Penalised Spline-Based Multinomial Restricted Unfolding: Methodology, Simulation, and Application

Barnes, Finn

Penalised Spline-Based Multinomial Restricted Unfolding: Methodology, Simulation, and Application

Finn George Kent Barnes

Master's Thesis Psychology,

Methodology and Statistics Unit, Institute of Psychology

Faculty of Social and Behavioral Sciences, Leiden University

Date: 09/04/2026

Student number: 4059379

Supervisor: Dr. Mark de Rooij (internal)

Abstract

Multinomial Restricted Unfolding (MRU) is a geometric framework for modelling categorical outcomes by embedding observations and response categories within a shared low-dimensional latent space, where class membership probabilities are determined by Euclidean distances. Although this formulation captures non-linear structure in the latent space, the mapping from observed predictors to latent coordinates is linear, limiting the model's ability to represent predictor–class relationships. This study extends the MRU framework by introducing B-spline transformations of the predictors, enabling flexible modelling of covariate effects. A ridge penalty is incorporated into the negative log-likelihood objective function to control model complexity and reduce over-fitting. The proposed approach is evaluated through a simulation study manipulating sample size and latent class probability sharpness, alongside a comparative application to the *lmap* liver data set. Simulation results show that the spline-based penalised MRU can recover non-linear predictor–class relationships to a moderate degree, with performance improving as sample size increases. In the comparison analysis, spline-based model achieved a fit and predictive accuracy comparable to the original MRU. Overall, the results demonstrate that penalised spline extension can enhance the representational capacity of MRU though effectiveness is moderated by their distributional properties and homogeneity of predictor effects.

1 Introduction

Multinomial Restricted Unfolding (MRU) is a relatively new statistical framework used to explore an observations probability of class membership to a categorical variable. This is achieved by placing both observations and classes within a shared dimensional space, where the Euclidean distance between an observation and a class point reflects the probability of membership; the closer the distance, the higher the probability. This emerging area of research is constructed from the well-established foundations laid down in multidimensional unfolding, and the logistic regression techniques. Multidimensional unfolding (MDU) was originally proposed in 1958 by Warren S. Torgerson as an extension to Multidimensional scaling (MDS), and describes the process of reducing data onto a shared low dimensional plane. Logistic regression similarly was introduced by Cox in 1958, rounding out a bumper year for statisticians. This model proposed that the probability of a categorical outcome variable could be predicted given the input predictors using a link function.

The first formal implementation of the Multinomial Restricted Unfolding (MRU) model was introduced by De Rooij and Busing (2024), who established the mathematical framework for modelling multinomial outcome data. The model operationalized a shared low-dimensional latent space in which the probability of class membership was determined by the Euclidean distances between observation points $U \in \mathbb{R}^D$ and class points $V \in \mathbb{R}^D$. Here, U denotes the matrix of latent coordinates for all n observations, and V represents the matrix of latent coordinates for the C outcome classes, each embedded in the same D-dimensional space. The use of the Euclidean distance metric was a key tenet of this approach, as it preserved isotropy (distances are treated equivalently in all directions) and adhered to the Pythagorean theorem. Consequently, the resulting visualizations remained geometrically interpretable, with shorter distances indicating a higher likelihood of class endorsement.

Let $u_i \in \mathbb{R}^D$ denote the latent coordinate of observation i, and let $v_c \in \mathbb{R}^D$ denote the latent coordinate of class c. The Euclidean distance between observation i and class c is then given by:

$$d_{ic} = \|u_i - v_c\|_2$$

and the model-predicted class probabilities are defined as:

$$\hat{G}_{ic} = \frac{\exp(-d_{ic})}{\sum_{c'=1}^{C} \exp(-d_{ic'})}.$$

The model deviance, defined as twice the negative log-likelihood, was given by:

$$\text{Deviance} = -2 \sum_{i=1}^{n} \sum_{c=1}^{C} G_{ic} \log(\hat{G}_{ic}),$$

where $G_{ic} \in \{0, 1\}$ represents the observed class membership of observation i.

In addition to estimating the latent class coordinates V, the MRU model links observed predictor variables to the latent configuration of observations. Specifically, each observation's latent position u_i is modeled as a linear function of its predictor vector $x_i \in \mathbb{R}^p$, so that

$U = XB,$

where $X \in \mathbb{R}^{n \times p}$ denotes the matrix of observed predictors and $B \in \mathbb{R}^{p \times D}$ is a matrix of regression weights mapping predictors into the D-dimensional latent space. In this way, the geometric representation is not estimated freely for each observation, but is instead structured by the observed predictor information.

To estimate the model parameters, De Rooij and Busing employed a majorization–minimization (MM) optimisation strategy, which iteratively updated the latent configuration to minimize the deviance defined above. At each iteration, a surrogate function was constructed that upper-bounded the objective, ensuring that the deviance decreased monotonically. This property provided stable convergence toward a minimum of the likelihood surface. The deviance therefore served as the optimisation criterion throughout estimation.

Although this formulation effectively captures non-linear relationships between observations (U) and classes (V) through its Euclidean distance structure, the mapping from the observed predictors (X) to the latent space remains inherently linear ($U = XB$), where B denotes the regression weight matrix. However, this linear mapping limits the model's ability to capture curved relationships among predictors, thereby constraining the model in settings where predictor–class relationships are governed by nonlinear structure.

This limitation is particularly salient given empirical evidence that many real-world human behavioral processes are fundamentally nonlinear. For instance, Kim et al. (2007) demonstrated that human decision-making follows intrinsically nonlinear patterns, and that models capable of capturing such non-linearity achieve higher predictive validity. These findings align with earlier and subsequent research highlighting the pervasiveness of nonlinear mechanisms in cognitive and judgmental processes (Einhorn, 1970; Piegat and Sałabun, 2012; Schweizer and Vahlne, 2022). In applied domains such as consumer choice (Vallarino, 2025) and psychological profiling (Failenschmid et al., 2024), nonlinear relationships between predictors and preferences are not the exception but the norm—underscoring the need for models that can accommodate such complex dependencies. Accordingly, in this study the MRU framework is extended by incorporating non-linear transformations of the predictors (X). This offers a principled means of enhancing its representational flexibility and broadening its utility across domains where relationships between covariates and class membership are inherently non-linear.

To address this limitation, the present study develops and implements a spline-augmented extension of the MRU model, in which spline-based basis expansions of the raw predictors (X) are integrated directly into the latent mapping function. This required the construction of a modified optimisation procedure in which B-spline transformations are embedded within the MRU likelihood and estimated jointly under ridge regularisation. This algorithmic extension enables the predictor-to-latent mapping to flexibly capture non-linear relationships while preserving the geometric interpretation central to the unfolding framework.

In principle, the implementation of splines is supported by prior literature (Lu and Clarkson,1999; Wezel and Kosters, 2004; Winsberg and Ramsay, 1980) within the broader category of dimension reduction and non-metric multidimensional techniques. Winsberg and Ramsay provide a strong theoretical basis for the use of smooth transformations in multidimensional scaling, illustrating how flexible basis functions, such as B-splines, can be employed to capture nonlinear structures in unfolding models. Lu & Clarkson (1999) further demonstrate the value of splines in enhancing the interpretability and fit of multidimensional preference data through penalized smoothing. Similarly, Wezel and Kosters (2004) explore spline-based transformations in the context of MDU on ordinal data, highlighting spline basis functions ability to preserve rank order information while accommodating complex patterns in the data. Collectively, these studies establish both the feasibility and theoretical justification for incorporating spline transformations within the MRU

framework, thereby advancing its capacity to model more complex relationships between predictors and class memberships.

However, although incorporating additional non-linear mechanics may enhance the expressive capacity of the MRU framework, it concurrently heightens the risk of over-parameterization and adds complexity to the optimization process, which may compromise the model's stability and generalizability. Theoretically, this issue reflects the classical bias–variance trade-off. The incorporation of additional non-linear mappings within the MRU framework reduces structural bias by allowing more flexible functional relationships between predictors and the latent space. However, this relaxation simultaneously increases variance, as the expanded parameter space enables the model to adapt, potentially excessively, to sample-specific fluctuations. As a result, while the model's apparent in-sample performance may improve, its out-of-sample predictive accuracy and overall stability can deteriorate.

To mitigate this risk, the present study adopts a penalized unfolding framework. Penalization imposes explicit constraints on the model parameters, thereby discouraging excessive flexibility through the shrinkage of large coefficient estimates toward zero. This regularization serves as a principled form of complexity control, sacrificing a small amount of local adaptation in exchange for substantially improved numerical stability and out-of-sample generalisation. Within the MRU context, this is implemented by augmenting the model's objective function with a ridge penalty term that constrains the magnitude of the regression weight matrix (B). In doing so, the penalized MRU preserves the interpretability and geometric coherence of the original model, while mitigating the over-fitting risk introduced by the additional non-linear mechanics

The proposed penalized spline-based MRU model is characterized by three primary tunable parameters: the spline degrees of freedom (df), which control the granularity of the non-linear transformation; the spline polynomial degree (pd), which determines the local curvature; and the regularization penalty (λ), which regulates the overall model complexity. To achieve an optimal balance between model flexibility and generalization, these parameters were selected using a nested cross-validation (CV) procedure. In the inner loop, candidate combinations of $(\text{df}, \text{pd}, \lambda)$ were evaluated by minimizing the mean penalized negative log-likelihood (PNLL) on validation folds, with classification accuracy serving as a secondary performance criterion. In the outer loop, the best-performing hyperparameters were applied to held-out data to obtain estimates of predictive performance. The search space spanned both parsimonious and flexible model configurations, with $\text{df} \in 4, \ldots, 10$, $\text{pd} \in 1, 2, 3$, and $\lambda \in 10^{-3,\ldots,1}$. The final model was then refitted using the optimal parameters identified from the outer loop and subsequently evaluated for both in-sample fit and out-of-sample generalization.

Finally, this study undertakes a comparative evaluation of the proposed penalized spline-based MRU model against the original MRU formulation. The primary goal is to determine whether the added flexibility of spline-based transformations improves model fit and recovery of predictor–class relationships, and to identify the data conditions under which such flexibility yields meaningful advantages. The comparison is designed to disentangle gains in flexibility from potential losses by examining model behaviour relative to the established approach in both empirical and controlled settings. A simulation study is implemented to assess how the proposed model performs across varying predictor structures, probabilistic discrimination between classes, and sample size, enabling direct evaluation of its ability to recover known predictor–class relationships and to generalise to unseen data. In combination, these complementary analyses clarify when spline-based extensions provide substantive improvements over the original MRU framework and differentiate the conditions under which such extensions operate most effectively.

Hypothesis 1: The penalized spline-based MRU model will improve model fit and more accurately recover predictor–class relationships compared with the originally formulated MRU.

Hypothesis 2: The spline extension will enable the modeling of non-linear effects of the predictors.

2 The Model

2.1 Spline Integration

To model the relationships between predictors and class membership within the Multinomial Restricted Unfolding (MRU) framework, each predictor $x_p \in \mathbb{R}^n$ was transformed via a B-spline basis expansion. This transformation enabled the model to approximate complex, non-linear response surfaces while maintaining sufficient smoothness to preserve interpretability within the optimization structure of the MRU framework. Basis-splines (B-splines) are a family of piece-wise polynomial functions that are locally defined over segments of the predictor domain (De Boor, 2001). These segments are partitioned by a sequence of knots, denoted:

$$\{\kappa_1, \kappa_2, \ldots, \kappa_K\},$$

which divide the range of x_p into contiguous intervals. Within each interval, a polynomial of degree pd is fitted, and the pieces are joined at the knot positions such that the resulting function is continuous and differentiable. The flexibility of a B-spline is therefore determined by both its pd (local curvature/flexibility) and df (granularity).

In the current model the determination and placement of internal knots K was governed by the relationship:

$$K = \mathrm{df} - \mathrm{pd} - 1,$$

This ensured that the total number of basis functions equals df. Thus, each predictor contributes df columns to the expanded design matrix.

To promote data lead flexibility, internal knots were placed according to empirical quantiles of each predictor x_p:

$$\kappa_j = Q_{x_p}\left(\frac{j}{K+1}\right), \qquad j = 1, \ldots, K,$$

where $Q_{x_p}(\cdot)$ is the empirical quantile function of x_p. This quantile-based knot placement ensured that regions with higher data density receive a greater concentration of knots, allowing the spline to capture subtle local curvature, while sparser regions are modeled more smoothly (Perperoglou et al., 2019; Harrell, 2015). This adaptive allocation of flexibility helps prevent over-fitting and promoted a parsimonious representation of non-linear effects.

The spline basis matrix was constructed using the df and pd. The corresponding B-spline basis functions were then evaluated at all n observations for each predictor x_p. The resulting basis matrix for predictor x_p, denoted S_{x_p}, was defined as follows:

$$S_{x_p} = \begin{bmatrix} \phi_1(x_{p1}) & \phi_2(x_{p1}) & \cdots & \phi_{\mathrm{df}}(x_{p1}) \\ \phi_1(x_{p2}) & \phi_2(x_{p2}) & \cdots & \phi_{\mathrm{df}}(x_{p2}) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(x_{pn}) & \phi_2(x_{pn}) & \cdots & \phi_{\mathrm{df}}(x_{pn}) \end{bmatrix} \in \mathbb{R}^{n \times \mathrm{df}},$$

where $\phi_j(\cdot)$ denotes the j-th B-spline basis function. Thus, S_{x_p} contained one column for each basis function of predictor x_p, and one row per observation. Although spline models can in principle assign a different number of basis functions df_p to each predictor, the present implementation imposed a common df across all predictors $p = 1, \ldots, P$. All spline basis matrices across predictors are then concatenated column-wise to form the full spline-expanded design matrix:

$$X_{\mathrm{spline}} = \left[\, S_{x_1}, S_{x_2}, \ldots, S_{x_P} \,\right] \in \mathbb{R}^{n \times P_s},$$

where $P_s = P \times \mathrm{df}$ represents the total number of basis-expanded predictor columns. Each column of X_{spline} corresponds to one basis function from the spline expansion of a predictor.

All columns of X_{spline} were then standardized to have zero mean and unit variance. This step improved numerical stability in the estimation of the regression coefficients (B), particularly because spline expansions may induce collinearity due to adjacent basis function overlap in their regions of support and thus co-active across subsets of observations (Liu, Nassar & Podgórski, 2020; De Boor, 2001). Additionally, standardization also ensured that all predictors contribute comparably to the latent unfolding process due to the mitigation of scale differences (Lee, 2015).

In summary, the spline integration process reformulates the original covariates x_p into smooth, data-adaptive basis functions $\phi(x_p)$, providing a flexible and stable representation of potentially non-linear relationships between predictors and class membership. The implementation with empirical knot placement and standardization meant the spline-expanded design matrix X_{spline} balanced local adaptability with global smoothness. This transformation enabled the MRU model to capture complex curvature in the predictor–class relationship that the previous formulation cannot represent, while preserving the additive and interpretable structure that underpins the unfolding framework.

2.2 Latent Space Representation

The spline-expanded predictor matrix X_{spline} is linearly mapped into a D-dimensional latent space through the regression weight matrix $B \in \mathbb{R}^{P_s \times D}$:

$$u_i = X_{\mathrm{spline},i} B = B^\top \phi(x_i),$$

where $\phi(x_i)$ denotes the i-th row vector of basis evaluations (so $X_{\mathrm{spline},i} = \phi(x_i)$), and $u_i \in \mathbb{R}^D$ represents the latent coordinates associated with observation i. The matrix B therefore defines a linear mapping from the spline-expanded covariate space $\mathbb{R}^{P_s}$ to the latent unfolding space $\mathbb{R}^D$. Collectively, these coordinates form the matrix $U = X_{\mathrm{spline}} B \in \mathbb{R}^{n \times D}$, which captures how individual observations were positioned within the latent space as a function of their predictors.

Likewise in this same latent space, each response category $c = 1, \ldots, C$ was represented by a class point $v_c \in \mathbb{R}^D$, with all class coordinates collected in the matrix $V \in \mathbb{R}^{C \times D}$. The relative positioning of U and V define the underlying geometric structure of the model.

2.3 Probabilistic Assignment in Latent Space

The MRU framework describes class probabilities as a function of the Euclidean distances between observation points (U) and class points (V) in the latent space (De Rooij & Busing, 2024). For observation i, smaller distances to a class point correspond to higher probabilities of class membership, while larger distances correspond to lower probabilities.

To convert distances into probabilities, the model uses an exponential distance-weighting function followed by normalisation across all classes. This formulation ensures that each row of $\hat{G}$ sums to one, so that classes closer in latent space receive higher predicted probabilities. The exponential weighting effectively implements a distance-based decay mechanism such that as the distance d_{ic} increases, the corresponding probability $\hat{G}_{ic}$ decreases exponentially.

This latent-space representation serves as a D-dimensional geometric embedding of the multinomial response structure. The positions of observations (u_i) and classes (v_c) are optimized jointly to minimize the discrepancy between the observed class memberships G and the modeled probabilities $\hat{G}$.

2.4 Majorization–Minimization Optimization

Model estimation proceeded by minimizing the penalized negative log-likelihood (PNLL) of the MRU model. This objective combined the multinomial log-likelihood with a ridge penalty on the regression weight matrix B:

$$\mathrm{PNLL}(B, V) = -\sum_{i=1}^{n}\sum_{c=1}^{C} G_{ic}\,\log\big(\hat{G}_{ic}\big) \;+\; \lambda\|B\|_F^2,$$

where $\hat{G}_{ic}$ denotes the model-predicted probability of class c for observation i, and $|\cdot|_F$ is the Frobenius norm. To minimise this objective, the model uses a majorization–minimization (MM) algorithm, which iteratively constructs and minimizes a sequence of quadratic surrogate functions that upper-bound the true objective. Each iteration guarantees a non-increasing PNLL, providing stable convergence even for complex loss functions.

At each iteration $t = 1, 2, \ldots$, the algorithm proceeds as follows. First, the current latent positions of observations are computed as $u_i^{(t)} = X_{\mathrm{spline},i}B^{(t)}$, while class points are represented by $v_c^{(t)}$. The Euclidean distances between observation and class points are then evaluated in the latent space.

These distances are subsequently converted into predicted class probabilities using an exponential distance-weighting function, such that observations are more likely to be assigned to nearby class points and less likely to distant ones.

A working response matrix $Z^{(t)}$ is then constructed to capture the local discrepancies between the observed and modelled class probabilities at the current estimate. Elementwise, this is defined as

$$Z_{ic}^{(t)} = d_{ic}^{(t)} + 4\big(\hat{G}_{ic}^{(t)} - G_{ic}\big),$$

where G is the observed class-indicator matrix and $d^{(t)} \in \mathbb{R}^{n\times C}$ is the matrix of pairwise Euclidean distances between the current latent observation coordinates and class points. In matrix form, written as $Z^{(t)} = d^{(t)} + 4(\hat{G}^{(t)} - G)$. Functionally, this matrix serves as a local approximation of the residual distances, fulfilling a role analogous to the working response in iteratively re-weighted least squares (IRLS).

To synthesise the quadratic majorizing function, two auxiliary matrices W and A were defined element-wise according to the sign patterns of $d^{(t)}$ and $Z^{(t)}$. These matrices determined the curvature and linear shift of the surrogate objective, respectively.

$$\begin{aligned}
&W_{ic} = \tfrac{1}{8}, \quad A_{ic} = \tfrac{1}{8}\,\frac{Z_{ic}}{d_{ic}} && \text{if } Z_{ic} > 0,\ d_{ic} > 0,\\
&W_{ic} = \tfrac{1}{8}, \quad A_{ic} = 0 && \text{if } Z_{ic} > 0,\ d_{ic} = 0,\\
&W_{ic} = \tfrac{1}{8}\,\frac{d_{ic} + |Z_{ic}|}{d_{ic}}, \quad A_{ic} = 0 && \text{if } Z_{ic} < 0,\ d_{ic} > 0,\\
&W_{ic} = \tfrac{1}{8}\,\frac{\varepsilon + Z_{ic}^2}{\varepsilon}, \quad A_{ic} = 0 && \text{if } Z_{ic} < 0,\ d_{ic} = 0.
\end{aligned}$$

Here, ε was a small positive constant to avoid division by zero. The weights W and A ensured that the surrogate objective always majorized the true penalized likelihood, guaranteeing a descent direction in every iteration.

Given the surrogate weights, the updated class configuration $V^{(t+1)}$ was obtained by solving the following weighted least-squares equation:

$$V^{(t+1)} = \big(\mathrm{diag}(W^\top \mathbf{1})\big)^{-1}\big(\mathrm{diag}(A^\top \mathbf{1})V^{(t)} - A^\top U^{(t)} + W^\top U^{(t)}\big),$$

where $\mathbf{1}$ denotes a column vector of ones.

This step updated the positions of the class points in the latent space such that they minimize the weighted discrepancy between observed and predicted distances, given the current configuration of U. The regression weight matrix B was updated next by minimizing the penalized quadratic surrogate with respect to B. This leads to the following system of penalized normal equations:

$$\Big(I_D \otimes \big(X_{\text{spline}}^\top D_W X_{\text{spline}} + \lambda I_{P_s}\big)\Big)\,\text{vec}\big(B^{(t+1)}\big) = \text{vec}\Big(X_{\text{spline}}^\top \big(D_A U^{(t)} - A V^{(t+1)} + W V^{(t+1)}\big)\Big).$$

where λ is the ridge regularization parameter, $\otimes$ denotes the Kronecker product, and $\text{vec}(\cdot)$ is the vectorization operator. Here, $D_W = \text{diag}(W\mathbf{1}_C)$ and $D_A = \text{diag}(A\mathbf{1}_C)$ denote diagonal matrices formed from the row sums of W and A, respectively. Solving this system yields $B^{(t+1)}$, which is then reshaped into its matrix form $P_s \times D$. The latent coordinates of all observations are then recomputed as:

$$U^{(t+1)} = X_{\text{spline}} B^{(t+1)}.$$

The introduction of spline-based transformations considerably expands the dimensionality of the predictor matrix X_{spline}, while also potentially inducing correlations among adjacent basis functions due to their overlapping regions of support. Consequently, the matrix $X_{\text{spline}}^\top W X_{\text{spline}}$ in the penalized normal equations may become ill-conditioned (i.e., its smallest eigenvalues approach zero), making numerical inversion more unstable, as small computational or sampling variations can produce disproportionately large changes in the estimated coefficients. In addition, the exponential transformation of Euclidean distances within the MM framework increases the overall non-convexity of the optimization surface, thereby heightening sensitivity to initialization and the risk of local minima. Together, these effects add both computational and numerical complexity to the estimation procedure.

The inclusion of the ridge penalty term, $\lambda\|B\|_F^2$, directly counteracts these issues by improving the conditioning of the covariance matrix and controlling the effective model complexity. The addition of λI to the weighted cross-product $X_{\text{spline}}^\top W X_{\text{spline}}$, inflates its smaller eigenvalues, preventing numerical singularities and reducing sensitivity to collinearity among spline basis functions. This regularization effectively shrinks the regression coefficients toward zero, discouraging overly flexible mappings that could overfit sample-specific variation. Thus, the ridge penalty stabilizes the covariance structure, mitigates over-parameterization, and promotes smoother, more reliable convergence across MM iterations.

Conceptually, the MM updates in the penalized spline-based MRU framework are closely related to the SMACOF algorithm used in multidimensional scaling (De Leeuw, 1977). In both approaches, each iteration minimizes a locally weighted least-squares objective derived from a majorizing function, ensuring a monotonic descent of the objective value. In the present context, this principle is extended to accommodate spline-transformed predictors and ridge regularization, where the surrogate weights W and A determine the local curvature of the penalized likelihood surface. Consequently, the algorithm inherits the interpretability and convergence guarantees of SMACOF-style majorization while adapting them to the high-dimensional, regularized setting required for stable estimation under spline-based expansions.

Finally, after each iteration, the penalized negative log-likelihood is evaluated. The algorithm terminates when both the absolute and relative changes in PNLL fall below a predefined tolerance δ, indicating convergence.

In summary, the MM optimization scheme iteratively refined the latent configuration of observations and classes by alternating between closed-form updates for V and B under a majorizing surrogate of the penalized log-likelihood. Each iteration ensured a monotonic decrease in the PNLL while accommodating spline-expanded predictors and ridge regularization. The use of surrogate weights W and A stabilized the optimization landscape, and the alternating least-squares structure enabled efficient updates of both latent coordinates and regression parameters.

The iterative MM scheme requires initial configurations for both the latent observation coordinates and class locations. These starting values are obtained using a basis function–adjusted discriminant analysis procedure, described next.

2.5 Basis Function–Adjusted Discriminant Analysis

The latent observation and class configurations were initialized through a Discriminant Analysis procedure. This approach extends a linear discriminant analysis–based initialization to incorporate the high-dimensional structure of the spline-expanded predictor matrix X_{spline}. The objective is to obtain informative starting values for the regression weight matrix B and the class configuration V, which together determine the initial latent representation of the observations and classes. This improves the convergence behaviour of the subsequent MM algorithm.

First, the class centroids were computed in the spline-expanded predictor space:

$$M = (G^\top G)^{-1} G^\top X_{\text{spline}} \in \mathbb{R}^{C \times P_s},$$

The projection matrix B_0 was then obtained by performing LDA on X_{spline} with class labels y. The LDA solution yields

$B_0 \in \mathbb{R}^{P_s \times D}$,

which identifies the directions of maximal class separation in the spline-expanded space. Projecting the observations onto these discriminant directions provides the initial latent observation coordinates:

$U_0 = X_{\text{spline}} B_0$.

The corresponding initial class coordinates are obtained by projecting the class centroids into the same discriminant space:

$V_0 = M B_0 \in \mathbb{R}^{C \times D}$.

In summary, U_0 and V_0 provide deterministic starting values for the MM algorithm. Since LDA accounts for covariance structure and class separation, this initialization provides a simple, efficient, and numerically stable alternative to random starts while remaining fully consistent with the spline-expanded representation.

3 Methods

3.1 Simulation Data generation

The simulation established a population with explicit linear and non-linear relationships between predictors and latent positions, allowing the spline-based penalized MRU model to be evaluated under conditions where the true underlying structure was known. Three continuous covariates were generated for each observation, with each predictor x_j independently drawn from a normal distribution with variance 1.5^2, that is $x_j \sim \mathcal{N}(0, 1.5^2)$ for $j = 1, 2, 3$. These predictors were transformed into a two-dimensional latent representation through a set of predefined functions intended to create a varied response surface.

The first predictor, x_1, was mapped using a "biloop" transformation designed to produce a two-lobed closed curve. This mapping used the intermediate function $\tau(x_1) = \tanh(x_1/4)$, from which the horizontal and vertical coordinates were computed as:

$$u(x_1) = \begin{cases} 4\left[1 + \cos(2\pi\tau(x_1) + \pi)\right], & x_1 \geq 0, \\ -4\left[1 + \cos(2\pi\tau(x_1) - \pi)\right], & x_1 < 0, \end{cases}$$

and

$$v(x_1) = \sin(2\pi\,\tau(x_1))\,.$$

These produce the latent coordinate $U_1(x_1) = [\,u(x_1),\ v(x_1)\,]^\top$, forming a smooth, loop-shaped structure where the direction of curvature depends on the sign of x_1.

The second predictor, x_2, was transformed using a "bowl" function that generated a concave, radially symmetric surface. For each value, its magnitude was taken as $r = |x_2|$ and re-scaled using $q = \sqrt{\chi^2_{0.9975,1}}$. A warped radius $u = \tanh(r/q)$ was then computed and used to define a radial weight $\omega(x_2) = 10\,u^2(1-u)^2$. The transformation yielded the latent coordinate:

$$U_2(x_2) = \begin{bmatrix} 10\,u^6(1-u)^2 \\ \omega(x_2)\,x_2 \end{bmatrix},$$

which produced a bowl-shaped trajectory that expands and contracts smoothly as $|x_2|$ increases.

The third predictor, x_3, contributed a simple linear component to the latent representation, defined as $U_3(x_3) = [\,0.5x_3,\ -0.7x_3\,]^\top$. The true latent position for each observation was then obtained as the additive combination of these three components, $U_i = U_1(x_{1i}) + U_2(x_{2i}) + U_3(x_{3i})$.

Class locations were not synthetically generated but instead taken from an MRU analysis of the liver dataset, providing empirically grounded class geometry. Letting $V = [v_1^\top, \ldots, v_C^\top] \in \mathbb{R}^{C\times 2}$ denote these class centroids. Class membership probabilities were then generated using the MRU exponential distance transformation,

$$p_{ic} = \frac{\exp(-d_{ic})}{\sum_{c'=1}^{C} \exp(-d_{ic'})},$$

Observed class labels were then sampled from a multinomial distribution with probabilities $p_{i1}, \ldots, p_{iC}$, yielding an indicator matrix $G \in \{0,1\}^{n\times C}$. This process produced complete data sets consisting of predictors, latent coordinates, true class probabilities p_{ic}, and sampled class assignments G.

Since class membership probabilities in MRU are determined by an exponential transformation of Euclidean distances followed by normalisation across classes, global scaling of these distances provides a principled mechanism for regulating the sharpness of the resulting class-probability surface. Importantly, this operation leaves the latent configuration unchanged and instead modifies the steepness of the probability mapping. Introducing a scaling factor $\alpha > 0$, the distance term is scaled as:

$$U_i^{(\alpha)} = \alpha\,U_i, \qquad V_c^{(\alpha)} = \alpha\,V_c,$$

which implies that all pairwise distances transform as

$$d_{ic}^{(\alpha)} = \|U_i^{(\alpha)} - V_c^{(\alpha)}\|_2 = \alpha\,d_{ic}.$$

Values of $\alpha < 1$ reduce the effective distances entering the exponential mapping, producing flatter and more ambiguous class-probability surfaces. In contrast, values of $\alpha > 1$ amplify these distances, yielding sharper probability surfaces and stronger probabilistic discrimination. Importantly, the geometric configuration of the latent space remains unchanged, only the mapping from distances to probabilities is modified. In this study, three scaling levels were examined ($\alpha = 0.5, 1, 2$), chosen to span regimes of diffuse, moderate, and sharp probabilistic discrimination without altering the underlying latent structure.

To examine the effect of sample size on estimation accuracy and generalisation, four training datasets were generated with sizes $n \in \{100, 250, 500, 1000\}$, each drawn independently from the same population model. A large held-out test set of 10,000 observations was also generated once and used for all predictive evaluations, ensuring stable accuracy estimates across conditions.

3.2 Simulation Design

The simulation was designed to evaluate how the spline-based penalized MRU model performed under varying data-richness and class-overlap conditions by combining controlled data generation with a structured model-selection and evaluation procedure.

Two experimental factors were manipulated. First, four training sample sizes ($n \in 100, 250, 500, 1000$) were examined, allowing model performance to be assessed under increasingly data-rich settings. Second, the latent class-probability distribution was uniformly scaled using factors $\alpha \in 0.5, 1, 2$, which produced low-contrast, baseline, and high-contrast probabilistic discrimination conditions, respectively. The underlying generative functions and empirical class locations were held constant across all conditions, such that any observed differences in performance could be attributed solely to changes in sample size and the degree of probabilistic scaling, rather than to alterations in the latent structure.

Model specification for each training dataset was determined through an explicit five-fold stratified cross-validation procedure. A full grid search was conducted over df, pd, and λ, ensuring that all aspects of the spline construction and regularization were tuned optimally. For each hyper-parameter combination, the spline-based penalized MRU model was fitted to four folds of the data, and predictive performance was evaluated on the remaining fold, rotating this role across all five partitions. Cross-validated deviance served as the primary selection criterion, with classification accuracy used as a secondary check to guard against overly complex models that reduce deviance without improving predictive discrimination. The full procedure is described below:

Input Criteria

- Predictor matrix X
- Response vector y
- Latent dimension D
- Hyperparameter grids:
  - `df_grid` (spline degrees of freedom)
  - `pd_grid` (spline polynomial degree)
  - `lambda_grid` (ridge penalty)
- Number of folds F

Algorithm

1. Stratify the response y into F approximately class-balanced folds.
2. For each hyperparameter combination $(\text{df}, \text{pd}, \lambda)$ in `df_grid × pd_grid × lambda_grid`:
   a. Ensure that df $>$ pd; otherwise skip this combination.
   b. For each fold $f = 1, \ldots, F$:
      i. Define the training set as all observations not in fold f, and the validation set as the observations in fold f.
      ii. Fit MRU on the training data using latent dimension D, spline df = df, spline polynomial degree = pd, and ridge penalty λ.
      iii. Transform the validation predictors using the spline basis learned from the training data.
      iv. Compute the predicted class probabilities for the validation set.
      v. Evaluate validation performance:

- Classification accuracy
- Negative log-likelihood (NLL)
- Deviance $= 2 \times NLL$

c. Aggregate performance across the F folds by computing, for each metric:

- NLL
- Deviance
- Classification accuracy

For each of these metrics, report the following summary statistics:

- Mean
- Standard deviation

3. Rank all hyperparameter combinations by either:
   - Lowest mean deviance
4. Return:
   - A summary table of all evaluated combinations and their cross-validated metrics.
   - The selected optimal hyperparameter combination $(\text{df}, \text{pd}, \lambda)$.

3.3 Materials

All simulations, model fitting procedures, and analyses were conducted in R (version 4.4.2) using custom code written specifically for this project. The implementation relied on several established R packages. Spline basis construction was carried out using the *splines* package. Visualisation of latent spaces, predictor trajectories, and performance curves was conducted using *ggplot*2.

In addition, the study made use of information from the liver dataset included in the lmap package. For the simulation study, only the class coordinates extracted from this dataset were used; these provided fixed latent class positions in the population-generating model. In contrast, the comparison study (Section X) made full use of the dataset, incorporating both the biochemical predictors (Aspartate Aminotransferase (AS), Alanine Aminotransferase (AL), and Glutamate Dehydrogenase (GD)) and the diagnostic group memberships (Acute Viral Hepatitis (AVH), Persistent Chronic Hepatitis (PCH), Aggressive Chronic Hepatitis (ACH), and Post-Necrotic Cirrhosis (PNC)) to evaluate model performance on real clinical data.

3.4 Simulation Procedure

The training data were generated using the population mechanism described in Section 3.1, yielding a sample of size n consisting of covariates X, their associated latent positions U, class coordinates V, and observed class memberships y.

Model selection was conducted over a comprehensive grid of candidate spline–MRU specifications. This grid spanned combinations of spline degrees of freedom (df), polynomial degree (pd), and ridge penalty strength (λ), thereby defining the full set of models considered during cross-validation.

For each candidate specification, five-fold stratified cross-validation was performed on the training data. The sample was partitioned into five approximately equal folds while preserving class proportions. Models were fitted on four folds and evaluated on the remaining fold. This procedure repeated until each fold had served

as the validation set once. Cross-validated deviance, averaged across folds, served as the primary criterion for hyper-parameter selection, with classification accuracy used as a supplementary diagnostic.

Following model selection, the spline-based penalized MRU model was refitted to the complete training dataset using the optimal hyperparameter combination. This refitting produced the final estimates of the regression matrix B, the class coordinate matrix V, and the fitted latent positions U for the given sample-size condition.

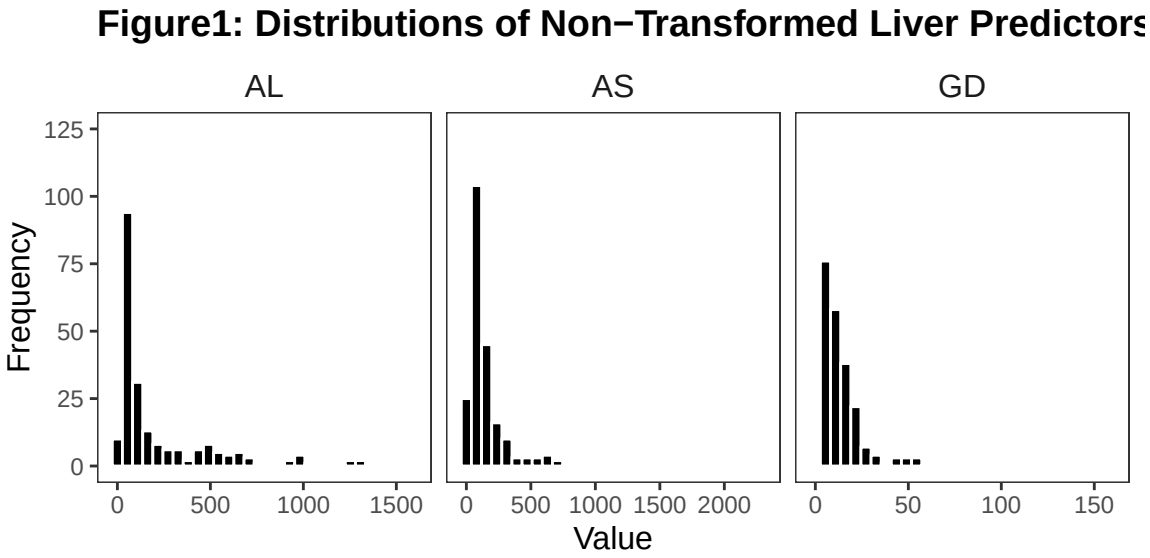
Finally, out-of-sample performance was assessed using a fixed held-out test set of 10,000 observations, generated once and excluded from all stages of model selection. The test predictors were standardised and transformed using the spline basis learned from the training data, after which test latent positions were computed as $U_{\text{test}} = X_{\text{test}}B$. Class probabilities were then obtained via the MRU distance kernel, and test deviance alongside classification accuracy were calculated to quantify generalisation performance.

3.5 Comparative Analysis Data

The liver dataset from the lmap package in R consists of biochemical liver function measurements and corresponding diagnostic group memberships. The predictor matrix X is a 218×3 matrix containing observed values on three liver function tests: Aspartate Aminotransferase (AS), Alanine Aminotransferase (AL), and Glutamate Dehydrogenase (GD). The response information is provided in two forms: a vector y of length 218 indicating the class label for each observation, and a 218×4 indicator matrix G encoding membership in one of four clinically defined liver condition groups; Acute Viral Hepatitis (AVH), Persistent Chronic Hepatitis (PCH), Aggressive Chronic Hepatitis (ACH), and Post-Necrotic Cirrhosis (PNC).

To contextualize the dataset, the three biochemical predictors index different aspects of liver function and cell damage. AS and AL are enzymes released into the blood when liver cells are irritated or damaged, meaning higher values often signal active inflammation. GD on the other hand is a mitochondrial enzyme that tends to rise when liver damage is more substantial. The four diagnostic groups also reflect a gradient of severity. AVH involves sudden but often reversible inflammation, PCH represents longer-lasting, lower-grade injury, ACH indicates more pronounced, ongoing inflammation and finally, PNC is the most advanced condition, marked by lasting structural damage.

3.5.1 Non-Transformed Predictors



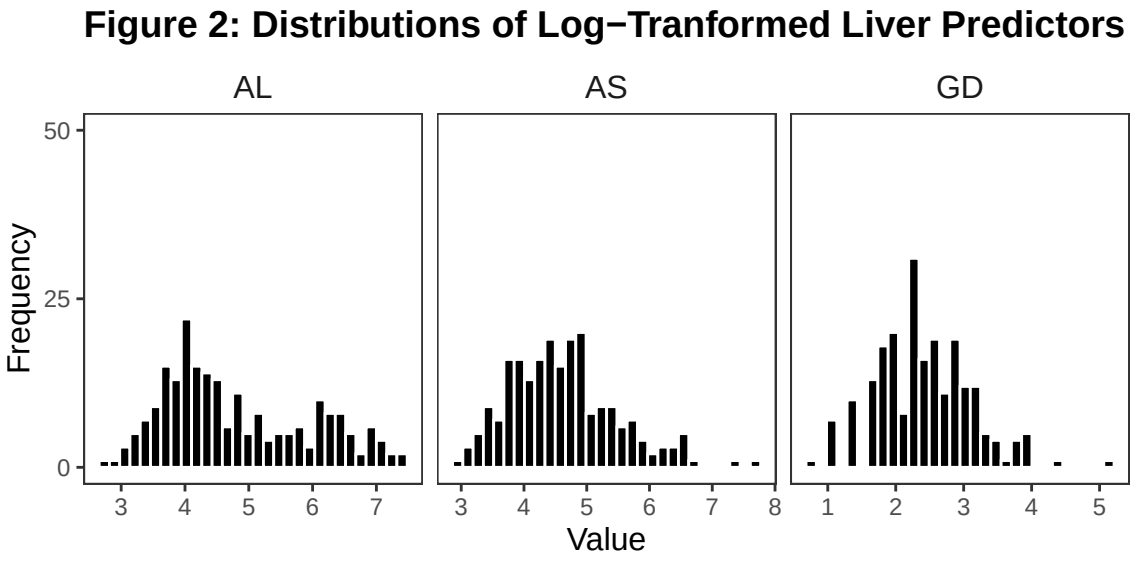
As shown in Figure 1, the three biochemical predictors in the liver dataset AL, AS, and GD exhibit substantial departures from normality in their non-transformed form. All three variables display pronounced right-

skewness and extremely heavy-tailed distributions, a pattern reflected in the descriptive statistics shown in Table 1. AS demonstrates the most severe distortion (skewness = 5.53; kurtosis = 42.08), with most observations falling below 200 units but several extreme outliers exceeding 2,000 units. AL shows a similar, though less extreme, distributional shape (skewness = 2.08; kurtosis = 4.08), with the majority of values below 400 units and a subset extending beyond 1,000 units. GD mirrors the behaviour of AS (skewness = 5.26; kurtosis = 43.54), with a dense concentration of values between 0 and 30 units and isolated cases surpassing 150. Despite differences in their absolute scales, all three predictors share the same underlying structure of strong right-skewness and heavy tails, indicating that a small number of extreme observations exert disproportionate influence on their overall distribution profiles.

Table 1: Descriptive Statistics for Non-Transformed Liver Dataset Predictors

Variable	Mean	SD	Min	Q1	Median	Q3	Max	Skewness	Kurtosis
AS	157.70	225.99	20	54	92.5	165.00	2298	5.53	42.08
AL	242.15	319.06	15	52	86.5	324.25	1595	2.08	4.08
GD	14.65	14.70	2	7	11.0	18.00	160	5.26	43.54

3.5.2 Log-Transformed Predictors



As shown in Figure 2, applying a logarithmic transformation substantially improved the distributional properties of the three liver predictors (AL, AS, and GD). The transformation markedly reduced the extreme right-skewness and heavy-tailed behavior observed in the raw data, resulting in distributions that are considerably more symmetric and compact. This improvement is reflected in the descriptive statistics presented in Table 2. For AS, skewness decreases from 5.53 in the raw scale to 0.66 in the log-transformed scale, and kurtosis falls from 42.08 to 0.46, producing a distribution that approximates normality. AL exhibits a similar shift, with skewness reduced to 0.53 and kurtosis to –0.85, indicating a slightly flattened but well-behaved distribution with most values clustered between 4 and 6 on the log scale. GD shows comparable gains, with skewness reduced to 0.40 and kurtosis to 0.38, and values tightly concentrated between 2 and 3. Overall, the log transformation effectively suppresses the influence of extreme outliers and stabilizes the variance.

Table 2: Descriptive Statistics for Log-Transformed Liver Dataset Predictors

Variable	Mean	SD	Min	Q1	Median	Q3	Max	Skewness	Kurtosis
AS	4.62	0.85	3.00	3.99	4.53	5.11	7.74	0.66	0.46
AL	4.80	1.14	2.71	3.95	4.46	5.78	7.37	0.53	-0.85
GD	2.42	0.70	0.69	1.95	2.40	2.89	5.08	0.40	0.38

3.6 Comparative Analysis Procedure

The comparative analysis began by preparing the liver dataset for modelling. The three biochemical predictors (AS, AL, GD) were extracted as the covariate matrix X, and diagnostic group memberships were represented using both the class-label vector y and indicator matrix G. For both models, the raw predictors and log-transformed predictors, the data were standardised using variable-specific means and standard deviations.

A full hyperparameter grid was defined for the spline-augmented model, covering candidate values for the spline df, pd, and λ. These combinations formed the set of possible spline-MRU model configurations to be evaluated during cross-validation.

Each model underwent five-fold stratified cross-validation using the algorithm
described in Section 3.2, which specifies the general cross-validation framework employed throughout this thesis. The data were partitioned into five approximately equal folds with preserved class proportions. For every hyper-parameter combination, the model was fitted on four folds and evaluated on the remaining fold, cycling across all five folds so that each subset served once as the validation set. Mean cross-validated deviance served as the primary selection criterion.

After identifying the optimal hyper-parameter combination for each predictor variant, the corresponding spline-based penalised MRU model was refitted to the full dataset. This produced the final estimates of the regression matrix (B), the class coordinate matrix (V), and the latent observations (U) under both the raw-predictor and log-predictor specifications.

Finally, model performance was compared directly between the fitted models. Deviance, accuracy, predictor trajectories, and latent-space visualizations were examined to assess whether the predictor transformation improved model fit, interpretability, and predictor effect recovery.

4 Simulation Results

4.1 Model Selection for Simulation Analysis

Across all conditions (N and α), model selection was performed using a grid search over the spline and regularisation hyper-parameters. The spline (df) were varied from 4 to 10, and the spline (pd) was varied from 1 to 3, allowing the basis functions to capture relationships ranging from linear to curved. The ridge penalty parameter λ was explored over a logarithmic sequence from 10^{-3} to 10^{1}, covering models from weakly to strongly regularised. Each combination of (df, pd, λ) constituted a candidate model, and its performance was evaluated using 5-fold stratified cross-validation. This grid spans a broad range of spline complexities and penalty magnitudes, ensuring that the selected hyperparameters reflect an appropriate trade-off between model flexibility and generalisation across all simulation conditions.

Table 3: Cross-validated Selected Hyperparameters for the Spline-based Penalised MRU Model

N	α	df	pd	λ
100	0.5	5	3	1.0
100	1.0	6	3	1.0
100	2.0	4	2	0.1
250	0.5	7	3	10.0
250	1.0	4	3	1.0
250	2.0	4	3	0.1
500	0.5	4	3	10.0
500	1.0	4	3	1.0
500	2.0	6	2	1.0
1000	0.5	7	3	10.0
1000	1.0	5	3	1.0
1000	2.0	4	2	0.1

Table 3 presents the hyperparameters selected through cross-validation for the spline-based penalised MRU model across all simulation conditions. The selected spline df, pd, and λ penalty values vary across the combinations of N and α, indicating that different configurations were preferred depending on the characteristics of the simulated data. In general, the selected spline bases fall within a relatively narrow range of complexity, with df between 4 and 7 and pd between 2 and 3. The penalty parameter λ also shows systematic variation across conditions, with values ranging from 0.1 to 10.

Across sample sizes, increasing N did not produce a clear directional change in the hyperparameters chosen by cross-validation. Although the selected configurations differed across conditions, the spline df and pd did not consistently increase or decrease with larger sample sizes. This suggests that, within the simulation design, sample size had a limited influence on the complexity of the spline basis favoured during model selection.

In contrast, the scaling factor α was associated with more consistent patterns in the selected hyperparameters. Smaller values of α were often paired with higher spline df and larger λ strengths, while larger α values tended to correspond to simpler spline bases and weaker λ. These patterns show that the hyper-parameter combinations identified by cross-validation varied systematically with the structure of the simulated latent space and were more strongly influenced by α than by N.

4.2 Descriptive Results

Table 4: Cross-validated out-of-sample performance (training data only) of spline-based penalised MRU

N	α	Mean Deviance	SD Deviance	Mean Accuracy	SD Accuracy
100	0.5	36.695	8.851	0.614	0.095
100	1.0	30.409	4.980	0.659	0.019
100	2.0	20.630	0.921	0.780	0.102
250	0.5	87.970	5.682	0.619	0.088
250	1.0	56.940	14.082	0.799	0.078
250	2.0	50.737	8.987	0.840	0.043
500	0.5	168.902	11.590	0.650	0.049
500	1.0	122.927	10.586	0.754	0.039
500	2.0	108.066	10.516	0.758	0.020
1000	0.5	358.596	14.389	0.603	0.009
1000	1.0	245.522	30.922	0.768	0.044
1000	2.0	183.643	16.320	0.829	0.023

Table 4 summarises the out-of-sample performance of the spline-based penalised MRU model across all simulation conditions. Mean deviance and accuracy vary systematically across combinations of N and α, reflecting differences in sample size and the degree of probabilistic discrimination induced by the scaling parameter. The accompanying accuracy standard deviations decrease with increasing N, indicating more stable cross-validation estimates in data-rich settings. Performance also improves consistently with larger values of α, where the class-probability surface is sharper and the classification task correspondingly less ambiguous.

With respect to sample size, increasing N primarily affected the stability of the performance metrics rather than their overall magnitude. As N increased from 100 to 1000, both deviance and accuracy exhibited smaller standard deviations, demonstrating more consistent performance across folds. When deviance was considered relative to sample size (i.e., deviance per N), the values remained broadly comparable across conditions, reinforcing the observation that larger data sets did not systematically improve or worsen the model's absolute fit. Instead, increasing N primarily enhanced the reliability of the estimates, indicating that sample size influenced the precision of the performance metrics more strongly than their overall level.

In contrast, variation in the scaling factor α showed a clear and systematic association with model performance. Higher α values, corresponding to greater contrast in the class-probability distributions generated by the exponential distance mapping, consistently produced lower mean deviances and higher accuracies across all sample sizes. Conversely, smaller α values, which induce flatter and more diffuse probability surfaces, resulted in degraded predictive performance and greater variability across folds. These patterns indicate that α served as the primary determinant of classification difficulty within the simulation, with sharper probabilistic discrimination yielding more accurate and stable outcomes.

Table 5: Test-set performance of the spline-based penalised MRU model

N	α	Deviance	Accuracy
100	0.5	22596.68	0.525
100	1.0	16033.17	0.692
100	2.0	14969.89	0.751
250	0.5	18931.28	0.585
250	1.0	13216.48	0.738
250	2.0	10902.93	0.795
500	0.5	18260.23	0.594
500	1.0	13086.54	0.741
500	2.0	11204.95	0.785
1000	0.5	17678.92	0.607
1000	1.0	12642.61	0.739
1000	2.0	9957.07	0.808

Table 5 reports the true out-of-sample performance of the spline-based penalised MRU model, evaluated on test sets of size $N_{\text{test}} = 10{,}000$. For each class-separation condition ($\alpha = 0.5, 1, 2$), a corresponding test set was generated under the same data-generating mechanism, ensuring that models were evaluated on data drawn from the same latent structure they were trained to approximate. Each row therefore corresponds to a model trained on one of four sample sizes ($N = 100, 250, 500, 1000$) and assessed on its matched α-specific test set. Importantly, N refers exclusively to the size of the training data used to estimate the model parameters; test-set size remained fixed at 10,000 for all conditions.

Probability sharpness (α) exerted the strongest and most consistent influence on model performance. Higher values of α produced lower test-set deviance and higher accuracy, reflecting reduced ambiguity in the induced class-probability distribution. Model performance improved smoothly as α increased, indicating that the spline-based MRU model is highly sensitive to the degree of probabilistic contrast generated by the scaling parameter.

The effect of increasing training-sample size N was also evident, though more gradual than that of α. As N increased from 100 to 1000, both deviance and accuracy steadily improve, demonstrating that the spline-based MRU estimator benefited from additional data when learning the predictor–class structure. Notably, gains in performance become more incremental after $N = 250$. The increase from 100 to 250 yielded substantial reductions in deviance and large improvements in accuracy, whereas the transitions from 250-to-500 and 500-to-1000 produce smaller but still refinements. This suggests that approximately 250 observations were sufficient for the model to capture the dominant structure of the relationships, with larger samples yielding diminishing, but still meaningful, returns in predictive performance.

4.3 Effects in the Predictors

4.3.1 Effect of Sample Size (N) on Predictor Representations

Figure 3.1: Predictor trajectories (N = 100; α = 1)

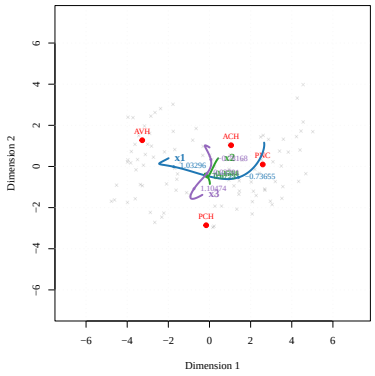


Figure 3.2: Predictor trajectories (N = 250; α = 1)

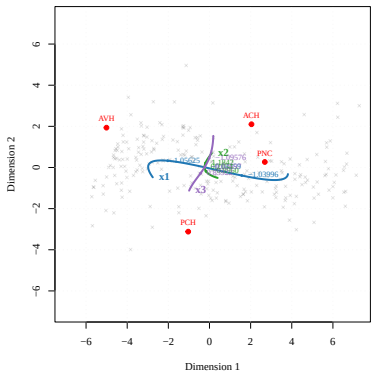


Figure 3.3: Predictor trajectories (N = 500; α = 1)

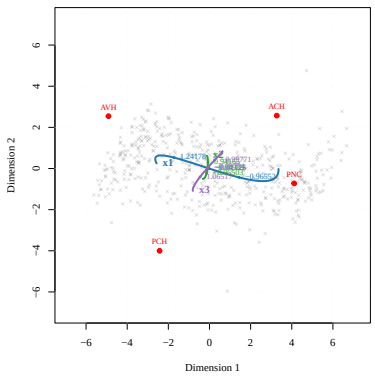
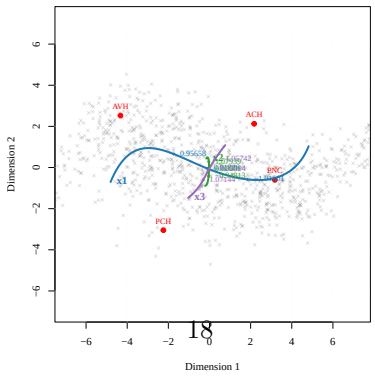


Figure 3.4: Predictor trajectories (N = 1000; α = 1)



Figures 3.1–3.4 depict how the spline-based penalised MRU model's latent-space representations evolve as the training sample size increases from $N = 100$ to $N = 1000$. Predictors x_1, x_2 and x_3 are displayed in blue, green and purple, respectively. For each predictor, selected values from the original data scale are mapped onto the latent space and displayed along their corresponding trajectories. These values correspond to the 25th, 50th, and 75th percentile.

The most notable departure from the overall pattern occurs at $N = 100$, where the estimated predictor trajectories are visibly different from renditions with larger sample sizes, exhibiting sharp bends and inconsistent curvature that may largely reflect sampling noise. At this smallest sample size, the class centres are only weakly separated, and the latent observations appear widely dispersed, suggesting that the spline basis is under-supported and the model lacks sufficient information to recover the underlying structure. This may explain why predictors x_1 (biloop), x_2 (bowel) and x_3 (straight line) trajectories bear little resemblance to the intended design.

When the sample size increases to $N = 250$, the predictor curves begin to converge and adopt a smoother, more interpretable shape. The trajectories show fewer abrupt deviations, and their overall geometry more closely resembled the underlying structure imposed in the data-generating process. The class points also become more distinctly positioned relative to one another, and the latent configuration as a whole forms a clearer, more interpretable spatial layout. For example, AVH and PNC occupy distinct regions of the x_1 trajectory, indicating differential sensitivity to this predictor. Likewise, ACH and PCH show distinct alignments along x_3, each class associating with a different segment of the curve.

Moreover, the x_1 trajectory is recovered relatively well, with its characteristic looping geometry becoming increasingly apparent. In contrast, the x_2 trajectory, intended to exhibit a bowl-shaped curvature, remains largely unrecognizable and fails to capture the intended concave structure imposed in the data-generating process. The x_3 trajectory is moderately recovered, with its straight-line structure becoming visible, exhibiting only minor curvature deviations.

At $N = 500$, the predictor trajectories have matured into consistent trajectories with only subtle shifts in the overall structure. The x_3 trajectory is further refined, exhibiting increasingly minor deviations from a straight line. x_1 trajectory becomes smoother but still does not fully capture the concave structure of a biloop. Similarly, little improvement is observed in the x_2 trajectory compared with smaller sample sizes, and it remains largely unrecognizable. Class separation however does become increasingly well-defined, with the class points forming a clearer geometric arrangement and the latent observation's clustering more distinctly.

A further increase in sample size ($N = 1000$) yields only incremental refinements, with the x_1 and x_3 trajectories showing marginal improvements toward their intended structures. In contrast, the x_2 trajectory remains largely unrecognizable, suggesting that the model is unable to recover its true geometry regardless of sample size.

4.3.2 Influence of Probabilistic Discrimination (α) on Predictor Representations

Figure 4.1: Predictor trajectories ($\alpha = 0.5$; N = 500)

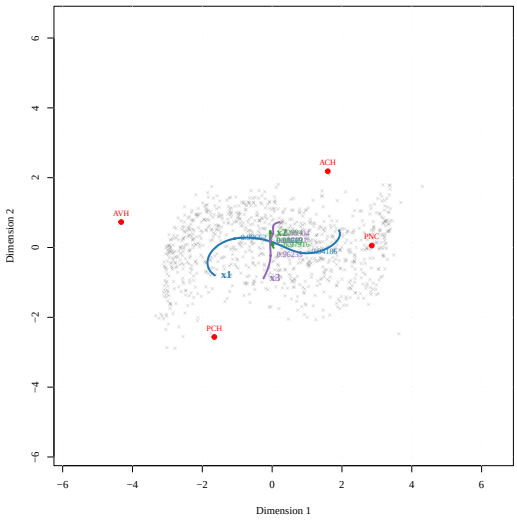


Figure 4.2: Predictor trajectories ($\alpha = 1$; N = 500)

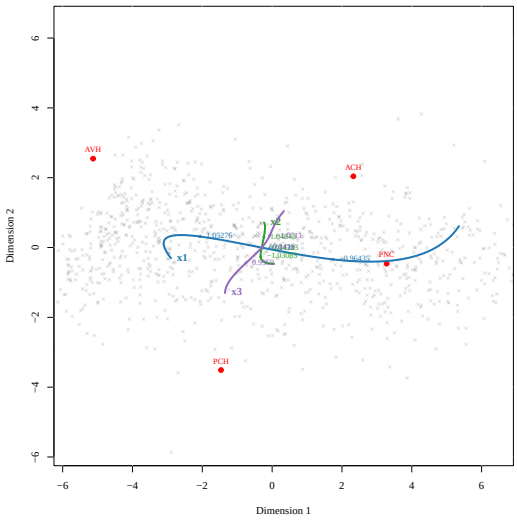
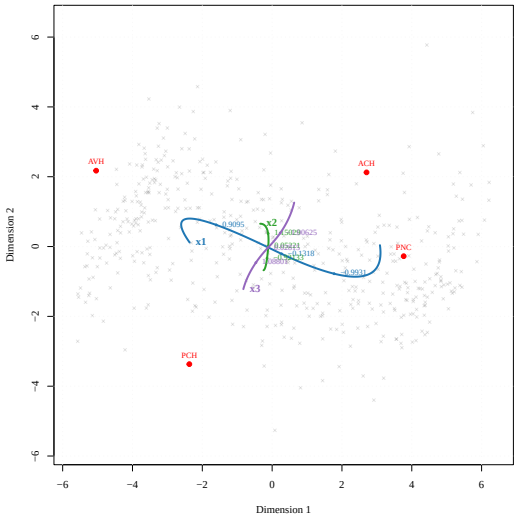


Figure 4.3: Predictor trajectories ($\alpha = 2$; N = 500)



Figures 4.1–4.3 illustrate how increasing the probability sharpness parameter α influences the latent landscape recovered by the spline-based penalized MRU model. Each panel displays the same latent configuration and predictor trajectories, while differing only in the scaling applied to the distance-to-probability mapping.

At the lowest sharpness level ($\alpha = 0.5$; Figure 4.1), the induced class-probability surface is relatively flat, resulting in diffuse and uncertain class assignments despite an unchanged geometric configuration. Although the latent class centres occupy the same positions across conditions, the probability contrast between them is weak, producing visually ambiguous group boundaries. As α increases (Figures 4.2 and 4.3), the probability surface becomes progressively steeper, yielding sharper transitions between classes and more decisive assignments. This increasing probabilistic discrimination provides a clear explanation for the improved classification performance observed at higher α levels.

Importantly, the predictor trajectories remain broadly similar in form across all values of α. Their overall shapes and directional trends are unchanged, indicating that the estimated relationship between predictors and latent dimensions is stable across probability-scaling regimes. Changes in α are therefore reflected not in the geometry of predictor effects, but in the sharpness with which those effects translate into class assignments. This further supports the conclusion that the spline-based MRU model captures predictor structure reliably, while α primarily governs classification entropy rather than latent geometry.

5 Comparative Analysis Results

5.1 Model Selection for Comparison Analysis

Table 6: Model parameters selected for the spline-based MRU models in the liver data analysis.

Model	Predictors	Transformation	Degrees.of.Freedom	Polynomial.Degree	Ridge.Penalty
Spl. MRU	log(X)	B-spline	6	2	0.1
Spl. MRU	X	B-spline	4	1	0.1

The nested CV grid search revealed meaningful differences in the optimal spline parameters selected for the log-transformed and raw predictor versions of the spline-based penalised MRU model. When the predictors were log-transformed, the model consistently chose a more flexible spline structure, with a higher number of degrees of freedom ($df = 6$) and a higher spline polynomial degree ($pd = 2$). In contrast, the model fitted to the raw predictors favoured a simpler configuration ($df = 4$, $pd = 1$). This divergence indicates that the two transformations support different levels of functional complexity, with the log-transformed predictors permitting smoother and more adaptable basis functions.

The underlying reason for this difference is that the log transformation substantially reduced skewness and compressed extreme values, stabilizing the predictor distributions (as shown in Figure 1 & 2). This stabilisation enabled the spline-based MRU model to estimate richer curvature without incurring over-fitting or producing numerical instability. In contrast, the raw predictors characterised by heavier tails and larger scale differences constrained the model to select a more conservative spline basis, as overly flexible functions would amplify noise driven by extreme observations.

Finally, both models settled on the same ridge penalty (0.1), implying that overall regularization strength was similar once the optimal spline structure was found. Collectively, these results suggest that applying a log transformation to the predictors improved the spline-based MRU ability to employ richer, smoother spline functions, thereby enhancing its ability to model subtle effects in the data.

5.2 Descriptive Results

Table 7: Comparison of Standard and Spline-Based MRU Models on the Liver Dataset

Model	Predictors	Transformation	Mean.Deviance	SD.Deviance	Mean.Accuracy	SD.Accuracy
Std. MRU	log(X)	None	47.35	5.49	0.798	0.022
Spl. MRU	log(X)	B-spline	39.70	5.30	0.822	0.045
Spl. MRU	X	B-spline	42.57	8.13	0.826	0.039

The spline-based models both achieved a lower cross-validated mean deviance than the standard MRU. In this context, the "mean deviance" denotes the average deviance computed across the K validation folds in the nested cross-validation procedure, rather than the deviance of a single model fitted to the full dataset. The corresponding standard deviations quantify the degree of variability in deviance across folds, providing an indication of how sensitive each model is to the particular composition of the training and validation subsets. The same interpretation applies to the reported accuracy values. The summation of performance in this out-of-sample manner, captured each model's expected generalization behaviour rather than its fit to a single sample partition, thereby offering a more robust basis for model comparison.

The Spline-Penalized MRU with log-transformed predictors achieved the lowest average deviance (39.7) with the smallest fold-to-fold variation (SD = 5.3), indicating both better fit and greater stability. The spline-MRU applied to the raw predictors also improved upon the standard MRU (mean deviance = 42.6 vs. 47.4), though with higher variability (SD = 8.1), suggesting that the untransformed predictors yielded less stable embeddings. Overall, these results indicate that incorporating non-linear transformation on predictors allowed the model to capture curvature in the predictor–class relationships that the orginal formulation of MRU could not fully represent.

Classification accuracy followed a similar pattern. The spline-based models achieved accuracies in the range of 0.82–0.83, compared to 0.80 for the standard MRU. Although, the improvement is modest, it was consistent across validation folds, indicating enhanced predictive generalization.

Finally, the log-transformed spline-MRU showed the most stable performance across folds, with the lowest standard deviation in deviance (5.3) and the most consistent median deviance. This suggests that applying a log transformation before spline expansion reduces skew and the influence of extreme values, enabling smoother and more reliable embeddings.

5.3 Effects of the predictors

5.3.1 Standard MRU - log(X)

Figure 5: Std. MRU − log Transformation

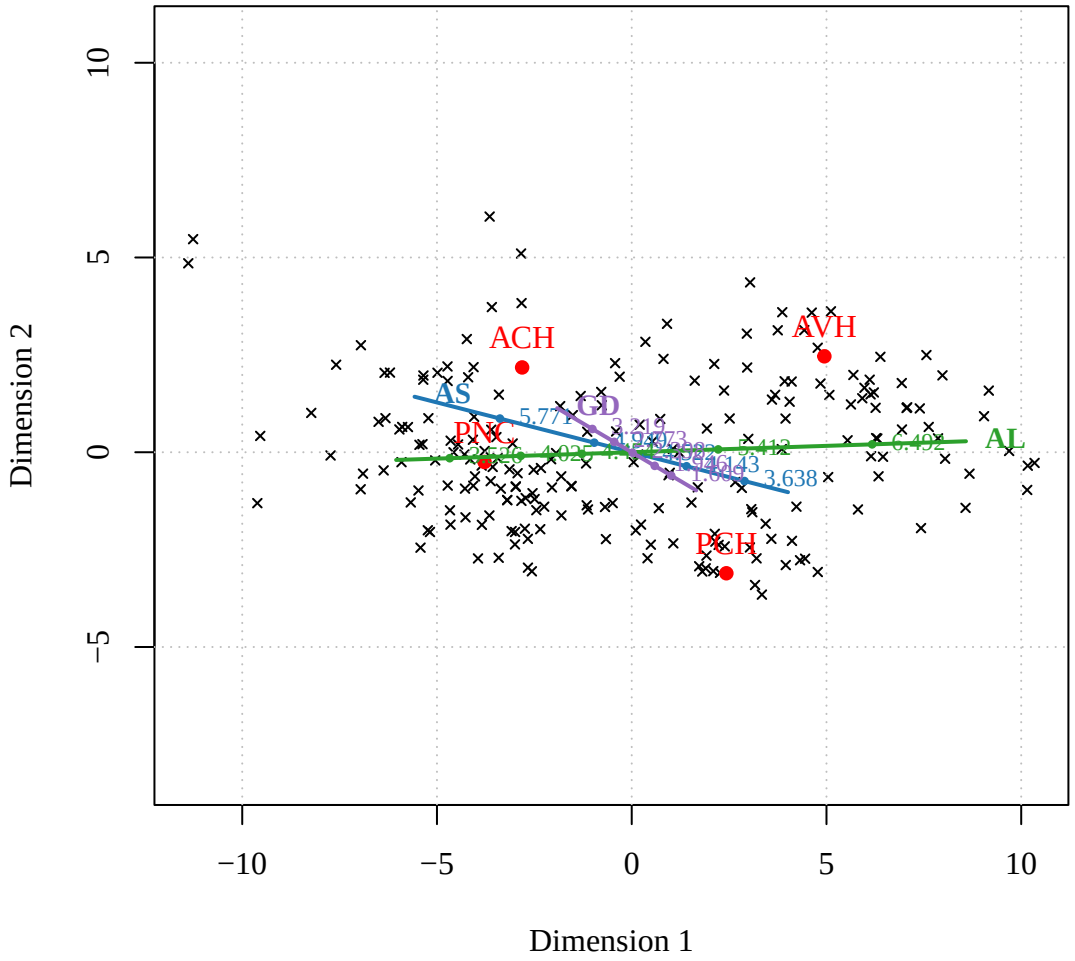


Figure 5 shows the two-dimensional latent-space configuration produced by the orginal formulated MRU model using log-transformed predictors. The black crosses indicate the projected locations of individual observations, while the red points mark the estimated class centroids for the four diagnostic groups (ACH, AVH, PCH, and PNC).

The distribution of observations revealed considerable overlap among classes, reflecting limited separability of the diagnostic groups based on these biochemical markers. However, the class centroids were positioned at clearly separated locations, suggesting that the MRU model was still able to extract meaningful differences between the diagnostic categories despite substantial within-class variability. Moreover, the predictor trajectories showed clear alignment with these class distinctions; for example, lower values of AL were associated with movement toward the PNC centroid, while higher AL values shifted observations in the direction of AVH. The linear form of these trajectories further aids interpretation, as each predictor contributes to class separation in a single dominant direction, making the underlying relationships between biochemical markers and diagnostic categories straightforward to interpret within the latent space.

Figure 6.1: AS effect on log−odds

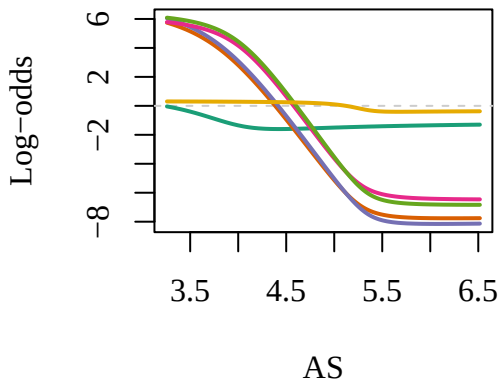


Figure 6.2: AL effect on log−odds

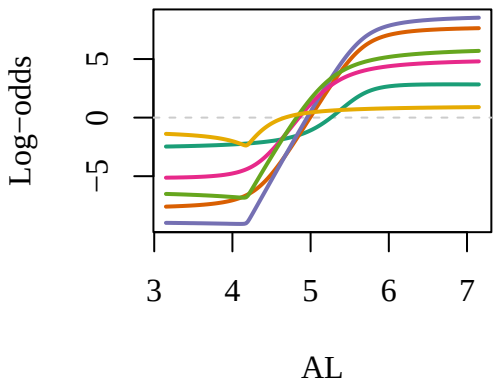
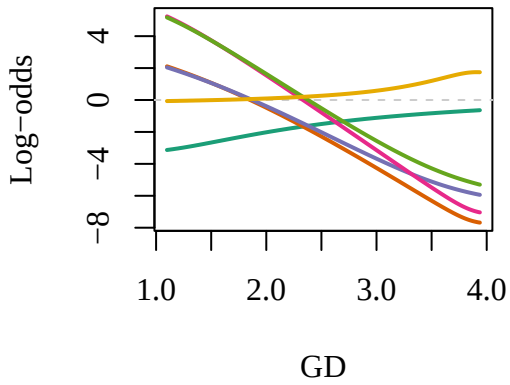
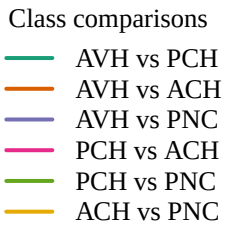


Figure 6.3: GD effect on log−odds





The combined log-odds plots in Figures 6.1-6.3 illustrate how each predictor; AS, AL, and GD, influences the relative probability of class membership under the Standard MRU model with log-transformed predictors. Each curve represents one of the six pairwise class comparisons, showing how the predicted log-odds of belonging to the first class versus the second change as the predictor varies, with all other predictors held constant.

A clear pattern emerges for AS (Figure 6.1). Across all six comparisons, the log-odds decrease steadily as AS increases. This may indicate that individuals with higher AS scores become progressively less likely to belong to the earlier-stage classes AVH and PCH, and become increasingly more likely to be assigned to ACH or PNC. The very steep declines in the AVH-ACH, AVH-PNC, and PCH-ACH trajectories further suggest that AS plays a particularly strong role in distinguishing the more severe diagnostic categories from the less severe ones. This smooth, monotonic behaviour reflects a globally consistent relationship.

By contrast, the curves for AL (Figure 6.2) reveal a distinctly different, more nuanced structure. Rather than exhibiting a uniform monotonic pattern, AL shows a marked non-linearity: many of the trajectories follow an S-shaped form, with a sharp transition occurring between AL levels of approximately 4.5 and 6.0. Below this interval, the log odds curves remain relatively flat, suggesting little discrimination between classes; however, once AL crosses the mid-range threshold, the probabilities shift rapidly, especially in comparisons

such as AVH-PCH, AVH-ACH, and PCH-ACH. This threshold-like behavior implies that AL influences class separation only once values enter a specific region, a pattern the linear MRU structure is not naturally designed to accommodate. In this sense, AL exposes a limitation of the original MRU model: its inability to flexibly represent relationships that change abruptly in specific predictor ranges.

A third pattern emerges for GD (Figure 6.3), whose trajectories are more heterogeneous and pair-specific. Some comparisons (e.g., AVH–PCH and PCH–ACH) display a clear downward trend, indicating that increases in GD decrease the likelihood of belonging to the first class. However, most notably ACH–PNC remained almost entirely flat across the predictor range, suggesting that GD provides little discriminatory information for those particular class boundaries. This mixed behavior indicates that GD does not exert a uniform global effect, but instead varies across pairwise class contrasts. Such variability highlights a limitation of linear transformations of predictors in unfolding models: while they adequately capture strong global trends (as seen with AS), they are less well suited to predictors such as GD and AL, whose effects are more localized or class-specific.

5.3.2 Spline Penalized MRU - log(X)

Figure 7: Spl. MRU − Log Transformation

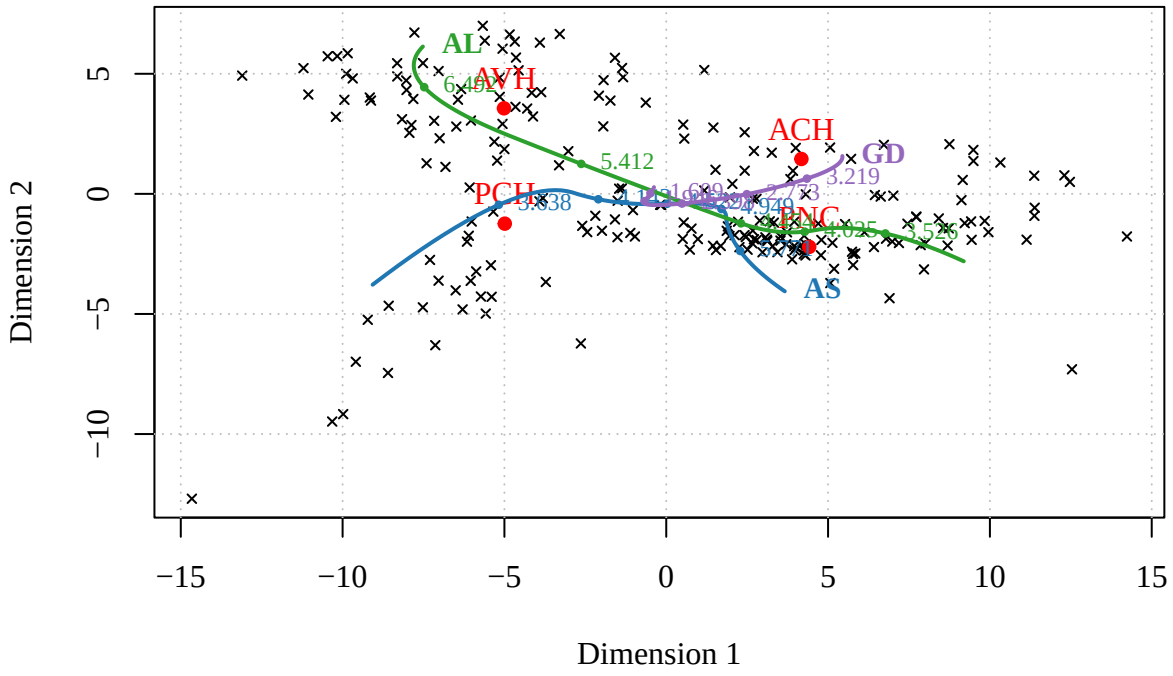


Figure 7 illistrates the two-dimensional latent-space configuration produced by the spline-based penalized MRU model using log-transformed predictors.

In contrast to the originally formulated MRU, these predictor trajectories are clearly curved, reflecting the non-linear transformations introduced by the spline basis expansion. This additional flexibility allows the trajectories to bend and adapt to local patterns in the data, capturing more nuanced relationships between the predictors and the diagnostic categories. Despite this nonlinearity, the trajectories remain interpretable, with each curve tracing a distinct path relative to the class centroids. For instance, the AL trajectory initially curves toward AVH before bending rightward, indicating a shift in class affinity across its range.

Furthermore, the spline-based penalised model produced a more structured arrangement of observations within the latent space, evident from the clearer clustering around their respective diagnostic categories. Although some overlap remains, particularly between ACH and PNC, the overall class structure is noticeably more distinct than in the original MRU solution, indicating that the spline transformation captured additional discriminative patterns in the data.

Figure 8.1: AS effect on log−odds

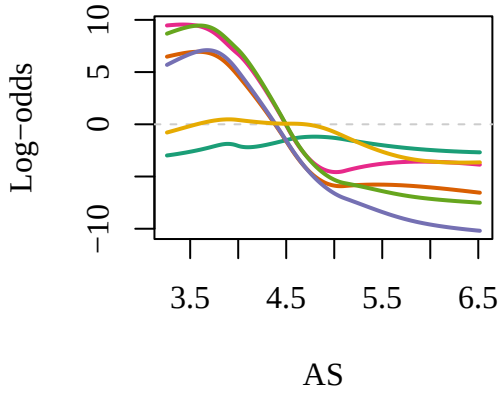


Figure 8.2: AL effect on log−odds

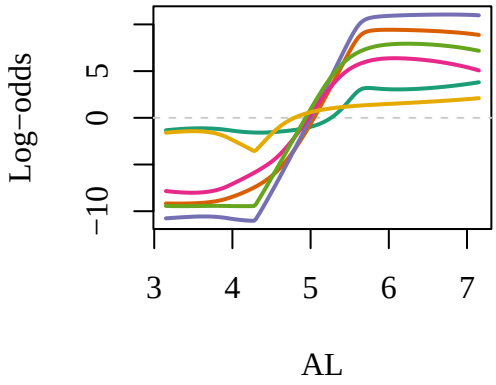
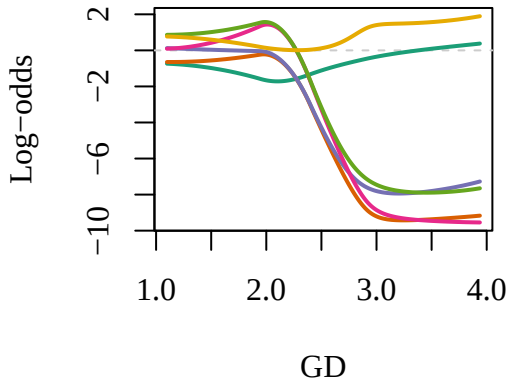
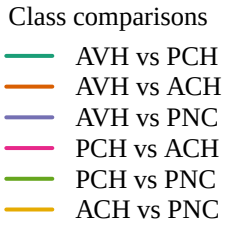


Figure 8.3: GD effect on log−odds





Figures 8.1–8.3 present the estimated log-odds functions for all pairwise class comparisons across the three predictors; AS, AL, and GD, derived from the Spline-Penalized MRU model with log-transformed predictors. Unlike the standard MRU model which assumes a strictly linear association between each predictor and the latent coordinates, the spline-augmented formulation relaxes this constraint by allowing the predictor–class relationships to vary smoothly across the predictor range. This added flexibility provides a more adaptable functional form, enabling the model to capture structured variation that would remain unrepresented under a less flexible specification.

Figure 8.1 shows the effect of AS on the log-odds across all six class contrasts. The overall pattern is similar to that of the standard MRU: as AS increases, the likelihood of AVH and PCH decreases, while the probability of ACH and PNC increases. However, the spline model revealed a noticeably richer structure. Many curves exhibit bends and/or plateaus, indicating that the discriminative influence of AS is not constant across its

range. For instance, in several contrasts such as AVH vs PCH and AVH vs PNC, the decline in log-odds is steep at moderate AS values but flattens at the higher end, suggesting diminishing marginal effect. Other contrasts display subtle undulations, implying local reversals or pockets of instability.

The corresponding patterns for AL in Figure 8.2 show the most nuanced relationships. Most curves begin relatively flat at low AL values, followed by a pronounced inflection between AL = 4.5 and 6.0, where the log-odds shift sharply. These transitions reflect threshold-like effects, where individuals with AL levels above a certain point become substantially more likely to be classified into the more severe liver condition groups.

In Figure 8.3, the predictor GD also displays meaningful curvature. Many pairwise comparisons show a steadily decreasing log-odds trajectory, but the rates of decline differ across the range of GD, again indicating non-constant effects. The comparison between ACH and PNC, for example, shows a shallow peak near GD = 2.2, suggesting a weak but genuine local deviation. Other contrasts, such as PCH vs ACH and AVH vs PNC, maintain a broadly monotonic pattern but still demonstrate subtle curvature that reveals more gradual rather than strictly linear transitions.

5.3.3 Spline Penalized MRU - X

Figure 9: SPMRU − No Transformation

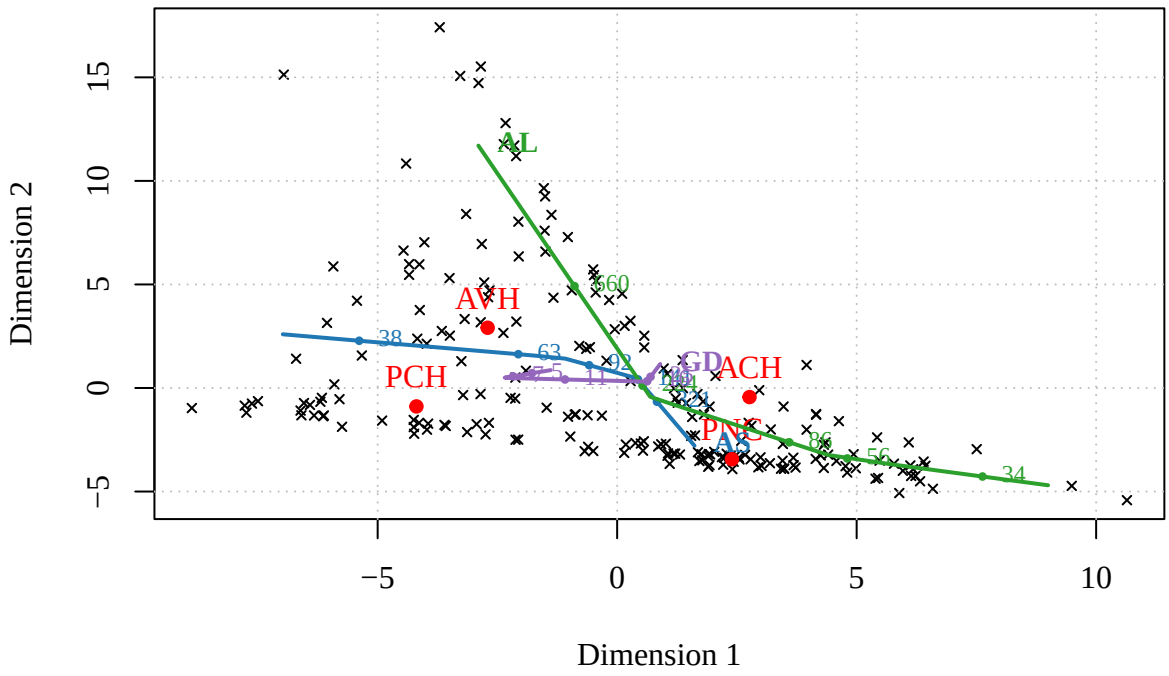


Figure 9 describes the two-dimensional latent-space configuration produced by the spline-based penalized MRU model using un-transformed predictors.

In contrast to the previous log-transformed models, these predictor trajectories appear less smooth and more erratic. This is particularly evident for GD, whose trajectory becomes highly irregular and difficult to interpret, showing abrupt changes in direction that do not correspond to meaningful class structure. This irregularity likely reflects the strong skew and differing numerical scales of the raw predictors, which heighten the model's sensitivity to local fluctuations and amplify distortions in the trajectories.

While the class centroids are separated, indicating that the model identified some degree of discriminatory structure, the extent of this separation is noticeably reduced, particularly between ACH and PNC. In addition, the observations themselves show weaker clustering around their respective centroids, forming more diffuse and overlapping clouds compared with the log-transformed spline model.

Figure 10.1: AS effect on log−odds

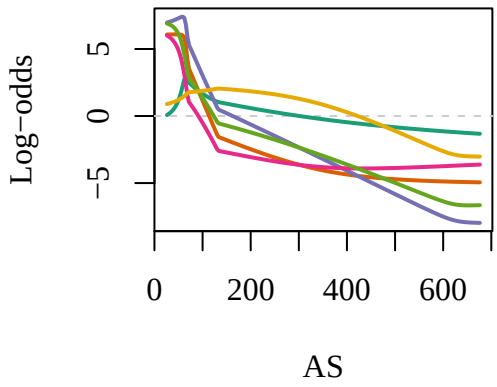


Figure 10.2: AL effect on log−odds

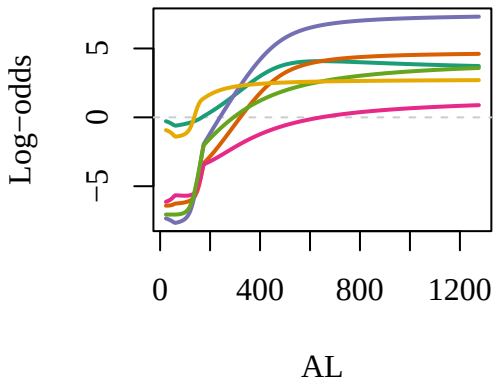
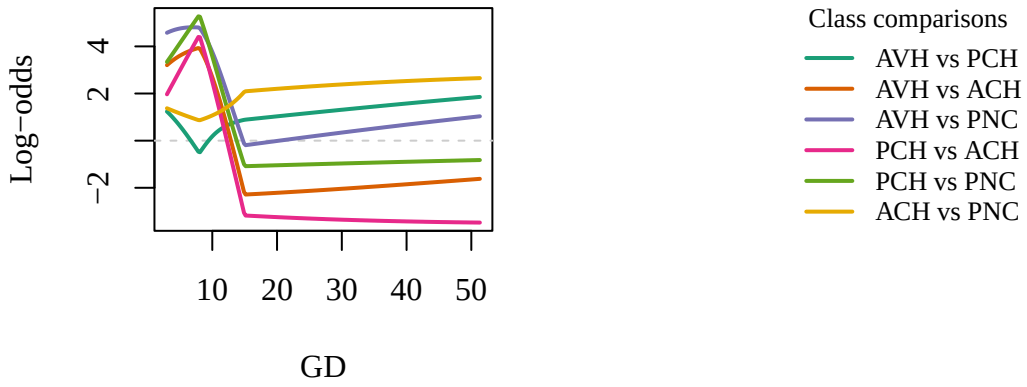


Figure 10.3: GD effect on log−odds



In Figure 10.1, the AS trajectories show a distinctive pattern characterized by sharp early spikes, followed by steep collapses across all six pairwise class comparisons. For every contrast, the log-odds begin at comparatively high levels when AS is close to zero, but fall abruptly within the first ~100-150 units. This initial drop is most pronounced in contrasts involving AVH, where the effect reverses direction rapidly moving from strongly favoring AVH to strongly dis-favouring it, within a very small portion of the AS range. Beyond this sharp transition zone, all curves stabilise into relatively flat, low-slope segments, indicating that once AS reaches a moderate level, additional increases contribute little further discrimination between classes. The pattern suggests that AS carries nearly all of its discriminatory value at the lower end of its range, with minimal impact thereafter.

The structure for AL in Figure 10.2 exhibit the most clearly organised and interpretable non-linear structure. Most curves remain relatively flat at low AL values, but around the mid-range (AL = 350–700), the log-odds undergo a pronounced upward shift. This steep transition in the middle band shift the probability of

belonging to one of the more severe groups (ACH, PCH, or PNC) dramatically. The curves then plateau at higher AL values, indicating that the discriminative strength of AL saturates beyond the upper end of this transition zone. AVH-related contrasts show steep slopes in this region, emphasising that AL is a strong driver of separation between AVH and the more advanced diagnostic categories. Overall, AL provides a comparatively structured signal with a consistent pattern across contrasts, reflecting its stable and interpretable effect on class membership.

Figure 10.3 displays the effect of GD and reveals more complex and heterogeneous trends. The most striking feature across all contrasts is the sharp early peak at very low GD values (around GD = 8–12), followed by an abrupt decline. This suggests that GD exerts an outsized influence near the lower end of its range, producing sudden changes in class separation. Beyond this initial spike, the trajectories diverge. These patterns indicate that GD influences different class boundaries in distinct ways, with some effects being monotonic and others displaying subtle non-monotonicity. Although the overall magnitude of GD's effects is smaller than that of AS or AL, the presence of sharp early transitions and varied long-range slopes underscores its role as a locally sensitive but globally modest predictor.

5.4 Inferential Results

Paired-sample t-tests were conducted to compare fold-wise deviance and classification accuracy between the spline-penalised MRU model and the standard MRU, using identical five-fold cross-validation splits. This pairing ensures that each deviance value from the spline models is directly matched to the corresponding held-out fold from the standard MRU, enabling a clean comparison of generalisation performance.

For the spline model applied to log-transformed predictors, the average deviance was modestly lower than that of the standard MRU (mean difference = -5.67), but this reduction was not statistically significant, $t(4)$ = -0.91, p = .42, 95% CI [-23.0, 11.7]. This reflects an inconsistent fold-wise pattern: the spline model outperformed MRU on some folds (folds 3–5), but underperformed on others (folds 1-2).

Paired-sample t-tests were also used to compare fold-wise classification accuracy under the same cross-validation splits. The spline model with log-transformed predictors achieved a slightly higher mean accuracy than the standard MRU (mean difference = +0.0247), but this improvement was not statistically significant, $t(4) = 0.60$, p = .58, 95% CI [-0.090, 0.140]. As with deviance, performance varied across folds, with no consistent advantage for the spline model.

6 Discussion

6.1 Simulation Discussion

The simulation study was designed to evaluate how effectively the spline-based penalized MRU model could recover latent predictor-class relationships under controlled conditions where the underlying generative structure was known. Across all combinations of sample size (N) and probability sharpness (α), several consistent themes emerged, illustrating both the strengths and limitations of the penalized spline-augmented MRU framework.

Overall, the spline-based MRU reliably recovered the underlying class structure, with performance strongly influenced by the sharpness of the class-probability surface, controlled through the scaling parameter α, and only modestly affected by sample size. Across both the training and test sets, increases in sample size moderately improved predictive accuracy and the stability of predictive performance (Tables 4). Increases in α systematically improved model fit at comparable sample sizes, producing lower deviances and higher classification accuracies. These findings are not unsurprising, given that α directly controlled the sharpness of the class-probability surface with higher values produce steeper mappings from latent distances to class probabilities and thereby increasing the relative ease of the classification task.

Regarding the recovery of predictor trajectories at small sample sizes (e.g., $N = 100$), latent representations and predictor curves displayed clear signs of noise sensitivity, including erratic curvature and weak class separation (Figures 2.1 and 3.1). These patterns indicate that the spline basis, although flexible, requires a sufficient number of observations to stabilise its local polynomial structure and avoid overfitting to sampling fluctuations. Once the sample size increased to $N = 250$, the estimated trajectories became markedly more consistent with those observed at larger sample sizes. As shown in Figure 3.2, abrupt bends in the predictor curves largely disappeared, producing a more interpretable output. The model achieved moderate success in recovering the underlying structures of x_1 and x_3, whereas the highly nonlinear x_2 trajectory remained largely unrecoverable. Further increases in sample size—to $N = 500$ and then $N = 1000$—produced only incremental refinements (Figures 3.3–3.4). By $N = 500$, the model had effectively converged on the dominant expressible predictor-class structure, with additional observations serving mainly to reduce minor fluctuations rather than alter the underlying estimates. This plateau indicates that the penalized spline-based MRU does not require very large samples to recover the dominant structure, but also highlights a broader limitation in expressing highly nonlinear predictor effects.

The inability to recover highly nonlinear predictor effects may stem from the structural composition of the algorithm. Notably, increasing the sample size did not substantially improve the recovery of these complex patterns, indicating that this limitation is intrinsic to the model rather than a consequence of insufficient data. A primary contributor is likely the imposed common degree of freedom across all predictors. This constraint forces the model to balance fit across both linear and highly nonlinear predictors, resulting in a compromise between the flexibility needed to capture nonlinear structures and the rigidity of linear patterns. Consequently, the model's capacity to fully recover either type of trajectory is limited.

The discriminative sharpness of the classes exerted little influence on the predictor trajectories, with the latent-space structure remaining largely consistent across changes in α. As visible in Figures 4.1-4.3, the shapes, orientations, and smoothness of the predictor curves showed minimal variation across sharpness conditions. This indicates that α influences the clarity with which the model translates latent distances into class assignments, but not the geometry of the latent space itself. Thus, while probabilistic sharpness determines how easily classes can be discriminated, it does not materially affect the recovery of predictor effects.

Finally, the simulation demonstrated the spline-based penalized MRU model's mixed ability to recover the nonlinear predictor relationships embedded in the data. The biloop transformation applied to x_1 produced a curvature-rich trajectory that the model captured reasonably well at moderate and large sample sizes. In contrast, the bowl-shaped transformation applied to x_2 could not be recovered, reflecting both the highly nonlinear nature of the trajectory and structural constraints within the algorithm. The linear effect of x_3 was consistently estimated, with accuracy improving as the sample size increased. These results indicate that the spline expansion enables MRU to approximate nonlinear predictor structures that the original formulation could not represent. However, they also highlight a limitation of the model when presented with heterogeneous predictor effects of varying complexity.

6.2 Comparative Analysis Discussion

The present comparison evaluated whether introducing spline-based transformations of predictors into the MRU framework improved interpretability and predictive performance when applied to the liver data-set, relative to the original formulation. In addition, the analysis examined how the underlying distributions of the predictor variables influenced both interpretability and model performance, particularly in terms of how the spline-based models handled to skewed or unevenly scaled predictors.

Quantitatively, no significant difference was observed in model performance. The added flexibility introduced via the B-spline expansion did not translate into improved generalization performance. Although the spline models did achieve numerically better fit and higher accuracy, these differences were marginal and inconsistent. An explanation for this outcome may rest in the predictive accuracy results which was high across all models, reflecting a strong signal-to-noise ratio. This left little room for further improvement,

meaning that increased model flexibility could refine the solution but was unlikely to yield substantial gains due to diminishing returns.

On the other hand, qualitative inspection showed that the spline-penalized MRU model produced richer representations of predictor effects. This was evident in the more nuanced relationships between predictors and class centroids, as well as in the clearer clustering of observations around their respective class locations. A likely explanation for this is the added flexibility allowed the model to effectively capture more subtle shifts in how predictor values relate to class locations, leading to more coherent trajectories and improved clustering around centroids. However, an important caveat is that this improvement was evident only when the predictors were log-transformed prior to modelling. The absence of this pre-processing revealed a key limitation in the model, namely its highly sensitive nature to predictor distribution. This highlights the importance of appropriate data pre-processing when using spline-based penalized MRU model, as the qualitative advantages depend strongly on the statistical properties of the input predictors.

In summary, the spline-based MRU models offered potential interpretive advantages, providing deeper insight into the relationships between predictor trajectories and class centroids. However, this did not extend to improved quantitative performance, indicating that the added flexibility enhanced interpretability but did not necessarily translate into better generalization of fit or predictive accuracy. It remains possible that data sets with a lower signal-to-noise ratio may show greater benefit from spline-based transformations in terms of both model performance, but confirming this would require further investigation.

6.3 General Discussion

The present study investigated whether extending the MRU framework with spline-based (B-spline) transformations and ridge regularisation enhances the model's ability to recover predictor–class relationships, and whether these additions improved interpretability, model fit, and predictive performance relative to the original formulation. This was examined through a controlled simulation study and real-world comparison using the *lmap* liver data set.

A primary finding was that the spline-based MRU meaningfully expands the representational capacity of the original framework. Across both studies, the spline extension proved effective at recovering nuanced predictor effects and producing interpretable latent representations. However, this increased flexibility also introduced an important practical consideration. The spline expansion is sensitive to the distributional properties of the predictors. This is likely because, although spline basis functions are locally supported, they still overlap with their neighbors, allowing observations in sparsely populated or heavy-tailed regions to exert disproportionate influence on the fitted function. When predictors display strong skewness or contain extreme values, the spline basis is forced to bend sharply in areas with limited data support, which introduces artificial curvature and local instability. These distortions are further compounded by the ridge penalty, which may over-shrink coefficients in some regions while under-shrinking them in others, thereby amplifying the imbalance created by the uneven distribution of the predictor. As a result, spline behavior becomes tightly coupled to the underlying distributional properties of the predictors, with skewness and extreme values amplifying instability through both the basis structure and the ridge penalty.

Moreover, the homogeneous or heterogeneous nature of predictor effects is another important consideration. The current findings seem to suggest that, for optimal model performance, a set of consistently nonlinear predictors may allow the spline-based MRU to function most effectively. In contrast, when linear and nonlinear predictors are combined, the trade-off inherent in the current model formulation becomes more pronounced, limiting the model's ability to allocate flexibility efficiently and fully capture complex predictor trajectories.

Another crucial finding was that the spline expansion, even when combined with regularization, did not improve quantitative model fit or predictive performance over the original formulation. Whilst the spline-based penalized MRU performed marginally better on average, these improvements were neither consistent nor substantial in magnitude. This indicates that linear combination of predictors coupled with the non-linear Euclidean distance structure already competently and accurately capture the underlying structures within

the data. Thus, introducing additional non-linear mechanics yields diminishing returns, as the majority of the relevant structure is already recovered by the original MRU framework. This outcome bolsters support for the original MRU formulation, indicating that its linear predictor mapping and Euclidean distance structure remain highly effective for modelling class–predictor relationships, particularly in contexts where simplicity of implementation and predictive accuracy are paramount.

6.3.1 Findings in Relation to Prior Research

Importantly, the current findings also contribute to the broader smoothing literature on nonlinear modelling. The improvements observed with the B-spline expansion were not uniform and were most pronounced when the underlying structure of the predictor distributions was itself well-behaved, either through log transformation or approximate normality. This pattern is consistent with longstanding work in the spline literature, which emphasized that the performance and stability of spline-based models depend critically on both the underlying functional form and the statistical properties of the predictors (Eilers and Marx, 1996; De Boor, 2001; Hastie, Tibshirani and Friedman, 2009).

Notably, the present results also echo a central conclusion of Eilers and Marx (1996), namely that B-splines combined with a smoothness penalty can provide an effective balance between flexibility and stability. In contrast to their approach, where the penalty was applied directly to the smoothness of the spline functions, the current model imposes the penalty on the regression weights, achieving a similar stabilization of estimated trajectories. Penalized B-splines (P-splines) mitigate the instability and over-fitting that arise from increasing knot density while still delivering smooth, well-behaved functional estimates, even under strongly nonlinear generative mechanisms. The application of this approach within the MRU framework extends earlier work on spline-based unfolding (Lu and Clarkson, 1999; Winsberg and Ramsay, 1980; Wezel and Kosters, 2004) by demonstrating that penalized B-splines, implemented here through a ridge penalty on the spline coefficient's constitute a theoretically justified and effective method for capturing non-linear predictor–class relationships in multinomial unfolding contexts.

6.3.2 Strengths

An important strength of the present study lies in its holistic approach to model evaluation. The balanced use of both quantitative and qualitative criteria provides complementary perspectives on model performance and utility. Quantitatively, model performance was assessed through cross-validated in-sample and out-of-sample deviance, as well as classification accuracy, offering robust evidence regarding the models' generalization capabilities. Qualitatively, the study examined latent-space geometry and the behavior of predictor trajectories, yielding insight into how the models organised class structure and represent predictor relationships. This multi-faceted approach enabled a more comprehensive understanding of model behavior, capturing both numerical performance and interpretive characteristics.

Moreover, a further strength of the present study was the comprehensive and reproducible design of the simulation component. The simulation framework incorporated multiple non-linear generative functions which allowed the models to be evaluated under a variety predictor-class relationships. Class separation was systematically varied, enabling assessment of how the spline-based MRU behaves across conditions ranging from highly overlapping to well-separated classes. Multiple sample sizes were also included, ensuring that conclusions were not tied to a particular level of data richness. In addition, the use of empirically informed class centroids derived from the liver data set provided a realistic foundation for the simulated latent structure. These elements contributed to a simulation design capable of spanning across a broad spectrum of plausible data-generation scenarios.

6.3.3 Limitations

A key limitation of the present research was the reliance on the *lmap* liver data set, both as the primary source of real-world validation and as the basis for defining class coordinates in the simulation study. As a

result, the findings may partly reflect idiosyncrasies of this specific data set rather than general properties of the spline-based penalized MRU framework. By deriving class centroids directly from the liver data set, the simulation study may have unintentionally reinforced patterns unique to this particular sample thereby constraining the range of latent structures the model was able to capture. Consequently, the generalisability of the findings to other data sets with different statistical characteristics may be limited.

Another limitation of the present study concerns the restricted dimensionality of the latent space. All analyses were conducted in two dimensions, which was appropriate for visualization and facilitates intuitive interpretation of predictor trajectories and class configurations. However, constraining the MRU framework to a 2D solution may limit its expressive capacity, particularly when modelling more complex relationships between predictors and class structure. Spline-based transformations may behave differently in higher-dimensional latent spaces, where additional degrees of freedom could allow the model to capture patterns that cannot be represented in two dimensions. As a result, the findings reported here may underestimate the potential benefits or risks of spline-based MRU when the latent space dimensionality is not restricted. Future studies should therefore investigate how model behavior changes when higher-dimensional embedding are included.

6.3.4 Future Directions

A key avenue for future research concerns the flexibility of the spline basis. Although spline models can, in principle, assign a different number of basis functions df_p to each predictor, the present implementation imposed a common degree of freedom across all predictors $p = 1, \ldots, P$. This simplification ensured comparability and reduced the complexity of the tuning procedure, but it may have also constrained the model's ability to represent heterogeneous predictor relationships. The allowance of df_p to vary by predictor would permit the MRU model to allocate modelling capacity where it is most needed. For example, assigning a richer basis to a predictor with strong curvature while using a less rich basis for a predictor that exhibits near-linear effects. This predictor-specific adaptation could be achieved through penalization scheme that encourage differential smoothness across predictors. Exploring this extensions would provide a more data-responsive implementation of spline-based MRU and may yield further improvements in interpretability, model fit and predictive accuracy.

A further direction for future research concerns the use of higher-dimensional latent spaces. The present study restricted all analyses to two dimensions. However, this constraint may have limited the expressive power of the spline-based MRU model, as performance in higher-dimensional embeddings could either improve by enabling the model to capture more complex structures or degrade if the added flexibility introduces instability and over-fitting. Investigating the effects of increased latent dimensionality may therefore provide a deeper understanding of both the potential benefits and risks of spline-based transformations. Future work should examine how model behaviour, stability, and interpretability evolve as the latent embedding is expanded.

6.3.5 Conclusion

In conclusion, the spline-based penalized MRU extends the original unfolding framework by enabling the model to capture richer predictor–class relationships while preserving its geometric interpretability. Across simulations and the real world application, the spline extension yielded more nuanced latent representations and improved recovery of predictor effects, particularly when predictor distributions were well-behaved or transformed to approximate normality. However, this method requires careful tuning, can be sensitive to data characteristics and does not improve quantitatively on model fit or predictive performance over the original formulation. Overall, the approach offers a flexible and interpretable alternative to the standard MRU model and provides a promising foundation for future developments in unfolding.

References

De Boor, C. (2001). *A practical guide to splines* (Rev. ed.). Springer. https://doi.org/10.2307/2006241

De Rooij, M., & Busing, F. M. T. A. (2024). Multinomial restricted unfolding. *Journal of Classification, 41*, 190–213. https://doi.org/10.1007/s00357-024-09465-3

Eilers, P. H. C., & Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science, 11(2)*, 89–121. https://doi.org/10.1214/ss/1038425655

Einhorn, H. J. (1970). The use of nonlinear, noncompensatory models in decision making. *Psychological Bulletin, 73(3)*, 221-230. https://doi.org/10.1037/h0028695

Failenschmid, J. I., Vogelsmeier, L. V. D. E., Mulder, J., & Jongerling, J. (2024). Modelling nonlinear psychological processes: Reviewing and evaluating non-parametric approaches and their applicability to intensive longitudinal data. PsyArXiv Preprints. https://doi.org/10.31234/osf.io/26mde

Harrell, F. E., Jr. (2015). *Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis* (2nd ed.). Springer. https://doi.org/10.1007/978-3-319-19425-7

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7

Kim, C. N., Yang, K. H., & Kim, J. (2008). Human decision-making behavior and modeling effects. *Decision Support Systems*, 45(3), 517–527. https://doi.org/10.1016/j.dss.2007.06.011

Lee, S. (2015). A note on standardization in penalized regressions. *Journal of the Korean Statistical Society, 26(2)*, 505–516. https://dx.doi.org/10.7465/jkdi.2015.26.2.505

Liu, X., Nassar, H., & Podgórski, K. (2020). Splinets: Efficient orthonormalization of the B-splines. *Applied Mathematics and Computation, 369*, 124870. https://doi.org/10.48550/arXiv.1910.07341

Lu, Z. Q., & Clarkson, D. B. (1999). *Monotone spline and multidimensional scaling.* Conference paper presented at the Interface Symposium on Computing Science and Statistics, Seattle, WA, United States.

Piegat, A., & Sałabun, W. (2012). Nonlinearity of human multi-criteria in decision-making. *Journal of Theoretical and Applied Computer Science, 6(3)*, 36–49. https://www.jtacs.org

Perperoglou, A., Sauerbrei, W., Abrahamowicz, M., & Schmid, M. (2019). A review of spline function procedures in R. *BMC Medical Research Methodology, 19(1)*, 46. https://doi.org/10.1186/s12874-019-0666-3

Schweizer, R., & Vahlne, J.-E. (2022). Non-linear internationalization and the Uppsala model – On the importance of individuals. *Journal of Business Research, 140*, 583–592. https://doi.org/10.1016/j.jbusres.2021.11.025

Vallarino, D. (2025). How do consumers really choose? Exposing hidden preferences with the Mixture of Experts model. *arXiv*. https://arxiv.org/abs/2503.05800

Van Wezel, M. C., & Kosters, W. A. (2004). Nonmetric multidimensional scaling: Neural networks versus traditional techniques. *Intelligent Data Analysis: An International Journal, 8 (6)*, 601–613. https://doi.org/10.3233/IDA-2004-8606

Winsberg, S., & Ramsay, J. O. (1980). Monotonic Transformations to Additivity Using Splines. *Biometrika*, 67(3), 669–674. https://doi.org/10.2307/2335137