AIMS Mathematics, 11(3): 6804–6833.
DOI: 10.3934/math.2026281
Received: 04 January 2026
Revised: 06 February 2026
Accepted: 09 March 2026
Published: 17 March 2026

https://www.aimspress.com/journal/Math

Research article

Process-augmented lifetime distributions: A generative Bayesian framework for latent structure inference in survival analysis

Xu Liu[1] **and Xufeng Niu**[2,*]

[1] Department of Statistics and Mathematics, Shandong University of Finance and Economics, No. 7366, East Second Ring Road, Jinan City,Shandong Province, China

[2] Department of Statistics, College of Art and Science, Florida State University, 117 N. Woodward Ave., Tallahassee, FL 32306-4330, United States of America

* **Correspondence:** Email: liuxu3298@gmail.com.

Abstract: In survival analysis, the predictive accuracy of lifetime distributions is frequently compromised since the continuous observation assumption is violated, resulting in discretized covariate data. Conventional survival analysis estimators typically address this limitation by treating the observed measurement as the true state or by assuming a mutually independent error structure, *complicating the functional predictor* that systematically induces discrepancy and obscures particle risk heterogeneity. In this article we propose an alternative Bayesian hierarchical framework, the process-augmented lifetime distribution, which explicitly models the representation gap between the true latent exposure and the observed measurement as a structured, recoverable stochastic process. We construct a generative chain that decomposes the hazard function into an explicit channel spanned by the measurement basis and an orthogonal implicit channel that captures the unobserved variance. This spectral decomposition ensures the identifiability of the latent signal without requiring external validation data by enforcing rigorous orthogonal constraints. We further develop a process augmentation scheme that introduces an intermediate auxiliary variable, transforming the intractable non-Gaussian likelihood into a conjugate structure applicable to an exact Gibbs sampler. This computational formulation permits the derivation of novel posterior functional measurements, which partition the predictive variance into irreducible failure noise and reducible measurement uncertainty. We propose a simulation study and an empirical analysis of high-frequency market stability to confirm that the proposed estimator successfully corrects discrepancy error and resolves intra-bin heterogeneity, providing a robust approach for quantifying structural decision risk in systems subject to severe observation imperfections.

Keywords: Bayesian survival analysis; process decomposition; process-augmented lifetime distribution; accelerated failure time; uncertainty quantification; exact Gibbs sampler
Mathematics Subject Classification: 62C10, 62N02, 62P20

1. Introduction

Time-to-event outcomes constitute a fundamental class of data in applied science, serving as the foundation for reliability engineering, biomedical survival analysis, and environmental risk assessment. Among this research, the analyst's target is not just concentrated on the binary classification of whether a failure event occurs; rather, the theoretical framework is to characterize the temporal evolution of risk, quantifying when a system is likely to fail and determining how exogenous stresses or intrinsic properties affect the lifetime distribution. Survival analysis provides the principled statistical module for this target, emphasizing the stochastic evolution from a well-defined origin time, such as the credit scoring [32], the diagnosis of a behavior [3], or the initiation of a financial stress test [28], to a terminal event of interest. While rigorously accommodating the incomplete observation structures inherent to longitudinal studies, most of them are notably right-censoring [30]. One practical difficulty in reliability studies is that, by the conclusion of the observation window, a non-negligible proportion of units will remain functioning, whether they will be mechanical components or financial subjects. Their event times are thus only partially observed, creating a right-censoring structure that cannot be appropriately predicted by standard regression or classification pipelines without explicit probabilistic modeling [30]. Consequently, survival analysis has become the major approach for learning event-time dynamics from censored data, allowing researchers to estimate critical reliability metrics such as the survival probability function, the mean residual life, and the time-dependent hazard rate [6].

However, the fidelity of these lifetime distributions is fundamentally constrained by the quality of the covariate data. In modern reliability engineering and prognostics, particularly within behavioral scoring and financial risk assessment, the primary conceptual deficit is often not just stochastic measurement noise, but a more structural phenomenon which we term representation loss [4, 22, 32]. When the recorded covariates (such as discretized health scores, administrative credit ratings) are low-fidelity measures for the complex, continuous latent states actually drive the failure dynamic. For instance, in analyzing technology credit scoring, the true tendency for failure is a continuous, high-dimensional process; yet, researchers often observe only aggregated usage cycles [38]. Similarly, in behavioral survival studies, a subject's true physiological or financial latent factor is considered continuous, but it is practically recorded via rounded indices [3]. Also, in quantitative finance, the intrinsic stability of the trading market is driven by continuous high-frequency microstructure noise, yet analysts and scholars often rely on integrated daily volatility regimes to predict crash events [10]. In such cases, the classical additive noise assumption in the signal noise module where the observed covariate is treated as the true exposure plus random Gaussian error, becomes insufficient because it ignores the structural gap between the complex reality of the risk factor and its compressed estimation [11, 19, 34]. We introduce representation loss as a specific form of discrepancy error that standard survival models fail to capture, leading to the hierarchical discrepancy work to amplify the underlying signal in noise [7, 8].

The theoretical support of survival analysis, formed by the survival and hazard functions, provides the intricate functional solutions to address the prediction obstacle, provided and assumed that the noise can be adequately extended [39]. Let T denote a continuous event time and C a censoring time, with observed time $Y = \min(T, C)$ and censoring indicator $\delta = \mathbf{1}\{T \leq C\}$. The survival function $S(t) = \Pr(T > t)$ summarizes the probability of a unit functioning beyond time t, while the hazard function characterizes the instantaneous failure rate conditional on survival up to t [30]. These

methods enable both descriptive and model-based analysis, supporting likelihood-based inference that naturally capture information asymmetry of censoring [6]. In many real-world scientific problems, time is recorded in discrete units (e.g., duty cycles, months of operation) [40], making discrete-time survival models operationally convenient. For instance, discrete-time formulations allow for a particle decomposition of risk that bridges the gap between continuous reliability theory and observational data constraints [9]. The flexibility to model in either continuous or discrete time is important for modern survival data analysis where measurement frequency, sensor sampling rates, and operational constraints differ substantially across engineering and biological domains.

Despite the sophistication of previous survival modeling schemes, spanning nonparametric estimators like the Kaplan–Meier [24], parametric accelerated failure time (AFT) models [32, 35], and the semiparametric Cox proportional hazards (PH) model [12] which rely on assumptions that can be restrictive in complex scientific applications. A critical limitation for those approaches is that they assume that the covariates entering the hazard function are precise measurements of risk-driving. In reliability risk assessment, one may question whether the observed stress factors fully capture the unobserved micro-structural driver leading to failure, or whether the covariate-to-risk relationship is adequately captured by a simple linear predictor while the underlying data properties imply non-linear saturation effects. Such concerns have motivated a substantial literature on improving classical survival formulations. Notable extensions include the mixture cure model, which explicitly allows for a subpopulation of immune or nonsusceptible units (e.g., components with zero manufacturing defects that will not fail due to a specific mode), thereby relaxing the "everyone fails eventually" premise [38]. Cure-model ideas have been further developed to incorporate macro-environmental factors and time-varying covariates, reflecting the reality that external operating conditions can shift baseline risk and modulate covariate impacts [14, 36]. Prediction-driven cure modeling methods that improve effect consistency provide additional modules for enhancing stability and interpretability in real-world applications (e.g., credit censoring) [21, 41]. Parallel strands of work also proposed flexible functional forms, such as spline-based hazard models and generalized additive structures, to capture nonlinearities and heterogeneities that linear predictors may miss [13,29]. To summarize, these extensions reflect the working endeavor to align survival models more closely with the complexity of modern data.

While these deterministic methods increase the flexibility of the hazard function for better prediction performance, they do not fundamentally check the inherent relationship between the observed data and the latent risk process. They effectively assume that if the function is complex enough, it can compensate for the fact that observation is a poor measure for the true exposure. We argue that this is not adequate in reasoning: If the measurement process involves structural information loss (e.g., quantization or censoring of the covariate), no amount of flexibility in the hazard function can recover the lost signal without an explicit generative model for the observation. To address this, we turn to the discrepancy error exploration which was considered in the spatial process. Recent work by [7] introduced a framework for hierarchical models where they assumed errors are correlated with the latent process, challenging the traditional assumption of independence between signal and noise. They demonstrated that by modeling the discrepancy error that determines the difference between the latent truth and the observed data as a structured random field, one can leverage the correlation structure to deconfound the covariance matrix between signal and noise, thus improving prediction and inference. Furthermore, [8] extended this into a deep hierarchical generalized transformation (DHGT) framework,

using a telescoping sum of discrepancy latent processes to construct a deep generative hierarchy that bridges the gap between coarse observations and fine-scale latent processes.

Conceptually, this work aligns with the broader methodological trend of relaxing overly strict assumptions in survival analysis. For example, previous literatures consider admitting cure fractions [38], allowing time-varying effects [14, 22], and improving functional flexibility [13, 29]. However, we target a distinct and practically consequential source of misspecification in survival analysis: measurement discrepancy. In this article, we propose a process-augmented lifetime distribution framework to address the discrepancy measure and relevant uncertainty quantification. The key idea is to consider the incomplete state space where we treat the observed covariate data not as fixed constants, but as imperfect measurements of an underlying state, propagating this uncertainty into the hazard model through an explicit augmentation layer. By explicitly modeling the discrepancy between the observation and the true exposure as a structured, recoverable component, our approach deviates from standard measurement error models which typically assume the error process is mutually independent of the signal.

The first contribution in this article is that we build a generative chain for representation loss: We formulate a hierarchical latent-process model, $X \to r \to Z$, which decomposes the hazard function into an explicit channel (spanned by the measurement instrument) and an orthogonal implicit channel (capturing the representation gap). We show that this decomposition prevents confounded estimation by enforcing orthogonality between the observable basis and the latent discrepancy, thereby ensuring the detection of latent structural risk factors without requiring external validation data.

The second contribution is that we construct the exact Bayesian inference via process augmentation: Analogous to computational strategies in spatial statistics [7], we augment the state space with an intermediate representation variable r following a log-normal distribution. This transformation converts the intractable non-Gaussian likelihood of the underlying process into a conjugate hierarchical structure, allowing for the derivation of an exact Gibbs sampler. This enables efficient posterior simulation without the need for Metropolis-Hastings approximations or asymptotic regressions.

Further, we also propose the particle uncertainty quantification: A new class of posterior functional measurement termed latent failure mode scores, the representation-gap signature and the implicit survival shift. We state that these metrics allow the researchers to decompose the predictive variance of the lifetime distribution into random failure and measurement limitation components. This provides a rigorous predictor for distinguishing hidden risk from random noise, offering a theoretical advantage for survival analysis in settings where data quality is compromised.

The remainder of this article is organized as follows. In Section 2 we introduce the proposed process-augmented lifetime distribution framework, defines the generative chain for representation loss with detailed mathematical support, and derives the exact Gibbs sampler used for implementation. This section also formalizes the posterior functional measurements used for uncertainty quantification, including the latent failure mode score. In Section 3 we present a survival data simulation study that evaluates the finite-sample performance of the proposed estimator against naive and infeasible benchmarks, illustrating the model's capacity to resolve latent heterogeneity. In Section 4, we provide an empirical study in high-frequency market stability using E-mini S&P 500 futures trading during 2020, proving the model's utility in recovering latent volatility risks from observed trading data. We conclude with a discussion of the methodology and directions for future research in Section 5.

2. Methods: Process-augmentation survival analysis

In reliability analysis and survival modeling, the observed lifetime distribution is often a compound result of observable stresses and hidden, structural degradation. Classical survival models typically conceptualize risk estimation through an additive noise model structure: The observed covariate is treated as the true latent exposure contaminated by random noise. These models focus on correcting discrepancy errors under parametric error assumptions. However, in many modern survival studies, particularly those involving high-dimensional biological data or administratively coarsened records, the primary conceptual deficit originates not from random noise, but representation loss. The recorded variable is often a compressed, discretized, or coarsened representation of a much richer latent exposure that actually drives the survival risk.

To address this, we propose a process-augmented lifetime distribution (PALD) framework. By extending the discrepancy error paradigms of spatio-temporal sequence and the deep hierarchical generalized transformation (DHGT) from [8], we introduce a generative chain that decomposes the hazard into two orthogonal channels:

(1) **The explicit channel:** The component of the exposure that is structurally representable by the measurement instrument, spanned by the explicit basis $\tilde{G}$ (e.g., the step-function approximation of a discretized observation).

(2) **The implicit channel:** An orthogonal discrepancy channel that captures the representation gap defined by the features of x that reside in the null space of $\tilde{G}$, which are systematically lost during the measurement process but are required to explain the survival outcome.

We formalize this via a hierarchical generative chain ($X \to r \to Z$). Following the discrepancy error augmenting scheme from process augmentation, we introduce an intermediate latent representation r that acts as a bridge between the ideal exposure and the observed data. By embedding the latent exposure x into a low-rank basis subspace (G) determined by the measurement, and augmenting it with an orthogonal complement (B), we effectively create a deep generative hierarchical state analogous to the layers in the DHGT model. In this section, we start from basic notation introduction, the background mathematical suppot illustration, and introduce the process augmented lifetime distribution framework and its exact inference as follows.

2.1. Model formulation

2.1.1. Data structure and notation

We consider a survival analysis setting involving a group of n subjects indexed by $i = 1, \ldots, n$. For each experimental unit, we observe the data space $(t_i, \delta_i, Z_i, \mathbf{w}_i)$, where $t_i \in \mathbb{R}^+$ denotes the observed time-to-event and $\delta_i \in \{0, 1\}$ serves as the event indicator. Specifically, $\delta_i = 1$ denotes a confirmed failure event, where $\delta_i = 0$ indicates right-censoring. The covariate of primary interest is $Z_i \in \mathbb{R}$, which acts as a resolution-limited measure for the true structural exposure, analogous to the observed dataset defined in [8]. We further account for exogenous confounding through a vector of error-free adjustment covariates $\mathbf{w}_i \in \mathbb{R}^p 4$. We define $\mathbf{W} \in \mathbb{R}^{n \times p}$ as the fixed-effects design matrix where the i-th row is given by $\mathbf{w}_i^\top$, and we denote the log-transformed observed times by $\ell_i := \log(t_i)$.

To utilize the process-augmentation scheme, we introduce a set of latent variables that bridge the topological gap between the observation and the underlying risk process. Let $y^\star = (y_i^\star)_{i=1}^n$ represent the latent augmented log-survival times, which correspond to the exact failure times for event-free observations. The true high-fidelity latent exposure driving the survival risk is denoted by the vector $x = (x_i)_{i=1}^n$. Within our generative framework, we further define $r = (r_i)_{i=1}^n$ as intermediate latent representation variable. This variable mediates the relationship between the true exposure x and Z in the generative chain, which forms the core innovation for this article. The latent exposure x is decomposed into explicit and implicit components parameterized by the coefficient vectors $\eta \in \mathbb{R}^K$ for representable basis features and $\xi \in \mathbb{R}^{d_\perp}$ for discrepancy features.

The model is governed by a set of regression and precision parameters. The vector $\boldsymbol{\theta} = (\beta_0, \beta_x, \boldsymbol{\gamma}^\top)^\top \in \mathbb{R}^{p+2}$ contains the regression coefficients linking the covariates to the survival outcome. The residual variability of the survival times on the accelerated failure time (AFT) scale is captured by the variance parameter $\sigma_y^2 > 0$. Finally, the fidelity of the generative chain is controlled by the precision parameters $\tau_c > 0$ and $\tau_b > 0$. These parameters are designed to quantify the robustness of the decomposition in the representation and measurement steps defined in the process model.

2.1.2. Decomposition of the latent exposure

One of our methodological contributions is that we reconceptualize the latent exposure x not just as a variable obscured by stochastic noise, but as a complex signal compressed structurally. We posit that the true exposure driving the survival risk exists in a high-dimensional space spanned by two orthogonal subspaces: The subspace explicit to the measurement instrument and the subspace implicit from it. To formalize this, we introduce the spectral decomposition:

$$x := G\eta + B\xi \in \mathbb{R}^n, \tag{2.1}$$

where the latent exposure is expressed as the linear combination of explicit features $G\eta$ and implicit discrepancy features $B\xi$, and $\xi \in \mathbb{R}^{n-K}$ is a stochastic process realization, not a fixed parameter vector. This formulation fundamentally extends classical measurement error modeling by shifting the focus from bias correction to feature recovery.

The matrix $G \in \mathbb{R}^{n\times K}$, termed as explicit basis matrix, defines the subspace of variation that is structurally representable by the observed measure Z. Unlike standard errors-in-variables models that treat the observation as a continuous variable with additive Gaussian noise, we construct G such that its column space strictly spans the information content captured by the measurement process (e.g., step-functions for discretized data or splines for smoothed trends). We enforce the orthonormal constraint $G^\top G = I_K$, ensuring that the coefficient vector η uniquely represents the magnitude of these explicitly representable features found within the observed data.

We construct the formal definition of the implicit basis matrix $B \in \mathbb{R}^{n\times(n-K)}$ as one of the key innovations. This matrix represents the orthogonal complement to G, satisfying the structural constraints $B^\top B = I_{n-K}$ and the orthogonal condition $G^\top B = 0$. In classical regression calibration or simulation-extrapolation (SIMEX) approaches, the component of x orthogonal to the observation is typically subsumed into the residual error term of the outcome model or assumed to be negligible. In contrast, our framework explicitly models this representation gap through the term $B\xi$, where ξ quantifies the magnitude of high-frequency or micro-scale variations that are systematically lost during the measurement process but remain effective predictors of the lifetime distribution. This

decomposition ensures that any variation in x that cannot be mathematically mapped onto the basis G is forced into the discrepancy channel $B\xi$, allowing us to isolate and analyze the survival impact of the latent structure.

2.1.3. Basis construction and rank selection

The spectral decomposition $x = G\eta + B\xi$ relies entirely on the premise that the basis matrix G accurately delineates the boundary between the explicit information content of Z and the implicit latent structure. Consequently, the construction of G and the determination of its rank K are not separate statistical choices but are coupled theoretical operations: The rank K is the intrinsic dimension of the functional subspace spanned by the measurement instrument.

To bridge abstract decomposition and practical estimation together, we first define the raw functional space induced by the measurement process. We posit that Z is a realization of the true exposure x passed through a lossy compression operator (e.g., rounding or smoothing). Therefore, the explicit basis G must be constructed to span the range of this operator. We operate this via a raw basis construction followed by a rigorous orthogonalized procedure to ensure identifiability against fixed effects X.

Definition 2.1 (The raw measurement subspace)**.** *Let $Z \in \mathbb{R}^n$ be the observed measurement vector. We postulate that Z is generated by applying a resolution-limited operator $\mathcal{T} : \mathbb{R} \to \mathbb{R}$ to the latent exposure (e.g., quantization, censoring, or rounding). Define the raw measurement basis $\tilde{G} \in \mathbb{R}^{n\times K_0}$ as the matrix of basis functions $\{\phi_k\}_{k=1}^{K_0}$ that span the range of this operator.*

$$\mathcal{S}_{raw} = span(\tilde{G}) \subset \mathbb{R}^n. \tag{2.2}$$

For discrete measurements, $\tilde{G}$ consists of indicator functions for the observation bins.

For continuous measurements, $\tilde{G}$ consists of the spanning basis (e.g., B-splines) of the smoothing kernel.

We now provide the main theoretical result that allows us to construct the orthonormal basis G and the implicit basis B such that the latent exposure is uniquely identified. This proposition establishes the algebraic mechanism for separating the explicit signal from both the confounding fixed effects and the implicit discrepancy.

Proposition 2.2 (Orthogonal decomposition and basis construction)**.** *Let $X \in \mathbb{R}^n$ denote the vector of the true latent exposure and $\mathcal{S}_{raw} \subset \mathbb{R}^n$ be the raw measurement subspace defined in Definition 2.1. Suppose that $\tilde{G}$ is the basis matrix for $\mathcal{S}_{raw}$. Then, there exists a unique orthogonal decomposition of the latent exposure given by:*

$$X = \tilde{G}\eta + B\xi,$$

where η represents the coefficient vector of the visible structure, B is a basis matrix spanning the orthogonal complement $\mathcal{S}_{raw}^{\perp}$ (the "implicit channel"), and ξ captures the unobserved granular heterogeneity. This proposition formally isolates the information lost during the coarsened measurement process, allowing the representation gap ($B\xi$) to be modeled as a distinct stochastic process rather than a simple error term.

Proof. We have the unique orthogonal decomposition of X with respect to the subspace $\mathcal{S}_{raw} = \text{span}(\tilde{G})$ given by:

$$X = P_{\mathcal{S}}X + P_{\mathcal{S}^\perp}X = \tilde{G}(\tilde{G}^\top\tilde{G})^{-1}\tilde{G}^\top X + (I_n - P_{\mathcal{S}})X = \tilde{G}\eta + B\xi,$$

where the structural coefficients are identified as

$$\eta = (\tilde{G}^\top\tilde{G})^{-1}\tilde{G}^\top X,$$
$$\xi = B^\top(I_n - \tilde{G}(\tilde{G}^\top\tilde{G})^{-1}\tilde{G}^\top)X.$$

The orthogonality of the decomposition follows immediately from the construction of the basis matrices $\tilde{G}$ and B

$$\langle\tilde{G}\eta, B\xi\rangle = \eta^\top\tilde{G}^\top B\xi = \eta^\top(\mathbf{0})\xi = 0,$$

since the columns of B span the null space of $\tilde{G}^\top$. Consequently, $\tilde{G}\eta$ represents the visible projection onto $\mathcal{S}_{raw}$, and $B\xi$ represents the orthogonal residual in $\mathcal{S}_{raw}^\perp$. □

In Proposition 2.2, we establish this orthogonal decomposition of the latent exposure, which provides the algebraic format to separate the observable signal from the representation gap. Specifically, the hierarchical prior $\xi \sim \mathcal{N}(0, I_{d_\perp})$ regulates the model scale. This unit-variance constraint alters the magnitude of the latent discrepancy and prevents the confounding of the spectral coefficients with the regression parameters. By fixing the precision of the implicit process, we ensure that the structural features recovered by the augmentation layer are identifiable from both the fixed effects and the residual failure noise.

Having established the construction of the explicit basis, we must address the implicit component. Constructing the matrix $B \in \mathbb{R}^{n\times(n-K)}$ explicitly is computationally prohibitive for large n. However, the inference algorithm requires only the projection of vectors onto the implicit subspace. The following corollary derives the analytic form of the implicit projector, enabling efficient computation.

Corollary 2.3 (The implicit projector)**.** *Let $M_{\tilde{G}} = I_n - \tilde{G}(\tilde{G}^\top\tilde{G})^{-1}\tilde{G}^\top$ denote the implicit projector onto the orthogonal complement $\mathcal{S}_{raw}^\perp$. Under the assumption that the latent heterogeneity ξ in Proposition 2.2 follows a centered Gaussian process with covariance $\sigma_\xi^2 I_{n-K_0}$, the residual vector $x_\perp = M_{\tilde{G}}X$ satisfies the following properties:*

(1) Idempotent variance filtering: The covariance structure of the residual restricts variation strictly to the null space of the measurement basis:

$$Var(r) = \sigma_\xi^2 M_{\tilde{G}}. \tag{2.3}$$

(2) Residual measure sufficiency: The quadratic form of the projected data constitutes a minimal sufficient statistic for the magnitude of the unobserved heterogeneity:

$$S_r^2 := X^\top M_{\tilde{G}}X \sim \sigma_\xi^2 \cdot \chi^2_{(n-K_0)}. \tag{2.4}$$

Proof. Recall that $M_{\tilde{G}}$ projects onto $\mathcal{S}_{raw}^\perp$. By construction, $M_{\tilde{G}}$ is symmetric and idempotent. We establish the annihilation property with respect to the visible component:

$$M_{\tilde{G}}\tilde{G} = (I_n - \tilde{G}(\tilde{G}^\top\tilde{G})^{-1}\tilde{G}^\top)\tilde{G} = \tilde{G} - \tilde{G}(\tilde{G}^\top\tilde{G})^{-1}(\tilde{G}^\top\tilde{G}) = \tilde{G} - \tilde{G} = \mathbf{0}.$$

Substituting the decomposition $X = \tilde{G}\eta + B\xi$ from Proposition 2.2 into the definition of r, and applying

$$r = M_{\tilde{G}}(\tilde{G}\eta + B\xi) = M_{\tilde{G}}\tilde{G}\eta + M_{\tilde{G}}B\xi = B\xi.$$

(Note that $M_{\tilde{G}}B = B$ because the columns of B lie entirely in the range of $M_{\tilde{G}}$). The expectation and variance are derived immediately:

$$\begin{aligned} E[r] &= BE[\xi] = \mathbf{0}, \\ Var(r) &= BVar(\xi)B^\top = B(\sigma_\xi^2 I_{n-K_0})B^\top = \sigma_\xi^2(BB^\top). \end{aligned}$$

Since $BB^\top$ represents the projection onto the span of B (which is exactly $\mathcal{S}_{raw}^{\perp}$), we have $BB^\top = M_{\tilde{G}}$. Thus, $Var(r) = \sigma_\xi^2 M_{\tilde{G}}$.

We then examine the residual S_r^2

$$\begin{aligned} X^\top M_{\tilde{G}} X &= (\tilde{G}\eta + B\xi)^\top M_{\tilde{G}}(\tilde{G}\eta + B\xi) \\ &= \xi^\top B^\top M_{\tilde{G}} B\xi \\ &= \xi^\top (B^\top B)\xi \quad (\text{given } B^\top M_{\tilde{G}} = B^\top) \\ &= \xi^\top I_{n-K_0}\xi \\ &= \sum_{i=1}^{n-K_0} \xi_i^2. \end{aligned}$$

Given the assumption $\xi \sim \mathcal{N}(0, \sigma_\xi^2 I)$, the standardized sum $\frac{1}{\sigma_\xi^2}\sum \xi_i^2$ follows a Chi-squared distribution with degrees of freedom equal to the rank of the hidden space, $n - K_0$.

$$\frac{X^\top M_{\tilde{G}} X}{\sigma_\xi^2} \sim \chi^2_{(n-K_0)}.$$

This confirms that $X^\top M_{\tilde{G}} X$ captures informative features regarding σ_ξ^2 available in the subspace orthogonal to $\tilde{G}$, satisfying the sufficiency condition. □

Motivated by the orthogonal decomposition in Proposition 2.2, we now specify the dimensionality of the subspace to ensure the identifiability of the latent components. Unlike standard dimension reduction techniques that select rank to maximize explained variance (e.g., principal component analysis), our framework requires the rank to be determined by the structural resolution of the measurement instrument itself. We specify the dimensions of the visible and implicit channels as follows:

$$\begin{aligned} K &\equiv \text{rank}(\tilde{G}) \\ d_\perp &\equiv \dim(\text{null}(\tilde{G}^\top)) = n - K. \end{aligned}$$

This deterministic specification plays two critical roles. First, setting K equal to the rank of the raw measurement basis $\tilde{G}$ forces the explicit channel to capture the maximal amount of variation representable by the coarsened measurement. Second, and more importantly for subsequent estimation, the dimension $d_\perp$ defines the volume of the augmentation space. By fixing these dimensions, we ensure that the representation gap is not treated as unstructured noise, but rather as a process with exactly $d_\perp$ degrees of freedom. In the next subsection, we will treat the $d_\perp$-dimensional vector $B\xi$ as auxiliary variables under the process augmentation scheme given this dimensional constraint.

2.1.4. Process augmentation and the generative chain

Previously, [7] and [8] originally introduced the process augmentation structure to resolve spatial confounding and improve computational tractability in high-dimensional settings. We further extend this framework here to explicitly address the representation gap identified in Proposition 2.2. Classical errors-in-variables models typically condition the true latent variable on the observation (e.g., $X \mid Z$), implicitly assuming that the two variables possess a topologically equivalent metric space separated only by additive noise. We argue that this assumption fails in the presence of resolution limits, where the observed algebra $\sigma(Z)$ is a strictly coarser filtration of the complete state space $\sigma(X)$. To recover information loss in the subspace $\mathcal{S}_{raw}^{\perp}$, we propose a generative chain that models the data generation process in its natural causal direction. Motivated by the orthogonal decomposition in Proposition 2.2, we define our specified process augmentation workflow in a hierarchical function module of Markov kernels:

$$\begin{aligned} \textbf{Process Model:} \quad & X(\eta, \xi) = \tilde{G}\eta + B\xi, \\ \textbf{Measurement Model:} \quad & Z \mid \eta \sim \mathcal{N}(\tilde{G}\eta, \sigma_{\epsilon}^2 I_n), \\ \textbf{Augmentation Prior:} \quad & \xi \sim \mathcal{N}(\mathbf{0}, \sigma_{\xi}^2 I_{d_{\perp}}). \end{aligned} \tag{2.5}$$

This hierarchical Eq (2.5) reformulates the intractable marginal likelihood $p(Z \mid X)$ by introducing the structural components as auxiliary variables. The process model governs the reconstruction of the complete state X. It explicitly combines the observed projection $\tilde{G}\eta$ with the augmented residual process $B\xi$, where the dimensionality of the augmentation is strictly determined by the rank deficiency $d_{\perp}$ defined in Corollary 2.3. This stage models the resolution constraints of the measurement instrument. We posit that the intermediate representation r is not merely a component of X, but a stochastic mapping of the complete state onto the representable manifold $\mathcal{S}_{raw}$. Unlike standard regression calibration, we model the discrepancy $X - r$ as a structured process governed by the implicit projector P_B defined in Corollary 2.3. We specify the transition density as a Gaussian kernel centered on the latent exposure:

$$\begin{aligned} p(r \mid X, \tau_c) &= \mathcal{N}(r \mid X, (\tau_c D_c)^{-1}) \\ &\propto |\tau_c D_c|^{1/2} \exp\left(-\frac{1}{2}(r - X)^{\top}(\tau_c D_c)(r - X)\right), \end{aligned} \tag{2.6}$$

where τ_c is the process precision parameter and D_c is a diagonal scaling matrix. This kernel formalizes the process loss. As $\tau_c \to \infty$, the probability mass function of the distribution converges on $r = X$. However, for finite τ_c, this model accommodates the non-zero spectral function of the implicit feature $B\xi$, effectively considering the high-frequency variation that the instrument cannot resolve.

The measurement model, on the other hand, works as a degenerate filter, observing only the projection of the state onto the raw measurement subspace $\mathcal{S}_{raw}$, contaminated by the noise ϵ. In this stage, we model the strictly telescoping error inherent in the recording mechanism. Given the structural representation r, the observed Z is generated via an independent Gaussian noise process:

$$\begin{aligned} p(Z \mid r, \tau_b) &= \mathcal{N}(Z \mid r, (\tau_b D_b)^{-1}) \\ &\propto |\tau_b D_b|^{1/2} \exp\left(-\frac{1}{2}(Z - r)^{\top}(\tau_b D_b)(Z - r)\right). \end{aligned} \tag{2.7}$$

Here, τ_b is denoted as the parameter of measurement precision. We establish the conditional independence structure $Z \perp X \mid r$, proposing that observation noise only contaminates the projected representation, not the true latent process. One key theoretical innovation of this framework is that the separation of process loss (implicit structure) from measurement noise (error term ϵ) can be derived analytically. We obtain the conditional likelihood $p(Z \mid X)$ by integrating out the augmented variable r into the state space $\mathbb{R}^n$. Applying the properties of convolved Gaussian kernels, we derive the marginal density as:

$$\begin{aligned} p(Z \mid X) &= \int_{\mathbb{R}^n} p(Z \mid r) p(r \mid X)\, dr \\ &= \int_{\mathbb{R}^n} \mathcal{N}(Z \mid r, \Sigma_b) \cdot \mathcal{N}(r \mid X, \Sigma_c)\, dr, \quad \text{where } \Sigma_b = (\tau_b D_b)^{-1}, \Sigma_c = (\tau_c D_c)^{-1} \\ &= \mathcal{N}(Z \mid X, \Sigma_b + \Sigma_c). \end{aligned} \tag{2.8}$$

The covariance structure $\Sigma_{total} = \Sigma_b + \Sigma_c$ provides the theoretical support for our augmentation strategy. By decomposing the total error variance, we can explicitly map the components to the subspaces defined in Proposition 2.2.

(1) Implicit channel variance: Σ_c captures the variance of the orthogonal discrepancy $B\xi$. It measures the irreducible uncertainty due to the coarseness of the basis $\tilde{G}$.

(2) Explicit channel variance: Σ_b captures the variance of the measurement noise ϵ.

From that, this variance decomposition and deconfounding module effectively transforms the representation gap in survival analysis from a constant, *confounding parameter* into a recoverable covariance component, allowing Monte Carlo Markov chain (MCMC) implementations (e.g., Gibbs sampler) to iteratively update the latent variables η and ξ *accordingly*.

2.1.5. Survival submodel: Log-normal accelerated failure time

In this subsection, we link the latent high-fidelity exposure X to the observed time-to-event outcomes through an accelerated failure time (AFT) mechanism. The selection of the log-normal as a special case is not merely for its parametric convenience, but because it preserves the exponential distribution form that preserves the Gaussian conjugacy required by the spectral decomposition defined in Section 2.1.3. We argue that the representation gap $B\xi$ can be recovered via exact inference function rather than approximation methods such as MCMC.

We define the systematic component of the survival process by projecting the log-lifetime onto the augmented state space. Let T_i denote the true survival time for subject i. We posit that the log-transformed lifetime $y_i^* = \log(T_i)$ is generated by a linear functional of the complete latent state, explicitly incorporating the orthogonal channels derived in Section 2.2:

$$\begin{aligned} y_i^* &= \mu_i(\eta, \xi) + \sigma_y \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, 1), \\ \mu_i(\eta, \xi) &= \beta_0 + \mathbf{w}_i^\top \boldsymbol{\gamma} + \beta_x(\tilde{\mathbf{g}}_i^\top \boldsymbol{\eta} + \mathbf{b}_i^\top \boldsymbol{\xi}). \end{aligned} \tag{2.9}$$

Here, term $\tilde{\mathbf{g}}_i^\top \boldsymbol{\eta} + \mathbf{b}_i^\top \boldsymbol{\xi}$ is the latent exposure and $\mathbf{w}_i$ represents fixed covariates with coefficients $\boldsymbol{\gamma}$. We expand the exposure coefficient β_x where $\beta_x(\tilde{\mathbf{g}}_i^\top \boldsymbol{\eta})$ captures the effect of the observed signal

(the smooth trend or binned average resolvable by augmentation). The term $\beta_x(\mathbf{b}_i^\top \boldsymbol{\xi})$ captures the effect of the particle heterogeneity (the high-frequency risk factors orthogonal to the measurement instrument). By structurally embedding ξ to the survival time, we force the survival likelihood to inform the estimation of the implicit process, effectively utilizing the unexplained variance in lifetimes to reconstruct the underlying information of the exposure.

Besides, given the structural predictor $\mu(\eta, \xi)$, the probability density function for the latent log-lifetimes is governed by the Gaussian kernel. The conditional density $f(y^* \mid \cdot)$ is defined below:

$$\begin{aligned} p(y^* \mid \boldsymbol{\beta}, \sigma_y^2, \eta, \xi) &= \prod_{i=1}^{n} \phi\left(\frac{y_i^* - \mu_i(\eta, \xi)}{\sigma_y}\right) \\ &\propto (\sigma_y^2)^{-n/2} \exp\left(-\frac{1}{2\sigma_y^2} \|\mathbf{y}^* - (\beta_0 \mathbf{1} + \mathbf{W}\boldsymbol{\gamma} + \beta_x(\tilde{G}\boldsymbol{\eta} + B\boldsymbol{\xi}))\|^2\right). \end{aligned} \tag{2.10}$$

This algebraic form highlights the collaboration with the process augmentation scheme in Section 2.1.4. Since the kernel is quadratic in $\boldsymbol{\xi}$, it combines with the augmentation prior $p(\boldsymbol{\xi}) \propto \exp(-\frac{1}{2}\boldsymbol{\xi}^\top \boldsymbol{\xi})$ analytically, allowing the posterior of the hidden structure to be derived in closed form.

Consider the connection between the latent process y^* and the observed data $\mathcal{D} = \{(t_i, \delta_i)\}_{i=1}^n$, we mediate them by a deterministic censorship filter. This is not merely a revision of the likelihood function (2.10), but a support constraint on the latent variable y^*. Let C_i denote the set of feasible log-lifetimes given the observation:

$$C_i = \begin{cases} \{y_i^* : y_i^* = \log t_i\} & \text{if } \delta_i = 1 \text{ (Event)}, \\ \{y_i^* : y_i^* > \log t_i\} & \text{if } \delta_i = 0 \text{ (Censored)}. \end{cases} \tag{2.11}$$

The joint data distribution is thus expressed as the product of the structural kernel and the indicator function of censorship:

$$p(\mathbf{t}, \boldsymbol{\delta} \mid \mathbf{y}^*) = \prod_{i=1}^{n} \mathbb{I}(y_i^* \in C_i). \tag{2.12}$$

Therefore, we obtain the complete joint generative distribution. By substituting the specific Gaussian kernels from Eq (2.10) and the augmentation priors, the full joint density $p(\mathbf{y}^*, Z, r, \eta, \xi \mid \Theta)$ decomposes as

$$\begin{aligned} p(\mathbf{y}^*, Z, r, \eta, \xi \mid \Theta) \propto & p(\mathbf{t}, \boldsymbol{\delta} \mid \mathbf{y}^*) \times p(y^* \mid \eta, \xi, \sigma_y^2) \times p(Z \mid r, \Sigma_b) \times p(r \mid \eta, \xi, \Sigma_c) \times p(\xi) \\ \propto & \left[\prod_{i=1}^{n} \mathbb{I}(y_i^* \in C_i)\right] \times (\sigma_y^2)^{-n/2} \exp\left(-\frac{1}{2\sigma_y^2} \|\mathbf{y}^* - \mu(\eta, \xi)\|^2\right) \\ & \times |\Sigma_b|^{-1/2} \exp\left(-\frac{1}{2}(Z - r)^\top \Sigma_b^{-1} (Z - r)\right) \\ & \times |\Sigma_c|^{-1/2} \exp\left(-\frac{1}{2}(r - (\tilde{G}\eta + B\xi))^\top \Sigma_c^{-1} (r - (\tilde{G}\eta + B\xi))\right) \\ & \times \exp\left(-\frac{1}{2}\xi^\top \xi\right), \end{aligned} \tag{2.13}$$

where $\mu(\eta, \xi) = \beta_0 \mathbf{1} + \mathbf{W}\boldsymbol{\gamma} + \beta_x(\tilde{G}\boldsymbol{\eta} + B\boldsymbol{\xi})$. Equation (2.13) explicitly demonstrates the paired role of the implicit feature $B\boldsymbol{\xi}$: It possesses quadratic structure in both the process fidelity kernel and the survival likelihood kernel. This mathematical structure proves that the representation gap is identifiable, as $B\boldsymbol{\xi}$ is triangulated by both the observation deviations ($r - \tilde{G}\eta$) and the unexplained variation in survival times ($y^* - \mu_{cov}$).

2.2. Posterior computation: The exact Gibbs sampler

The generative function we constructed assumes that the observed survival outcomes are driven by the latent $X = \tilde{G}\eta + B\xi$. While it is straightforward to simulate this process, the inverse problem remains unaddressed: How do we obtain the sample draws of ξ, process parameters from the observation Z, and the censored lifetimes (t, δ). In survival studies, the representation gap implies that the likelihood surface for the true exposure is multimodal or flat; single point estimation would likely collapse the implicit channel $B\xi$ to zero. Inspired by [7], we can provide a Bayesian implementation to quantify uncertainty of the latent risk effects for reliable failure prediction by examining the posterior distribution of the latent process.

2.2.1. Prior specification and identifiability

One common challenge in process-augmented models is to tractably distinguish the structural implicit signals from measurement noise. In standard errors-in-variables models [11, 16, 34], these components are often confounded. However, our orthogonal decomposition is capable of strictly identifying structural information through the imposition of hierarchical regularization constraints.

To prevent overfitting in the implicit channel, we assign standard Gaussian priors to the spectral coefficients:

$$\eta \sim \mathcal{N}(\mathbf{0}, \sigma_\eta^2 I_K),$$
$$\xi \sim \mathcal{N}(\mathbf{0}, I_{d_\perp}).$$

The unit variance on ξ follows a Ridge-type regularization. This forces $B\xi$ to extract only those deviations that are structurally orthogonal to $\tilde{G}$ and utilize sufficient signal measurements to alter the survival likelihood. If the data does not support hidden heterogeneity, the posterior mass of ξ will naturally converge to 0, whereas the explicit signal η varies according to the data scale. For the fixed effects in the AFT submodel, we assume a weakly informative multivariate Gaussian prior:

$$\boldsymbol{\gamma} \sim \mathcal{N}(\mathbf{b}_0, \mathbf{V}_0),$$

where $\mathbf{V}_0$ is a diagonal matrix with large variance terms (e.g., scale of 10^4 or larger). The scale parameters governing the hierarchy are assigned priors from the inverse-Gamma family to maintain conjugacy with the Gaussian kernels derived in Eq (2.10):

$$\sigma_y^2 \sim \mathcal{IG}(a_y, b_y),$$
$$\Sigma_b \equiv \sigma_b^2 I_n, \quad \sigma_b^2 \sim \mathcal{IG}(a_b, b_b)$$
$$\Sigma_c \equiv \sigma_c^2 I_n, \quad \sigma_c^2 \sim \mathcal{IG}(a_c, b_c).$$

This prior selection ensures that the full conditional distributions for all parameters belong to exponential distributional families (Gaussian, inverse-Gamma, or truncated normal), allowing the implementation of an exact MCMC type sampler. Unlike Metropolis-Hastings approximations, we prevent the convergence issues associated in high-dimensional latent variable settings.

2.2.2. The full joint posterior kernel

Given the observed data $\mathcal{D} = \{Z, \mathbf{t}, \boldsymbol{\delta}, \mathbf{W}\}$, we need to extract the amplified latent structure by inverting this generative chain. The target is to obtain the joint posterior distribution of the complete parameter space $\Omega = \{y^\star, r, \eta, \xi, \boldsymbol{\gamma}, \beta_x, \sigma_y^2, \Sigma_c, \Sigma_b\}$, contains the information for both the representation gap and the survival function. Following the hierarchical formulation established in Eq (2.5), we decompose the global statistical model into three distinct conditional layers: the data model of observed dataset, the process model of latent structural functions, and the parameter model which contains all prior regularization.

We propose the target posterior distribution as the product of the following conditional densities:

$$\begin{aligned} \textbf{Data Model:} \quad & p(\mathcal{D} \mid \Omega) = p(Z \mid r, \Sigma_b) \times \mathbb{I}_C(y^\star; \mathbf{t}, \boldsymbol{\delta}) \\ \textbf{Process Model:} \quad & p(\text{Latent} \mid \Theta) = p(r \mid \tilde{G}\eta + B\xi, \Sigma_c) \times p(y^\star \mid \mu(\eta, \xi), \sigma_y^2) \\ \textbf{Parameter Model:} \quad & \pi(\Theta) = \pi(\boldsymbol{\gamma}, \beta_x)\pi(\sigma_y^2)\pi(\Sigma_c)\pi(\Sigma_b)\pi(\eta)\pi(\xi). \end{aligned} \tag{2.14}$$

The derivation of the posterior distribution $p(\Omega \mid \mathcal{D})$ proceeds by systematically substituting the Gaussian kernels derived in Eq (2.10) and priors and combining them to reveal the conjugate structure.

$$\begin{aligned} p(\Omega \mid \mathcal{D}) &\propto \pi(\Theta) \times p(\text{Latent} \mid \Theta) \times p(\mathcal{D} \mid \Omega) \\ &\propto [\pi(\boldsymbol{\gamma}, \beta_x) \dots \pi(\xi)] \\ &\quad \times \left[|\Sigma_c|^{-1/2} \exp\left(-\frac{1}{2}(r - (\tilde{G}\eta + B\xi))^\top \Sigma_c^{-1} (r - (\tilde{G}\eta + B\xi)) \right) \right] \\ &\quad \times \left[(\sigma_y^2)^{-n/2} \exp\left(-\frac{1}{2\sigma_y^2} \|y^\star - \mu(\eta, \xi)\|^2 \right) \right] \\ &\quad \times \left[|\Sigma_b|^{-1/2} \exp\left(-\frac{1}{2}(Z - r)^\top \Sigma_b^{-1} (Z - r) \right) \right] \times \mathbb{I}_C(y^\star; \mathbf{t}, \boldsymbol{\delta}) \\ &\propto \mathbb{I}_C(y^\star; \mathbf{t}, \boldsymbol{\delta}) \\ &\quad \times (\sigma_y^2)^{-(\frac{n}{2} + a_y + 1)} \exp\left[-\frac{1}{\sigma_y^2} \left(b_y + \frac{1}{2} \|y^\star - (\beta_0 \mathbf{1} + \mathbf{W}\boldsymbol{\gamma} + \beta_x(\tilde{G}\eta + B\xi))\|^2 \right) \right] \\ &\quad \times |\Sigma_c|^{-(\frac{1}{2} + a_c + 1)} \exp\left[-\text{tr}\left(\Sigma_c^{-1} \left(b_c I_n + \frac{1}{2}(r - (\tilde{G}\eta + B\xi))(r - (\tilde{G}\eta + B\xi))^\top \right) \right) \right] \\ &\quad \times |\Sigma_b|^{-(\frac{1}{2} + a_b + 1)} \exp\left[-\text{tr}\left(\Sigma_b^{-1} \left(b_b I_n + \frac{1}{2}(Z - r)(Z - r)^\top \right) \right) \right] \\ &\quad \times \exp\left[-\frac{1}{2}(\boldsymbol{\gamma} - \mathbf{b}_0)^\top \mathbf{V}_0^{-1} (\boldsymbol{\gamma} - \mathbf{b}_0) \right] \exp\left[-\frac{1}{2}\eta^\top \eta \right] \exp\left[-\frac{1}{2}\xi^\top \xi \right]. \end{aligned} \tag{2.15}$$

From Eq (2.15), it is observed that the implicit structural component $B\xi$ appears quadratically in both the survival term (via μ) and the process term (via the mean of r). Also, the theoretical conjugacy for all parameters validates the implementation of an exact Gibbs sampler, ensuring efficient parameter convergence in high-dimensional augmentation scenario and providing a computationally efficient module in large survival data.

2.2.3. The exact Gibbs sampler algorithm

Given the conjugacy of the joint posterior kernel derived in Eq (2.15), the high-dimensional inference problem reduces to a sequence of exact conditional updates. We exploit the exact sampling of hierarchical conditional independence structure: The representation variable r acts as a stochastic latent process, augmenting the decomposition for complex measurement error from the survival observations.

The sampling procedure proceeds in four functional blocks. First, we proceed the augmentation stage to impute the censored lifetimes $y^\star$ from a truncated Gaussian distribution restricted by the observed event history. Then, we reconstruct the process to update the representation by computing the precision-weighted average of the observed Z and the current latent exposure estimate. After that, we can provide the latent structure inference with joint sampling explicit (η) and implicit (ξ) basis coefficients, effectively extracting and measuring the true exposure X using gradients from both the survival residuals and the process fidelity term. Finally, we update the estimation of the regression coefficients and variance components via standard Bayesian linear regression and inverse-Gamma/Gamma conjugate prior distribution family.

The operational procedure is detailed in Algorithm 1.

Algorithm 1 Exact Gibbs sampler for process-augmented lifetime distributions

1: Initialize iteration counter $b = 1$ and set initial values for latent variables $y^{\star[0]}, r^{[0]}, \eta^{[0]}, \xi^{[0]}$ and parameters $\boldsymbol{\gamma}^{[0]}, \beta_x^{[0]}, \sigma_y^{2[0]}, \tau_c^{[0]}, \tau_b^{[0]}$.
2: Sample augmented survival times $y^{\star[b]}$ from the conditional truncated normal distribution:

$$y^{\star[b]} \sim \mathcal{TN}_{[\log t,\infty)}(\mu^{[b-1]}, \sigma_y^{2[b-1]}).$$

3: Sample the representation variable $r^{[b]}$ from the conditional Gaussian distribution $\mathcal{N}(m_r^{[b-1]}, Q_r^{-1[b-1]})$, updating the bridge between exposure and observation.
4: Sample the basis coefficients $\eta^{[b]}$ and $\xi^{[b]}$ jointly from the multivariate Gaussian $\mathcal{N}(m_{lat}^{[b-1]}, Q_{lat}^{-1[b-1]})$, and update the latent exposure $x^{[b]} = \tilde{G}\eta^{[b]} + B\xi^{[b]}$.
5: Sample covariate coefficients $\boldsymbol{\gamma}^{[b]}$ and exposure sensitivity $\beta_x^{[b]}$ from the standard Bayesian linear regression posterior conditional on $y^{\star[b]}$ and $x^{[b]}$.
6: Sample variance and precision parameters $\sigma_y^{2[b]}, \tau_c^{[b]}, \tau_b^{[b]}$ from their respective inverse-Gamma and Gamma conditional distributions.
7: Set $b \leftarrow b + 1$.
8: Repeat steps 2–7 until $b = B$ for a prespecified number of iterations B.

2.3. Discrepancy signatures and posterior functional measurements

The implementation of the exact Gibbs sampler (Algorithm 1) yields a sequence of dependent draws from the high-dimensional joint posterior $p(\Omega \mid \mathcal{D})$. While standard errors-in-variables literature typically treats the deviation between the true exposure and the latent process as noise component, where this unstructured term is marginalized and ignored, the process augmentation framework we put forward reinterprets this deviation fundamentally. Motivated by the orthogonal decomposition property, we argue that the implicit component $B\xi$ represents structured latent heterogeneity: The physical features of the exposure that are systematically orthogonal to the measurement $\tilde{G}$, significant

in driving the survival outcome.

We formalize the extraction of this hidden structure through the definition of discrepancy signatures. These are posterior functional measurements designed to quantify the representation gap, which is the specific scale of risk information loss due to the resolution limits of the latent process. Since the analytic derivation of these marginal expectations involves intractable integrals over the augmented state space of censored lifetimes $y^\star$ and intermediate representations r, we rely on the computational results of the MCMC sampling. By the ergodic theorem, the sampled averages of the posterior mean that the retained iterations converge almost surely to the true posterior expectations. In this section, we define three distinct functionals that utilize the recovered latent chains to characterize the geometry of the information loss and its impact on system reliability.

2.3.1. The latent failure mode score ($\delta^{(\perp)}$)

The first discrepancy signature we designed quantifies the magnitude of the risk heterogeneity that resides in the null space of the measurement instrument. While the term $\tilde{G}\eta$ captures the variation structurally representable by the latent process (e.g., the step-function approximation of a discretized covariate in Definition 2.1), standard survival estimators implicitly force the orthogonal complement $B\xi$ into the error term $\sigma_y\varepsilon$. We argue that in the presence of resolution limits, this orthogonal component is not noise, but structured latent exposure—specific physical dynamics (such as intra-bin fluctuations in medical imaging) that the instrument cannot resolve but which drive the failure outcome.

We propose the latent failure mode score, $\delta^{(\perp)}$, as the posterior expectation of the orthogonal projection of the exposure, scaled by its prognostic relevance. This functional measurement isolates the specific features of the hazard function that are invisible to the observer but extractable via the survival likelihood. Let $\mathcal{H}_\perp$ denote the Hilbert subspace spanned by the basis B in Definition 2.1. The latent failure mode score is defined as the posterior mean of the exposure's projection onto $\mathcal{H}_\perp$:

$$\delta_i^{(\perp)} := \mathbb{E}_{\Omega|\mathcal{D}}\left[\mathbf{b}_i^\top \boldsymbol{\xi}\right] = \int_{\mathbb{R}^{d_\perp}} (\mathbf{b}_i^\top \boldsymbol{\xi})\, p(\boldsymbol{\xi} \mid \mathcal{D})\, d\boldsymbol{\xi}. \tag{2.16}$$

Equation (2.16) represents the expected deviation of the true exposure from the observed Z_i. When $\delta_i^{(\perp)} > 0$, this implies that the measurement instrument systematically underestimates the true risk factor of the subject (e.g., a hidden stress outlier), while $\delta_i^{(\perp)} < 0$ implies an overestimation. Since the analytical integration in (2.16) over the high-dimensional posterior is intractable, we approximate $\delta_i^{(\perp)}$ using the ergodic averages of the MCMC chain generated by algorithm 1:

$$\hat{\delta}_i^{(\perp)} \approx \frac{1}{M}\sum_{m=1}^{M}\left(\sum_{k=1}^{n-K} B_{i,k}\xi_k^{(m)}\right), \tag{2.17}$$

where $\{\boldsymbol{\xi}^{(m)}\}_{m=1}^M$ is the sequence of M retained draws of the implicit coefficients after burn-in. We apply the projector property $B\xi = (I - P_{\tilde{G}})X$ instead of constructing the dense matrix B explicitly. Thus, for each iteration, the score is computed as the residual of the visible projection: $\hat{\delta}^{(m)} = X^{(m)} - \tilde{G}\eta^{(m)}$.

Another derivative for uncertainty quantification we designed is the signal-to-noise ratio (SNR) of $\delta^{(\perp)}$. We consider the second moment of the posterior chain to quantify whether the recovered feature

is a structural reality or a standard artifact of the prior:

$$\mathrm{SNR}_i := \frac{|\hat{\delta}_i^{(\perp)}|}{\sqrt{\mathrm{Var}(\delta_i^{(\perp)} \mid \mathcal{D})}},$$

where the posterior variance is computed as:

$$\mathrm{Var}(\delta_i^{(\perp)} \mid \mathcal{D}) \approx \frac{1}{M-1} \sum_{m=1}^{M} \left((\mathbf{b}_i^\top \boldsymbol{\xi}^{(m)}) - \hat{\delta}_i^{(\perp)} \right)^2.$$

In the additive form where we mostly focus on amplifying the noise , the subject with a high score $\hat{\delta}_i^{(\perp)}$ but low SNR suggests that the representation gap is uninformative (dominated by σ_y noise). Conversely, a high SNR indicates a statistically significant latent failure mode where a risk factor is effectively disclosed by augmentation that was absent in the original observed data.

2.3.2. The representation-gap signature ($\Delta^{(rx)}$)

We also design the representation-gap signature to reveal the statistical mechanics of the generative chain in process augmentation. We posit that the mapping from the ideal exposure X to the intermediate representation r is governed by a stochastic degradation process. Consequently, the vector difference $\boldsymbol{\Delta}^{(rx)} = r - X$ quantifies the fidelity loss (the specific information decay occurs before the signal) and is even subjected to measurement noise. In our hierarchical formulation, the posterior distribution of this gap captures the generative connection between the survival data (which requires X to explain the observed failure times) and the latent process (which constrains r to the observed data). We formalize this connection through the representation gap signature $\Delta^{(rx)}$, defined as the posterior expectation of the process residual.

We define $\Delta_i^{(rx)}$ as the ergodic average of the deviation between the bridge variable r and the complete latent state X:

$$\Delta_i^{(rx)} := \mathbb{E}_{\Omega|\mathcal{D}}\left[r_i - X_i\right] = \mathbb{E}_{\Omega|\mathcal{D}}\left[r_i - (\tilde{\mathbf{g}}_i^\top \boldsymbol{\eta} + \mathbf{b}_i^\top \boldsymbol{\xi})\right]. \tag{2.18}$$

Mathematically, Eq (2.18) assesses the topological incompatibility between the inherent characters of the exposure (driven by biological/ medical factor) and the representation structure (driven by the instrument). If the process precision $\tau_c \to \infty$, the representation r converges to X almost surely, and $\Delta_i^{(rx)}$ converges to zero in probability. Otherwise, a nonzero $\Delta_i^{(rx)}$ indicates a significant posterior divergence between the exposure implied by the censoring distribution $\mathbb{I}_C(y^\star; \mathbf{t}, \boldsymbol{\delta})$ and the exposure implied by the measurement kernel $p(Z \mid r, \Sigma_b)$. For tractability, we also use the joint posterior samples to compute the signature via the difference of the retained chains. Let $\{(r^{(m)}, \eta^{(m)}, \xi^{(m)})\}_{m=1}^{M}$ denote the MCMC draws. The estimation is approximated:

$$\hat{\Delta}_i^{(rx)} \approx \frac{1}{M} \sum_{m=1}^{M} \left(r_i^{(m)} - (\tilde{\mathbf{g}}_i^\top \boldsymbol{\eta}^{(m)} + \mathbf{b}_i^\top \boldsymbol{\xi}^{(m)}) \right). \tag{2.19}$$

A value of $\hat{\Delta}_i^{(rx)} \approx 0$ implies a consistent generative chain. The observed Z and the survival outcome t possess compatible values for the exposure, meaning the instrument's logic is sufficient to represent

the underlying process. A significant magnitude $|\hat{\Delta}_i^{(rx)}| \gg 0$ implies that the survival outcome demands a value for X that is fundamentally distinct from what the representation r can support. This allows researchers to identify specific subjects or subpopulations where the measurement protocol separated from either random noise or orthogonal features fails to maintain signal fidelity.

2.3.3. The implicit survival shift ($\mathcal{S}^{(\text{imp})}$)

In this section, we formalize the translation of the representation gap into the temporal aspect of the log-normal distribution special case through the designed implicit survival shift. This functional measurement isolates the magnitude of the log-lifetime perturbation attributable to the structural features that the measurement instrument fails to resolve.

Recall the log-normal AFT specification in Eq (2.9), where the log-lifetime $y_i^\star$ is generated by the combined linear predictor $\mu_i(\eta, \xi)$. We decompose the conditional expectation of the log-lifetime into explicit and implicit components:

$$\mathbb{E}[y_i^\star \mid \eta, \xi] = [\beta_0 + \mathbf{w}i^\top \boldsymbol{\gamma} + \beta_x(\tilde{\mathbf{g}}i^\top \boldsymbol{\eta})] + [\beta_x(\mathbf{b}i^\top \boldsymbol{\xi})]\,. \tag{2.20}$$

Here, $\beta_x(\mathbf{b}_i^\top \boldsymbol{\xi})$ represents the systematic deviation in expected survival driven by the orthogonal latent process. We define the implicit survival shift, $\mathcal{S}_i^{(\text{imp})}$, as the posterior expectation of the posterior correlation between the exposure sensitivity β_x and the latent coordinates ξ:

$$\begin{aligned} \mathcal{S}_i^{(\text{imp})} &= \int \Omega\beta_x(\mathbf{b}i^\top \boldsymbol{\xi}),\, p(\Omega \mid \mathcal{D}),\, d\Omega \\ &\approx \frac{1}{M} \sum m = 1^M \beta_x^{(m)} \Big(\sum k = 1^{d_\perp} B_{i,k} \xi_k^{(m)} \Big)\,. \end{aligned} \tag{2.21}$$

Equation (2.21) highlights the core mechanism of the process augmentation framework where the explicit recovery of information is lost during the measurement process. In standard survival analysis, the variation driven by the unmeasured features is mathematically indistinguishable within random noise, forcing the estimator to absorb structural defects into the residual variance σ_y^2. By augmenting the survival likelihood with the orthogonal implicit process $B\xi$, we build a theoretical separation between the reducible variance related to the observation's resolution limit and the irreducible randomness of the failure. We formally state that by proving that the total predictive uncertainty is not an integrated noise, but a sum of identifiable structural components.

Proposition 2.4 (Reliability variance decomposition)**.** *Let $y^\star$ denote the latent log-lifetime of a generic unit with structural coordinates $\tilde{\mathbf{g}}$ and $\mathbf{b}$. Under the spectral construction $\mathcal{S}_{raw} \perp \mathcal{S}_{raw}^\perp$ (Proposition 2.2) and the independence of the generative priors $\eta \perp \xi$, the total predictive variance of $y^\star$ decomposes additively:*

$$\text{Var}(y^\star) = \beta_x^2 \text{Var}(\tilde{\mathbf{g}}^\top \boldsymbol{\eta}) + \beta_x^2 \text{Var}(\mathbf{b}^\top \boldsymbol{\xi}) + \sigma_y^2. \tag{2.22}$$

Proof. We apply the law of total variance conditioning on the latent parameter set $\Theta = \{\eta, \xi, \beta_x, \sigma_y\}$. The derivation proceeds by expanding the structural variance of the linear predictor.

$$\begin{aligned}\mathrm{Var}(y^\star) &= \mathbb{E}_\Theta[\mathrm{Var}(y^\star \mid \Theta)] + \mathrm{Var}_\Theta(\mathbb{E}[y^\star \mid \Theta]) \\ &= \mathbb{E}_\Theta[\sigma_y^2] + \mathrm{Var}_\Theta(\mu(\eta, \xi)) \\ &= \sigma_y^2 + \mathrm{Var}_\Theta\left(\beta_0 + \mathbf{w}^\top \boldsymbol{\gamma} + \beta_x(\tilde{\mathbf{g}}^\top \boldsymbol{\eta} + \mathbf{b}^\top \boldsymbol{\xi})\right).\end{aligned}$$

Treating the covariate terms as fixed constants, we expand the variance of the composite exposure term using the bilinearity of the covariance operator:

$$\begin{aligned}\mathrm{Var}_\Theta(\mu) &= \beta_x^2 \mathrm{Var}(\tilde{\mathbf{g}}^\top \boldsymbol{\eta} + \mathbf{b}^\top \boldsymbol{\xi}) \\ &= \beta_x^2 \left[\mathrm{Var}(\tilde{\mathbf{g}}^\top \boldsymbol{\eta}) + \mathrm{Var}(\mathbf{b}^\top \boldsymbol{\xi}) + 2\mathrm{Cov}(\tilde{\mathbf{g}}^\top \boldsymbol{\eta}, \mathbf{b}^\top \boldsymbol{\xi})\right].\end{aligned}$$

By the construction of the subspaces and the independence of the priors, we get

$$\mathrm{Cov}(\tilde{\mathbf{g}}^\top \boldsymbol{\eta}, \mathbf{b}^\top \boldsymbol{\xi}) = \tilde{\mathbf{g}}^\top \mathbb{E}[\boldsymbol{\eta}\boldsymbol{\xi}^\top]\mathbf{b} - (\tilde{\mathbf{g}}^\top \mathbb{E}[\boldsymbol{\eta}])(\mathbf{b}^\top \mathbb{E}[\boldsymbol{\xi}]) = 0.$$

Thus we have

$$\mathrm{Var}(y^\star) = \beta_x^2 \mathrm{Var}(\tilde{\mathbf{g}}^\top \boldsymbol{\eta}) + \beta_x^2 \mathrm{Var}(\mathbf{b}^\top \boldsymbol{\xi}) + \sigma_y^2$$

which is composed by the structural variance of the explicit signal, the reducible uncertainty arising from the information gap, and the irreducible latent process noise. □

To operate this theoretical decomposition for reliability decision-making in survival analysis, we define latent risk probability (LRP), a Bayesian posterior probability estimate used for hypothesis testing of a directional inequality constraints (specifically, $H_0 : \mathcal{S}_i^{(\mathrm{imp})} \geq 0$ vs $H_1 : \mathcal{S}_i^{(\mathrm{imp})} < 0$). LRP determines whether the estimated survival shift is statistically significant relative to the irreducible latent process noise:

$$\mathrm{LRP}_i := \Pr(\mathcal{S}i^{(\mathrm{imp})} < 0 \mid \mathcal{D}) \approx \frac{1}{M} \sum m = 1^M \mathbb{I}\left(\beta_x^{(m)}(\mathbf{b}_i^\top \boldsymbol{\xi}^{(m)}) < 0\right).(\beta_x > 0). \tag{2.23}$$

Statistically, when $\mathrm{LRP}_i > 0.95$ it means we have strong evidence for a latent failure mode. It identifies specific units where the representation gap includes a structural effect that actively shortens the expected lifetime, identifying potential high-risk factors for observed Z_i. This measure serves as the primary post estimation metric for empirical study in Section 4, allowing us to separate hidden high and low risk market dynamics from stable observed volatility sequence.

3. Simulation study: Validating process augmentation

In the method section, we provide standard survival estimators are systematically biased when the latent exposure is observed only through a coarse, low-fidelity observation. To empirically validate this hypothesis and demonstrate the finite-sample performance of the proposed process-augmented lifetime distribution, we conducted a Monte Carlo simulation study. The primary objectives of this numerical experiment are to: (1) quantify the decay bias inherent in standard AFT models under severe representation loss, (2) verify the ability of the proposed generative chain ($X \to r \to Z$) with augmentation to recover the unbiased effect size, and (3) demonstrate the capacity of the model to resolve irreducible uncertainty within discrete observation bins, identifying latent risk factors that are implicit to the observation.

3.1. Generative mechanism and experimental design

We simulate a reliability surveillance scenario where the true stress acting on a component is continuous, but the recording instrument provides only a categorical count. This setting explicitly tests the model's ability to bridge the representation gap where information is structurally rebuilt rather than merely perturbed by white noise. For each of $n = 2,000$ subjects in one group, we draw the true, high-fidelity latent exposure x_i from a standard normal distribution:

$$x_i \sim \mathcal{N}(0, 1), \quad \text{for } i = 1, \ldots, n. \tag{3.1}$$

The standard x_i aligns with the identifiability constraint imposed by the ridge-type prior in our Bayesian specification (Section 2.2.1), ensuring that the scale of the recovered latent variable is identifiable relative to the regression coefficients.

For coarse measurement instrument, we apply a non-linear quantization operator $\mathcal{T}(\cdot)$ to the true exposure. Specifically, we discretize X_i into a four-level ordinal covariate Z_i using quartile-based width. This transformation reforms the continuous scenario into a discrete step function. The explicit basis $\tilde{G}$ (in 2.1) calculates each bin mean, while the orthogonal implicit process $B\xi$ contains the intra-bin variation.

We construct the failure times T_i followed by the log-Normal accelerated failure time (AFT) mechanism driven by the true latent exposure x_i:

$$\log(T_i) = \beta_0 + \beta_x x_i + \sigma_y \epsilon_i, \tag{3.2}$$

where the parameters are set as follows: $\beta_0 = 3.0$ initializes the baseline lifetime, $\beta_x = -0.8$ initializes the true structural effect of the stress predictor, indicating that higher stress reduces survival, and $\sigma_y = 0.6$ initializes the irreducible failure noise ($\epsilon_i \sim \mathcal{N}(0, 1)$). At last, we generate censoring times $C_i \sim \text{Exp}(\lambda = 0.05)$ with an approximate censoring rate of 30%.

3.2. Implementation and estimator specification

We implement the proposed process-augmentation estimator using the exact Gibbs sampling scheme detailed in Algorithm 1. The explicit basis matrix G is constructed via orthonormal step functions corresponding to the discrete levels of Z. We perform 50 independent Monte Carlo replicates. In each replicate, the Gibbs sampler was run for 2,500 iterations with a burn-in period of 500 iterations. We compare model performance with two benchmark models: the naive AFT Model (fitted on Z) and the truth estimator (fitted on the unobservable x).

3.3. Performance assessment: Bias and consistency

Table 1 summarizes the performance of the estimators across the 50 replicates. The results confirm the theoretical expectations regarding measurement error in survival analysis.

Table 1. Exact Gibbs sampler simulation results ($R = 50$ replicates) comparison of parameter estimation bias and MSE for the exposure coefficient β_x (Truth = -0.8).

Model	Mean Estimate ($\hat{\beta}_x$)	Mean Bias	MSE
Naive AFT (Benchmark)	-0.6603	0.1397	0.0201
Proposed (Process-Aug)	**-0.8145**	**-0.0145**	**0.0092**
Truth Estimator	-0.7992	0.0008	0.0004

The **naive AFT model** exhibits severe discrepancy (Bias $\approx$ 0.14), consistently underestimating the magnitude of the exposure. This occurs because the observation Z lacks the variance of the true exposure, degenerating the signal-to-noise ratio. In contrast, the proposed process augmentation estimator has a value of -0.8145, which is statistically indistinguishable from the true value of β_x given the sampling variability. The significant reduction in mean squared error (MSE) demonstrates that the generative chain bridges the representation gap, augmenting the observation with a valid latent predictor.

3.4. Feature discovery and uncertainty quantification

In the simulation study, our framework also provides its ability to resolve latent heterogeneity. We plot the posterior predictive distributions from the simulation run in Figure 1:

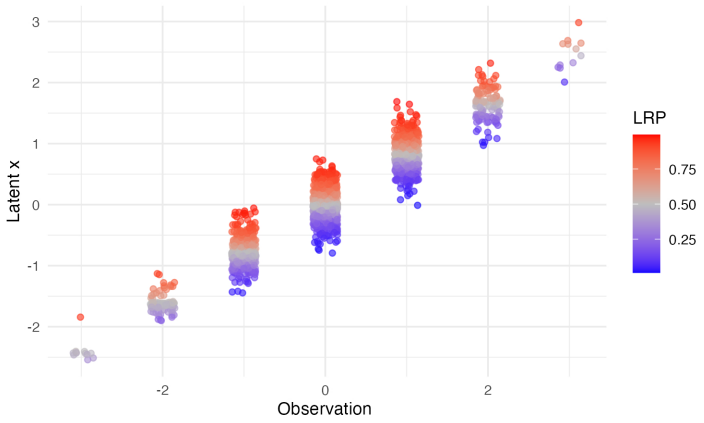

Figure 1. The scatter plot displays the ergodic averages of the recovered latent exposure ($\hat{x}$) conditioned on the coarse discretized observation Z. While standard regression approaches assume homoscedasticity within each discrete bin (e.g., $Z = 0$), the proposed generative chain recovers the intra-bin variance via the orthogonal implicit channel ($B\xi$). The chromatic scale indicates the latent risk probability. Red points ($\text{LRP}_i > 0.95$) identify specific subjects where the model statistically confirms a positive structural divergence from the observation.

Figure 1 visualizes the fundamental capacity of the process-augmentation framework to decompose latent process structure that is underlying survival observations. While standard regression approaches enforce the homoscedastic assumption within each observation bin, effectively collapses all risk variation to the bin centroid, our generative augmented chain decomposes the particle intra-bin heterogeneity, performed as the vertical dispersion of the latent exposure x orthogonal to the discretization observation. This decomposition allows for the precise classification of risk using the latent risk probability (LRP). Subjects clustered in red (LRP $>$ 0.95) are statistically identified as possessing a hidden negative indicator, representing a structural anomaly in observation where the implicit process $B\xi$ actively decomposes reliability; conversely, blue subjects (LRP $<$ 0.05), the hidden positive indicator, indicate inherent random noise that remains unextracted compared to the informative components.

Figure 2 demonstrates the translation of these implicit features into the reliability test by contrasting

the posterior predictive survival functions of two observational indicators. Where the benchmark estimator plots a single marginal prediction (dashed line), treating the representation gap as irreducible noise, our proposed framework stratifies the group into 2 risk lines with credible intervals. The subject with a hidden negative indicator (red line) and the subject with a hidden positive indicator (blue line) compose the whole survival probability dynamic, confirming the distinct impact of the implicit survival shift $\mathcal{S}^{(\text{imp})}$. Crucially, the 95% Bayesian credible intervals validates the variance decomposition established in Proposition 2.4: The detected divergence is not residual reducible noise, but resolved irreducible uncertainty is capable of converting measurement uncertainty into tractable risk decomposition in survival analysis.

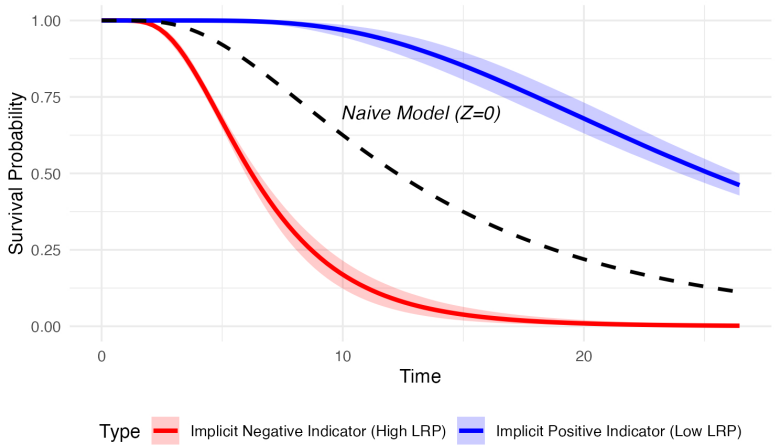

Figure 2. Comparison plot of estimated survival functions $\hat{S}(t)$ for two observationally equivalent subjects ($Z = 0$). The naive AFT estimator (black dashed line) generates a single marginal prediction for both subjects, implicitly assuming 0 measurement variance and treating the representation gap as random noise. Our proposed process-augmented framework in this study stratifies the subjects into two classes: implicit negative indicator (red, high LRP) and implicit positive indicator (blue, low LRP).

4. Empirical study: High-frequency volatility and market stability

To demonstrate the capability of our proposed PALD model in a real data analysis with severe structural discrepancy, in this section we analyze the duration of market stability during the 2020 financial crisis (see Figure 3). In high-frequency trading market, the true latent risk exposures are usually defined as the continuous integrated volatility of the price process, is unobservable and must be estimated from observed discrete price data. Given that the price sequence is contaminated by microstructure noise, this creates a representation gap where the observed covariate Z fails to capture the high-frequency liquidity shocks X that drive sudden market dynamics.

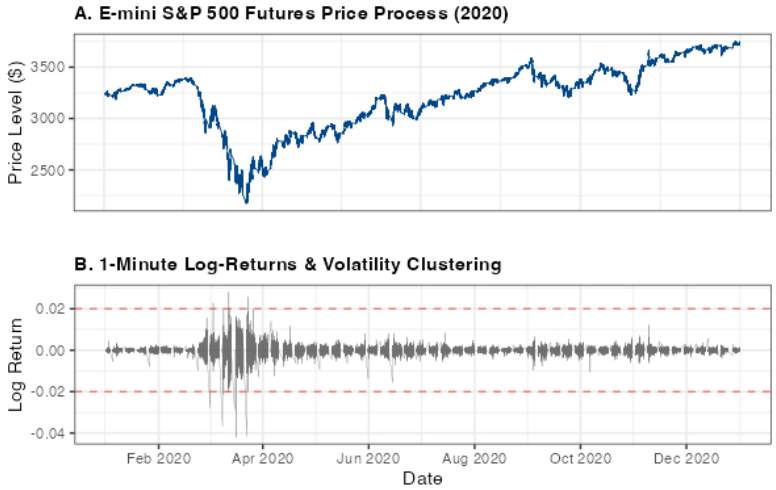

Figure 3. High-frequency market dynamics (2020). (A) E-mini S&P 500 futures price process showing the structural break during the onset of the COVID-19 pandemic. (B) The 1-minute intraday log-return sequence. The red dashed lines indicate the ±2% volatility threshold; trading sessions where the cumulative returns breach this envelope are classified as 'failure events' (T_i) in the survival analysis framework.

4.1. Data description and preprocessing

We utilize the high-frequency E-mini S&P 500 (ES) futures dataset from the year 2020. This specific trading window was selected to capture the transition from the low-volatility regime of early 2020 to the extreme liquidity crisis of March 2020 due to the pandemic (See Figure 3). The full dataset includes $N = 253$ trading days and each day has 23 hours trading activity; we focus our analysis on a subset of $n = 72$ trading days centered on the volatility clustering to rigorously test the model's stability to recover latent signals during sudden systematic breaks. For each trading day $i = 1, \ldots, n$, we observe the 1-minute logarithmic returns $r_{i,t}$. We frame the analysis as a first-hitting-time (FHT) problem [18, 27], where the goal is to predict the persistence of market stability given the coarsened pre-market risk signals.

To process the survival outcome from the sequence, we consider threshold regression studies in finance [1,15,17], which define the failure event not as a biological death, but as an exceeding volatility event which is the moment when the cumulative intraday stress crosses a critical stability barrier [31]. Let $P_{i,t}$ denote the price at minute t on day i. We define the failure time T_i as the minute when the cumulative absolute log-return exceeds the threshold γ:

$$T_i = \inf\left\{t > 0 : \left|\sum_{s=1}^{t} r_{i,s}\right| \geq 0.02\right\}, \tag{4.1}$$

where the threshold is set at ±2%, representing an extreme intraday crash. If the cumulative stress never exceeds this boundary by the end of the trading session (standard trading window $t = 390$), the observation is treated as right-censored ($\delta_i = 0$), implying the future market survived the day.

To construct the generative chain ($X \to r \to Z$) in Section 2, we denote the observed covariate Z_i by applying a quantity operator $\mathcal{T}(\cdot)$ to the previous day's realized volatility. We discretize continuous

volatility into quintiles, following an ordinal covariate $Z_i \in \{1, \ldots, 5\}$ representing volatility from low to extreme. While Z_i collapses all days within a quintile to a single risk level, the implicit channel ($B\xi$) is tasked with recovering intra-bin heterogeneity, which is the specific microstructure noise distinct to day i causing the crash time T_i.

4.2. Empirical analysis and results

The left panel in Figure 4 presents the Kaplan-Meier estimates stratified by volatility. While a global trend is visible ($p < 0.01$), the survival curves exhibit significant crossing hazards, particularly where the low volatility exhibits a steeper drop in stability than the medium during mid-session trading. This non-monotonic phenomenon can also be found in Table 2. Besides, Figure 4 illustrates the dynamics of the failure event at the right panel. One stable trading day (Jan 06) floats within ±2% bounds, while the failure day (Feb 25) exhibits a persistent accumulation of stress that undermines the survival probability over time. If analysts use a standard AFT estimator, which relies solely on the covariate, they treat the intra-bin heterogeneity in Figure 4 as irreducible noise (σ_y). In contrast, our framework utilizes the implicit survival shift ($S^{(imp)}$) in Section 2 to reconstruct these high-frequency trading sequences, recovering the structural risk information lost during the discretized process.

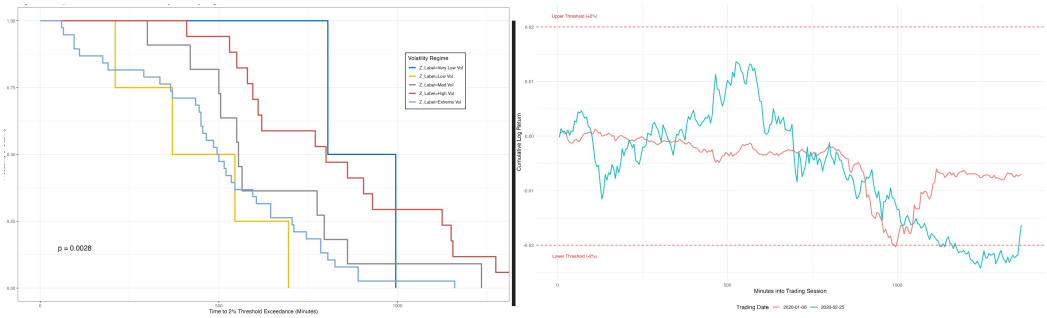

Figure 4. The left panel show the Kaplan-Meier estimates of stability duration stratified by volatility (Z). The crossing hazards between the low (yellow) and medium (grey) lines indicate that the observed covariate fails to order the latent risk correctly, indicating the implicit channel may exist. The right panel states the price intraday cumulative return paths illustrating the first-hitting-time (FHT) event. The green line (Feb 25) demonstrates a latent failure process accumulating stress until crossing the −2% bound, which is usually ignored in the daily price observation.

Table 2. Number of trading days at risk across volatility over time.

Z	t_0	t_{250}	t_{500}	t_{750}	t_{1000}	t_{1250}
1: Very Low	15	15	14	10	2	0
2: Low	14	12	8	3	1	0
3: Med	14	14	12	8	3	1
4: High	15	15	15	12	5	2
5: Extreme	14	12	10	6	2	0

To validate the predictive performance of the proposed process-augmented module, we use two established survival estimators: the semi-parametric Cox proportional hazards (PH) model [26] and the parametric Weibull accelerated failure time (AFT) model [20, 23] as benchmark models. The Cox PH estimator serves as a robust baseline, relaxing distributional assumptions on the hazard function to test whether the latent risk signals recovered by our model persist under semi-parametric conditions. Conversely, the Weibull AFT provides a direct parametric benchmark, testing the necessity of the log-normal specification for modeling market stability duration. We use the trade-off between out-of-sample fit and model complexity as the performance indicies. Since our proposed Bayesian estimator incorporates high-dimensional latent states $(x_1, \ldots, x_n)$, standard likelihood ratio tests are inappropriate. Instead, we choose the deviance information criterion (DIC) [37] as the Bayesian counterpart index to the Akaike information criterion (AIC) (see [2, 25] for standard reference) for the benchmark models. These penalized deviance measures are defined as

$$\begin{aligned} \text{AIC} &= -2\ln(\hat{L}) + 2k, \\ \text{DIC} &= \bar{D} + p_D = \mathbb{E}_{\Omega|\mathcal{D}}[-2\ln L(\Omega)] + p_D, \end{aligned} \tag{4.2}$$

where $\hat{L}$ denotes the maximum likelihood with k fixed parameters, and $\bar{D}$ represents the posterior mean deviance. The term p_D quantifies the effective number of parameters after the Bayesian sampler, allowing us to directly compare whether the information gained from the implicit channel ($B\xi$) balanced the computational complexity of the augmentation layer. The empirical results are shown below.

Table 3 presents the goodness-of-fit assessment across the candidate estimators. The DIC of the proposed process-augmented framework is 383.5, representing relatively competitive performance among the full-likelihood parametric estimators (log-normal AFT: AIC=425, Weibull AFT: AIC=430.4). This substantial deviance confirms that the inclusion of the implicit channel ($B\xi$) indeed recovers structural information that standard parametric models discard as residual noise. Although the Cox PH model also exhibits a relatively low AIC (387.5), this performance is achieved by the partial likelihood function, which does not consider the baseline hazard.

Table 3. Model performance comparison

Model	Estimator	Metric	Value
	Proposed (Process-Aug)	**DIC**	**383.5***
Parametric	Naive Log-Normal AFT	AIC	425.0
(Full Likelihood)	Weibull AFT	AIC	430.4
Semi-Parametric	Cox Proportional Hazards	AIC	387.5

*

The structural recovery can be seen in Figure 5, which contrasts the predictive dynamics for a representative trading session in the medium volatility regime ($Z = 3$). Standard estimators, including the naive AFT and Cox PH (dotted black), producing a single deterministic survival flow that conflates all sessions within the dynamic to a group average, cause a representation loss. The process-augmented estimator resolves the intra-bin heterogeneity, recovering a risk envelope bounded by hidden high-risk (solid red) and hidden low-risk (solid blue) latent states. This separation reveals that observations

*Note that Cox PH AIC is based on partial likelihood.

classified as observationally equivalent by the discrete covariate actually possess distinct failure modes: The high risk latent state identifies trading days where latent liquidity stress accelerates the time-to-crash, which is a crucial signal that remains unobserved in standard statistical benchmarks.

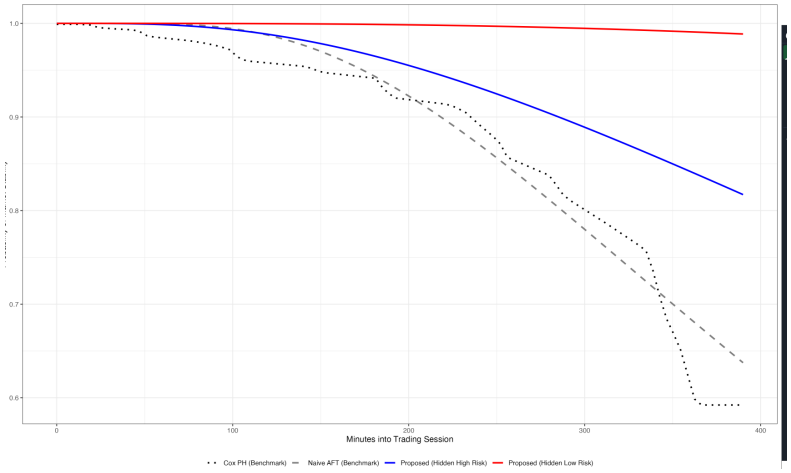

Figure 5. Standard benchmarks (grey and black dotted lines) produce single deterministic curve for the medium volatility ($Z = 3$). The process-augmented estimator resolves the hidden heterogeneity containing both hidden high risk (red line) and hidden low risk (blue line) dynamics.

4.3. Post-estimation in quantifying representation loss

In this section, we formalize implicit survival shift ($\mathcal{I}_i$) measure as the functional deviation from the average hazard, estimated by the benchmark, to the specific hazard recovered by the augmentation layer. To calculate, this metric measures the Euclidean distance from the survival line (dashed gray line) and the process-augmented line (solid red/blue lines) observed in Figure 5. We quantify the risk classification for individual trading sessions using the LRP in Section 2.3.3, defined as the posterior probability that a specific observation belongs to the hidden high risk:

$$\text{LRP}_i = P(S_i^{(imp)} < 0 \mid \mathcal{D}). \tag{4.3}$$

In Figure 5, the LRP acts as the probabilistic selection criterion that stratifies observationally equivalent volatility ($Z = 3$) into the distinct upper and lower bounds of the risk envelope. Specifically, the metric identifies that 7.1% of the sessions in this regime exhibit an LRP of 0.954, which separates them as hidden high risk events. This classification allows analysts to isolate the implicit channel of the trading sequence where latent liquidity stress accelerates the time-to-crash ($LRP > 0.95$), thus enriching the coarse volatility signal provided by the observed sequence with volatile, risk-sensitive structural information.

5. Discussion and conclusions

In this article, we address a fundamental confounding in reliability engineering and survival analysis: The topological incompatibility between continuous latent states and the resolution-limited discrete observations used to record them. While the vanilla statistical literatures found one solution

for this functional misspecification through flexible hazard estimators, which range from spline-based smoothing to deep learning extensions, these methodologies are highly assumed that the observed covariates are mathematically equivalent to the latent risk factors. We have argued that when the measurement instrument induces structural operations (e.g., discretization or censoring), standard estimators suffer from irreducible decay bias and a complete loss of intra-bin variance. To resolve this, we introduced the process-augmented lifetime distribution. By converting the discrepancy error decomposition of spatial statistics to the temporal domain, we constructed a hierarchical generative chain ($X \to r \to Z$) that redefines the representation gap not just as inherent noise, but as a structured, orthogonal signal extractable via Bayesian implementation.

Our theoretical contribution centers on the spectral decomposition of the latent exposure into explicit and implicit subspaces (Proposition 2.2). This formulation enables the derivation of the implicit survival shift ($\mathcal{S}^{(\text{imp})}$) and the latent risk probability (LRP), analytic posterior functionals that translate abstract matrix orthogonality into tractable reliability metrics. The simulation study empirically validated the efficacy of this framework, demonstrating that the proposed augmented estimator corrects the bias, reduces the mean squared error of the exposure comparing to the naive benchmark, and bridges the continuous latent states from observed discrete bins. Figures 1 and 2 highlight the practical implication of this augmentation: the ability to differentiate subjects possessing both negative and positive indicators within a group of observationally equivalent sample draws. Further, our empirical study in high-frequency data demonstrates the model's capacity to recover structural discrepancy in real-world scenarios, identifying a risk envelope of hidden liquidity stress within integrated volatility dynamics (Figure 5). By explicitly decomposing the total predictive variance into irreducible uncertainty and reducible uncertainty via Proposition 2.4, the framework provides a rigorous framework for distinguishing explicit measurement variance from irreducible failure randomness.

While this work establishes a rigorous mechanism for process-augmented lifetime distributions, the current formulation operates under specific constraints. First, the proposed Bayesian hierarchical model depends on isotropic Gaussian priors to maintain conjugacy while the log-normal AFT kernel ensures exact inference but limits the modeling of leptokurtic or asymmetric latent stresses often encountered in extreme-value reliability survival studies. Future research will discover scale-mixture process priors, such as the horseshoe or Dirichlet process mixtures, to accommodate non-Gaussian latent augmenting process. Second, the computational cost of the exact Gibbs sampler are affected by the dimension of the implicit subspace. To enhance scalability for high-dimensional sensor arrays, future work will transform exact MCMC type models to deterministic approximation approaches, such as variational Bayes (VB) or Hamiltonian Monte Carlo (HMC), as one potential methodological evolution. Finally, the generative chain of process augmentation is topologically flexible; future work could aim to embed this architecture within semi-parametric Cox proportional hazards or generalized Gamma frameworks, thereby extending the utility of discrepancy signatures to a broader class of biomedical and engineering inverse problems in survival analysis.

Author contributions

Xu Liu was responsible for constructing the PALD framework and the formulation of the generative Bayesian hierarchical model, designed and implemented the exact Gibbs sampler. He conducted the

Monte Carlo simulation study to validate estimator consistency and performed the empirical analysis on high-frequency market stability using E-mini S&P 500 futures data. Furthermore, he drafted the original manuscript, including the derivation of posterior functional measurements like the latent failure mode score.

Xufeng Niu assisted in the conceptualization of the representation loss framework. He contributed to the rigorous mathematical deduction of the spectral decomposition and the orthogonal projection properties. Also, he provided detailed revisions to the experimental design and reviewed the final manuscript to ensure theoretical alignment with current literature. All authors have read and approved the final version of the manuscript for publication.

Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

The authors declare no conflict of interest.

References

1. Y. Aït-Sahalia, J. Yu, High frequency market microstructure noise estimates and liquidity measures, *Ann. Appl. Stat.,* **3** (2009), 422–457. http://dx.doi.org/10.1214/08-AOAS200
2. H. Akaike, A new look at the statistical model identification, *IEEE T. Automat. Contr.,* **19** (1974), 716–723. http://dx.doi.org/10.1109/TAC.1974.1100705
3. B. Alves, J. Dias, Survival mixture models in behavioral scoring, *Expert Syst. Appl.,* **42** (2015), 3902–3910. http://dx.doi.org/10.1016/j.eswa.2014.12.036
4. J. Banasik, J. Crook, L. Thomas, Not if but when will borrowers default, *J. Oper. Res. Soc.,* **50** (1999), 1185–1190. http://dx.doi.org/10.1057/palgrave.jors.2600851
5. T. Bellotti, J. Crook, Forecasting and stress testing credit card default using dynamic models, *Int. J. Forecasting,* **29** (2013), 563–574. http://dx.doi.org/10.1016/j.ijforecast.2013.04.003
6. N. Benítez-Parejo, M. Rodríguez del Águila, S. Pérez-Vicente, Survival analysis and Cox regression, *Allergol. Immunopath.,* **39** (2011), 362–373. http://dx.doi.org/10.1016/j.aller.2011.07.007
7. J. Bradley, C. Wikle, S. Holan, Hierarchical models for spatial data with errors that are correlated with the latent process, *Stat. Sinica,* **30** (2020), 81–109.
8. J. Bradley, S. Zhou, X. Liu, Deep hierarchical generalized transformation models for spatio-temporal data with discrepancy errors, *Spat. Stat.,* **55** (2023), 100749. http://dx.doi.org/10.1016/j.spasta.2023.100749
9. J. Breeden, A. Bellotti, Y. Leonova, Instabilities in Cox proportional hazards models in credit risk, *J. Credit Risk,* **19** (2023), 29–55. http://dx.doi.org/10.21314/JCR.2022.014
10. S. Chaker, The signal and the noise volatilities, *Res. Int. Bus. Financ.,* **50** (2019), 79–105. http://dx.doi.org/10.1016/j.ribaf.2019.04.008

11. C. L. Cheng, J. W. Van Ness, *Statistical regression with measurement error*, Kendall's library of statistics, vol. 6, Arnold, London; co-published by Oxford University Press, New York, 1999. MR1719513.
12. D. Cox, Regression models and life-tables, *J. Roy. Stat. Soc. B*, **34** (1972), 187–202. http://dx.doi.org/10.1111/j.2517-6161.1972.tb00899.x
13. V. Djeundje, J. Crook, Identifying hidden patterns in credit risk survival data using generalised additive models, *Eur. J. Oper. Res.,* **277** (2019), 366–376. http://dx.doi.org/10.1016/j.ejor.2019.02.006
14. L. Dirick, T. Bellotti, G. Claeskens, B. Baesens, Macro-economic factors in credit risk calculations: Including time-varying covariates in mixture cure models, *J. Bus. Econ. Stat.,* **37** (2019), 40–53. http://dx.doi.org/10.1080/07350015.2016.1260471
15. R. F. Engle, J. R. Russell, Autoregressive conditional duration: A new model for irregularly spaced transaction data, *Econometrica,* **66** (1998), 1127–1162. http://dx.doi.org/10.2307/2999632
16. J. J. Hanfelt, K. Y. Liang, Approximate likelihoods for generalized linear errors-in-variables models, *J. Roy. Stat. Soc. B*, **59** (1997), 627–637. https://doi.org/10.1111/1467-9868.00087
17. P. R. Hansen, A. Lunde, Realized variance and market microstructure noise, *J. Bus. Econ. Stat.,* **24** (2006), 127–161. http://dx.doi.org/10.1198/073500106000000071
18. X. He, M. L. T. Lee, First-hitting-time based threshold regression, *Int. Encycl. Stat. Sci.*, Springer, Berlin, Heidelberg, 2011, 523–524. http://dx.doi.org/10.1007/978-3-642-04898-2_252
19. T. A. Hsieh, H. Choi, M. Kim, Multimodal representation loss between timed text and audio for regularized speech separation, *arXiv preprint*, 2024.
20. M. Ibrahim, H. Goual, M. K. Khaoula, A. H. Al-Nefaie, A. M. AboAlkhair, H. M. Yousof, A novel accelerated failure time model with risk analysis under actuarial data, censored and uncensored application, *Stat. Optim. Inform. Comput.*, **14** (2025), 1198–1225. http://dx.doi.org/10.19139/soic-2310-5070-2627
21. C. Jiang, Z. Wang, H. Zhao, A prediction-driven mixture cure model and its application in credit scoring, *Eur. J. Oper. Res.,* **277** (2019), 20–31. http://dx.doi.org/10.1016/j.ejor.2019.01.072
22. Y. Ju, S. Jeon, S. Sohn, Behavioral technology credit scoring model with time-dependent covariates for stress test, *Eur. J. Oper. Res.,* **242** (2015), 910–919. http://dx.doi.org/10.1016/j.ejor.2014.10.054
23. J. D. Kalbfleisch, R. L. Prentice, *The statistical analysis of failure time data*, Wiley Ser. Prob. Stat., John Wiley & Sons, 2Eds., 2002. http://dx.doi.org/10.1002/9781118032985.fmatter
24. E. Kaplan, P. Meier, Nonparametric estimation from incomplete observations, *J. Am. Stat. Assoc.,* **53** (1958), 457–481. http://dx.doi.org/10.2307/2281868
25. S. Konishi, G. Kitagawa, *Information criteria and statistical modeling*, Springer Series in Statistics, Springer, New York, 2008. http://dx.doi.org/10.1007/978-0-387-71887-3
26. I. Kuitunen, V. T. Ponkilainen, M. M. Uimonen, A. Eskelinen, A. Reito, Testing the proportional hazards assumption in cox regression and dealing with possible non-proportionality in total joint arthroplasty research: Methodological perspectives and review, *BMC Musculoskel. Dis.,* **22** (2021), 489. http://dx.doi.org/10.1186/s12891-021-04379-2
27. M. L. T. Lee, G. A. Whitmore, Threshold regression for survival analysis: Modeling event times by a stochastic process, *Stat. Sci.,* **21** (2006), 501–513. http://dx.doi.org/10.1214/088342306000000330

28. Z. Li, J. Crook, G. Andreeva, Y. Tang, Predicting the risk of financial distress using corporate governance measures, *Pac.-Basin Financ. J.,* **68** (2021), 101334. http://dx.doi.org/10.1016/j.pacfin.2020.101334
29. S. Luo, X. Kong, T. Nie, Spline based survival model for credit risk modeling, *Eur. J. Oper. Res.,* **253** (2016), 869–879. http://dx.doi.org/10.1016/j.ejor.2016.02.050
30. D. Machin, Y. Cheung, M. Parmar, *Survival analysis: A practical approach*, John Wiley & Sons, Ltd, 2006. http://dx.doi.org/10.1002/0470034572
31. G. Marinos, M. Karvounis, I. N. Athanasiadis, *An innovative framework for threshold exceedance forecasting in timeseries using survival analysis*, Knowledge Discovery, Knowledge Engineering and Knowledge Management (Communications in Computer and Information Science), **2454** (2025), 296–316. http://dx.doi.org/10.1007/978-3-031-87569-4_14
32. T. Moon, S. Sohn, Survival analysis for technology credit scoring adjusting total perception, *J. Oper. Res. Soc.,* **62** (2011), 1159–1168. http://dx.doi.org/10.1057/jors.2010.80
33. B. Narain, *Survival analysis and the credit granting decision*, In: L. Thomas, D. Edelman, J. Crook (eds), Readings in Credit Scoring: Foundations, Developments, and Aims, Chapter 16, Oxford University Press, 1992, 236–247. http://dx.doi.org/10.1093/oso/9780198527978.003.0022
34. L. H. Nghiem, M. C. Byrd, C. J. Potgieter, Estimation in linear errors-in-variables models with unknown error distribution, *Biometrika,* **107** (2020), 841–856. https://doi.org/10.1093/biomet/asaa025
35. H. Noh, T. Roh, I. Han, Prognostic personal credit risk model considering censored information, *Expert Syst. Appl.,* **28** (2005), 753–762. http://dx.doi.org/10.1016/j.eswa.2004.12.032
36. M. Othus, B. Barlogie, M. L. LeBlanc, J. J. Crowley, Cure models as a useful statistical tool for analyzing survival, *Clin. Cancer Res.,* **18** (2012), 3731–3736. 10.1158/1078-0432.CCR-11-2859
37. D. J. Spiegelhalter, N. G. Best, B. P. Carlin, A. van der Linde, Bayesian measures of model complexity and fit, *J. Roy. Stat. Soc. B,* **64** (2002), 583–639. http://dx.doi.org/10.1111/1467-9868.00353
38. E. Tong, C. Mues, L. Thomas, Mixture cure models in credit scoring: If and when borrowers default, *Eur. J. Oper. Res.,* **218** (2012), 132–139. http://dx.doi.org/10.1016/j.ejor.2011.10.007
39. P. Wang, Y. Li, C. Reddy, Machine learning for survival analysis: A survey, *ACM Comput. Surv.,* **51** (2019), 110. http://dx.doi.org/10.1145/3214306
40. Z. Wang, C. Jiang, Y. Ding, X. Lyu, Y. Liu, A novel behavioral scoring model for estimating probability of default over time in peer-to-peer lending, *Electronic Commer. Res. Appl.,* **27** (2018), 74–82. http://dx.doi.org/10.1016/j.elerap.2017.12.006
41. C. Zheng, J. Zhu, X. Fan, S. Chen, Z. Zhang, Promoting variable effect consistency in mixture cure model for credit scoring, *Discrete Dyn. Nat. Soc.,* **2022** (2022), 3112987. http://dx.doi.org/10.1155/2022/3112987

AIMS Press