CURVES, NOT POINTS: STRUCTURED CURVE APPROXIMATION FOR ENGAGEMENT DYNAMICS IN SHORT-FORM VIDEO AND INTERACTIVE LIVE STREAMING

Anonymous authors
Paper under double-blind review

ABSTRACT

Audience engagement with video is inherently a trajectory, yet it is commonly reduced to scalar quantities such as watch time or completion rate. We instead model engagement as a structured function of time using low-dimensional parametric curve approximators.

We first study live-stream engagement in KuaiLive-M3, where viewers may enter and leave throughout a room. We model occupancy as $O = J - X$, using structured approximators for cumulative joins and exits, with a causal GRU predicting compact exit-curve parameters from evolving room representations. The resulting model achieves 55.9% I_5, where I_5 is the fraction of rooms whose whole-trajectory occupancy IAE is at most 5%, compared with 49.0% for the strongest independent pointwise neural baseline.

We then study cross-audience retention transfer on KuaiRec-small and KuaiRec-big, which contain the same videos viewed by disjoint populations. We factor the retention representation into a video-specific parameter and behavioral watch-style parameters shared across videos. The resulting model achieves 75.3% C_5, where C_5 is the fraction of test videos whose retention curve has sup-norm error at most 5%, compared with 68.3% for the strongest neural baseline.

Together, these results show that structured engagement curves provide compact functional representations that support accurate prediction, composition, and cross-audience transfer across distinct video-engagement regimes.

1 INTRODUCTION

Online video has become a high-stakes attention economy in which the value of content depends not only on whether viewers start watching, but on whether they continue watching. For subscription streaming platforms, engagement is directly connected to commercial outcomes: Netflix describes viewing engagement as its best proxy for member enjoyment and links greater viewing to retention, acquisition, revenue, and profit Netflix (2025). These stakes are particularly high for professionally produced television and streaming content, where high-end drama is expensive and risky to produce and successful programs may attract audiences at global scale Dunleavy (2026). Engagement is also fundamental to video discovery: large platforms such as YouTube have long deployed personalized video recommendation systems Davidson et al. (2010), while recent work on industrial video recommendation explicitly treats watch-time prediction as an important learning problem Zhao et al. (2024). Predicting how an audience evolves over the course of a video is therefore not merely a descriptive analytics task: it can inform content evaluation, recommendation, and decisions about how effectively costly video content sustains viewer attention.

Interactive live video: KuaiLive-M3. We study complete engagement trajectories in interactive live video as our primary prediction problem (Fig. 1). In ***KuaiLive-M3***, a live-stream room evolves continuously as viewers arrive and depart and as the streamed content changes over time. For room r, let $J_r(u)$ and $X_r(u)$ denote the normalized cumulative fractions of joins and exits by normalized time $u \in [0, 1]$; the resulting audience occupancy is $O_r(u) = J_r(u) - X_r(u)$. The dataset additionally provides time-indexed room representations, and we write $e_{r,\leq u}$ for the sequence of representations available no later than u. This enables a genuine held-out-room prediction problem: a shared

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

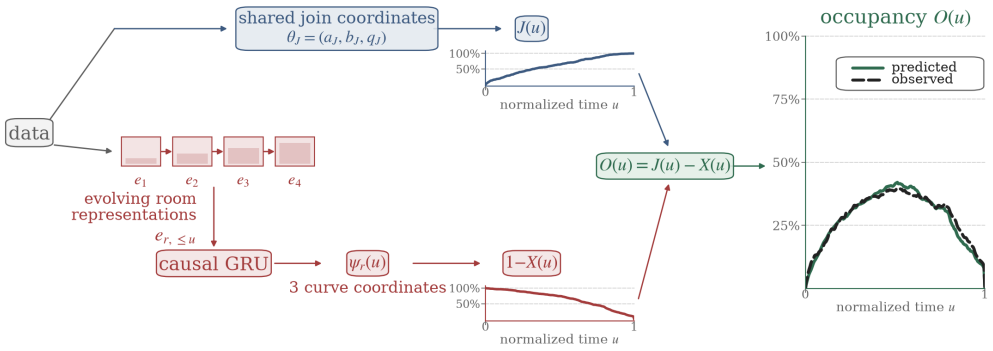


Figure 1: **KuaiLive-M3 Interactive Live Videos (Rooms):** A causal GRU predicts compact exit-curve coordinates from evolving room representations, which combine with joining dynamics to predict occupancy $O_r(u) = J_r(u) - X_r(u)$.

neural encoder can consume the causal history of an evolving room and predict a low-dimensional representation of its future audience trajectory.

We represent the target trajectory through a small number of structured functional coordinates rather than predicting an unconstrained value independently at every time point. For KuaiLive-M3, model selection identifies the 3-parameter **cure-Kumaraswamy approximator**:

$$K(u; a, b, q) = q + (1 - q)(1 - u^a)^b.$$

Joining is represented by a shared coordinate vector $\theta_J = (a_J, b_J, q_J)$, while the more dynamic exit process is conditioned on the evolving room: a shared causal GRU g_ϕ maps $e_{r,\leq u}$ to three time-dependent curve coordinates $\psi_r(u)$. We model the complementary exit curve $1 - X_r$, whose decreasing orientation matches the cure-Kumaraswamy form, and compose the two processes to obtain occupancy of interactive live videos from KuaiLive-M3:

$$\widehat{J}_r(u) = 1 - K(u; \theta_J)\,,$$

$$1 - \widehat{X}_r(u) = K(u; \psi_r(u))\,, \qquad \psi_r(u) = g_\phi(e_{r,\leq u}) + b',$$

$$\widehat{O}_r(u) = \widehat{J}_r(u) - \widehat{X}_r(u).$$

Thus the GRU need not directly emit an entire discretized trajectory; its output interface consists of only three coordinates that define a complete functional curve.

Short-form video: KuaiRec. We complement the live-stream setting with short-form video on *KuaiRec* (Fig. 2), where the same videos are observed under two disjoint viewer populations, KuaiRec-small and KuaiRec-big. For video v and audience a, the retention curve $\hat{S}_{v,a}(u)$ gives the fraction of viewers who watch through normalized progress u. Viewers are partitioned into $K = 5$ behavioral *watch-style groups*, and $\pi_{v,a,c}$ denotes the fraction of the audience for video v belonging to group c.

Model selection yields a **Burr XII approximator with fixed Gaussian and sinusoidal residual terms**:

$$\widetilde{S}_{v,a}(u) = \sum_{c=1}^{K} \pi_{v,a,c} \left[\left(1 + \left(\frac{u}{\lambda_v} \right)^{c_c} \right)^{-k_c} + \alpha_c \mathcal{N}\left(u; \mu, \sigma^2\right) + \beta_c \sin(3\pi u) \right].$$

Here, λ_v is video-specific, while $(c_c, k_c, \alpha_c, \beta_c)$ are shared within watch-style group c.

For cross-audience transfer, the watch-style parameters are learned from KuaiRec-small training videos and then frozen. For a held-out video, λ_v is estimated only from its KuaiRec-small retention curve, while the target KuaiRec-big mixture weights $\pi_{v,\text{big},c}$ are observed audience covariates. The

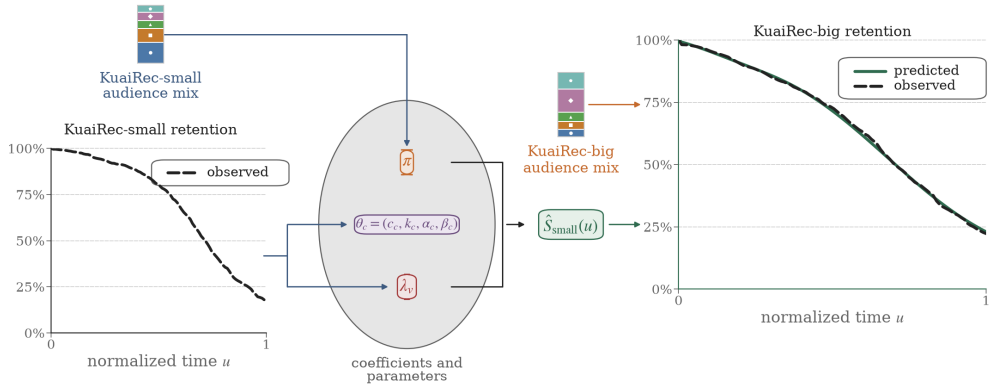


Figure 2: **KuaiRec Short Form Videos:** Watch-style parameters learned from KuaiRec-small and a video-specific parameter inferred from the source retention curve are combined with the KuaiRec-big audience mixture to predict cross-audience retention.

same learned conditional structure is therefore transferred to a different audience without fitting to the KuaiRec-big retention trajectory.

The fitted families in both datasets are used as compact curve approximators rather than exact survival distributions. The two settings therefore test complementary forms of generalization: held-out live-room prediction from evolving representations and cross-audience transfer for the same short-form content.

Contributions. Our main contributions are:

- We introduce a structured functional output formulation for engagement-trajectory prediction, representing complete retention and occupancy curves through a small number of curve coordinates rather than independent predictions at discretized time points.
- On KuaiLive-M3, we model dynamic audience occupancy as $O = J - X$, use a causal GRU to predict compact exit-curve coordinates from evolving room representations, and jointly optimize the composed occupancy trajectory.
- On KuaiRec, we study cross-audience generalization for the same videos under disjoint viewer populations and introduce a video/watch-style factorization that reduces KuaiRec-small-to-KuaiRec-big transfer IAE from 6.71% to 1.90%.
- Across both datasets, structured curve prediction outperforms independent pointwise neural predictors, shared neural networks that directly predict the full trajectory, time-series forecasting baseline such as DeepTime Woo et al. (2023), and event-time methods such as DSM Nagpal et al. (2021) and Dynamic-DeepHit Lee et al. (2019), while using a substantially smaller functional output space.

2 Related Work

Among methods closest to our prediction setting, we compare against three representative time-series and event-time models. DeepTime Woo et al. (2023) is a time-series forecasting method that represents a signal using a learned neural basis and adapts series-specific coefficients to predict future values. Deep Survival Machines (DSM) Nagpal et al. (2021) models conditional time-to-event distributions using covariate-dependent mixtures of parametric survival distributions. Dynamic-DeepHit Lee et al. (2019) uses recurrent representations of longitudinal covariates to dynamically predict event-time risks. Beyond these direct baselines, several bodies of work are closely related to our setting but address different prediction targets. In video recommendation, watch time is commonly used as a

supervision or ranking signal Zhao et al. (2019), with subsequent work developing structured watch-time predictors and behavioral models Lin et al. (2023); Anwer et al. (2025); because raw watch time is confounded by video duration and other exposure effects, related methods have studied deconfounding Zhan et al. (2022), watch-time-gain objectives Zheng et al. (2022), explicit correction for video-length effects Quan et al. (2023), causal labels Zhang et al. (2023), noisy-feedback modeling Zhao et al. (2023), counterfactual watch time Zhao et al. (2024), and more recent approaches to debiased or invariant engagement modeling Liu et al. (2026); Jin et al. (2026). These methods primarily use watch behavior to improve preference estimation, ranking, or debiased engagement prediction rather than predicting the complete audience trajectory considered here. Adjacent work studies audience-retention curves as an empirical measurement object on YouTube, including the comparison and design of retention metrics Altman and Jiménez (2019), and earlier studies characterize how abandonment and viewing behavior vary with user and delivery conditions Krishnan and Sitaraman (2012); Balachandran et al. (2002); at a broader methodological level, functional data analysis studies complete functions as statistical objects and develops low-dimensional representations of curves Ramsay and Silverman (1997). Our retention formulation also draws on classical survival-analysis ideas. Kaplan–Meier provides a nonparametric estimator of survival curves Kaplan and Meier (1958), while broader statistical treatments develop general survival and event-history models Andersen et al. (2012); Klein and Moeschberger (2003); Lin and Zelterman (2002), and the Weibull family provides a compact parametric model of survival-shaped decay Weibull (1951). These works provide statistical foundations for survival modeling; for predictive comparison, we include the modern neural event-time methods DSM and Dynamic-DeepHit described above. Our fitted curve families instead serve as low-dimensional functional approximators, with their parameters estimated using nonlinear least-squares optimization related to the Levenberg–Marquardt family of methods Marquardt (1963); Kaltenbach (2022). Finally, prior live-streaming research studies evolving audiences primarily in service of recommendation: bi-directional models jointly reason about user preference and future streamer audiences Zhang et al. (2021), dynamic representation learning captures changing live-room states as viewers enter, remain, and leave Gao et al. (2023), and recent work models evolving user–streamer–room interactions and room-level collaboration Guo et al. (2026b); KuaiLive-M3 itself provides the multimodal live-stream setting used in our experiments Guo et al. (2026a).

3 METHODOLOGY

KuaiLive-M3 Interactive live videos. We first consider interactive live-stream rooms from KuaiLive-M3 Guo et al. (2026a). For a room (interactive video) r of duration L_r, physical time t is mapped to normalized progress $u = \frac{t}{L_r} \in [0, 1]$. Let N_r denote the number of distinct viewers that ever enter the room. For viewer i, let u_{ir}^{in} be the first entry normalized time and u_{ir}^{out} the final departure normalized time, then the empirical cumulative joining J, cumulative exit X and occupancy O are

$$J_r(u) = \frac{1}{N_r} \sum_{i=1}^{N_r} \mathbf{1}\left[u_{ir}^{\text{in}} \le u\right],$$

$$X_r(u) = \frac{1}{N_r} \sum_{i=1}^{N_r} \mathbf{1}\left[u_{ir}^{\text{out}} \le u\right],$$

$$O_r(u) = J_r(u) - X_r(u)$$

Equivalently, $O_r(u)$ is the number of viewers active at time u divided by the total number of viewers that ever entered the room. Thus the live-stream curve need never attain 100% occupancy.

We treat the fitted functions as approximators of empirical audience trajectories rather than as exact probabilistic survival laws. To identify compact functional forms for the joining and exit processes, we search a library of approximately 30 low-dimensional closed-form curve families defined on normalized progress u (see Appendix B). Candidate joining models use one parameter vector shared across rooms, whereas candidate exit models predict the parameters of the decreasing complementary trajectory $1 - X_r$ from the causal history of the evolving room. Families and architectures are selected using validation sup-norm error. The selected family for both components is the 3-parameter cure-Kumaraswamy approximator

$$K(u; a, b, q) = q + (1 - q)(1 - u^a)^b, \qquad a > 0,\ b > 0,\ q \in [0, 1] \tag{1}$$

For joining, a single coordinate vector $\theta_J(a_J, b_J, q_J)$ is shared across rooms, giving

$$\widehat{J}_r(u) = 1 - K(u; \theta_J)$$

For exits, KuaiLive-M3 provides 128-dimensional time-indexed representations of the evolving room. Let $e_{r,\le u}$ denote the causal sequence of representations available up to progress u. A shared GRU maps this sequence to the three curve coordinates $\psi_r(u) = g_\phi(e_{r,\le u}) + b$, from which we predict

$$1 - \widehat{X}_r(u) = K(u; \psi_r(u))$$

Thus the GRU predicts a three-dimensional functional representation of the full exit behavior. We first fit the joining and exit components separately and compose them as $\widehat{J}_r - \widehat{X}_r$. After selecting their functional structures on validation data, we jointly fine-tune the pair for occupancy $\widehat{O}_r(u) = \widehat{J}_r(u) - \widehat{X}_r(u)$ using the loss function

$$\mathcal{L}_{\text{joint}} = \mathcal{L}_O + \lambda_J \mathcal{L}_J + \lambda_X \mathcal{L}_X + \lambda_{\text{ord}} \mathcal{L}_{\text{ord}} \tag{2}$$

where $\mathcal{L}_O$, $\mathcal{L}_J$, and $\mathcal{L}_X$ are the mean absolute trajectory losses for occupancy, joining, and exiting, respectively, and $\mathcal{L}_{\text{ord}} = \mathbb{E}_{r,u}\left[\max\{0, \widehat{X}_r(u) - \widehat{J}_r(u)\}\right]$ penalizes cumulative exits exceeding cumulative joins. The jointly trained model is selected using validation mean occupancy sup-norm error.

We compare against three classes of predictors of the same occupancy target. Independent-point baselines train separate MLP, RNN, or GRU predictors for each requested location on a shared temporal grid, so parameters grow with the number of prediction points. Shared along-curve baselines instead use a single MLP, RNN, or GRU across the complete trajectory. We additionally compare against DeepTime Woo et al. (2023), DSM Nagpal et al. (2021), and Dynamic-DeepHit Lee et al. (2019), adapting each method to predict a complete occupancy trajectory on the same normalized clock.

KuaiRec Short-form videos. We next consider short-form retention using KuaiRec Gao et al. (2022). For video v of duration L_v, normalized time is again $u = t/L_v$. If viewer i watches for normalized duration u_{iv}, then for audience a the empirical retention curve is

$$S_{v,a}(u) = \frac{1}{N_{v,a}} \sum_{i=1}^{N_{v,a}} \mathbf{1}\left[u_{iv} \ge u\right], \qquad u \in [0, 1].$$

Thus the vertical axis is the fraction of the observed audience that reaches progress u, and every empirical retention curve begins at 100%.

KuaiRec contains two datasets, KuaiRec-small and KuaiRec-big, which observe approximately $3.3K$ of the same videos under different viewer populations. This overlap lets us study whether a retention representation learned from one audience transfers to another.

We first consider a video-specific alternative in which all curve coordinates are fitted from a video's KuaiRec-small retention trajectory and then frozen when evaluating the corresponding KuaiRec-big trajectory. Our proposed model instead separates video-specific variation from viewer watch style. Viewers are clustered into $K = 5$ watch-style groups using K-means on training-interaction statistics describing early skipping, completion, normalized watch duration, rewatching, and watch-time variability. Let $\pi_{v,a,c}$ denote the fraction of viewers of video v in audience a assigned to group c, with $\sum_{c=1}^{K} \pi_{v,a,c} = 1$. For each candidate family, every coordinate is designated either video-specific (V) or watch-style-specific (C). Family and V/C-allocation choices are made using validation small-to-big transfer: the C-coordinates are learned from training videos, while only the V-coordinates of a held-out video are inferred from its KuaiRec-small trajectory before replacing the source mixture by the KuaiRec-big mixture. The selected approximator is a Burr XII base with Gaussian and sinusoidal residuals fit using a standard nonlinear least squares (NLS) loss function:

$$\widehat{S}_{v,a}(u) = \sum_{c=1}^{K} \pi_{v,a,c} \left[\left(1 + \left(\frac{u}{\lambda_v}\right)^{c_c}\right)^{-k_c} + \alpha_c \mathcal{N}\big(u; 0.20, 0.12^2\big) + \beta_c \sin(3\pi u) \right] \tag{3}$$

Here, $\mathcal{N}(u; \mu, \sigma^2) = \exp\big(-(u-\mu)^2/(2\sigma^2)\big)$. The Gaussian mean, standard deviation and the sinusoidal frequency are fixed constants selected by ablation on the validation set. We next determine

whether each fitted parameter is video-specific or shared within a watch-style group. For the Burr parameters (λ, c, k), we first search all 8 possible video/watch-style allocations. After validation selects VCC, we keep this allocation fixed and search the 4 possible allocations of the two residual amplitudes. Validation selects $(\lambda, c, k, \alpha, \beta) = (V, C, C, C, C)$. Thus, only λ_v is video-specific, whereas c_c, k_c, α_c, and β_c are shared within watch-style group c. The resulting mixture is projected to $[0, 1]$ and to a nonincreasing orientation before evaluation.

At transfer time, the watch-style coordinates are frozen, λ_v is estimated only from the held-out video's KuaiRec-small trajectory, and the KuaiRec-big mixture weights are substituted:

$$\widehat{S}_{v,\mathrm{big}}(u) = \sum_{c=1}^{K} \pi_{v,\mathrm{big},c} \left[\left(1 + \left(\frac{u}{\lambda_v} \right)^{c_c} \right)^{-k_c} + \alpha_c \mathcal{N}\big(u; 0.20, 0.12^2\big) + \beta_c \sin(3\pi u) \right]$$

No parameter is fitted to the held-out video's KuaiRec-big retention curve, ensuring generalizability: the target-audience quantities $\pi_{v,\mathrm{big},c}$ are observed covariates analogous to target-domain mixture proportions in domain-adaptation settings, rather than parameters learned from the target retention trajectory. We compare against the same three baseline paradigms as in KuaiLive-M3: independent-point predictors, shared along-curve neural models, and the DeepTime, DSM, and Dynamic-DeepHit temporal/event-time baselines.

Curve-level Error Metrics. For an individual trajectory Y with prediction $\widehat{Y}$, we define the integrated absolute error (IAE) (see Appendix A for a proof that E_{IAE} equals the average absolute error) and sup-norm error as

$$E_{\mathrm{IAE}} = 100 \int_0^1 \left| Y(u) - \widehat{Y}(u) \right| \, du \tag{4}$$

$$E_{\infty} = 100 \sup_{u \in [0,1]} \left| Y(u) - \widehat{Y}(u) \right| \tag{5}$$

Both are per-datapoint (video) errors. We report their means and standard deviations over the evaluation set. We additionally report two additional dataset-level error metrics:

$$I_\tau = \frac{100}{|\mathcal{D}|} \sum_{d \in \mathcal{D}} \mathbf{1}\left[E_{\mathrm{IAE},d} \le \tau \right] \tag{6}$$

$$C_\tau = \frac{100}{|\mathcal{D}|} \sum_{d \in \mathcal{D}} \mathbf{1}\left[E_{\infty,d} \le \tau \right] \tag{7}$$

Thus, I_τ is the percentage of datapoints whose average absolute trajectory error is at most $\tau\%$, while C_τ is the percentage whose entire predicted curve remains within $\tau\%$ of the target. We generally report I_τ; when it approaches saturation and becomes less discriminative, we then report the more stringent C_τ.

4 Experiments and Results

We evaluate the structured-curve formulation in two complementary settings. Section 4.1 evaluates complete occupancy-trajectory prediction on KuaiLive-M3 and cross-audience retention transfer on KuaiRec compared against neural and time-series baselines. Section 4.2 studies the role of author classes and behavioral watch-style groups across both datasets. Section 4.3 measures the approximation ceiling for KuaiLive-M3 when the live-stream exit representation is optimized directly rather than predicted by the GRU. Section 4.4 evaluates online extensions of structured curve prediction for KuaiLive-M3, where the length of the interactive video is not known apriori. Section 4.5 examines violations of the intended survival-style curve geometry. Additional family-selection, reconstruction, composition, parameter-allocation, residual, sensitivity, information-budget, cluster-count, and duration-robustness experiments are reported in Appendix C.

Table 1: Comparison of structured curve prediction with neural, time-series, and event-time baselines on KuaiLive-M3. Lower mean E_{IAE} and mean E_{sup} are better; higher I_5 is better.

Group	Method	mean $\mathbf{E}_{\mathrm{IAE}}$	mean $\mathbf{E}_{\mathrm{sup}}$	$\mathbf{I_5}$	# Params.
Independent pointwise neural network	MLP ($N = 10$)	16.20%	26.20%	12.0%	83,890
	RNN ($N = 10$)	6.40%	15.80%	43.0%	125,490
	GRU ($N = 10$)	6.00%	14.90%	49.0%	373,810
	GRU ($N = 100$)	6.00%	19.60%	48.0%	3,738,100
	GRU ($N = 500$)	6.10%	22.40%	46.0%	18,690,500
	GRU ($N = 3000$)	6.40%	25.80%	37.0%	112,143,000
Shared neural network along curve	MLP	16.26%	26.10%	12.3%	8,389
	RNN	6.80%	19.30%	33.5%	12,549
	GRU	6.39%	18.20%	41.4%	37,381
Time-series & event-time	DeepTime (Woo et al. (2023))	8.02%	15.24%	34.5%	6,922
	DSM (Nagpal et al. (2021))	10.27%	26.83%	8.2%	6,728
	Dynamic-DeepHit (Lee et al. (2019))	12.59%	28.84%	1.2%	44,163
Structured curves (ours)	cure-Kumaraswamy	**5.53%**	**13.70%**	**55.9%**	37,848

Table 2: Comparison of structured curve transfer with direct, time-series, and event-time baselines on KuaiRec. Lower mean E_{IAE} is better; higher C_τ is better. C_τ is used because I_τ is saturated.

Group	Method	mean $\mathbf{E}_{\mathrm{IAE}}$	$\mathbf{C_{10}}$	$\mathbf{C_5}$	# Params./k
Independent pointwise prediction	Ridge (N = 100)	2.15%	95.0%	63.9%	1,000
	KNN (N = 100)	2.21%	93.2%	60.3%	2,323,000
	MLP (N = 100)	2.33%	10.8%	5.7%	275,300
	RNN (N = 100)	2.10%	94.1%	52.7%	140,900
	GRU (N = 100)	2.05%	95.2%	58.0%	416,100
Shared neural network along curve	MLP	2.01%	96.6%	68.3%	2,753
	RNN	2.23%	95.1%	59.9%	1,409
	GRU	2.08%	96.0%	67.9%	4,161
Time-series & event-time	DeepTime (Woo et al. (2023))	13.43%	23.2%	5.3%	6,922
	DSM (Nagpal et al. (2021))	15.73%	0.0%	0.0%	4,744
	Dynamic-DeepHit (Lee et al. (2019))	27.28%	0.0%	0.0%	42,179
Structured curves (ours)	Burr XII + G20 + sine3	**1.90%**	**98.7%**	**75.3%**	1

4.1 Structured Curves vs. Neural and Time-Series Baselines Across Engagement Settings

We compare structured curve prediction against three groups of baselines across KuaiLive-M3 and KuaiRec: *independent pointwise* models, which use a separate ML model/neural network to predict each of N fixed locations along the curve; *shared along-curve* models, which use one neural network as a parameterization of the occupancy or survival curve; and *time-series/event-time* algorithms. As shown in Table 1, the structured model achieves the lowest mean E_{IAE} and mean E_{sup}, and the highest I_5, on KuaiLive-M3. Table 2 shows the same advantage on KuaiRec, where the structured model attains the lowest mean E_{IAE} and the highest C_5. Training-seed stability for the neural baselines on both datasets is reported in Appendix C.1.

4.2 The Role of Class Structure in Engagement Modeling

We compare three ways of introducing shared population structure: *author classes*, where rooms are grouped by author metadata; *shared*, where all rooms use the same class-level parameters; and *watch-style classes*, where rooms are grouped by behavioral viewing patterns. As shown in Table 3, KuaiLive-M3 is largely insensitive to this choice: author, shared, and watch-style models achieve nearly identical occupancy accuracy. In contrast, Table 4 shows that KuaiRec transfer depends strongly on the class definition: behavioral watch-style groups retain good cross-audience transfer, while video-content and content-affinity classes collapse toward the raw small–big audience gap.

Table 3: KuaiLive-M3 occupancy prediction under alternative class structures. Lower mean E_{IAE} and mean E_{sup} are better; higher I_5 is better.

Class structure	mean $\mathbf{E}_{\mathrm{IAE}}$	mean $\mathbf{E}_{\mathrm{sup}}$	$\mathbf{I_5}$
Author	5.50%	13.70%	56%
Shared	5.50%	13.60%	56%
Watch style	**5.40%**	**13.40%**	**56%**

Table 4: KuaiRec class-structure comparison on the locked small-to-big transfer test set. Lower mean E_{IAE} is better; higher C_τ is better.

Class structure	mean $\mathbf{E}_{\mathrm{IAE}}$	$\mathbf{C_{10}}$	$\mathbf{C_5}$
Watch style	**1.90%**	**98.7%**	**75.3%**
Video L1	6.80%	24.9%	0.0%
Content affinity	6.90%	23.9%	0.0%

Table 5: KuaiLive-M3 performance with the selected cure-Kumaraswamy template under the standard and ideal-room-representation settings. Lower mean E_{IAE} and mean E_{sup} are better; higher I_5 and I_2 are better.

Room representation	mean $\mathbf{E}_{\mathrm{IAE}}$	mean $\mathbf{E}_{\mathrm{sup}}$	$\mathbf{I_5}$	$\mathbf{I_2}$
Fixed LLM representation	5.53%	13.70%	55.9%	5.1%
Ideal room representation	**2.23%**	**8.94%**	**99.2%**	**45.0%**

We hypothesize that this occurs because KuaiLive room dynamics exhibit substantial heterogeneity even within the same class, whereas KuaiRec shift is driven directly by changes in the composition of behavioral subpopulations across audiences. Additional component-level, oracle, VCC-transfer, class-count, and category-wise analyses are reported in Appendix C.3.

4.3 Oracle Performance with Ideal Curve Coordinates

The KuaiLive-M3 dataset provides fixed room representations generated by a closed-source LLM that is not fine-tuned for engagement prediction. To estimate the performance ceiling under ideal task-specific representations, we replace the GRU-predicted curve parameters with free learnable parameters optimized directly for each test room under the occupancy objective. This experiment is therefore not intended to measure generalization, but rather the best reconstruction performance the structured model could attain with perfectly adapted room representations. As shown in Table 5, mean occupancy E_{IAE} drops from about 5.6% to about 2.2%, indicating substantial headroom from improving the room representation.

4.4 Prediction Without Final Room Duration

The primary KuaiLive-M3 model uses normalized progress $u = t/L$, which requires knowing the final room duration L. To evaluate the setting where L is unknown at prediction time, we replace u with online clocks based only on elapsed time: absolute time t and the logarithmic clock

$$z(t) = \frac{\log(1 + t/\tau)}{\log(1 + 190\,\mathrm{h}/\tau)}, \qquad \tau = 3.42\,\mathrm{h},$$

where τ is the median training-room duration. As shown in Table 6, normalized progress is substantially better: mean occupancy IAE increases from 5.5% to 7.5% when the final duration is unavailable. This suggests that, in truly online settings, accurately estimating the eventual room duration may be more important than the precise choice of elapsed-time clock.

4.5 Geometric Validity of the Fitted Curves

The fitted functions are compact approximators rather than exact survival laws, so we check whether their raw outputs violate the intended range or monotonicity constraints. As shown in Table 7, the learned KuaiLive-M3 join and exit components exhibit no measured range or monotonicity violations, while the resulting occupancy curve is negative on only 1.25% of valid ticks. Monotonicity is not applicable to occupancy, since $O = J - X$ need not be monotone. The selected KuaiRec transfer curve likewise shows only rare raw violations before projection. Thus, the predictive gains are not driven by large departures from the intended curve geometry; the fitted families remain well behaved in practice despite being used as flexible curve approximators rather than exact survival distributions.

Table 6: KuaiLive-M3 occupancy prediction with and without access to the final room duration. Lower mean E_{IAE} and mean E_{sup} are better; higher I_5 is better.

Clock	**mean $\mathbf{E}_{\mathrm{IAE}}$**	**mean $\mathbf{E}_{\mathrm{sup}}$**	**$\mathbf{I}_5$**
Normalized	**5.50%**	**13.70%**	**55.9%**
Absolute	7.50%	20.00%	32.2%
Logarithmic	7.50%	19.40%	35.3%

Table 7: Geometric violations of the selected KuaiLive-M3 and KuaiRec approximators.

Curve	**<0 frac.**	**>1 frac.**	**Nonmono. frac.**
KuaiLive-M3 J	0.00%	0.00%	0.00%
KuaiLive-M3 $1-X$	0.00%	0.00%	0.00%
KuaiLive-M3 occupancy O	1.25%	0.00%	N/A
KuaiRec-big transfer S_{big}	0.06%	0.00%	0.83%

5 Conclusion

We presented structured functional models for complete engagement trajectories in two complementary settings. On KuaiLive-M3, compact curve parameterizations outperform pointwise neural, time-series, and event-time baselines for occupancy prediction, while class structure has only a minor effect and most remaining headroom appears to come from improving task-specific room representations. Performance also degrades when final room duration is unavailable, highlighting the value of relative progress information. On KuaiRec, structured modeling enables accurate cross-audience retention transfer, but only when the representation explicitly separates video-specific effects from behavioral watch-style variation; content-based class structures do not provide the same transfer benefit. Across both datasets, the fitted curves remain geometrically well behaved despite being used as flexible approximators rather than exact survival models. Overall, these results support structured curve prediction as a compact and interpretable alternative to independent pointwise forecasting for modeling audience engagement.

Data and Code Availability

An anonymized repository containing the code required to reproduce the reported experiments is present here.

AI Disclosure

Generative AI tools were used to assist with code development and manuscript writing and editing. The authors reviewed and verified all AI-assisted code and text. Generative AI tools were not used to make experimental conclusions or independently validate results. The authors take responsibility for the final content of the paper.

REFERENCES

Eitan Altman and Tania Jiménez. Measuring audience retention in youtube. In *Proceedings of the 12th EAI international conference on performance evaluation methodologies and tools*, pages 79–85, 2019.

Per K Andersen, Ornulf Borgan, Richard D Gill, and Niels Keiding. *Statistical models based on counting processes*. Springer Science & Business Media, 2012.

Zunaira Anwer, Shahnawaz Qureshi, Syed Muhammad Zeeshan Iqbal, Ali Zia, and Sajid Anwer. Predicting user behavior on video streaming by using watch-time duration analysis. *Knowledge-Based Systems*, page 114779, 2025.

Anand Balachandran, Geoffrey M Voelker, Paramvir Bahl, and P Venkat Rangan. Characterizing user behavior and network performance in a public wireless lan. In *Proceedings of the 2002 ACM SIGMETRICS international conference on Measurement and modeling of computer systems*, pages 195–205, 2002.

James Davidson, Benjamin Liebald, Junning Liu, Palash Nandy, Taylor Van Vleet, Ullas Gargi, Sujoy Gupta, Yu He, Mike Lambert, Blake Livingston, et al. The youtube video recommendation system. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 293–296, 2010.

Trisha Dunleavy. Creative innovation, global engagement and public outcomes: The significance of netflix's adolescence. *Critical Studies in Television*, page 17496020261455492, 2026.

Chongming Gao, Shijun Li, Wenqiang Lei, Jiawei Chen, Biao Li, Peng Jiang, Xiangnan He, Jiaxin Mao, and Tat-Seng Chua. Kuairec: A fully-observed dataset and insights for evaluating recommender systems. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pages 540–550, 2022.

Ge Gao, Hongyan Liu, and Kang Zhao. Live streaming recommendations based on dynamic representation learning. *Decision Support Systems*, 169:113957, 2023.

Ke Guo, Changle Qu, Jiayaqi Cheng, Xiao Zhang, Shijun Wang, Xiaoyu Zhang, Xueliang Wang, Le Zhang, Lantao Hu, and Jun Xu. Kuailive-m3: A multi-modal, multi-domain, and multi-feedback dataset for live streaming recommendation. *arXiv preprint arXiv:2607.24862*, 2026a.

Ke Guo, Changle Qu, Xiao Zhang, Liqin Zhao, Shijun Wang, Yanan Niu, and Jun Xu. Room matters: Dynamic room-level collaboration information modeling for live streaming recommendation. In *Proceedings of the ACM Web Conference 2026*, pages 6045–6056, 2026b.

Chenghou Jin, Yixin Ren, Hongxu Ma, Yewei Xia, Yi Guan, Hao Zhang, Jiandong Ding, Jihong Guan, and Shuigeng Zhou. Invariant feature learning for counterfactual watch-time prediction in video recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 14964–14972, 2026.

Marius Kaltenbach. *The Levenberg-Marquardt Method and its Implementation in Python*. PhD thesis, Universität Konstanz, 2022.

Edward L Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481, 1958.

John P Klein and Melvin L Moeschberger. *Survival analysis: techniques for censored and truncated data*, volume 1230. Springer, 2003.

S Shunmuga Krishnan and Ramesh K Sitaraman. Video stream quality impacts viewer behavior: inferring causality using quasi-experimental designs. In *Proceedings of the 2012 Internet Measurement Conference*, pages 211–224, 2012.

Changhee Lee, Jinsung Yoon, and Mihaela Van Der Schaar. Dynamic-deephit: A deep learning approach for dynamic survival analysis with competing risks based on longitudinal data. *IEEE Transactions on Biomedical Engineering*, 67(1):122–133, 2019.

Haiqun Lin and Daniel Zelterman. Modeling survival data: extending the cox model, 2002.

Xiao Lin, Xiaokai Chen, Linfeng Song, Jingwei Liu, Biao Li, and Peng Jiang. Tree based progressive regression model for watch-time prediction in short-video recommendation. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 4497–4506, 2023.

Emily Liu, Kuan Han, Minfeng Zhan, Bocheng Zhao, Guanyu Mu, and Yang Song. Relative advantage debiasing for watch-time prediction in short-video recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 15296–15305, 2026.

Donald W Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics*, 11(2):431–441, 1963.

Chirag Nagpal, Xinyu Li, and Artur Dubrawski. Deep survival machines: Fully parametric survival regression and representation learning for censored data with competing risks. *IEEE Journal of Biomedical and Health Informatics*, 25(8):3163–3175, 2021.

Netflix. What we watched the second half of 2024. *Netflix News*, 2025.

Yuhan Quan, Jingtao Ding, Chen Gao, Nian Li, Lingling Yi, Depeng Jin, and Yong Li. Alleviating video-length effect for micro-video recommendation. *ACM Transactions on Information Systems*, 42(2):1–24, 2023.

James O Ramsay and Bernard W Silverman. *Functional data analysis*. Springer, 1997.

Waloddi Weibull. A statistical distribution function of wide applicability. *Journal of applied mechanics*, 1951.

Gerald Woo, Chenghao Liu, Doyen Sahoo, Akshat Kumar, and Steven Hoi. Learning deep time-index models for time series forecasting. In *International Conference on Machine Learning*, pages 37217–37237. PMLR, 2023.

Ruohan Zhan, Changhua Pei, Qiang Su, Jianfeng Wen, Xueliang Wang, Guanyu Mu, Dong Zheng, Peng Jiang, and Kun Gai. Deconfounding duration bias in watch-time prediction for video recommendation. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*, pages 4472–4481, 2022.

Shuai Zhang, Hongyan Liu, Jun He, Sanpu Han, and Xiaoyong Du. A deep bi-directional prediction model for live streaming recommendation. *Information Processing & Management*, 58(2):102453, 2021.

Yang Zhang, Yimeng Bai, Jianxin Chang, Xiaoxue Zang, Song Lu, Jing Lu, Fuli Feng, Yanan Niu, and Yang Song. Leveraging watch-time feedback for short-video recommendations: A causal labeling framework. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pages 4952–4959, 2023.

Haiyuan Zhao, Lei Zhang, Jun Xu, Guohao Cai, Zhenhua Dong, and Ji-Rong Wen. Uncovering user interest from biased and noised watch time in video recommendation. In *Proceedings of the 17th ACM Conference on Recommender Systems*, pages 528–539, 2023.

Haiyuan Zhao, Guohao Cai, Jieming Zhu, Zhenhua Dong, Jun Xu, and Ji-Rong Wen. Counteracting duration bias in video recommendation via counterfactual watch time. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 4455–4466, 2024.

Zhe Zhao, Lichan Hong, Li Wei, Jilin Chen, Aniruddh Nath, Shawn Andrews, Aditee Kumthekar, Maheswaran Sathiamoorthy, Xinyang Yi, and Ed Chi. Recommending what video to watch next: a multitask ranking system. In *Proceedings of the 13th ACM conference on recommender systems*, pages 43–51, 2019.

Yu Zheng, Chen Gao, Jingtao Ding, Lingling Yi, Depeng Jin, Yong Li, and Meng Wang. Dvr: micro-video recommendation optimizing watch-time-gain under duration bias. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 334–345, 2022.

A Integrated Absolute Error as Average Error Under a Uniform Distribution

For a target trajectory $Y : [0,1] \to \mathbb{R}$ and prediction $\widehat{Y} : [0,1] \to \mathbb{R}$, define the pointwise absolute error as $e(u) = |Y(u) - \widehat{Y}(u)|$. The integrated absolute error is

$$E_{\mathrm{IAE}} = \int_0^1 e(u)\, du.$$

Consider a uniform partition of $[0,1]$ into N intervals of width $\Delta u = 1/N$, and choose one point u_j^* from each interval. The corresponding Riemann sum is $\sum_{j=1}^N e(u_j^*)\Delta u = \frac{1}{N}\sum_{j=1}^N e(u_j^*)$. Hence,

$$E_{\mathrm{IAE}} = \lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} \left| Y(u_j^*) - \widehat{Y}(u_j^*) \right|.$$

Therefore, on the normalized domain $[0,1]$, IAE is exactly the expected absolute pointwise error when $u \sim \mathrm{Uniform}[0,1]$, i.e., $E_{\mathrm{IAE}} = \mathbb{E}_{u\sim\mathrm{Uniform}[0,1]}[|Y(u) - \widehat{Y}(u)|]$. On a finite uniform evaluation grid $u_1, \ldots, u_N$, it is approximated by

$$E_{\mathrm{IAE}} \approx \frac{1}{N} \sum_{j=1}^{N} \left| Y(u_j) - \widehat{Y}(u_j) \right|.$$

Thus IAE is the average absolute trajectory error over normalized progress. When reported as a percentage, the quantity above is multiplied by 100.

B Library of Curve Families.

We consider a library of 30 low-dimensional template families for a remaining-mass curve $F(u)$, where $u \in [0,1]$ denotes normalized time. The families are grouped below according to their functional form. The identifier shown in parentheses after each family name is the corresponding term used in the implementation and in the reported results.

Weibull (`weibull`). The two-parameter Weibull family, with $\lambda > 0$ and $k > 0$, is

$$F(u) = \exp\left[-\left(\frac{u}{\lambda}\right)^k \right]. \tag{A.1}$$

Log-logistic (`loglogistic2`). The two-parameter log-logistic family, with $\lambda > 0$ and $k > 0$, is

$$F(u) = \frac{1}{1 + \left(\frac{u}{\lambda}\right)^k}. \tag{A.2}$$

Burr XII (`burr3`). The three-parameter Burr XII family, with $\lambda, c, k > 0$, is

$$F(u) = \left[1 + \left(\frac{u}{\lambda}\right)^c \right]^{-k}. \tag{A.3}$$

Exponentiated Weibull (`expweibull3`). The three-parameter exponentiated Weibull family, with $\lambda, k, \alpha > 0$, is

$$F(u) = 1 - \left[1 - \exp\left(-\left(\frac{u}{\lambda}\right)^k \right) \right]^{\alpha}. \tag{A.4}$$

Generalized gamma (`gengamma3`). Let $\Gamma(\alpha) = \int_0^\infty z^{\alpha-1} e^{-z}\, dz$. The generalized-gamma family, with $\lambda, \alpha, p > 0$, is

$$F(u) = \frac{1}{\Gamma(\alpha)} \int_{(u/\lambda)^p}^{\infty} z^{\alpha-1} e^{-z}\, dz. \tag{A.5}$$

Modified Weibull (`modweibull3`). The modified Weibull family is

$$F(u) = \exp\left[-au^b \exp(cu)\right], \qquad a > 0, \quad b > 0. \tag{A.6}$$

Cure-Kumaraswamy (`cure_kumar3`). The cure-Kumaraswamy family, with $a, b > 0$ and cure fraction $q \in (0, 1)$, is

$$F(u) = q + (1 - q)(1 - u^a)^b. \tag{A.7}$$

Cure-Beta (`cure_beta3`). Let $B(a, b) = \int_0^1 z^{a-1}(1 - z)^{b-1}\, dz$. The cure-Beta family, with $a, b > 0$ and $q \in (0, 1)$, is

$$F(u) = q + (1 - q)\left[1 - \frac{1}{B(a, b)} \int_0^u z^{a-1}(1 - z)^{b-1}\, dz\right]. \tag{A.8}$$

Degree-three Bernstein (`bernstein3`). Let $\alpha_0 = 1$ and $\alpha_i = \prod_{j=1}^{i} \sigma_j$, where $\sigma_j \in (0, 1)$. The degree-three Bernstein family is

$$F(u) = (1 - u)^3 + 3u(1 - u)^2\alpha_1 + 3u^2(1 - u)\alpha_2 + u^3\alpha_3. \tag{A.9}$$

Degree-four Bernstein (`bernstein4`). Using the same cumulative-product parameterization $\alpha_i = \prod_{j=1}^{i} \sigma_j$, the degree-four Bernstein family is

$$F(u) = (1 - u)^4 + 4u(1 - u)^3\alpha_1 + 6u^2(1 - u)^2\alpha_2 + 4u^3(1 - u)\alpha_3 + u^4\alpha_4. \tag{A.10}$$

Weibull with one damped sine (`w1ds`). A damped-sine component has the form $a_\ell e^{-d_\ell u} \sin(\omega_\ell u + \phi_\ell)$, with $d_\ell \geq 0$. The one-sine family is

$$F(u) = \exp\left[-\left(\frac{u}{\lambda}\right)^k\right] + a_1 e^{-d_1 u} \sin(\omega_1 u + \phi_1). \tag{A.11}$$

Weibull with two damped sines (`w2ds`). The two-sine family is

$$F(u) = \exp\left[-\left(\frac{u}{\lambda}\right)^k\right] + \sum_{\ell=1}^{2} a_\ell e^{-d_\ell u} \sin(\omega_\ell u + \phi_\ell). \tag{A.12}$$

Weibull with three damped sines (`w3ds`). The three-sine family is

$$F(u) = \exp\left[-\left(\frac{u}{\lambda}\right)^k\right] + \sum_{\ell=1}^{3} a_\ell e^{-d_\ell u} \sin(\omega_\ell u + \phi_\ell). \tag{A.13}$$

Weibull with four damped sines (`w4ds`). The four-sine family is

$$F(u) = \exp\left[-\left(\frac{u}{\lambda}\right)^k\right] + \sum_{\ell=1}^{4} a_\ell e^{-d_\ell u} \sin(\omega_\ell u + \phi_\ell). \tag{A.14}$$

Two damped sines with one Gaussian (`w2ds_g1`). A Gaussian component has the form $A_\ell \exp(-((u - \mu_\ell)/\sigma_\ell)^2)$, with $\mu_\ell \in [0, 1]$ and $\sigma_\ell > 0$. The one-Gaussian variant is

$$F(u) = \exp\left[-\left(\frac{u}{\lambda}\right)^k\right] + \sum_{\ell=1}^{2} a_\ell e^{-d_\ell u} \sin(\omega_\ell u + \phi_\ell) + A_1 \exp\left[-\left(\frac{u - \mu_1}{\sigma_1}\right)^2\right]. \tag{A.15}$$

Two damped sines with two Gaussians (`w2ds_g2`). The two-Gaussian variant is

$$F(u) = \exp\left[-\left(\frac{u}{\lambda}\right)^k\right] + \sum_{\ell=1}^{2} a_\ell e^{-d_\ell u} \sin(\omega_\ell u + \phi_\ell) + \sum_{\ell=1}^{2} A_\ell \exp\left[-\left(\frac{u - \mu_\ell}{\sigma_\ell}\right)^2\right]. \tag{A.16}$$

Two damped sines with three Gaussians (`w2ds_g3`). The three-Gaussian variant is

$$F(u) = \exp\left[-\left(\frac{u}{\lambda}\right)^k\right] + \sum_{\ell=1}^{2} a_\ell e^{-d_\ell u}\sin(\omega_\ell u + \phi_\ell) + \sum_{\ell=1}^{3} A_\ell \exp\left[-\left(\frac{u-\mu_\ell}{\sigma_\ell}\right)^2\right]. \quad \text{(A.17)}$$

Direct basis expansions (`pca2`, `pca3`, `pca4`, `pca5`). Let $\mu(u)$ denote a reference curve and $\phi_1, \ldots, \phi_r$ denote basis functions. The rank-r direct basis family is

$$F(u) = \mu(u) + \sum_{j=1}^{r} c_j \phi_j(u), \qquad r \in \{2, 3, 4, 5\}. \quad \text{(A.18)}$$

The choices $r = 2, 3, 4, 5$ correspond respectively to `pca2`, `pca3`, `pca4`, and `pca5`.

Weibull residual basis (`weibull_respc1`, `weibull_respc2`). Let $\mu_R(u)$ denote a reference residual and ψ_1, ψ_2 denote residual basis functions. The rank-r family is

$$F(u) = \exp\left[-\left(\frac{u}{\lambda}\right)^k\right] + \mu_R(u) + \sum_{j=1}^{r} c_j \psi_j(u), \qquad r \in \{1, 2\}. \quad \text{(A.19)}$$

The choices $r = 1$ and $r = 2$ correspond respectively to `weibull_respc1` and `weibull_respc2`.

Monotone Weibull residual basis (`weibull_respc1_monotone`, `weibull_respc2_monotone`). The monotone variants use the same underlying residual-basis curve followed by the projection

$$F_{\text{mono}}(u_i) = \min_{j \le i} \text{clip}(F(u_j), 0, 1)\,, \qquad r \in \{1, 2\}. \quad \text{(A.20)}$$

The rank-one and rank-two variants correspond respectively to `weibull_respc1_monotone` and `weibull_respc2_monotone`.

Weibull hazard basis (`weibull_hazpc1`, `weibull_hazpc2`). Let $r_j = \exp(\mu_\rho(u_j) + \sum_{m=1}^{r} c_m \rho_m(u_j))$ be a positive multiplier on the interval $[u_j, u_{j+1}]$. For $u \in [u_j, u_{j+1}]$, the rank-r family is

$$F(u) = \exp\left\{-\sum_{i<j} r_i\left[\left(\frac{u_{i+1}}{\lambda}\right)^k - \left(\frac{u_i}{\lambda}\right)^k\right] - r_j\left[\left(\frac{u}{\lambda}\right)^k - \left(\frac{u_j}{\lambda}\right)^k\right]\right\}, \qquad r \in \{1, 2\}. \quad \text{(A.21)}$$

The choices $r = 1$ and $r = 2$ correspond respectively to `weibull_hazpc1` and `weibull_hazpc2`.

Log-hazard basis (`loghazard_pc2`, `loghazard_pc3`, `loghazard_pc4`). Let $\delta_j = \exp(\mu_H(u_j) + \sum_{m=1}^{r} c_m \eta_m(u_j))$. For $u \in [u_j, u_{j+1}]$, define $\gamma = (u - u_j)/(u_{j+1} - u_j)$. The rank-$r$ family is

$$F(u) = \exp\left[-\sum_{i<j} \delta_i - \gamma\delta_j\right], \qquad r \in \{2, 3, 4\}. \quad \text{(A.22)}$$

The choices $r = 2, 3, 4$ correspond respectively to `loghazard_pc2`, `loghazard_pc3`, and `loghazard_pc4`.

C Additional Experiments and Ablations

This appendix reports the remaining experiments used for model selection, ablation, robustness analysis, and interpretation.

Table 8: Training-seed stability on the locked KuaiLive-M3 test set. Values are mean ± sample standard deviation over three independent training seeds.

Model	mean $\mathbf{E}_{\mathrm{IAE}}$	$\mathbf{I_5}$
Structured Curves	$\mathbf{5.53\% \pm 0.01\%}$	$\mathbf{56.0\% \pm 0.1\%}$
GRU along u	$6.37\% \pm 0.04\%$	$41.0\% \pm 0.5\%$
RNN along u	$6.82\% \pm 0.02\%$	$33.7\% \pm 1.4\%$
MLP along u	$13.33\% \pm 5.07\%$	$16.7\% \pm 7.5\%$

Table 9: Training-seed stability on the locked KuaiRec test set. Values are mean ± sample standard deviation over three independent training seeds.

Model	mean $\mathbf{E}_{\mathrm{IAE}}$	$\mathbf{C_5}$
MLP along u	$1.99\% \pm 0.02\%$	$68.4\% \pm 0.3\%$
GRU per bin	$2.05\% \pm 0.00\%$	$57.9\% \pm 0.2\%$
RNN per bin	$2.10\% \pm 0.00\%$	$52.7\% \pm 0.3\%$
GRU along u	$2.12\% \pm 0.04\%$	$63.8\% \pm 4.3\%$
RNN along u	$2.21\% \pm 0.02\%$	$60.4\% \pm 2.7\%$
MLP per bin	$2.25\% \pm 0.10\%$	$19.6\% \pm 15.1\%$

C.1 Training-Seed Stability

We retrain the learned neural models with three independent random seeds on the same locked test splits. On KuaiLive-M3, Table 8 shows that the structured model is highly stable across retrainings, while the MLP baseline exhibits substantially larger variation. On KuaiRec, Table 9 shows that the selected Burr+G20+sine3 representation is deterministic once fitted; among the learned baselines, the per-bin MLP exhibits particularly high variability in C_5.

C.2 Transfer Across Alternative Structured Curve Families

We compare alternative structured curve families under the same cross-audience transfer protocol. Each family divides its coordinates between video-specific and watch-style-specific parameters. The *mask* records this allocation: V denotes a video-specific coordinate estimated for each held-out video from its KuaiRec-small retention curve, while C denotes a coordinate shared within each watch-style group and learned from the KuaiRec-small training set. For residual-augmented families, the terms before and after $+$ give the allocation for the base family and residual parameters, respectively. The column k is the number of video-specific coordinates and n_{hyp} is the number of coordinates associated with each watch-style group. At transfer time, all learned watch-style parameters are frozen and only the observed KuaiRec-big audience mixture is changed.

As shown in Table 10, the selected Burr+G20+sine3 VCC+CC representation achieves the best mean E_{IAE} and C_5, reaching 1.90% mean E_{IAE}, 75.3% C_5, and 98.7% C_{10}. The plain VCC Burr model reaches 2.27% mean E_{IAE} and about 62% C_5, indicating that the Gaussian and sinusoidal residual terms provide additional transfer accuracy beyond the base Burr factorization.

C.3 Additional Class-Structure and Transfer Analyses

This appendix provides the supporting analyses for the class-structure experiments in Section 4.2. We report component-level and occupancy-level ablations on KuaiLive-M3, a direct transfer of the KuaiRec VCC parameter-sharing construction to live streams, an oracle experiment that removes the GRU predictor, sensitivity to the number of watch-style groups, and category-wise analyses on KuaiRec.

For KuaiLive-M3, we first compare author-indexed and completely shared cure-Kumaraswamy models for the component trajectories J and $1 - X$. Here, mean E_{IAE} is the average integrated absolute error over rooms, mean E_{sup} is the average sup-norm error, I_5 is the fraction of rooms with occupancy IAE at most 5%, and # Params. is the number of learned model parameters. As shown in Table 11, the differences between author-specific and shared component models are small.

Table 10: KuaiRec watch-style mixture transfer across alternative structured curve families. In the mask, V denotes a video-specific coordinate and C a watch-style-specific coordinate; k is the number of video-specific coordinates and n_{hyp} the number of coordinates associated with each watch-style group.

Family	**Mask**	k	n_{hyp}	**mean** $\mathbf{E}_{\mathrm{IAE}}$	$\mathbf{C_5}$	$\mathbf{C_{10}}$	$\mathbf{C_2}$
Burr+G20+sine3	VCC+CC	1	4	**1.90%**	**75%**	99%	5%
Exp-Weibull+resPC2	VCC+VC	2	3	1.92%	74%	98%	6%
Gen-Gamma+G20+ramp	VCC+CV	2	3	1.97%	75%	99%	5%
Gen-Gamma	VCC	1	2	2.25%	64%	98%	1%
W+1DS	VCVCCC	2	4	2.25%	60%	97%	2%
Burr	VCC	1	2	2.27%	62%	98%	1%
Exp-Weibull	VCC	1	2	2.28%	62%	98%	1%
Weibull	VC	1	1	2.30%	59%	97%	1%
Log-logistic	VC	1	1	3.59%	30%	92%	0%
Bernstein-4	CVVC	2	2	4.10%	36%	81%	0%
Mod-Weibull	CCV	1	2	4.26%	8%	41%	0%
Cure-Beta	CVV	2	1	4.70%	3%	72%	0%
Bernstein-3	VCV	2	1	4.86%	7%	43%	0%
Cure-Kumaraswamy	CVV	2	1	4.92%	1%	67%	0%

Table 11: KuaiLive-M3 class ablation for the component trajectories. Lower mean E_{IAE} and mean E_{sup} are better; higher I_5 is better.

Method	**mean** $\mathbf{E}_{\mathrm{IAE}}$	**mean** $\mathbf{E}_{\mathrm{sup}}$	$\mathbf{I_5}$	**# Params.**
Author-class J	7.50%	17.10%	38%	201
Shared J	7.60%	17.20%	38%	3
Author-class $1-X$	7.00%	17.30%	40%	37,647
Shared $1-X$	6.90%	17.30%	41%	37,449

We next evaluate the same class choices after composing joins and exits into occupancy $O = J - X$. In Table 12, O-IAE and O-sup denote occupancy IAE and sup-norm error; J-IAE and X-IAE are the corresponding component errors; and $X > J$ counts evaluated points at which the predicted cumulative exits exceed predicted cumulative joins. "Plugin" composes independently trained join and exit models, "joint" optimizes the occupancy objective end-to-end, and "oracle" replaces the learned exit predictor with a freely optimized representation. The degraded collapsed-plugin result is an initialization artifact: retraining the shared J and $1 - X$ models returns performance to the same range as the author-class model.

We then construct $K = 5$ behavioral watch-style groups using only training-room interactions. Each user is represented by five features: early-skipping probability $P(\mathrm{stay}/L < 0.20)$, completion probability $P(\mathrm{stay}/L \geq 0.90)$, median normalized stay $\min(\mathrm{stay}/L, 1)$, re-entry probability $P(n_{\mathrm{sessions}} > 1)$, and the standard deviation of normalized stay. Clustering 15,661 users from 20,214 training rooms and assigning each room by its majority user group produces three nonempty room groups with 11,558, 6,701, and 1,955 rooms. Table 13 compares the component trajectories under author, shared, and watch-style parameterization.

Table 14 gives the corresponding occupancy results. Watch-style grouping improves the plugin occupancy IAE from 0.071 for author classes to 0.067, but the difference largely disappears after joint optimization: author, shared, and watch-style models reach 0.055, 0.055, and 0.054 occupancy IAE, respectively.

We also transplant the KuaiRec VCC parameter-sharing construction to KuaiLive-M3. For the three cure-Kumaraswamy coordinates (a, b, q), VCC means that one coordinate is room-specific (V) while the remaining two are shared within the behavioral class (C). Table 15 reports the resulting approximation for J, X, and O. The "class mixture" rows use only transferable class-level structure,

Table 12: KuaiLive-M3 author-class versus shared occupancy models. O-IAE and O-sup are occupancy errors; J-IAE and X-IAE are component errors; $X > J$ counts points violating the cumulative join–exit ordering.

Method	O-IAE	O-sup	I_5	J-IAE	X-IAE	$X > J$	**# Params.**
Author plugin	0.071	0.158	40%	0.075	0.070	1394	37,848
Collapsed plugin	0.075	0.168	38%	0.083	0.070	2488	37,452
Retrained shared plugin	0.073	0.161	37%	0.076	0.069	657	37,452
Author joint	0.055	0.137	56%	0.075	0.075	826	37,848
Shared joint	0.055	0.136	56%	0.076	0.078	861	37,452
Author oracle	0.025	0.107	95%	0.078	0.078	1873	23,356,122
Shared oracle	0.025	0.107	94%	0.079	0.078	1853	23,355,924

Table 13: KuaiLive-M3 component prediction using author classes, no classes, or behavioral watch-style classes.

Method	**mean $\mathbf{E}_{\mathrm{IAE}}$**	**mean $\mathbf{E}_{\mathrm{sup}}$**	$\mathbf{I_5}$	**# Params.**
Author J	7.50%	17.10%	38%	201
Shared J	7.60%	17.20%	38%	3
Watch-style J	7.30%	16.70%	39%	15
Author $1 - X$	7.00%	17.30%	40%	37,647
Shared $1 - X$	6.90%	17.30%	41%	37,449
Watch-style $1 - X$	7.00%	17.30%	41%	37,461

whereas the held-out-room rows assign a free coordinate to each test room and therefore measure reconstruction capacity rather than deployable prediction. In the table, Target identifies the trajectory being approximated and # Params. is the total number of fitted parameters.

To check whether the weak class effect is merely being hidden by the capacity of the GRU, we remove the GRU and directly optimize the exit representation for every room. This "oracle" setting measures the best reconstruction attainable when the representation itself is not predicted. Table 16 shows that even in this setting the class effect remains small: watch-style grouping reaches 0.023 occupancy IAE, compared with 0.025 for both author and shared parameterizations.

Finally, we vary the number K of KuaiLive watch-style clusters. Here K is the requested number of behavioral user clusters before rooms are assigned by majority membership. As shown in Table 17, performance changes very little across $K = 2$ to $K = 10$: mean occupancy IAE remains between 5.383% and 5.430%. For $K = 8$ and $K = 10$, several user clusters receive no majority-assigned rooms, but this does not materially change the result.

For KuaiRec, we first compare within-audience reconstruction with cross-audience transfer for the selected Burr XII + G20 + sine3 model. In Table 18, n is the number of videos, mean E_{IAE} and mean E_{sup} are video-averaged IAE and sup-norm errors, I_5 is the fraction of videos with IAE at most 5%, and C_5 is the fraction whose sup-norm error is at most 5%. The two validation rows measure within-audience reconstruction, while the final row evaluates locked KuaiRec-small-to-KuaiRec-big transfer.

We next break the locked transfer result down by KuaiRec L1 content category. Table 19 includes every category with at least 30 test videos; n is the number of videos in that category and C_5 is the fraction whose sup-norm transfer error is at most 5%. No sufficiently large category returns to the roughly 6.8% video-only transfer gap. Pets, dance, and sports are among the strongest categories, while food and comedy are among the weaker ones.

Finally, we compare behavioral watch-style classes with content-based class definitions under the same locked transfer protocol. Table 20 reports the structured family used for each class scheme, mean transfer E_{IAE}, and C_{10} and C_5, the fractions of test videos whose sup-norm errors are at most 10% and 5%, respectively. Replacing watch-style groups with video L1 categories or content-affinity

Table 14: KuaiLive-M3 occupancy under alternative class structures. O-IAE and O-sup are occupancy errors; J-IAE and X-IAE are component errors; $X > J$ counts ordering violations.

Method	O-IAE	O-sup	I_5	J-IAE	X-IAE	$X > J$	**# Params.**
Author plugin	0.071	0.158	40%	0.075	0.070	1394	37,848
Shared plugin	0.073	0.161	37%	0.076	0.069	657	37,452
Watch-style plugin	0.067	0.153	42%	0.073	0.070	1258	37,476
Author joint	0.055	0.137	56%	0.075	0.075	826	37,848
Shared joint	0.055	0.136	56%	0.076	0.078	861	37,452
Watch-style joint	0.054	0.134	56%	0.073	0.073	823	37,476

Table 15: KuaiLive-M3 VCC parameter-sharing experiment.

Regime	**Target**	**mean $\mathbf{E}_{\mathrm{IAE}}$**	**mean $\mathbf{E}_{\mathrm{sup}}$**	**$\mathbf{I_5}$**	**# Params.**
Class mixture	J	7.70%	16.90%	38%	10
Class mixture	X	8.00%	16.90%	33%	10
Class mixture	O	7.90%	14.50%	33%	10
Held-out room coordinate	J	3.30%	9.40%	86%	4,053
Held-out room coordinate	X	3.90%	10.30%	77%	4,053
Held-out room coordinate	O	2.90%	7.90%	92%	4,053

classes removes most of the transfer advantage: mean E_{IAE} returns to 6.8%–6.9% and C_5 falls to zero.

Taken together, the additional analyses reinforce the main-text result: class structure has only a weak effect on KuaiLive-M3, including at the component level, under direct representation optimization, and across different numbers of behavioral clusters. In KuaiRec, however, behavioral watch-style groups capture the relevant population shift much more effectively than content-based alternatives. Further KuaiRec analyses of parameter sensitivity, V/C allocation, residual terms, target-audience information budget, and the number of watch-style groups are reported in Appendix C.7.

C.4 KuaiLive-M3: Approximator and Composition Selection

The main text evaluates the final KuaiLive-M3 model against neural, time-series, and event-time baselines. Here we report the model-selection experiments used to choose the structured join and complementary-exit approximators and to determine how they should be composed into occupancy.

Join-family selection. A single join parameter vector is shared across rooms. We search roughly 30 candidate unit-interval families and select using validation mean sup-norm error. In Table 21, k denotes the number of parameters in the corresponding curve family, mean E_{IAE} is the average integrated absolute error across rooms, mean E_{sup} is the average sup-norm error, and I_5 is the fraction of rooms with IAE at most 5%. Although W+2DS+G1 gives the best validation sup-norm error, the leading families are tightly clustered on test. We therefore retain the compact three-parameter cure-Kumaraswamy family, which also matches the selected complementary-exit family.

Complementary-exit family selection. For $1 - X$, a shared GRU maps the causal room-representation history available up to time u to the coordinates of each candidate family. Table 22 reports the relevant noncollapsed families. Here, # Params. is the total number of learned model parameters, including the GRU and the curve-coordinate output layer. Roughly 20 additional families collapse to a dummy solution with IAE near 0.46. Among the viable models, cure-Kumaraswamy performs best, with the next family already increasing mean IAE from 6.9% to 7.1%.

Plugin composition. We next freeze the independently trained join and complementary-exit models and construct occupancy as $O = J - X$, without any additional occupancy training. Table 23 reports this plugin composition. O-IAE and O-sup are occupancy errors, J-IAE and X-IAE are the corresponding component errors, and $X > J$ counts evaluated points at which the predicted

Table 16: KuaiLive-M3 class structure when the exit representation is optimized directly rather than predicted by the GRU.

Method	O-IAE	O-sup	I_5	J-IAE	X-IAE	**# Params.**
Author oracle	0.025	0.107	95%	0.078	0.078	23,356,122
Shared oracle	0.025	0.107	94%	0.079	0.078	23,355,924
Watch-style oracle	0.023	0.102	98%	0.076	0.074	23,355,936

Table 17: KuaiLive-M3 sensitivity to the number K of watch-style groups.

K	**mean** $\mathbf{E}_{\mathrm{IAE}}$	**mean** $\mathbf{E}_{\mathrm{sup}}$	$\mathbf{I_5}$
2	5.383%	13.219%	58.5%
3	5.418%	13.335%	57.7%
5	5.430%	13.388%	56.5%
8	5.384%	13.204%	57.7%
10	5.387%	13.162%	57.0%

cumulative exits exceed predicted cumulative joins. The retained join families all give similar plugin occupancy accuracy, with mean occupancy IAE around 7.2%–7.3%.

Joint occupancy optimization. Finally, we fine-tune each retained join/exit pair directly for occupancy, using the occupancy loss together with light J- and X-auxiliary losses and an $X > J$ ordering penalty. Table 24 reports the resulting test metrics together with validation occupancy sup-norm error, denoted Val. O-sup, which is the quantity used for model selection. Joint optimization substantially improves occupancy relative to plugin composition, reducing mean occupancy IAE from roughly 7.3% to 5.6%, while the strongest families remain very close to one another. Validation selects cure-Kumaraswamy for J, yielding the same compact three-parameter family for both J and $1 - X$.

Overall, the family search shows that several structured approximators perform similarly for joins, while cure-Kumaraswamy is more clearly preferred for the complementary exit process. Once occupancy is optimized jointly, the leading join families again become nearly indistinguishable, making cure-Kumaraswamy a natural final choice because it is compact and provides a common parameterization for both components.

C.5 KuaiLive-M3: Duration-Stratified Robustness

We evaluate whether occupancy prediction quality changes systematically with room duration by stratifying the locked KuaiLive-M3 test set using the test-set duration mean $\mu = 6.39$ h and standard deviation $\sigma = 13.88$ h. Each row in Table 25 reports the fraction of test rooms in the corresponding duration bucket, together with mean occupancy IAE (O-IAE), mean occupancy sup-norm error (O-sup), I_5, the fraction of rooms with occupancy IAE at most 5%, and the component errors J-IAE and X-IAE for cumulative joins and exits. Since $\mu - \sigma/2 = -0.55$ h, the lowest nominal duration bucket contains no rooms.

Table 25 shows that occupancy prediction is reasonably stable over most duration ranges. The shortest nonempty bucket, corresponding to rooms shorter than roughly 2.9 h, has the largest occupancy error (0.060 IAE), while the central duration ranges achieve approximately 0.051 IAE. For longer rooms, occupancy error increases only modestly even though the separate join and exit errors rise more clearly with duration, suggesting that errors in the two components partially cancel when composed into occupancy.

C.6 KuaiRec: Within-Audience Reconstruction and Family Selection

The main paper evaluates cross-audience transfer on the locked 995-video test set. All family-selection and parameter-allocation choices are made without using this test set. The data split contains 2,157

Table 18: KuaiRec selected-model reconstruction and transfer. The final column reports IAE coverage I_5 followed by sup-norm coverage C_5.

Setting	n	**mean** $\mathbf{E}_{\mathrm{IAE}}$	**mean** $\mathbf{E}_{\mathrm{sup}}$	I_5/C_5
Small reconstruction (val)	166	0.89%	2.90%	100% / 99%
Big reconstruction (val)	166	1.18%	3.10%	100% / 90%
Small→big transfer (test)	995	1.90%	4.20%	99% / 75%

Table 19: KuaiRec small-to-big transfer by L1 content category ($n \geq 30$).

Category	n	**mean** $\mathbf{E}_{\mathrm{IAE}}$	$\mathbf{C_5}$
Pets	42	1.68%	90%
Dance	35	1.69%	89%
Sports	43	1.70%	88%
Looks / appearance	101	1.76%	79%
Photography	44	1.76%	70%
Celebrity / entertainment	72	1.82%	74%
Other	59	1.93%	75%
Livelihood / news	189	1.93%	75%
Food	36	1.94%	56%
Fashion	32	2.05%	72%
Comedy	42	2.20%	60%

training videos, 166 held-out validation videos, and 995 locked transfer-test videos; the reconstruction experiments below use only the held-out validation set.

Video-specific reconstruction. We first allow every coordinate of each candidate family to be video-specific and fit each audience independently. In Table 26, k is the number of free curve parameters per video, Small/Big IAE are the mean integrated absolute errors on KuaiRec-small and KuaiRec-big, and Small/Big C_τ denote the fractions of validation videos whose sup-norm error is at most $\tau\%$. Increasing k substantially improves within-audience reconstruction, with the highest-capacity families approaching near-exact fits.

Table 26 therefore shows that within-audience reconstruction error can be driven very low simply by increasing video-specific capacity. As shown by the transfer experiments in the main paper, however, this additional capacity does not by itself produce cross-audience generalization.

Watch-style-mixture reconstruction. We next reconstruct each audience using the same video/watch-style factorization used for cross-audience transfer. In Table 27, the *mask* specifies which coordinates are video-specific (V) and which are shared within a watch-style group (C); for residual-augmented families, the terms before and after $+$ describe the allocations of the base family and residual coordinates, respectively. The column k is the number of video-specific coordinates, while n_{hyp} is the number of coordinates associated with each watch-style group. Small/Big IAE and Small/Big C_τ are defined as above for the two audiences.

The reconstruction results in Table 27 are intentionally weaker than those of the highest-capacity fully video-specific models. For example, W+1DS with all-video-specific parameters reaches $0.0049/0.0042$ IAE on KuaiRec-small/big, whereas the strongest watch-style-mixture reconstructions are approximately $0.008/0.010$. The benefit of the watch-style factorization is therefore not superior within-audience reconstruction; it is the ability to separate video-specific and population-specific variation in a form that transfers across audiences.

C.7 KuaiRec: Parameterization and Residual Ablations

The selected short-form model uses one video-specific Burr coordinate, λ, and four watch-style-specific coordinates: c, k, α_{G20}, and β_{sine3}. We analyze the sensitivity of these parameters, the video/watch-style allocation, and the contribution of the residual terms.

Table 20: KuaiRec class-taxonomy comparison on the locked transfer test set.

Class structure	Structured family	mean $\mathbf{E}_{\mathrm{IAE}}$	$\mathbf{C_{10}}$	$\mathbf{C_5}$
Watch style	Burr XII + G20 + sine3	**1.90%**	98.7%	**75.3%**
Video L1	Burr XII + G20 + sine3	6.80%	24.9%	0.0%
Content affinity	Burr XII + G20 + sine3	6.90%	23.9%	0.0%
Watch style	Gen-Gamma + G20 + ramp	2.00%	**98.8%**	74.7%
Watch style	W+1DS	2.30%	97.3%	59.8%

Table 21: KuaiLive-M3 join approximator sweep. Lower mean E_{IAE} and mean E_{sup} are better; higher I_5 is better.

Family	k	mean $\mathbf{E}_{\mathrm{IAE}}$	mean $\mathbf{E}_{\mathrm{sup}}$	$\mathbf{I_5}$
W+2DS+G1	13	7.7%	17.2%	37%
Bernstein-4	4	7.6%	17.3%	38%
Cure-Kumaraswamy	3	7.6%	17.2%	38%
Cure-Beta	3	7.6%	17.2%	38%
Bernstein-3	3	7.6%	17.5%	38%
PCA-3	3	7.7%	17.3%	37%
W+1DS	6	7.6%	17.2%	38%
PCA-4	4	7.6%	17.3%	38%
Exp-Weibull	3	7.7%	17.4%	37%
Weibull	2	8.1%	18.7%	31%
Burr	3	8.2%	19.0%	28%
Log-logistic	2	9.0%	20.9%	15%

One-at-a-time parameter sensitivity. Starting from a validation reconstruction IAE of 0.0103, we perturb each natural parameter in both directions and record the largest symmetric shift x for which the IAE increases by at most 5%. Table 28 reports both the absolute shift and the normalized shift x/IQR, where IQR is the empirical interquartile range of that parameter. The Burr coordinates are much more tightly identified than the residual amplitudes, especially the sinusoidal term.

λ-bootstrap sensitivity. We next isolate uncertainty in the video-specific scale coordinate λ. With all watch-style parameters frozen, we bootstrap the video-specific estimate 100 times and evaluate the induced validation-curve IAE. In Table 29, the median and 90th-percentile columns summarize the distribution of IAE across bootstrap samples for each audience.

Video/watch-style allocation. We next determine which Burr coordinates should be video-specific and which should be shared within a watch-style group. In the masks below, V denotes a video-specific coordinate and C a watch-style-specific coordinate, in the natural Burr parameter order (λ, c, k). Table 30 evaluates all principal allocations on validation transfer. The VCC allocation is best, meaning that only λ is video-specific while c and k are shared by watch-style group.

Holding the Burr base fixed to VCC, we then allocate the Gaussian and sinusoidal residual amplitudes. In Table 31, the two mask entries correspond to $(\alpha_{\mathrm{G20}}, \beta_{\mathrm{sine3}})$. Sharing both residual amplitudes within watch-style groups (CC) gives the best validation transfer result, although all four residual allocations are relatively close.

The resulting shortlist is then checked on the locked transfer test set. Table 32 reports transfer IAE together with C_5 and C_{10}, the fractions of test videos whose sup-norm error is at most 5% and 10%, respectively. The full VCC+CC allocation performs best. Changing only the residual allocation has a small effect, whereas making the Burr k coordinate video-specific (VCV+CC) reduces C_5 sharply. Fully video-specific (VVV+VV) and fully watch-style-specific (CCC+CC) representations transfer poorly.

Residual ablation and Gaussian width. We next isolate the contribution of the Gaussian and sinusoidal residual terms while keeping the Burr base fixed to VCC. In Table 33, k is the number of video-specific coordinates, n_{hyp} is the number of watch-style-specific coordinates, IAE and Sup

Table 22: KuaiLive-M3 complementary-exit family sweep with a GRU coordinate predictor.

Family	**mean $\mathbf{E}_{\mathrm{IAE}}$**	**mean $\mathbf{E}_{\mathrm{sup}}$**	**$\mathbf{I_5}$**	**# Params.**
Cure-Kumaraswamy	6.9%	17.3%	41%	37,449
Bernstein-3	7.1%	17.5%	39%	37,449
Cure-Beta	7.2%	17.4%	38%	37,449
Mod-Weibull	7.3%	18.5%	36%	37,449
Exp-Weibull	7.3%	18.5%	35%	37,449
Burr	7.3%	19.2%	36%	37,449
Bernstein-4	7.4%	17.7%	37%	37,516
Gen-Gamma	7.3%	19.1%	33%	37,449
Weibull	7.5%	19.2%	32%	37,382
Log-logistic	7.5%	19.7%	31%	37,382
PCA-3	46.0%	100.0%	0%	37,449
W+2DS	46.0%	100.0%	0%	37,918

Table 23: KuaiLive-M3 plugin occupancy for retained join families paired with a cure-Kumaraswamy GRU exit model.

Join family	O-IAE	O-sup	I_5	J-IAE	X-IAE	$X > J$	**# Params.**
W+2DS+G1	0.0722	0.1622	37.8%	0.0766	0.0694	1958	37,462
Bernstein-4	0.0721	0.1616	38.3%	0.0761	0.0694	39	37,453
Bernstein-3	0.0722	0.1634	37.6%	0.0763	0.0694	88	37,452
W+1DS	0.0724	0.1613	37.8%	0.0764	0.0694	2622	37,455
PCA-4	0.0727	0.1618	37.2%	0.0765	0.0694	2911	37,453
Cure-Beta	0.0730	0.1613	37.2%	0.0762	0.0694	657	37,452
Cure-Kumaraswamy	0.0730	0.1614	37.1%	0.0762	0.0694	679	37,452
Exp-Weibull	0.0732	0.1624	36.7%	0.0769	0.0694	3000	37,452
PCA-3	0.0733	0.1620	36.6%	0.0769	0.0694	3103	37,452

are the mean integrated absolute and sup-norm errors, and C_2, C_5, and C_{10} are the fractions of test videos with sup-norm error below the corresponding threshold.

Table 33 shows that G20 provides most of the residual improvement, increasing C_5 from 62.4% for plain Burr to 73.4%. Sine3 alone has little effect, but adding it on top of G20 raises C_5 further to 75.3%. The selected Gaussian width $\sigma = 0.12$ also performs better than the tested narrower and wider alternatives.

Residual-atom catalog. Finally, we report the residual search from which G20 and sine3 were selected. All Gaussian atoms in this search use $\sigma = 0.12$, and the search is performed on the 2,323 training+validation videos rather than on the locked transfer test set. In Table 34, Mask gives the best V/C allocation for each individual residual atom, while C_5 and IAE measure transfer quality under that allocation.

We then evaluate representative two-atom dictionaries. In Table 35, the two mask entries give the V/C allocation of the two listed residual atoms in order. G20 is the strongest singleton, while sine3 is comparatively weak by itself but becomes the strongest second atom when paired with G20.

Overall, the ablations support the selected decomposition: the Burr shape parameters are tightly constrained, λ is the coordinate that should remain video-specific, c and k should be shared by watch-style group, and most of the residual gain comes from the G20 component, with sine3 providing a smaller complementary improvement.

C.8 KuaiRec: Target-Audience Information Budget

The main KuaiRec transfer experiment uses the empirical target-audience watch-style mixture π_{big}. To measure how much target-domain information is actually required, we instead estimate this mixture from only m target-audience watches per video whose watch-style classes are known. We compare

Table 24: KuaiLive-M3 joint occupancy optimization. Val. O-sup is validation mean occupancy sup-norm error and is used for model selection.

Join family	O-IAE	O-sup	I_5	J-IAE	X-IAE	$X > J$	**# Params.**	**Val. O-sup**
Cure-Kumaraswamy	0.0558	0.1359	55.7%	0.0763	0.0764	757	37,452	0.1353
Cure-Beta	0.0557	0.1362	55.9%	0.0763	0.0765	710	37,452	0.1356
W+2DS+G1	0.0558	0.1362	55.9%	0.0763	0.0763	1641	37,462	0.1357
Exp-Weibull	0.0559	0.1366	55.3%	0.0770	0.0767	628	37,452	0.1356
PCA-3	0.0559	0.1364	55.6%	0.0769	0.0769	984	37,452	0.1358
W+1DS	0.0558	0.1364	56.0%	0.0764	0.0764	1365	37,455	0.1361
PCA-4	0.0558	0.1365	55.8%	0.0765	0.0764	1081	37,453	0.1361
Bernstein-4	0.0561	0.1376	55.1%	0.0761	0.0758	495	37,453	0.1370
Bernstein-3	0.0568	0.1398	54.1%	0.0763	0.0756	483	37,452	0.1391

Table 25: KuaiLive-M3 performance stratified by room duration.

Duration bucket	**Test fraction**	O-IAE	O-sup	I_5	J-IAE	X-IAE
$L < \mu - \sigma/2$	0.0%	–	–	–	–	–
$\mu - \sigma/2 \le L < \mu - \sigma/4$	37.2%	0.060	0.148	49%	0.070	0.072
$\mu - \sigma/4 \le L < \mu$	45.6%	0.051	0.129	60%	0.073	0.073
$\mu \le L < \mu + \sigma/4$	7.0%	0.051	0.127	62%	0.082	0.079
$\mu + \sigma/4 \le L \le \mu + \sigma/2$	2.7%	0.056	0.134	62%	0.091	0.087
$L > \mu + \sigma/2$	7.4%	0.058	0.138	57%	0.097	0.091
All test	100%	0.055	0.137	56%	0.075	0.075

this sampled estimate with three reference settings: the full target mixture, a uniform mixture, and reuse of the source-audience mixture π_{small}. Sampling experiments use five random seeds and report mean $\pm$ standard deviation.

In Table 36, IAE is the mean transfer integrated absolute error, while C_5 and C_{10} are the fractions of test videos whose sup-norm transfer error is at most 5% and 10%, respectively. The sampled rows use exactly m known-class target watches per video to estimate the mixture.

Table 36 shows that transfer quality improves smoothly as more target-audience mixture information becomes available. Reusing the source mixture performs poorly, confirming that the audience-composition shift is essential to the transfer problem. Even a uniform mixture provides a substantial improvement, while $m = 50$ target watches already outperforms that approximation. With $m = 250$, transfer IAE reaches 0.0197, within 0.001 of the full target-mixture result. Importantly, estimating π_{big} uses only watch-style labels and never uses the held-out KuaiRec-big retention trajectory; users without a known watch-style class are excluded from the mixture estimate.

C.9 KuaiRec: Number of Watch-Style Groups

We study the sensitivity of cross-audience transfer to the number K of behavioral watch-style groups. The selected Burr+G20+sine3 VCC+CC parameterization is kept fixed, and no new V/C mask search is performed for each K. The training population contains 7,176 users.

In Table 37, k is the number of video-specific coordinates, n_{hyp} is the number of watch-style-specific coordinates per group, and Total class coords is $K\, n_{\mathrm{hyp}}$. IAE is the mean transfer integrated absolute error, while C_5 and C_{10} are the fractions of test videos whose sup-norm transfer error is at most 5% and 10%, respectively. The final column reports the numbers of training users assigned to the resulting watch-style groups.

Table 37 shows that coarse partitions are insufficient: $K = 2$ and $K = 3$ substantially reduce transfer accuracy. In contrast, $K = 5, 8, 10$ form a stable regime, with IAE between 0.0181 and 0.0190 and C_5 between 75.3% and 77.2%. Thus the selected $K = 5$ is not a fragile choice; finer partitions provide only modest additional gains while increasing the number of class-specific coordinates.

Table 26: KuaiRec video-specific reconstruction on held-out validation videos. Lower IAE is better; higher C_τ is better.

Family	k	**Small IAE**	**Big IAE**	**Small C_5**	**Big C_5**	**Small C_2**	**Big C_2**
Weibull	2	0.0154	0.0165	95%	75%	6%	8%
Log-logistic	2	0.0168	0.0223	58%	29%	4%	7%
Burr	3	0.0140	0.0161	95%	76%	6%	8%
Exp-Weibull	3	0.0137	0.0120	95%	94%	10%	18%
Gen-Gamma	3	0.0137	0.0120	96%	93%	10%	19%
Mod-Weibull	3	0.0102	0.0092	95%	95%	37%	51%
Cure-Beta	3	0.0109	0.0094	94%	93%	15%	33%
Cure-Kumaraswamy	3	0.0110	0.0093	95%	95%	15%	32%
Bernstein-3	3	0.0273	0.0171	45%	86%	34%	44%
Bernstein-4	4	0.0194	0.0122	78%	89%	34%	61%
W+1DS	6	0.0049	0.0042	100%	98%	75%	78%
W+2DS	10	0.0031	0.0025	100%	100%	94%	95%
W+2DS+G1	13	0.0024	0.0021	100%	100%	98%	98%
W+3DS	14	0.0023	0.0020	100%	100%	99%	97%
W+2DS+G2	16	0.0020	0.0019	100%	100%	100%	99%
W+4DS	18	0.0020	0.0018	100%	100%	99%	99%
W+2DS+G3	19	0.0018	0.0017	100%	100%	100%	100%

Table 27: KuaiRec watch-style-mixture reconstruction on held-out validation videos. V denotes a video-specific coordinate and C a watch-style-specific coordinate.

Family	**Mask**	k	n_{hyp}	**Small IAE**	**Big IAE**	**Small C_5**	**Big C_5**	**Small C_2**	**Big C_2**
Exp-Weibull+resPC2	VCC+VC	2	3	0.0081	0.0102	99%	91%	23%	33%
Gen-Gamma+G20+ramp	VCC+CV	2	3	0.0085	0.0108	99%	92%	21%	23%
Burr+G20+sine3	VCC+CC	1	4	0.0090	0.0124	98%	89%	18%	17%
Gen-Gamma	VCC	1	2	0.0102	0.0130	99%	89%	14%	12%
Exp-Weibull	VCC	1	2	0.0102	0.0137	99%	89%	12%	9%
Burr	VCC	1	2	0.0103	0.0138	100%	89%	12%	8%
Weibull	VC	1	1	0.0118	0.0131	93%	91%	13%	11%
W+1DS	VCVCCC	2	4	0.0121	0.0113	87%	94%	12%	24%
Log-logistic	VC	1	1	0.0139	0.0237	94%	54%	5%	0%
Cure-Beta	CVV	2	1	0.0156	0.0232	75%	39%	1%	0%
Cure-Kumaraswamy	CVV	2	1	0.0179	0.0212	71%	67%	1%	2%
Mod-Weibull	CCV	1	2	0.0371	0.0317	6%	21%	0%	1%
Bernstein-3	VCV	2	1	0.0391	0.0246	44%	65%	0%	1%
Bernstein-4	CVVC	2	2	0.0412	0.0301	25%	74%	5%	6%

Table 28: One-at-a-time sensitivity of Burr+G20+sine3. x is the largest symmetric perturbation whose reconstruction IAE increases by at most 5%; x/IQR normalizes this shift by the empirical interquartile range of the parameter.

Parameter	**Maximum shift x**	$\mathbf{x}/\mathrm{IQR}$
λ	0.0059	0.0051
c	0.0150	0.0081
k	0.0503	0.031
α_{G20}	0.0068	0.27
β_{sine3}	0.0048	1.72

Table 29: Bootstrap sensitivity of the video-specific λ coordinate using 100 bootstrap samples.

Audience	**Median IAE**	**90th-percentile IAE**
KuaiRec-big	0.0103	0.0189
KuaiRec-small	0.0087	0.0119

Table 30: Validation transfer IAE for Burr-coordinate allocations. The three mask entries correspond to (λ, c, k), with V denoting video-specific and C watch-style-specific parameters.

Mask	Validation transfer IAE
VCC	0.0224
VCV	0.0299
CCV	0.0431
VVV	0.0716
CCC	0.1325

Table 31: Validation transfer IAE for the two residual-coordinate allocations. The mask entries correspond to $(\alpha_{\mathrm{G20}}, \beta_{\mathrm{sine3}})$.

Residual mask	Validation transfer IAE
CC	0.0187
CV	0.0192
VC	0.0199
VV	0.0201

Table 32: Locked-test comparison of Burr+G20+sine3 V/C allocations.

Mask	Transfer IAE	$\mathbf{C_5}$	$\mathbf{C_{10}}$
VCC+CC	0.0190	75.3%	98.7%
VCC+CV	0.0195	73.0%	98.8%
VCC+VC	0.0202	72.3%	98.7%
VCC+VV	0.0205	71.5%	98.6%
VCV+CC	0.0241	49.0%	96.3%
CCV+CC	0.0411	12.6%	66.2%
VVV+VV	0.0672	0.0%	22.7%
CCC+CC	0.1346	0.6%	19.7%

Table 33: Burr residual and Gaussian-width ablation on the locked transfer test set.

Model	Mask	k	n_{hyp}	IAE	Sup	$\mathbf{C_2}$	$\mathbf{C_5}$	$\mathbf{C_{10}}$
Burr only	VCC	1	2	0.0227	0.049	0.5%	62.4%	98.0%
Burr+G20	VCC+C	1	3	0.0195	0.043	3.3%	73.4%	98.7%
Burr+sine3	VCC+C	1	3	0.0226	0.049	0.6%	62.3%	98.3%
Burr+G20+sine3	VCC+CC	1	4	0.0190	0.042	5.2%	75.3%	98.7%
Burr+G20 ($\sigma = 0.04$)	VCC+C	1	3	0.0203	0.046	1.0%	68.8%	98.4%
Burr+G20 ($\sigma = 0.40$)	VCC+C	1	3	0.0200	0.045	1.7%	73.2%	98.6%
Burr+G20 ($\sigma = 0.80$)	VCC+C	1	3	0.0211	0.047	1.0%	69.2%	98.4%

Table 34: Best single residual atoms and their best V/C allocations.

Atom	Mask	$\mathbf{C_5}$	IAE
G20	C	75.3%	0.0192
sine2	C	72.9%	0.0204
sine1	C	67.5%	0.0214
G50	V	66.3%	0.0215
G80	C	64.8%	0.0228
sine3	V	64.3%	0.0225

Table 35: Representative two-atom residual dictionaries and their best V/C allocations.

Pair	**Mask**	$\mathbf{C_5}$	**IAE**
G20+sine3	CC	77.5%	0.0187
G20+sine1	CV	76.2%	0.0188
G20+sine2	CV	75.2%	0.0192
G50+sine3	VC	67.8%	0.0211
G80+sine3	VV	67.6%	0.0220

Table 36: KuaiRec transfer as a function of target-audience mixture information. Sampled-mixture rows report mean $\pm$ standard deviation over five seeds.

Mixture information	**IAE**	$\mathbf{C_5}$	$\mathbf{C_{10}}$
Full target mixture	0.0190	75.3%	98.7%
Uniform mixture	0.0259	56.9%	96.7%
Source (π_{small})	0.0671	0.1%	28.1%
$m = 10$	0.0347 ± 0.0005	$33.2\% \pm 1.8\%$	$84.9\% \pm 0.6\%$
$m = 25$	0.0263 ± 0.0004	$49.9\% \pm 0.9\%$	$95.2\% \pm 0.8\%$
$m = 50$	0.0228 ± 0.0003	$60.7\% \pm 0.6\%$	$97.4\% \pm 0.2\%$
$m = 100$	0.0211 ± 0.0001	$67.2\% \pm 1.0\%$	$98.3\% \pm 0.1\%$
$m = 250$	0.0197 ± 0.0002	$72.3\% \pm 1.0\%$	$98.5\% \pm 0.1\%$
$m = 500$	0.0194 ± 0.0001	$73.6\% \pm 0.8\%$	$98.6\% \pm 0.2\%$

Table 37: KuaiRec sensitivity to the number of watch-style groups.

K	k	n_{hyp}	**Total class coords**	**IAE**	$\mathbf{C_5}$	$\mathbf{C_{10}}$	**Training-user group sizes**
2	1	4	8	0.0413	3.5%	78.9%	4115, 3061
3	1	4	12	0.0295	32.2%	95.4%	2283, 1991, 2902
5	1	4	20	0.0190	75.3%	98.7%	1539, 1166, 650, 2032, 1789
8	1	4	32	0.0186	76.0%	98.7%	864, …, 430, …, 1168
10	1	4	40	0.0181	77.2%	98.8%	783, …, 214, …, 498