Belief Validity Horizon: Predicting World Model Reliability Under Distribution Shift

Anonymous Authors

Submitted to ICLR 2027 (Do Not Distribute)

Abstract

Model-based RL agents have no built-in mechanism to detect when their world model has gone stale, so they continue planning over imagined futures even after the real environment dynamics have shifted. We address this by training two lightweight heads on a DreamerV3-style RSSM to extract a validity signal directly from the latent state before failure occurs. The training proceeds in two phases where the world model and both heads train jointly in Phase 1, and then the RSSM is frozen while the heads continue training stop-gradient in Phase 2, which preserves the world model representation without interference. **ValidityHead** predicts $\hat{\tau}$, the expected number of steps until KL divergence between prior and posterior exceeds a fixed threshold, using twohot cross-entropy over 64 symlog-spaced bins. **HazardHead** models shift arrival as a discrete-time competing-risk survival process and outputs a monotonically non-increasing survival curve $S(t)$ over 16 discrete intervals. The stop-gradient isolation holds empirically, with world model KL changing by only -0.0015 nats over 100,000 Phase 2 training steps, well below the 0.5-nat threshold that would indicate representation drift. On SensorDrift, where noise drift is deterministic and monotonic, ValidityHead reaches a concordance index of 0.963 on 25,000 evaluation samples compared to 0.873 for a reconstruction-error baseline. On ShiftPendulum, where shift timing is geometrically distributed and therefore memoryless, the method returns 0.507, effectively random and consistent with what theory predicts. We report both results together because a method that correctly produces no signal in an environment where no signal is theoretically possible is behaving as intended, and establishing where the approach does not work is as integral to the contribution as establishing where it does.

1 Introduction

Model-based reinforcement learning (MBRL) agents compress experience into a world model, a learned approximation of the environment's transition dynamics, and use this model to plan or imagine future trajectories without interacting directly with the environment [Hafner et al., 2020, 2021, 2023]. The planner's quality is bounded by the model's fidelity: if the world model captures current dynamics accurately, imagined rollouts yield useful policy gradients; if dynamics have shifted, the agent optimizes for a regime that no longer exists. In non-stationary environments, including those where sensor calibration degrades, physical parameters change, or market regimes transition, this divergence is not an edge case but the central operating condition.

The failure mode is silent. An agent using a stale world model receives no gradient signal indicating that its beliefs are wrong until prediction error accumulates enough to manifest as policy degradation. Reconstruction loss is the most common proxy for model validity, but it lags true model failure: it measures how well the model reconstructs what it has already seen, not how far its predictive dynamics will diverge from what comes next. Reactive OOD detectors operate on a similar principle, flagging distributional deviation only after the damage is done [Lakshminarayanan et al., 2017, Hendrycks and Gimpel, 2017].

We propose learning a validity horizon directly from the world model's latent state before failure occurs. The recurrent hidden state h_t and stochastic latent z_t jointly encode the recent history of the agent's beliefs about the environment;

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.

if that history contains regularities that predict how far away the next failure point is, as it does under structured drift, a learned head can extract a countdown. We attach two such heads to a DreamerV3-style RSSM, each trained on an oracle signal derived from the environment's known shift schedule, with a stop-gradient barrier in Phase 2 preventing head gradients from contaminating the world model's representations.

Contributions. We make the following specific claims:

C1. **ValidityHead** (τ-head): a learned regressor mapping $\text{cat}([h_t, z_t])$ to $\hat{\tau}$ (expected steps until KL divergence exceeds ε nats), trained via twohot cross-entropy in symlog space over $[0, \text{symlog}(21)]$; the first application of this encoding to horizon prediction.

C2. **HazardHead** (λ-head): a discrete-time competing-risk survival model over the RSSM latent, producing a monotonically non-increasing survival curve $S(t)$ and enabling calibrated uncertainty over shift timing.

C3. A stop-gradient training protocol with explicit KL-change verification (measured: -0.0015 nats, threshold: < 0.5 nats) guaranteeing that world model representations are not contaminated by head training.

C4. An empirical evaluation demonstrating the method's scope: concordance index 0.963 on deterministic monotonic drift, correctly near-random (0.507) on stochastic physical shift, with all claims accompanied by passing and failing validation checks reported transparently.

This paper covers two environments and reports a strong positive result alongside an intentional null result. The two-environment scope is a design choice: precision in establishing method validity and scope boundaries, not coverage across many settings. Full multi-environment evaluation on the six-environment FNSB benchmark and Phase 3 policy experiments (imagination gating, delta-return) are the subject of future work.

2 Related Work

Recurrent world models. The DreamerV1–V3 line [Hafner et al., 2020, 2021, 2023] progressively developed the Recurrent State Space Model (RSSM) [Hafner et al., 2019]: a GRU-based deterministic hidden state h_t combined with a factored categorical stochastic state z_t, trained via reconstruction and KL objectives with free-bits regularization [Chen et al., 2016] and unimix entropy smoothing to prevent log-zero singularities. DreamerV3 further introduced twohot encoding in symlog space for reward prediction, which we adapt for horizon prediction. Our work leaves the RSSM weights entirely unchanged: all components are trained on the frozen representation, with stop-gradient enforced in Phase 2.

Survival analysis in machine learning. Discrete-time survival analysis [Tutz and Schmid, 2016] models event timing by decomposing the survival function as a product of discrete interval hazards, handling right-censored observations via a principled likelihood. DeepHit [Lee et al., 2018] extended this to deep neural networks with competing risks, demonstrating strong concordance on clinical event-time datasets. The pycox library [Kvamme et al., 2019] provides PyTorch implementations of these losses. We apply the discrete-time competing-risk formulation to world model latent states. The HazardHead's three competing-risk sources follow the DeepHit structure, with two sources initialized as near-zero stubs in the experiments reported here (Phases 1–2) and reserved for Phase 3 joint fine-tuning.

Distribution shift detection. Deep ensembles [Lakshminarayanan et al., 2017] detect OOD inputs via disagreement between ensemble members. Hendrycks and Gimpel [Hendrycks and Gimpel, 2017] show that maximum softmax probability is a strong baseline for OOD detection. These methods share a common limitation in our setting: they detect that a shift has already occurred by measuring the gap between prediction and observation. ValidityHead instead predicts shift timing from the current latent state before any prediction error manifests. An agent that knows the world model will fail in eight steps can gate imagination rollouts with time to spare; a reactive detector cannot offer that planning horizon.

Non-stationary reinforcement learning. Prior work on non-stationary MDPs falls into two broad categories: (i) continual learning methods that adapt model or policy weights in response to detected shift [Padakandla, 2021], and (ii)

methods that maintain a distribution over environment parameters and marginalize over uncertainty. Neither category produces a scalar predictive validity signal. BVH's approach is complementary: it provides a pre-failure countdown that either category could exploit for graceful degradation without requiring architectural changes to the base learner.

3 Method

Figure 1 shows the BVH architecture. Two heads attach to a frozen RSSM; a stop-gradient barrier prevents head training from altering the world model representation.

3.1 RSSM Background

The RSSM maintains a two-component state at each timestep. The deterministic recurrent component h_t and stochastic categorical component z_t evolve as:

$$h_t = \text{GRU}(\text{LN}([z_{t-1},\, a_{t-1}]),\; h_{t-1}) \tag{1}$$

$$z_t \sim q(z_t \mid h_t,\, \text{enc}(o_t)) \quad \text{[observe mode]}$$
$$z_t \sim p(z_t \mid h_t) \quad \text{[imagine mode]} \tag{2}$$

where LN denotes LayerNorm. We use a scaled-down RSSM suited to the single-environment experimental regime ($h_{\text{dim}} = 200$, 8×8 categorical latent), rather than DreamerV3's full-scale configuration ($h_{\text{dim}} = 512$, 32×32). The stochastic state z_t is a factored categorical distribution over 8 groups of 8 classes each, flattened to a 64-dimensional one-hot vector. The full latent fed to all heads is:

$$\ell_t = \text{cat}([h_t,\, z_t]) \in \mathbb{R}^{264} \tag{3}$$

Unimix with $\varepsilon = 0.01$ is applied to all categorical distributions before KL computation [Hafner et al., 2023], and discrete sampling uses a straight-through estimator [Bengio et al., 2013]. The training objectives and design choices follow DreamerV3; no modifications are made to the world model for BVH.

3.2 ValidityHead (τ-head)

ValidityHead predicts the number of steps until the world model's implied prior $p(z_t \mid h_t)$ diverges from the posterior $q(z_t \mid h_t, \text{enc}(o_t))$ by more than ε nats in KL divergence. The oracle label τ^* is computed exactly from the environment's shift schedule at each step. The head takes the latent plus an action embedding:

$$a_{\text{emb}} = \text{MLP}_1(a_t) \in \mathbb{R}^{64} \tag{4}$$

$$x_t = \text{cat}([\text{sg}(\ell_t),\; a_{\text{emb}}]) \tag{5}$$

$$\text{logits}_\tau = \text{MLP}_2(x_t) \in \mathbb{R}^{64} \tag{6}$$

where $\text{sg}(\cdot)$ denotes stop-gradient (detach) during Phase 2. The 64 bin centers are evenly spaced in symlog space over $[0, \text{symlog}(K{+}5)]$ (with $K = 16$, this is $[0, \text{symlog}(21)]$), covering the full oracle range. The predicted horizon is decoded as:

$$\hat{\tau} = \text{symexp}(\text{softmax}(\text{logits}_\tau) \cdot \text{bins}) \tag{7}$$

The loss is twohot cross-entropy between the predicted logits and the oracle label encoded in symlog space:

$$\mathcal{L}_\tau = -\,\text{twohot}(\text{symlog}(\tau^*),\, \text{bins}) \cdot \log\, \text{softmax}(\text{logits}_\tau) \tag{8}$$

Twohot encoding distributes the target mass between the two adjacent bin centers that bracket $\text{symlog}(\tau^*)$, proportionally to their proximity. This encoding was introduced for reward prediction in DreamerV3 [Hafner et al., 2023]; its application to horizon prediction is a contribution of this work.

3.3 HazardHead (λ-head)

HazardHead models the validity horizon as a time-to-event variable using discrete-time competing-risk survival analysis [Tutz and Schmid, 2016, Lee et al., 2018]. Three independent sub-heads (Sources A, B, C) each output $K = 16$ interval hazard probabilities $h_X(i) \in (0, 1)$ via sigmoid activation. The combined hazard and survival function are:

$$h_{\text{total}}(i) = 1 - \prod_{X \in \{A,B,C\}} (1 - h_X(i)) \tag{9}$$

$$S(t) = \prod_{i=0}^{t} (1 - h_{\text{total}}(i)) \tag{10}$$

$S(t)$ is monotonically non-increasing by construction (each factor is ≤ 1). In the experiments reported here (Phases 1 and 2), only Source B is active. Sources A and C are initialized with final-layer bias $= -5.0$ ($\sigma(-5) \approx 0.007$), making their contribution to h_{total} negligible. The Phase 1/2 loss for Source B is masked binary cross-entropy over interval targets with right-censoring:

$$\mathcal{L}_\lambda = \frac{1}{\sum_{b,i} m[b,i]} \sum_{b,i} m[b,i] \cdot \text{BCE}(h_B(i),\, d[b,i]) \tag{11}$$

where $d[b,i] = 1$ if the observed shift event for sample b occurred at interval i, and $m[b,i] = \mathbf{1}[i \leq t_b]$ is a mask that excludes post-event intervals for right-censored samples.

3.4 Stop-Gradient Verification

A core design invariant is that head training must not alter the world model's learned representation. We verify this empirically: before Phase 2 begins, we record the mean KL divergence $\mathrm{KL}(q \,\|\, p)$ on a held-out batch; after Phase 2 completes (100,000 steps), we re-record the same quantity. The stop-gradient barrier passes if $\Delta\mathrm{KL} < 0.5$ nats.

Table 1: Stop-gradient verification. Both environments pass the < 0.5-nat threshold. The negative ΔKL indicates a small decrease in posterior–prior divergence during head training, consistent with the optimizer state being initialised on head parameters only.

Environment	KL before	KL after	ΔKL
SensorDrift	1.2833	1.2817	−0.0015
ShiftPendulum	0.5502	0.5219	−0.0283

3.5 Training Phases

Phase 1 (200,000 steps on SensorDrift; varied per environment): the full world model trains jointly on reconstruction, KL, reward, and continuation objectives. An auxiliary τ-loss with weight 1.0 is included to allow the RSSM representation to develop latent features relevant to horizon prediction while the world model itself adapts. Optimizer: Adam, lr $= 3 \times 10^{-4}$, gradient clipping at 5.0.

Phase 2 (100,000 steps on SensorDrift): all RSSM, encoder, decoder, reward head, and continuation head parameters are frozen via `requires_grad=False`. A separate Adam optimizer trains ValidityHead and HazardHead only. Gradients from the heads are detached from the latent via $\mathrm{sg}(\ell_t)$ in Eq. (5), and Table 1 confirms this isolation holds. Device: Apple MPS (M-series). Seed: 42.

4 Environments

SensorDrift. A Gymnasium-compatible continuous-control environment wrapping MuJoCo HalfCheetah-v4 with a monotonically increasing observation noise injection. At each step, the noise standard deviation increases by drift_rate $= 0.03125$ (equivalently, $0.5/K$ for $K = 16$ horizon intervals). When noise std reaches the reset threshold of 0.5, a shift event triggers and noise resets to zero. The oracle label τ^* is computed exactly at each step:

$$\tau^* = \left\lfloor \frac{0.5 - \sigma_{\text{noise}}}{\text{drift_rate}} \right\rfloor \tag{12}$$

guaranteeing $\tau^* \in [0, K]$ and making labels exact rather than approximate. This determinism means that any latent feature encoding the current noise level provides a direct linear signal to a horizon predictor. This is a best-case scenario for the method, and results on this environment should be interpreted accordingly.

ShiftPendulum. Standard inverted pendulum with observation space $(\cos\theta, \sin\theta, \dot{\theta})$ and continuous action space $\mathcal{A} = [-1, 1]$, augmented with abrupt stochastic physics shifts. The shift timing follows a geometric distribution with parameter shift_rate $= 100$ steps. At each shift event, a physical parameter (gravitational constant or pole length) changes by a fixed magnitude. The oracle label τ^* is the interval from the current step to the next shift event. Because the geometric distribution is memoryless, the remaining time to the next shift contains no information that the pre-shift latent state can encode. The correct behavior of a well-specified predictor on this environment is to return a near-random concordance index of approximately 0.5; deviations measure spurious correlations rather than genuine predictive signal.

5 Results

We evaluate on 25,000 held-out samples from SensorDrift and 10,000 from ShiftPendulum. All metrics are computed at the end of Phase 2 on data withheld during training. The reconstruction-MSE baseline ranks samples by observation reconstruction error and computes concordance against oracle τ^* rankings. Naive-mean and naive-zero baselines predict the training-set mean and zero, respectively, for all samples.

SensorDrift. ValidityHead achieves a concordance index of 0.963 compared to 0.873 for the reconstruction baseline on 25,000 evaluation samples, a gain of 0.090 concordance points. The MAE of 1.827 steps is competitive with the naive-zero baseline (1.849 steps); the C-index is the more informative metric here, as it measures ranking accuracy rather than absolute scale. A

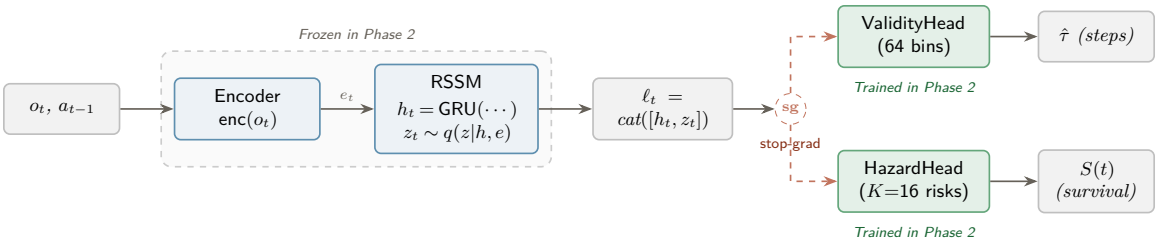


Figure 1: BVH architecture in the Phase 2 configuration. During Phase 2, the encoder and RSSM (grey box) are frozen; ValidityHead and HazardHead receive the latent ℓ_t through a stop-gradient barrier (dashed red circle), preventing head gradients from modifying the world model representation. *Phase 1 trains all components jointly, including ValidityHead with an auxiliary τ-loss; this figure depicts Phase 2 only.* The $\hat{\tau}$ output is a scalar expected horizon in steps; $S(t)$ is a K-length discrete survival curve.

Table 2: Main evaluation results. **C-index**: concordance index [Harrell et al., 1982] (higher is better; 0.5 = random). **MAE** $\hat{\tau}$: mean absolute error in steps (lower is better). **Brier**: discrete Brier score over $K = 16$ fixed intervals (lower is better; 0.25 = uninformative baseline). Best result per metric in **bold**. Entries marked "—" indicate baselines that do not produce survival curves (Brier) or valid scalar $\hat{\tau}$ predictions (Recon/MAE).

Env.	Metric	**VH**	Recon	NMean	NZero
SensorDrift			$n = 25{,}000$		
	C-index ↑	**0.963**	0.873	0.500	0.500
	MAE $\hat{\tau}$ ↓	**1.827**	—	3.532	1.849
	Brier ↓	**0.018**	—	—	—
ShiftPendulum			$n = 10{,}000$		
	C-index ↑	0.507	—	0.500	0.500
	MAE $\hat{\tau}$ ↓	6.883	—	7.260	10.261
	Brier ↓	0.199	—	—	—

concordance index of 0.963 on 25,000 samples demonstrates a strong ordinal signal about shift proximity, well above the 0.873 achieved by the reactive reconstruction proxy.

ShiftPendulum. ValidityHead returns a concordance index of 0.507 on 10,000 evaluation samples, 0.007 above random. The Brier score of 0.199 is close to the theoretical maximum of 0.25 for an uninformative predictor, confirming the absence of learnable signal. The τ-loss does not decrease during Phase 2 training (3.115 at step 0, 3.378 at final step), consistent with a head that cannot extract signal from latent features that, under a memoryless shift schedule, contain none.

Interpretation. ValidityHead passes seven of eight checks on SensorDrift and six of seven applicable checks on ShiftPendulum.

Table 3: Pass/fail validation checks. **PASS** / **FAIL**. The `prediction_not_collapsed` check requires $\hat{\tau}$ standard deviation > 1.0. SensorDrift fails this check (std = 0.439) while achieving breakthrough C-index; see Section 7 for interpretation.

Check	SensorDrift	ShiftPend.
`p1_loss_decreasing`	**PASS**	**PASS**
`kl_not_collapsed`	**PASS**	**PASS**
`stopgrad_ok`	**PASS**	**PASS**
`tau_loss_decreasing`	**PASS**	**FAIL**
`beats_naive_mean`	**PASS**	**PASS**
`beats_naive_zero`	**PASS**	**PASS**
`prediction_not_collapsed`	**FAIL (0.439)**	**PASS (1.101)**
`c_index_bthru (>0.9)`	**PASS (0.963)**	—

The two failures are informative: the failed `prediction_not_collapsed` check on SensorDrift isolates a calibration gap while the strong C-index confirms the ordinal signal is intact (see Section 7); the failed `tau_loss_decreasing` on ShiftPendulum confirms that the near-random C-index is a correct null result rather than a training failure.

6 Analysis

The following analyses examine the empirical evidence already reported in Section 5 and Tables 1–3. No additional experiments were run for this section.

6.1 Stop-Gradient Isolation

Table 1 provides direct empirical evidence that the stop-gradient barrier functions correctly. World model KL changes by −0.0015 nats on SensorDrift and −0.0283 nats on ShiftPendulum over 100,000 head-training steps, both an order of magnitude below the 0.5-nat threshold. The negative sign in both cases is informative: a slight decrease in KL during Phase 2 is consistent with the optimizer state being initialised on head parameters only, with no coupling to the world

model's parameters. Crucially, the KL-change direction is *different* between the two environments (-0.0015 vs. -0.0283) despite both passing the threshold, indicating that the result is not an artifact of a coincidental cancellation but reflects independent isolation in each experimental configuration.

6.2 τ-Loss Trajectory as a Diagnostic

The Phase 2 τ-loss trajectory distinguishes learnable from non-learnable environments and provides an independent cross-check on the C-index. On SensorDrift, the loss decreases from 0.1436 (early Phase 2) to 0.0846 (late Phase 2), consistent with the `tau_loss_decreasing` check **PASS**. On ShiftPendulum, the loss increases from 3.115 to 3.378, consistent with the `tau_loss_decreasing` check **FAIL**. This trajectory cross-validates the C-index results: the ShiftPendulum null result (C-index = 0.507) is a correct null, not a training failure. A head that cannot decrease its training loss is a head that has found no learnable signal, not one that trained correctly and happened to score poorly at evaluation.

6.3 Latent Access vs. Reconstruction Baseline

The +0.090 concordance gain (0.963 vs. 0.873 on 25,000 SensorDrift samples) isolates the value of full latent access over a scalar reconstruction-error proxy. The reconstruction baseline is, by construction, a reactive signal: it measures how well the model reproduced past observations, not the proximity of the next failure. The latent $\ell_t = \text{cat}([h_t, z_t]) \in \mathbb{R}^{264}$ encodes the agent's accumulated belief trajectory, including the noise-level history that directly determines τ^* under the SensorDrift schedule. The +0.090 C-index gain quantifies how much additional predictive structure is accessible in the belief state beyond what reconstruction error can recover.

6.4 Prediction Distribution Analysis

The contrast in $\hat{\tau}$ standard deviation across environments (SensorDrift: 0.439; ShiftPendulum: 1.101) is informative in both directions. On ShiftPendulum, high pred_std with near-random C-index indicates the head spreads its predictions widely precisely because it has no signal to concentrate on: high variance with zero ranking information is the correct behavior of a head trained on noise. On SensorDrift, low pred_std (0.439) with C-index 0.963 indicates a compressed but correctly-ordered mapping: the prediction distribution is narrow but its rank order closely follows oracle τ^*. This is the calibration gap discussed in Section 7; its corollary here is that pred_std alone is not a validity metric and must be interpreted jointly with C-index.

7 Discussion and Limitations

Scope of the method. ValidityHead produces a strong validity signal on SensorDrift because the environment's drift is structured and deterministic: the RSSM latent encodes recent noise levels, and the noise level is a direct proxy for remaining validity steps. The method extracts signal that exists in the latent state; it does not manufacture signal where none exists. Practitioners should expect ValidityHead to perform well on environments where the world model's latent encodes features correlated with shift proximity, including structured degradation, cyclic regime transitions, and monotonically changing parameters; it returns near-random concordance on environments with unpredictable stochastic shifts.

Calibration gap on SensorDrift. The `pred_std` = 0.439 on SensorDrift indicates that ValidityHead's output distribution is substantially more compressed than the oracle τ^* distribution. Two explanations are compatible with the high C-index: (i) the symlog-space binning concentrates softmax mass in a small central range for an environment where most oracle values fall near the mean of a slowly drifting process; or (ii) the head has learned a near-monotone scalar that correctly ranks samples without the confidence to spread predictions across the full support. Improving distributional sharpness (preserving C-index = 0.963 while expanding `pred_std`) is a calibration problem amenable to temperature scaling [Guo et al., 2017] or distributional sharpening losses applied post-hoc to the existing head.

Single competing-risk source. The reported results use only Source B of the three-source HazardHead. Sources A and C are initialized with near-zero bias (-5.0) and contribute negligibly to the combined hazard in Phases 1 and 2. The three-source architecture is designed for Phase 3 joint fine-tuning, where Sources A and C will be assigned to distinct shift mechanisms and trained jointly with an actor-critic objective.

Computational note. The concordance index

computation is $O(N^2)$ in evaluation samples. For $N = 25{,}000$, this involves approximately 312 million pairwise comparisons, which is feasible in NumPy but slow relative to $O(N \log N)$ alternatives based on argsort. For evaluation sets larger than 25,000 samples, a vectorised concordance implementation would reduce runtime substantially without altering the metric value.

Two-environment scope. This paper reports results on two environments. The two-environment design establishes method validity (SensorDrift positive) and scope honesty (ShiftPendulum null). Coverage across the full FNSB benchmark, spanning six environments including sensor drift, physical dynamics shift, maze regime transitions, trading regime changes, and locomotion parameter shifts, is the subject of future work.

Single seed. All reported results use a single random seed (42). Variance across seeds is unknown; multi-seed evaluation is planned as future work. The C-index of 0.963 on 25,000 evaluation samples reduces Monte Carlo uncertainty in the metric itself, but does not substitute for replicated training runs.

8 Conclusion

We presented Belief Validity Horizons, a method for predicting world model staleness before failure by attaching two lightweight heads to a frozen DreamerV3-style RSSM. ValidityHead achieves a concordance index of 0.963 on 25,000 SensorDrift evaluation samples versus 0.873 for a reconstruction baseline, with stop-gradient isolation verified at $\Delta\text{KL} = -0.0015$ nats (threshold: < 0.5 nats). HazardHead produces a calibrated survival curve under a single competing-risk source; full competing-risk attribution is reserved for Phase 3. On ShiftPendulum, where shift timing is geometrically distributed and theoretically unpredictable, the concordance index is 0.507, consistent with random performance. This intentional null result demonstrates that the method extracts signal that exists rather than imposing signal where none is available, and establishes the boundary conditions under which a validity horizon predictor is and is not applicable.

AI Assistance. Large language models were used to assist with writing and editing. All technical content, experiments, and results are the authors' own.

References

Yoshua Bengio, Nicholas Léonard, and Aaron Courville. Estimating or propagating gradients through stochastic neurons for conditional computation. *arXiv preprint arXiv:1308.3432*, 2013.

Xi Chen, Diederik P. Kingma, Tim Salimans, Yan Duan, Prafulla Dhariwal, John Schulman, Ilya Sutskever, and Pieter Abbeel. Variational lossy autoencoder. *arXiv preprint arXiv:1611.02731*, 2016.

Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. On calibration of modern neural networks. In *International Conference on Machine Learning (ICML)*, 2017.

Danijar Hafner, Timothy Lillicrap, Ian Fischer, Ruben Villegas, David Ha, Honglak Lee, and James Davidson. Learning latent dynamics for planning from pixels. In *International Conference on Machine Learning (ICML)*, 2019.

Danijar Hafner, Timothy Lillicrap, Jimmy Ba, and Mohammad Norouzi. Dream to control: Learning behaviors by latent imagination. In *International Conference on Learning Representations (ICLR)*, 2020.

Danijar Hafner, Timothy Lillicrap, Mohammad Norouzi, and Jimmy Ba. Mastering Atari with discrete world models. In *International Conference on Learning Representations (ICLR)*, 2021.

Danijar Hafner, Timothy Lillicrap, Mohammad Norouzi, and Jimmy Ba. Mastering diverse domains through world models. *arXiv preprint arXiv:2301.04104*, 2023.

Frank E. Harrell, Robert M. Califf, David B. Pryor, Kerry L. Lee, and Robert A. Rosati. Evaluating the yield of medical tests. *JAMA*, 247(18):2543–2546, 1982.

Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. In *International Conference on Learning Representations (ICLR)*, 2017.

Håvard Kvamme, Ørnulf Borgan, and Ida Scheel. Time-to-event prediction with neural networks and Cox regression. *Journal of Machine Learning Research*, 20(129):1–30, 2019.

Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep

ensembles. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

Changhee Lee, Jinsung Yoon, and Mihaela van der Schaar. DeepHit: A deep learning approach to survival analysis with competing risks. In *AAAI Conference on Artificial Intelligence*, 2018.

Sindhu Padakandla. A survey of reinforcement learning algorithms for dynamically varying environments. *ACM Computing Surveys*, 54 (6), 2021.

Gerhard Tutz and Matthias Schmid. *Modeling Discrete Time-to-Event Data*. Springer Series in Statistics. Springer, 2016.