by $\sqrt{N/500}$ in both the independent and ten-percent common-factor designs. Inactive vectors are zero. The generating updates are given in (40) and (41). The full-precision coefficients and replication-level results are recorded in the companion repository.

S4.2 M5 compared with M3

Table S1 compares M5 with M3 after fitting M1–M5 in each replication. This comparison evaluates the forecast gains from adding tail-intensity dynamics to the fitted dynamic-t reference, as in the empirical analysis. Percentage improvements use the definitions in Section 4.4 of the main paper, with M3 replacing M2 as the benchmark.

Both mean and median improvements in log score and one-percent quantile loss are positive in all 18 cells of cases 2–4. Cases 0 and 1 show the results when tail-intensity dynamics are absent from generation. All entries use 100 paired replications with 642 OOS dates.

S5 Empirical data and PCA construction

S5.1 Constituent histories and observed returns

Daily prices are obtained from Yahoo Finance. We use adjusted closing prices when available and otherwise reported closing prices, and calculate daily log returns. The S&P 500 membership histories

Table S1: OOS improvements of M5 relative to M3: mean / median

Case	Dependence	N	Log score (%)	1% quantile loss (%)
0: Static	Independent	20	-0.010 / -0.008	-0.084 / -0.077
	Common 10%	20	-0.012 / -0.010	-0.113 / -0.089
	Independent	50	-0.005 / -0.004	-0.045 / -0.034
	Common 10%	50	-0.006 / -0.005	-0.059 / -0.047
	Independent	100	-0.002 / -0.001	-0.016 / -0.017
	Common 10%	100	-0.005 / -0.004	-0.042 / -0.033
1: η only	Independent	20	-0.007 / -0.006	-0.048 / -0.013
	Common 10%	20	-0.012 / -0.008	-0.081 / -0.059
	Independent	50	-0.004 / -0.003	-0.030 / -0.032
	Common 10%	50	-0.007 / -0.004	-0.069 / -0.037
	Independent	100	-0.001 / -0.001	-0.011 / -0.011
	Common 10%	100	-0.004 / -0.003	-0.044 / -0.039
2: Even only	Independent	20	+0.038 / +0.026	+0.396 / +0.271
	Common 10%	20	+0.038 / +0.029	+0.397 / +0.370
	Independent	50	+0.032 / +0.030	+0.300 / +0.276
	Common 10%	50	+0.034 / +0.030	+0.367 / +0.332
	Independent	100	+0.026 / +0.023	+0.251 / +0.275
	Common 10%	100	+0.041 / +0.029	+0.417 / +0.300
3: η and even	Independent	20	+0.035 / +0.029	+0.343 / +0.323
	Common 10%	20	+0.023 / +0.023	+0.169 / +0.105
	Independent	50	+0.027 / +0.024	+0.292 / +0.229
	Common 10%	50	+0.034 / +0.029	+0.349 / +0.301
	Independent	100	+0.029 / +0.026	+0.279 / +0.247
	Common 10%	100	+0.042 / +0.035	+0.402 / +0.350
4: η, even, and odd	Independent	20	+0.041 / +0.033	+0.361 / +0.246
	Common 10%	20	+0.041 / +0.029	+0.408 / +0.344
	Independent	50	+0.034 / +0.028	+0.349 / +0.312
	Common 10%	50	+0.044 / +0.040	+0.434 / +0.338
	Independent	100	+0.034 / +0.033	+0.305 / +0.273
	Common 10%	100	+0.064 / +0.046	+0.510 / +0.342

Percentage improvements over 100 paired replications with M1–M5 estimated. Each replication has 642 OOS dates. Positive values improve on M3. Means use the exponential of the mean log-score difference or the reduction in paired mean losses. Medians use replication-specific percentage improvements.

use `fja05680/sp500`, taking the most recent snapshot dated on or before each trading date. The final snapshot, June 30, 2026, is carried forward to July 24, 2026. Nikkei 225 histories are reconstructed from the official constituent set as of August 31, 2026 and the official component-change ledger, replayed over the sample.

US sector assignments use public GICS Git snapshots and an audited backcast for historical constituents. Japanese sectors combine official current classifications with audited classifications for former constituents. On each date, only constituents with observed returns and estimable training PC1 loadings enter the raw and residual samples. Table S2 reports the daily sample-size ranges and training and OOS date counts for all 19 panels.

Table S2: The 19 constituent-return cross sections

Panel	N_t range	Median N_t	Training dates	OOS dates
S&P 500	403–482	455	1886	642
Nikkei 225	212–223	218	1829	624
US Energy	17–23	19	1886	642
US Materials	19–27	25	1886	642
US Industrials	54–77	70	1886	642
US Consumer Discretionary	43–69	59	1886	642
US Consumer Staples	30–36	32	1886	642
US Health Care	41–63	56	1886	642
US Financials	55–79	58	1886	642
US Information Technology	52–72	61	1886	642
US Communication Services	2–21	18	1886	642
US Utilities	27–30	29	1886	642
US Real Estate	26–30	30	1844	642
JP Technology	52–59	56	1829	624
JP Financials	18–21	20	1829	624
JP Consumer Goods	27–36	33	1829	624
JP Materials	48–60	57	1829	624
JP Capital Goods/Others	35–36	36	1829	624
JP Transportation and Utilities	19–21	19	1829	624

US and JP denote S&P 500 and Nikkei 225 sectors. Raw and PC1 residual use the same constituent–dates, restricted to constituents with estimable training PC1 loadings. The broad-market panels contain constituent returns, rather than the return on an index. The Real Estate panel starts on September 1, 2016; the other panels start on July 5, 2016.

S5.2 Training-fixed PCA

For each panel, we estimate covariance PCA from the training returns through December 29, 2023. Missing training returns are replaced by their constituent's observed training mean for this calculation only. The centered matrix therefore has zero entries at missing observations. Returns are not standardized by volatility. Constituents with no observed training returns or no positive centered training variation are excluded. The leading eigenvector has unit norm, and its sign is chosen to give a nonnegative sum of loadings. Means and loadings remain fixed in training and OOS evaluation.

The daily factor score and residuals are computed from observed returns using the formula in Section 5.4 of the main paper. A date must have at least two eligible observed constituents and a squared loading norm of at least 10^{-12}. No imputed return enters density estimation or forecast evaluation. Raw returns and PC1 residuals use the same constituent–dates. The reported pairwise correlations use matched training observations, and the PC1 shares are the leading eigenvalues divided by the training covariance traces.