Bayesian Adaptive Sparse Copula[†]

Martin Burda[a*] Artem Prokhorov[b]

January 3, 2025

Abstract

Bayesian nonparametric density estimation procedures are typically based on single-scale priors, such as Dirichlet process mixtures. Alternative multiscale density priors built on deep decision trees have a number of well-known advantages, including the ability to characterize abrupt local changes and to provide an estimate with a desired level of resolution. Despite their theoretical appeal, multiscale methods have typically been developed in the literature as univariate. Their multivariate versions are generally very costly to implement in practical applications, rendering such methods infeasible in many empirical cases of interest. One of the key reasons is the rapidly increasing number of multiscale mixture components required to represent dependence structures in higher dimensions. In this paper, we propose a random Bernstein polynomial prior on the unit hypercube of arbitrary dimension with a spike-and-slab shrinkage structure. The polynomial components with near-zero posterior weights are shrunk towards zero and thus omitted from posterior sampling. This results in posterior sparsity of the multiscale decision tree, alleviating the curse of dimensionality. We embed the proposed model in the form of a copula link function along with nonparametric marginals in a composite prior over general spaces of densities. We provide conditions for posterior consistency under the weak topology. We further illustrate the practical use of the model in an application to forecasting the Value at Risk and Expected Shortfall of a financial portfolio in a scenario where sampling from the non-sparse posterior would be infeasible.

JEL: C11, C14, C15, C58
Keywords: nonparametric copula, multiscale estimation, Bernstein polynomial

[†]We would like to thank the participants of the Midwest Econometrics Group meetings at Michigan State University, 2022, and the Canadian Econometrics Study Group meetings at McMaster University, 2023, for insightful comments and suggestions. Financial support was provided by the University of Sydney - University of Toronto Partnership Collaboration Awards. Prokhorov's research was supported by research grants from Australian Research Council and Russian Science Foundation.

[a]Department of Economics, University of Toronto, 150 St. George St., Toronto, ON M5S 3G7, Canada; Email: martin.burda@utoronto.ca

[*]Corresponding author.

[b]Discipline of Business Analytics & CEBA & CIREQ, University of Sydney, H70 Abercrombie Building, Sydney, NSW 2006, Australia; E-mail: artem.prokhorov@sydney.edu.au

1 Introduction

Nonparametric Bayesian estimation of multivariate densities has become a popular modeling tool in the literature due to its versatility of usage and relatively weak assumptions on the underlying model structure. Models based on infinite mixtures express a distribution as a composite object made of simpler distributions without a-priori restricting the number of data-driven mixture components. In many contexts, a countably infinite mixture is also a more realistic model than a mixture with a small fixed number of components.

Areas of application include treatment effect estimation (Chib and Hamilton, 2002), autoregressive panel data models (Hirano, 2002), financial econometrics (Jensen and Maheu, 2010), latent heterogeneity in discrete choice models (Kim et al., 2004; Burda et al., 2008), contingent valuation models (Fiebig et al., 2009), and instrumental variables models (Conley et al., 2008). The bulk of the available Bayesian nonparametric density estimation methods, including the popular Dirichlet Process mixtures, are single-scale approaches. Yet, alternative multiscale models, such as deep decision tree structures, feature many advantages. These include adaptability to abrupt local variation and the ability to adjust the estimate locally to the desired degree of resolution.

In a seminal paper for multiscale density estimation, Canale and Dunson (2016), henceforth CD, proposed a univariate multiscale Bernstein polynomial (msBP) prior on the unit interval based on infinite binary trees. However, a direct multivariate generalization of msBP (Burda and Prokhorov, 2024) is very costly to implement in practical applications, rendering such approach infeasible in many empirical cases of interest. Indeed, we show that in as few as three dimensions, the number of polynomial components to be evaluated in the non-sparse posterior is higher than several billion in tree scales higher than 10, exceeding the memory capacity of a typical high-end workstation.

In this paper, we propose a multivariate multiscale shrinkage Bernstein polynomial (SBP) prior on the unit hypercube, based on a spike-and-slab structure. The prior induces posterior tree sparsity as polynomial components with near-zero posterior weights are shrunk towards zero. We embed the SBP as a copula link function with Dirichlet Process mixtures for marginals, inducing a prior over general density spaces. We illustrate the empirical use of the SBP prior in an application to forecasting the Value at Risk and Expected Shortfall of a financial portfolio.

The practical appeal of our approach is that the number of shrunk polynomial components is very large relative to the number of terms retained in the tree. The sparse posterior thus requires only a fraction of the run time and memory size relative to its non-sparse counterpart. This makes our approach suitable for multivariate applications with more than a few dimensions, and usage within wider structural models.

We further provide conditions for posterior consistency under the weak topology, based on the feature of the SBP that it becomes asymptotically dense in the multivariate msBP. This result is valuable, since Bayesian nonparametric models can be inconsistent even with seemingly natural priors (Ghosal and van der Vaart, 2017). If the prior is not correctly specified or is too diffuse (e.g. allowing for too many components) then the corresponding posterior may not concentrate on the true data-generating process as the sample size increases. In practical terms, this result allows our approach to be used by analysts who are not Bayesian, as it provides a frequentist justification of its asymptotic validity.

In related literature, Burda and Prokhorov (2014) analyze a single-scale multivariate random Bernstein polynomial for the copula link function, resulting in a flexible but non-sparse model. The current contribution generalizes this approach to a multiscale tree-based structure with posterior sparsity. The motivation comes in part from the feature of the copula density function that concentrates the bulk of the probability mass in specific regions of the unit hypercube corresponding to the dependence structure of the data, typically around the diagonals and in corners. Such intrinsic sparsity with a high degree of local detail on the unit cube is directly amenable to representation with the SBP prior.

The remainder of the paper is organized as follows. In Section 2 we first directly extend the univariate msBP model of CD into higher dimensions, and then propose a sparse alternative based on a spike-and-slab prior structure. We compare the full and sparse tree size, showing the obstacles to practical use of the former. In Section 3 we link the latter with a nonparametric marginal density model for general sparse adaptive multivariate density estimation. Section 4 discusses the conditions for posterior consistency. An empirical application to forecasting the Value at Risk and the Expected Shortfall in Section 5 demonstrates practical relevance of the approach. The application code is freely available in the GitHub repository SBP. Section 6 concludes.

2 Shrinkage Bernstein Polynomial Prior

2.1 Multiscale Bernstein Mixture

In this Section we describe the Multiscale Bernstein Mixture of CD, following their notation. Let Y be a random variable defined over a unit interval $[0,1]$ with density f, which is assumed to follow a multiscale mixture of Bernstein polynomials (msBP) process:

$$f(y) = \sum_{s=0}^{\infty} \sum_{h=1}^{2^s} \pi_{s,h} Be(y; h, 2^s - h + 1) \tag{1}$$

where $Be(a,b)$ denotes the beta density with mean $a/(a+b)$, and $\{\pi_{s,h}\}$ are random weights. The framework can be represented as a binary tree in which each layer is indexed by a scale $s = 0, \ldots, \infty$, and a node index $h = 1, \ldots, 2^s$ within the scale. Thus, each h is implicitly indexed by s, though we omit this indexing for notational convenience where possible. Each node (s,h) in the tree corresponds to a $Be(y; h, 2^s - h + 1)$ density in (1), weighted by $\pi_{s,h}$. Thus, the multiscale mixture in (1) includes 2^s Bernstein polynomial basis densities at each scale s. A scheme of the binary tree with nodes (s,h) is presented in Figure 1.

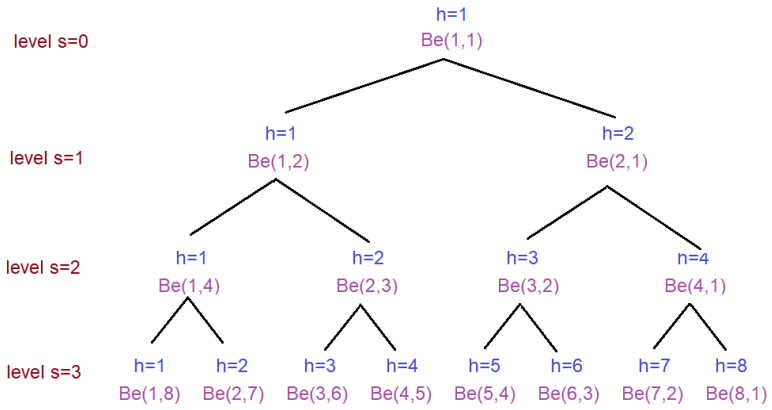

Figure 1: Binary Tree Representation a Multiscale Mixture of Bernstein Polynomials

At each node (s,h), the independent random variables

$$S_{s,h} \sim Be(1, a) \tag{2}$$

$$R_{s,h} \sim Be(b, b) \tag{3}$$

denote the probability of stopping at (s,h) conditionally on reaching (s,h), and taking the path to the right

conditionally on not stopping at (s,h), respectively. The tree node prior weights are then specified as

$$\pi_{s,h} = S_{s,h} \prod_{r<s} (1 - S_{r,g_{shr}})\, T_{shr} \tag{4}$$

where $g_{shr} = \lceil h/2^{s-r} \rceil$ is the node traveled through at scale r on the way to node h at scale s with $\lceil \cdot \rceil$ denoting the ceiling function, $T_{shr} = R_{r,g_{shr}}$ if $(r+1, g_{shr+1})$ is the right daughter of node (r, g_{shr}), and $T_{shr} = 1 - R_{r,g_{shr}}$ if $(r+1, g_{shr+1})$ is the left daughter of node (r, g_{shr}). The infinite tree of probability weights follows the stick-breaking process representation of the Dirichlet process (Sethuraman, 1994). Starting from a "stick" of length one, each time the stick is broken, it is consequently randomly divided in two parts: one for the probability of going right, the remainder for the probability of going left, before the next break. The individual pieces thus add up back to one.

In applications, the msBP process in (1) is approximated by fixing an upper bound s_{max} for the depth of the tree, yielding the scale s_{max} approximation

$$f_{s_{max}}(y) = \sum_{s=0}^{s_{max}} \sum_{h=1}^{2^s} \widetilde{\pi}_{s,h} Be(y; h, 2^s - h + 1). \tag{5}$$

$\widetilde{\pi}_{s,h}$ is identical to $\pi_{s,h}$ except that all the stopping probabilities at scale s_{max} are set to equal to one so that the weights $\widetilde{\pi}_{s,h}$ in (5) sum to one.

Conditionally on a data sample $\mathbf{y} = \{y_i\}_{i=1}^n$, the posterior draws iterate over two Gibbs blocks:

(a) Conditionally on the current values of the probabilities $\{\pi|\mathbf{y}\}$, allocate each y_i to a node (s,h).

(b) Conditionally on the node allocations, update the probabilities $\{\pi|\mathbf{y}\}$ with

$$\pi_{s,h}|\mathbf{y} = S_{s,h}|\mathbf{y} \prod_{r<s} (1 - S_{r,g_{shr}}|\mathbf{y})\, T_{shr}|\mathbf{y} \tag{6}$$

using posterior draws of $S_{s,h}|\mathbf{y}$ and $R_{s,h}|\mathbf{y}$.

The Beta density in (5) serves as a conjugate prior for the binomial conditional likelihood of data allocation to node (s,h). Consequently, the posterior draws of $S_{s,h}$ and $R_{s,h}$ in (b) are taken from

$$S_{s,h}|\mathbf{y} \sim Be(1 + n_{s,h}, a + v_{s,h} - n_{s,h}) \tag{7}$$

$$R_{s,h}|\mathbf{y} \sim Be(b + r_{s,h}, b + v_{s,h} - n_{s,h} - r_{s,h}) \tag{8}$$

where $n_{s,h}$ is the number of data points allocated to node (s,h), $v_{s,h}$ is the number of data points that "pass through" node (s,h) (i.e. are allocated to (s,h) or a node that overlaps (s,h) at a higher scale), and $r_{s,h}$ is the number of data points that proceed down to the right at node (s,h) (i.e. are allocated to a node that overlaps the right daughter node of (s,h) at a higher scale). $v_{s,h}$ has the interpretation of the total number of binomial trials for the node (s,h). In the posterior (7) if we take $n_{s,h}$ as the number of "success" outcomes then $v_{s,h} - n_{s,h}$ is the number of "failure" outcomes. Similarly, in the posterior (8) if we take $r_{s,h}$ as the number of "success" outcomes then $v_{s,h} - n_{s,h} - r_{s,h}$ is the number of "failure" outcomes. The parameters of both posteriors reflect the standard beta-binomial conjugate updating of prior beliefs about the probability of "success".

2.2 Tree Sparsity with Spike and Slab Prior

In this Section, we extend the univariate CD setup by introducing a shrinkage spike-and-slab prior (SSP) for sparsity on the posterior tree weights $\{\pi|\mathbf{y}\}$. This will serve as a stepping stone for the multivariate case that will be used in the subsequent copula dependence analysis.

The SSP is a hierarchical model in which a parameter κ either attains some fixed value κ_0 with non-zero probability, called "the spike", or is drawn from some other prior $p(\kappa)$ called "the slab". When $\kappa_0 = 0$ the SSP can induce posterior sparsity, offering a principled probabilistic alternative to penalty-based regularizers (Bai et al., 2022). The SSP was originally proposed as a tool for selection of subsets of predictor variables in a linear regression model (Lempers, 1971; Mitchell and Beauchamp, 1988; Ishwaran and Rao, 2005). Since its inception, the SSP has been extended to a variety of contexts, including generalized linear models, factor analysis, graphical models, non-parametric regression, and function selection in Structured Additive Regression (Scheipl et al., 2012).

An SSP prior for a parameter κ is constructed by introducing a latent variable γ such that

$$\gamma \sim Bernoulli(\psi) \tag{9}$$

$$\kappa|\gamma = 0 \sim \delta_{\{\kappa_0\}} \tag{10}$$

$$\kappa|\gamma = 1 \sim p(\kappa) \tag{11}$$

where $\delta_{\{0\}}$ is the Dirac measure of unit mass concentrated at κ_0, and ψ is a parameter. Marginalizing over

the distribution of γ, the SSP takes the form of a mixture model

$$\kappa \sim \psi p(\kappa) + (1-\psi)\delta_{\{\kappa_0\}}. \tag{12}$$

We change the prior for $S_{s,h}$ in (2) to

$$S_{s,h} \sim Be(\kappa, a) \tag{13}$$

and for the parameter κ we specify the SSP hierarchical structure (9)-(12) with $\kappa_0 = 0$. This reflects the belief of the analyst that conditional on reaching the node (s,h) the probability of stopping there is κ, drawn from (12). Thus, $S_{s,h}$, $R_{s,h}$, and by extension $\pi_{s,h}$ are endowed with a proper prior with full support over the decision tree. The posterior draws of $S_{s,h}$, instead of (7), are now taken according to

$$S_{s,h}|\mathbf{y} \sim Be(\kappa + n_{s,h}, a + v_{s,h} - n_{s,h}) \tag{14}$$

The degree of approximation smoothness can be controlled by the parameter s_{max} in (5) and by a further optional hyperprior $p(s)$ on s supported over the set of integers $1, \ldots, s_{max}$.

Let us introduce the following definition.

Definition 1 *For a binary decision tree node (s,h), let $\left[\frac{h_s-1}{2^s}, \frac{h_s}{2^s}\right)$ denote its grid cell obtained by partitioning of the unit interval at level s and node index h. Then, a posteriori, the node is called dormant if $y_i \notin \left[\frac{h_s-1}{2^s}, \frac{h_s}{2^s}\right)$ for all $i = 1, \ldots, n$, that is, if it does not contain any data point within its grid cell. Conversely, the node is called active if $y_i \in \left[\frac{h_s-1}{2^s}, \frac{h_s}{2^s}\right)$ for some $i = 1, \ldots, n$, that is, if it contains at least one data point within its grid cell.*

In order to achieve posterior tree sparsity, our goal is to obtain posterior tree node weights with the property $\pi_{s,h}|\mathbf{y} < \varepsilon$ for dormant nodes (s,h). Here ε denotes a "machine epsilon" which is a very small number indistinguishable from zero by computer floating-point arithmetic[1]. Such nodes can then be omitted from posterior sampling in the implementation, yielding posterior tree sparsity.

In the SSP specification (9)-(11), if we let $\psi \to 0$ then $Pr(\gamma = 0) \to 1$ and therefore there exists a value ψ^* close to 0 for which the property above will be satisfied. In the SSP (9) we set ψ to equal such value ψ^*.

[1] In the commonly used double precision real number representation in our application (64-bit IEEE 754 standard), the machine epsilon is approximately 1×10^{-16}.

Due to the SSP structure, the hyperparameter ψ can be interpreted in the prior for κ (12) as the weight the analyst assigns to $p(\kappa)$ as opposed to $\delta_{\{\kappa_0\}}$. In the posterior for $S_{s,h}$ (14), ψ reflects the influence of $p(\kappa)$ as opposed to the data information contained in $v_{s,h}$ and $n_{s,h}$. Thus, our SSP specification minimizes the influence of the prior $p(\kappa)$ and places heavy emphasis on data information. The exact functional form of $p(\kappa)$ is immaterial.

The posterior sparsity scheme is illustrated in Figure 2 for $n = 1$, with active nodes marked in red, while the remainder of the tree is dormant.

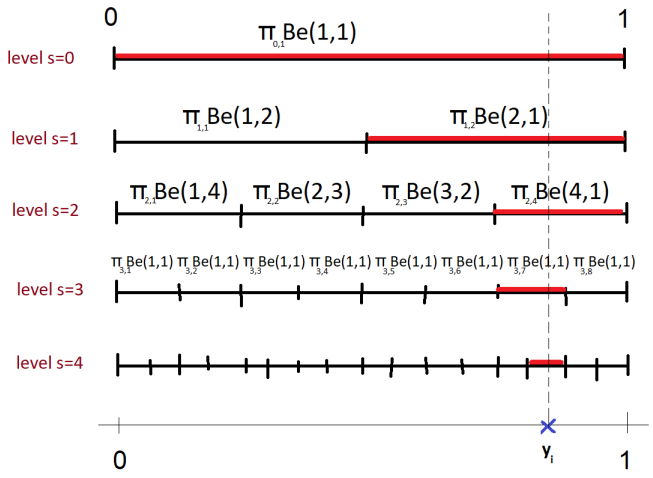

Figure 2: Posterior Tree Sparsity in One Dimension

For medium to large sample sizes, we expect most parts of the univariate tree to be active for nodes at the lower levels of approximation. However, sparsity becomes progressively more pronounced in higher dimensions, as we detail below. A key feature of a sparse tree that brings significant savings in posterior computation time is that evaluations of $\widetilde{\pi}_{s,h}$ only need to be performed along tree paths that contain at least one data point in their grid cell, up to $s = s_{max}$, i.e. the red nodes in the example in Figure (2).

2.3 Multivariate Multiscale Bernstein Mixture

We first restate the generalization the univariate msBP process of CD to higher dimensions d, as in Burda and Prokhorov (2024), and then introduce a sparse alternative based on a multivariate version of the shrinkage SSP prior. All intuition from the univariate case carries over to the multivariate scenario; instead of Beta densities on $[0, 1]$ we will use the Dirichlet density on the d-dimensional probability simplex with the SSP

hierarchical prior on its concentration parameter.

Let $\boldsymbol{Y}$ be a random vector defined over a unit hypercube $[0,1]^d$ with density $f(\boldsymbol{y})$, which is assumed to follow a multivariate multiscale mixture of Bernstein polynomials process

$$f(\boldsymbol{y}) = \sum_{s=0}^{\infty} \sum_{\mathbf{h}=(1,\ldots 1)}^{(2^s,\ldots,2^s)} \pi_{s,\mathbf{h}} \prod_{j=1}^{d} Be(y_j; h_j, 2^s - h_j + 1) \tag{15}$$

where y_j is the jth coordinate of $\boldsymbol{y} \in [0,1]^d$ and h_j is the node index within each scale s along the dimension j, with $\mathbf{h} = (h_1, \ldots, h_d)$. Similarly to the univariate case, $\mathbf{h}$ is implicitly indexed by its corresponding s though we omit this indexing for notational convenience. A scheme of a tree for $d = 2$ is presented in Figure 3 up to $s = 2$.

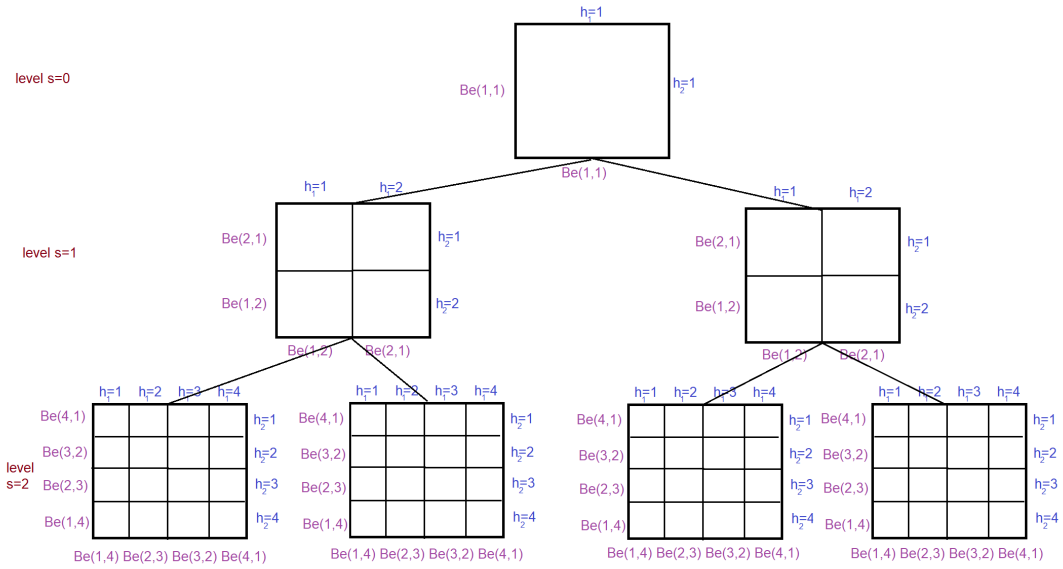

Figure 3: Two-dimensional Tree Representation of a Bivariate Multiscale Mixture of Bernstein Polynomials

Conditionally on reaching node $(s, \mathbf{h})$, the random variable

$$S_{s,\mathbf{h}} \sim Be(1, a) \tag{16}$$

denotes the probability of stopping at $(s, \mathbf{h})$. Let $(s, \mathbf{H}) = \{(s+1, \mathbf{h}_1), \ldots, (s+1, \mathbf{h}_{2^d})\}$ denote all the daughter nodes of $(s, \mathbf{h})$. Conditionally on not stopping at $(s, \mathbf{h})$, the 2^d-dimensional random vector

$$\mathbf{Q}_{s,\mathbf{H}} \sim Dir(\boldsymbol{b}) \tag{17}$$

where $\mathbf{Q}_{s,\mathbf{H}} = (Q_{s+1,\mathbf{h}_1}, \ldots, Q_{s,\mathbf{h}_{2^d}})$ then denotes the probabilities of advancing at level $s+1$ to any one of the 2^d daughter nodes of the node $(s,\mathbf{h})$. $Dir(\boldsymbol{b})$ stands for the Dirichlet distribution with $\boldsymbol{b} = (b_1, \ldots, b_{2^d})$. The tree node prior weights are given by

$$\pi_{s,\mathbf{h}} = S_{s,\mathbf{h}} \prod_{r=0}^{s-1} \left(1 - S_{r,\boldsymbol{g}_{s\mathbf{h}r}}\right) Q_{r,\boldsymbol{g}_{s\mathbf{h}_r}} \tag{18}$$

where $\boldsymbol{g}_{s\mathbf{h}r} = (g_{sh_1r}, \ldots, g_{sh_dr})$, with $g_{sh_jr} = \lceil h_j/2^{s-r} \rceil$, is the node traveled through at scale r on the way to node $\mathbf{h}$ at scale s. The prior (16)-(18) is the multivariate counterpart of (2)-(4).

The scale s_{max} approximation of (15) is then

$$f_{s_{max}}(\boldsymbol{y}) = \sum_{s=0}^{s_{max}} \sum_{\mathbf{h}=(1,\ldots 1)}^{(2^s,\ldots,2^s)} \widetilde{\pi}_{s,\mathbf{h}} \prod_{j=1}^{d} Be(y_j; h_j, 2^s - h_j + 1) \tag{19}$$

where $\widetilde{\pi}_{s,\mathbf{h}}$ is identical to $\pi_{s,\mathbf{h}}$ except that all the stopping probabilities at scale s_{max} are set to equal to one so that the weights $\widetilde{\pi}_{s,h}$ in (19) sum to one. The univariate msBP process (1) and its scale s_{max} approximation (5) are special cases of (15) and (19), respectively, for $d = 1$.

Similarly to the univariate case, conditionally on a data sample $\mathbf{y} = \{\boldsymbol{y}_i\}_{i=1}^n$, the posterior draws iterate over two Gibbs blocks:

(a) Conditionally on the current values of the probabilities $\{\pi|\mathbf{y}\}$, allocate each $\boldsymbol{y}_i$ to a node $(s,\mathbf{h})$.

(b) Conditionally on the node allocations, update the probabilities $\{\pi|\mathbf{y}\}$ with

$$\pi_{s,\mathbf{h}}|\mathbf{y} = S_{s,\mathbf{h}}|\mathbf{y} \prod_{r=0}^{s-1} \left(1 - S_{r,\boldsymbol{g}_{s\mathbf{h}r}}|\mathbf{y}\right) Q_{r,\boldsymbol{g}_{s\mathbf{h}_r}}|\mathbf{y} \tag{20}$$

using posterior draws of $S_{s,\mathbf{h}}|\mathbf{y}$ and $\mathbf{Q}_{s,\mathbf{H}}|\mathbf{y}$.

The Dirichlet distribution in (17) serves as a conjugate prior for the multinomial conditional likelihood of data allocation to node $(s,\mathbf{h})$. Consequently, the posterior draws of in (b) are taken from

$$S_{s,\mathbf{h}}|\mathbf{y} \sim Be(1 + n_{s,\mathbf{h}}, a + v_{s,\mathbf{h}} - n_{s,\mathbf{h}}) \tag{21}$$

$$\mathbf{Q}_{s,\mathbf{H}}|\mathbf{y} \sim Dir(\mathbf{b} + \mathbf{q}_{s,\mathbf{H}}, \mathbf{b} + v_{s,\mathbf{h}} - n_{s,\mathbf{h}} - \mathbf{q}_{s,\mathbf{H}}) \tag{22}$$

where $n_{s,\mathbf{h}}$ is a number of data points allocated to node $(s,\mathbf{h})$, $v_{s,\mathbf{h}}$ is a number of data points that pass through node $(s,\mathbf{h})$, and $\mathbf{q}_{s,\mathbf{H}}$ is a vector of the numbers of data points that proceed down to the respective

daughter nodes $(s, \mathbf{H})$. The Dirichlet-multinomial conjugate updating in (21) and (22) generalizes to d dimensions the intuition detailed above for (7) and (8).

2.4 Multivariate Posterior Tree Sparsity

We now change the prior for $S_{s,\mathbf{h}}$ in (3) to

$$S_{s,\mathbf{h}} \sim Be(\kappa, a) \tag{23}$$

and for the parameter κ we specify the SSP hierarchical structure (9)-(12) with $\kappa_0 = 0$ and $\psi = \psi^*$. The posterior draws of $S_{s,\mathbf{h}}$, instead of (21), are now taken according to

$$S_{s,\mathbf{h}}|\mathbf{y} \sim Be(\kappa + n_{s,\mathbf{h}}, a + v_{s,\mathbf{h}} - n_{s,\mathbf{h}}) \tag{24}$$

This approach induces posterior tree sparsity by the same principle as detailed for the univariate case. A node $(s, \mathbf{h})$ is active if the event

$$\left\{ y_i \in \left[\frac{h - 1/2}{2^s}, \frac{h}{2^s} \right) \text{ for some } i = 1, \ldots, n \right\}$$

is true, and dormant otherwise. The sparsity scheme is illustrated in Figure 4 for one data point, $n = 1$, with active nodes shaded while the remainder of the tree is dormant.

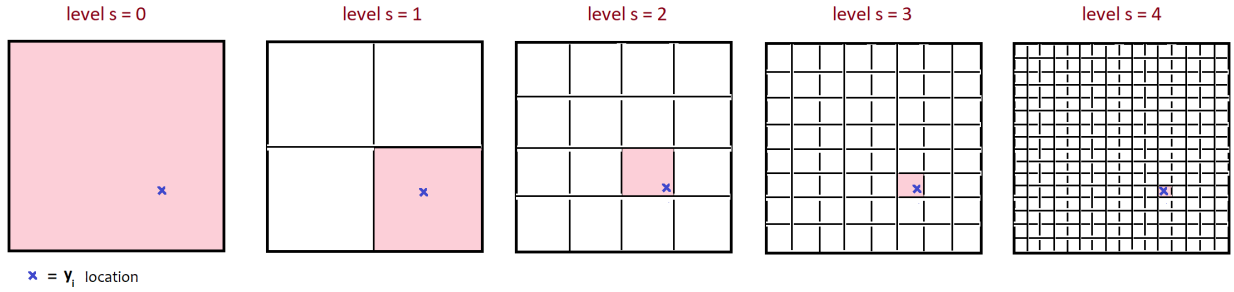

Figure 4: Tree Sparsity in 2 Dimensions

2.5 Comparison

In the univariate case of CD, and its non-sparse multivariate extension, in the node allocation step (a) a new Markov Chain (MC) update s_i is drawn for each i from the multinomial with probabilities proportional to

the total posterior probability mass associated with the given s,

$$Pr(s_i = s|\boldsymbol{y}_i) = \sum_{\mathbf{h}=(1,\ldots,1)}^{(2^s,\ldots,2^s)} \bar{\pi}_{s,\mathbf{h}} \prod_{j=1}^{d} Be(y_{ij}; h_j, 2^s - h_j + 1)$$

where $\bar{\pi}_{s,\mathbf{h}} = \pi_{s,\mathbf{h}}/\pi_s$ with

$$\pi_s = \sum_{\mathbf{h}=(1,\ldots 1)}^{(2^s,\ldots,2^s)} \pi_{s,\mathbf{h}}$$

if π_s exceeds a threshold $u_i \sim U(0, \pi_{s_i})$ from the previous MC draw.

Next, a new MC update of $\mathbf{h}$, for each i and the corresponding s_i, requires a draw from a multinomial over the support $\mathbf{h} = (1, \ldots, 1), \ldots, (2^{s_i}, \ldots, 2^{s_i})$ with posterior probability

$$Pr(\mathbf{h}_i = \mathbf{h}|\boldsymbol{y}_i, s_i) = \bar{\pi}_{s_i,\mathbf{h}} \prod_{j=1}^{d} Be(y_{ij}; h_j, 2^{s_i} - h_j + 1)$$

The size of the support of this multinomial grows exponentially in d and rapidly increases with s. For example, for $d = 3$ and $s_{max} = 10$ the tree size exceeds 1 billion nodes, with as many weights $\pi_{s,\mathbf{h}}$ required for evaluation of $f_{s_{max}}(y)$ in (19). These updates and approximation evaluations are feasible to implement in real time only in very few dimensions for relatively shallow trees (CD used $s_{max} = 4$ in their univariate application with $d = 1$).

Nonetheless, for the non-sparse tree, even for moderate s the vast majority of the posterior sampling probabilities for $\mathbf{h}$ will be very close to zero, since the product of the unimodal $Be(y_i; h_j, 2^{s_i} - h + 1)$ functions peaks over the node $(s, \mathbf{h})$ that contains $\boldsymbol{y}_i$ and drops sharply to zero over nodes that cover the support of $[0, 1]^d$ for values that move away from $\boldsymbol{y}_i$. This reflects the inherent sparsity property of the copula functions mentioned in the Introduction. These negligible posterior mass nodes are very unlikely to be allocated with an $\mathbf{h}$ update and hence the prior over them will only rarely be updated by the data in forming the posterior. Our SBP approach shrinks such weights below computer zero and eliminates them from evaluation for the MC updates of the posterior.

The ramifications of the curse of dimensionality in higher dimensions are similar in the tree weights update step (b), which requires $\sum_{s=0}^{s_{max}} 2^{d\times s}$ evaluations of $n_{s,\mathbf{h}}$, $v_{s,\mathbf{h}}$, and draws from (21) and (22), which quickly reach billions of operations in just a handful of dimensions for moderate tree depths. Our SBP approach allows us to perform the required updates only along the active paths through sparse trees, rendering their implementation feasible in real time. In the sparse tree there can be at most $s_{max}N$ active nodes which is

substantially less than $2^{d \times s_{max}}$ in the non-sparse tree. For example, for $d = 3$ and $s_{max} = 10$ the sparse tree will have around $10,000$ active nodes while the size of the non-sparse tree exceeds 1 billion nodes.

Which nodes are active in the posterior and hence the exact size of the sparse tree depends on the sample $\{\boldsymbol{y}\}_{i=1}^{n}$. For illustration, we compare the sizes of the sparse versus the non-sparse tree in Table 1 below, based on a random sample drawn from $U[0,1]^d$. In just a few dimensions for a relatively small scale s the size of the non-sparse tree exceeded the size of the sparse tree by orders of magnitude. In three dimensions for $s > 10$ the non-sparse tree was not feasible to store in the memory of our Dell Precision T7960 workstation.

	$d=1$		$d=2$		$d=3$	
s_{max}	Non-sparse	Sparse	Non-sparse	Sparse	Non-sparse	Sparse
1	3	2	5	2	9	2
2	7	6	21	10	73	18
3	15	10	85	18	585	34
4	31	14	341	26	4,681	50
5	63	22	1,365	57	37,449	119
6	127	37	5,461	119	299,593	212
7	255	63	21,845	212	2,396,745	312
8	511	108	87,381	311	19,173,961	412
9	1,023	175	349,525	411	153,391,689	512
10	2,047	253	1,398,101	511	1,227,133,513	612
11	4,095	346	5,592,405	611	-	712
12	8,191	443	22,369,621	711	-	812
13	16,383	543	89,478,485	811	-	912
14	32,767	643	357,913,941	911	-	1,012
15	65,535	743	1,431,655,765	1,011	-	1,112
16	131,071	843	-	1,111	-	1,212
17	262,143	943	-	1,211	-	1,312
18	524,287	1,043	-	1,311	-	1,412
19	1,048,575	1,143	-	1,411	-	1,512
20	2,097,151	1,243	-	1,511	-	1,612

Table 1: Summary of Estimated Coefficients

Figure 5 shows the location of active nodes in a tree for a simulated sample of 100 observations shaded in black while dormant nodes remain white.

3 Copula Link with Marginal Distributions

Due to its support on a $d-$dimensional hypercube, the mixture of Bernstein polynomials process (15) can be used to represent a non-parametric copula density model. We will link it with marginal distributions supported on the real line to obtain a general joint dependence structure for a real-valued vector of random variables. This approach extends Burda and Prokhorov (2014) who considered a random Bernstein polynomial prior without multiscale adaptivity and sparsity. Where applicable, we will use the notation and

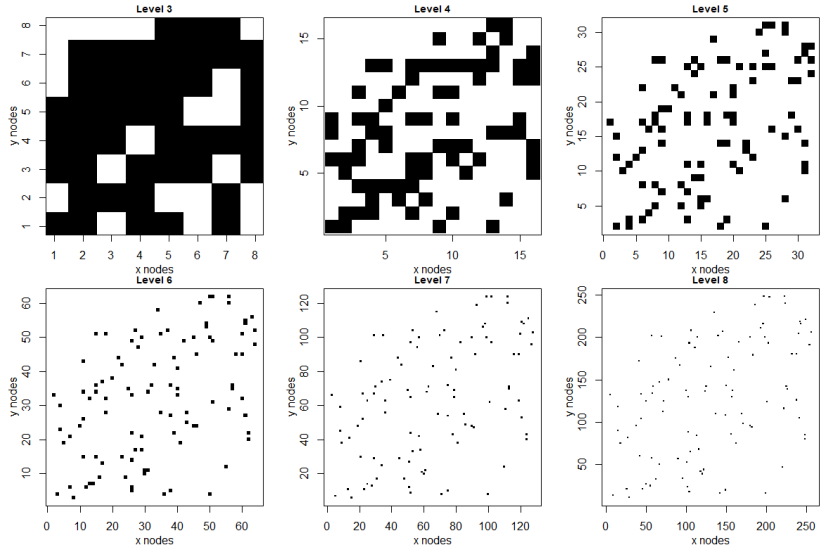

Figure 5: Active Nodes

terminology of Wu and Ghosal (2008), henceforth WG, whose results we use to establish the conditions for posterior consistency.

Let $\mathcal{X} \subseteq \mathbb{R}^d$ be a sample space with elements $\boldsymbol{x}$, Θ the space of a mixing parameter θ, and Φ the space of a hyper parameter ϕ. Let $\mathcal{D}(\mathcal{X})$ denote the space of probability measures F on $\mathcal{X}$. Denote by $\mathcal{M}(\Theta)$ the space of probability measures on Θ and let P be the mixing distribution on Θ with density p and a prior Π on $\mathcal{M}(\Theta)$ with weak support $supp(\Pi)$. Denote the prior for ϕ by μ, and the support of μ by $supp(\mu)$, with μ independent of P. Let $K(x;\theta,\phi)$ be a kernel on $\mathcal{X} \times \Theta \times \Phi$, such that $K(\boldsymbol{x};\theta,\phi)$ is a jointly measurable function with the property that for all $\theta \in \Theta$ and $\phi \in \Phi$, $K(\cdot;\theta,\phi)$ is a probability density on $\mathcal{X}$. Π, μ and $K(\boldsymbol{x};\theta,\phi)$ induce a prior on $\mathcal{D}(\mathcal{X})$ via the map

$$(\phi, P) \mapsto f_{P,\phi}(\boldsymbol{x}) \equiv \int K(\boldsymbol{x};\theta,\phi)dP(\theta) \tag{25}$$

Denote such composite prior by Π^*.

Let

$$\begin{aligned} \boldsymbol{y} &= F(\boldsymbol{x};\theta_m,\phi_m) \\ &= \int_{-\infty}^{\boldsymbol{x}} K_m(\boldsymbol{t};\theta_m,\phi_m)d\boldsymbol{t} \end{aligned} \tag{26}$$

where F is a vector of marginal cdfs of $\boldsymbol{x}$ obtained by integrating a vector of marginal density kernels K_m, θ_m is a marginal mixing parameter, and ϕ_m is a marginal hyperparameter. Denote the copula kernel in (15)

by

$$K_c(\boldsymbol{y}; \theta_c, \phi_c) = f(\boldsymbol{y}) \tag{27}$$

where θ_c is a copula mixing parameter and ϕ_m is a copula hyperparameter. The joint kernel in (25) can now be expressed as

$$K(\boldsymbol{x}; \theta, \phi) = K_c(F(\boldsymbol{x}; \theta_m, \phi_m); \theta_c, \phi_c) K_m(\boldsymbol{x}; \theta_m, \phi_m) \tag{28}$$

where

$$K_m(\boldsymbol{x}; \theta_m, \phi_m) = \prod_{j=1}^{d} K_{mj}(x_j; \theta_{mj}, \phi_{mj}). \tag{29}$$

With the view to our financial application, we specify the prior structure on the marginal kernel as in Ausín et al. (2014), with a dynamic Dirichlet Process mixture model:

$$\begin{aligned} x_t &= \gamma + h_t^{1/2} \varepsilon_t \\ h_t &= \omega + \alpha x_{t-1} + \beta h_{t-1} \\ \varepsilon_t | \lambda_t &\sim N(0, \lambda_t^{-1}) \\ \lambda_t | G &\sim G \\ G | \nu, G_0 &\sim DP(\nu, G_0) \end{aligned}$$

where $DP(\nu, G_0)$ is the Dirichlet process with concentration parameter ν and base measure G_0.

4 Posterior Consistency

In this Section we will establish the conditions for posterior consistency of the SBP. Schwartz (1965) proved a seminal result on Bayesian consistency showing that posterior consistency at a "true density" f_0 holds if the prior assigns positive probabilities to a specific type of neighborhoods of f_0 defined by Kullback-Leibler divergence measure (the so-called Kullback-Leibler property) and the size of the model is restricted in an appropriate sense. For the weak topology, the size condition holds automatically (Ghosal, Ghosh, and Ramamoorthi, 1999, Theorem 4.4.2) and hence proving the Kullback-Leibler (K-L) property also proves weak posterior consistency.

Denote by $\mathcal{F}$ the set of all possible joint densities with respect to probability measures in $\mathcal{D}$. Define a K-L neighborhood of a density $f \in \mathcal{F}$ of size ε by

$$\mathcal{K}_\varepsilon(f) \equiv \{g \in \mathcal{F} : \mathcal{K}(f;g) < \varepsilon\}$$

where

$$\mathcal{K}(f;g) = \int f \log(f/g)$$

is the K-L divergence between f and g. By convention, we say that the K-L property holds at $f_0 \in \mathcal{F}$ or f_0 is in the K-L support of Π^*, and write $f_0 \in \mathrm{KL}(\Pi^*)$ if $\Pi^*(\mathcal{K}_\varepsilon(f_0)) > 0$ for every $\varepsilon > 0$.

WG specified high-level conditions under which the K-L property holds for a mixture density $f_{P,\phi}(\boldsymbol{x})$ of the form (25) for a general kernel $K(\boldsymbol{x};\theta,\phi)$ with priors μ and Π for ϕ and P, respectively, and the prior Π^* induced by μ and Π on $\mathcal{D}(\mathcal{X})$. Specifically, if for any $\varepsilon > 0$, there exists P_ε, ϕ_ε, $A \subset \Phi$ with $\mu(A) > 0$ and $\mathcal{W} \subset \mathcal{M}(\Theta)$ with $\Pi(\mathcal{W}) > 0$, such that:
A1. $\int f_0 \log\left(\frac{f_0(\boldsymbol{x})}{f_{P_\varepsilon,\phi_\varepsilon}(\boldsymbol{x})}\right) < \varepsilon$;
A2. $\int f_0 \log\left(\frac{f_{P_\varepsilon,\phi_\varepsilon}(\boldsymbol{x})}{f_{P_\varepsilon,\phi}(\boldsymbol{x})}\right) < \varepsilon$ for every $\phi \in A$,
A3. $\int f_0 \log\left(\frac{f_{P_\varepsilon,\phi}(\boldsymbol{x})}{f_{P,\phi}(\boldsymbol{x})}\right) < \varepsilon$ for every $P \in \mathcal{W}$, $\phi \in A$,
then $f_0 \in \mathrm{KL}(\Pi^*)$ (WG, Theorem 1).

WG further showed that these Conditions A1–A3 were satisfied under specific low-level conditions for certain kernel types, such as the location-scale kernel, gamma kernel, random histogram, and the Bernstein polynomial kernel. Using a toolkit similar to WG, we provide the low-level conditions under which Conditions A1 and A3 and hence weak posterior consistency holds for our sparse multiscale Bernstein copula kernel (28). Since we do not specify hyperparameter priors in the copula or marginal kernels, ϕ_c and ϕ_m are vacuous in our case and we can drop them from further notation, rendering Condition A2 redundant.

Condition A1 was established to hold for a variant of (28) with a non-sparse single-level copula kernel $K_c(\cdot)$ in Theorem 1 of Burda and Prokhorov (2014) under the following Conditions:

B1. For some $0 < \overline{f} < \infty$, $0 < f_0(\boldsymbol{x}) \leq \overline{f}$ for all $\boldsymbol{x}$;
B2. For some $0 < \overline{p} < \infty$, $0 < p(\theta) < \overline{p}$ for all θ, where $p(\theta)$ is the density with respect to P;
B3. $K_m(\boldsymbol{x};\theta_m)$ is continuous in $\boldsymbol{x}$, positive, bounded and bounded away from zero everywhere;
B4. $K_c(\cdot)$, $K_m(\cdot)$, $\log f_{P_\varepsilon}(x)$, $\log K_c(\cdot)K_m(\cdot)$, and $\inf_{\theta\in D} K_c(\cdot)K_m(\cdot)$ are f_0-integrable, the latter for some

closed $D \supset supp(P_\varepsilon)$;

B5. For some $0 < \overline{K} < \infty$, $\int K_c \left(\int_{-\infty}^{\boldsymbol{x}} K_m(\boldsymbol{t}; \theta_m) d\boldsymbol{t}; \theta_c) \right) K_m(\boldsymbol{x}; \theta_m) d\theta = \overline{K}$ for all x;

B6. The weak support of Π is $\mathcal{M}(\Theta)$.

Conditions B1 and B2 require the true density and mixing prior density to be bounded and bounded away from zero. Condition B3 specifies regularity and boundedness conditions on the marginal kernel. Conditions B4 and B5 provide regularity and f_0-integrability conditions for both the copula and marginal kernels and their integrals. Condition B6 is relatively weak and does not make any specific assumptions on Π other than requiring that it has large weak support. Thus, Π accommodates a wide class of priors including the Dirichlet process.

Our SBP copula kernel $K_c(\cdot)$ that appears in Conditions B4 and B5 differs from Burda and Prokhorov (2014) while the other objects in B1–B6 are the same. Condition B4 is an assumption on f_0 that we maintain as being satisfied. What remains to be verified here is that B5 holds in our case under low-level assumptions. For any multiscale tree node $(s, \mathbf{h})$ the Beta density is continuous over the compact unit interval and so is their product over the unit hypercube, which is then bounded by the Extreme Value Theorem. Its weighted average $K_c \left(\int_{-\infty}^{\boldsymbol{x}} K_m(\boldsymbol{t}; \theta_m) d\boldsymbol{t}; \theta_c) \right)$ in (27) is then also bounded and merely rescales a Gaussian marginal density kernel by a finite number. The Gaussian mixture in Condition B5 therefore remains bounded.

Condition A3 was assumed to hold in Burda and Prokhorov (2014) by assuming that the sufficient conditions hold in WG Lemma 3 (updated in Wu and Ghosal, 2009). While its first two boundedness Assumptions A7 and A8 are directly satisfied by Conditions B3 and B4, here we newly show that the uniform equicontinuity Assumption A9 is also satisfied for our SBP copula kernel. Assumption A9 absent of hyperparameters stipulates that for a compact $C \subset \mathcal{X}$, there exists E containing D in its interior such that the family of maps $\{\theta \mapsto K(\boldsymbol{x}; \theta), \boldsymbol{x} \in C\}$ is uniformly equicontinuous on D. In our case, from (26)–(29),

$$K(\boldsymbol{x}; \theta) = (2\pi)^{-1/2} \sum_{s=0}^{\infty} \sum_{\mathbf{h}=(1,\ldots,1)}^{(2^s,\ldots,2^s)} \pi_{s,\mathbf{h}} \prod_{j=1}^{d} Be(y_j(x_j); h_j, 2^s - h_j + 1) \sigma_j^{-1} \exp\left(-(x_j - \gamma_j)^2/(2\sigma_j^2)\right) \tag{30}$$

with $\theta_c = \{\pi_{s,\mathbf{h}}\}$, $\theta_m = \{\gamma_j, \sigma_j^2\}_{j=1}^d$, and $\theta = \{\theta_m, \theta_c\}$. By the definition of uniform equicontinuity, we need to show that for any $\varepsilon > 0$, there exists $\delta > 0$ such that for all $x \in C$ and all $\theta, \theta' \in D$ with $\|\theta - \theta'\| < \delta$ we have $|K(x; \theta) - K(x; \theta')| < \varepsilon$. The product of beta densities over the unit cube is bounded, and so is its product with the Gaussian kernel for $\|\theta_m\| < \infty$. By the stick-breaking construction of (18), the tree weights $\pi_{s,\mathbf{h}}$ sum up to one with $\sum_{\mathbf{h}} \pi_{s,\mathbf{h}} \xrightarrow{s \to \infty} 0$ and therefore $c \equiv \sup_{\theta \in D} K(\boldsymbol{x}; \theta) < \infty$. Let $\delta = \varepsilon / c$, then

for $\|\theta - \theta'\| < \delta$ we have $|K(x;\theta) - K(x;\theta')| < \varepsilon$ which satisfied the definition of uniform equicontinuity for $K(\boldsymbol{x};\theta)$ on D.

5 Application to Value at Risk and Expected Shortfall

We apply our approach to one-day-ahead Value at Risk (VaR) and Expected Shortfall (ES) prediction for a portfolio consisting of several assets. We chose the AOR iShares Portfolio that has been named by Forbes Advisor as the best core balanced Exchange-Traded Fund (ETF) of April 2024 (Friedberg and Adams, 2024). The portfolio is composed of seven assets as detailed in Table 2. We omitted cash holdings with less than 0.5% portfolio weight.

Ticker	Name	Sector	Asset Class	Weight (%)
IVV	iShares core S&P 500 ETF	Corporates	Equity	34.5
IUSB	iShares core total USD bond market	Corporates	Fixed Income	33.1
IDEV	iShares core MSCI int devel ETF	Corporates	Equity	17.0
IEMG	iShares core MSCI emerging markets	Corporates	Equity	6.0
IAGG	iShares core Intl Aggregate Bnd ETF	Corporates	Fixed Income	5.8
IJH	iShares core S&P mid-cap ETF	Corporates	Equity	2.0
IJR	iShares core S&P small-cap ETF	Corporates	Equity	0.9

Table 2: AOR iShares Portfolio

Our data contains daily log returns of these assets from January 3, 2022 to May 31, 2024, obtained from Yahoo Finance (https://finance.yahoo.com), with total sample size $T = 606$ observations. The data series are plotted in Figure 9 in the Appendix, along with the numerical implementation algorithm for our SBP copula.

For sampling of the marginal distributions we follow Ausín et al. (2014). With 7 assets in our portfolio there are 21 pairwise combinations for visual inspection of the nonparametric copula dependence structure so we chose the mid-cap and small-cap equities, IJH and IJR, that are closely correlated, for illustration. For these two assets, Figure 6 shows the copula density, a heatmap overlaid with marginal cdfs evaluated at individual data points (left), and a 3D surface plot (right).

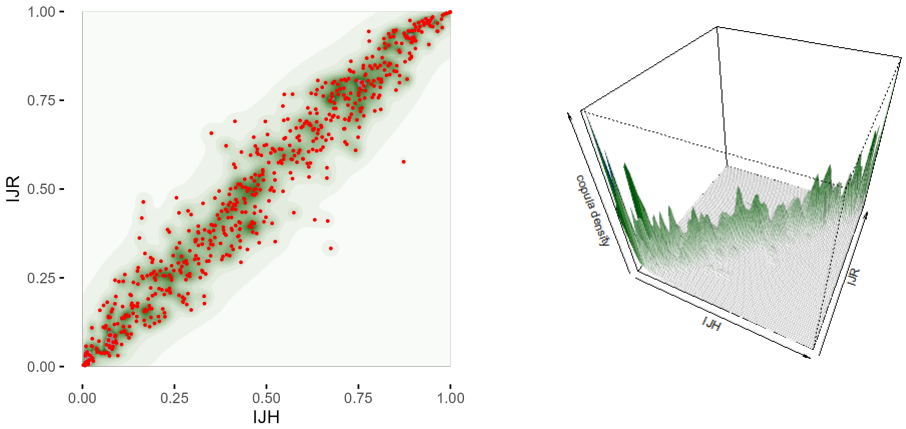

Figure 6: Copula Density

Starting from about 60% of the data set as a training sample spanning June 3, 2022 to May 31, 2023, we ran[2] the implementation algorithm detailed in the Appendix for $1,000$ iterations following a 100 burn-in section for each trading day of the following year, June 1, 2023 to May 31, 2024. At each day and iteration, we drew a simulated value from the one-day-ahead predictive copula density for the next trading day. The algorithm was initialized at the modal parameter values for the marginals and a random draw from the prior for the copula. A dot plot of the copula predictive density draws (blue) overlaid with the marginal cdfs evaluated at individual data points (red) is visualized in Figure 7.

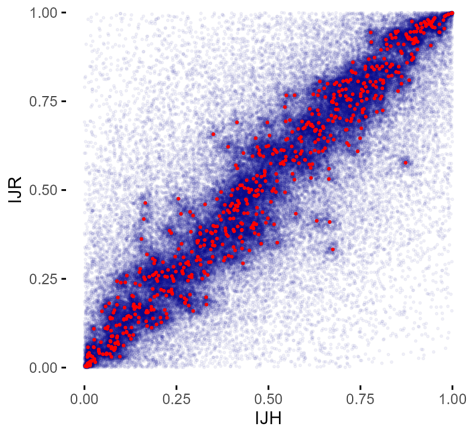

Figure 7: Simulated Draws from Copula Density

[2]Computations were performed in Modern Fortran on a Dell 7960 workstation with Windows Subsystem for Linux (WSL2) using the freely available Nvidia High Performance Computing (HPC) Software Development Kit (SDK). The marginal parameter sampling was implemented in parallel via Message Passing Interface (MPI), with root sampling of the copula structure. The full simulation run took about 1 hour to complete.

We inverted each simulated draw through the corresponding marginal densities, and calculated the resulting simulated portfolio value. From the $1,000$ draws for each day we obtained the corresponding quantile and average to obtain the prediction of VaR and ES, respectively. The daily predictions are shown in Figure 8. We observe that the predicted VaR and ES adapt to changes in volatility. We note that such computations for a portfolio of this dimension would be infeasible using standard nonparametric methods.

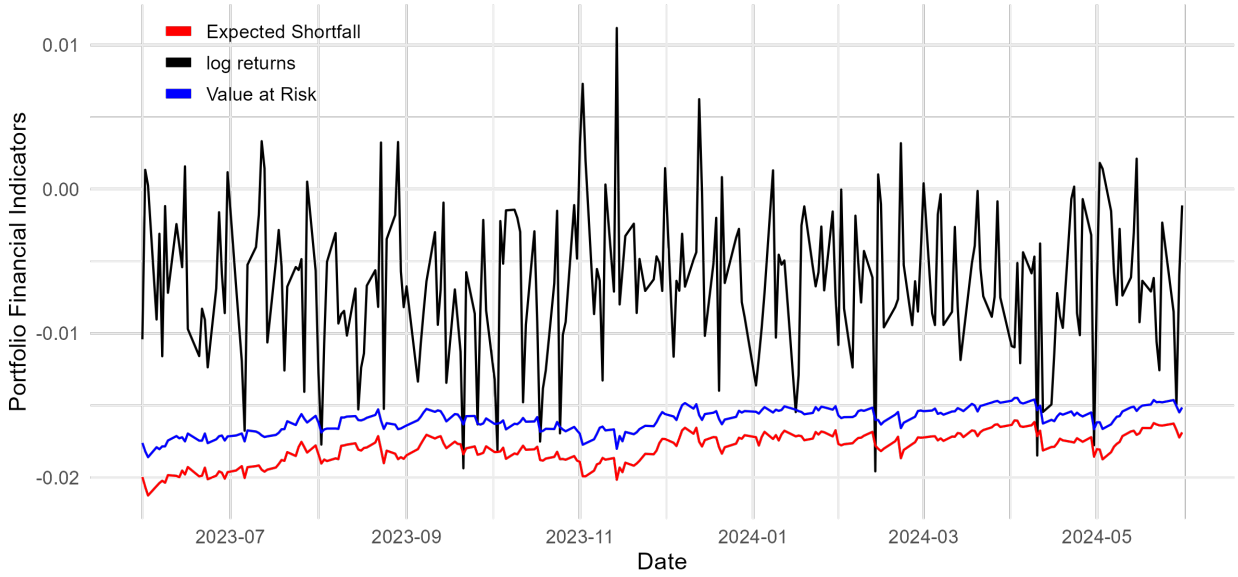

Figure 8: Value-at-Risk and Expected Shortfall

6 Conclusion

In this paper we propose a multivariate multiscale Bernstein polynomial (SBP) model based on a shrinkage spike-and-slab prior structure. Bernstein polynomial functional approximation components with relatively high posterior probability mass are retained in the multivariate tree while components with small weights are dropped. We use the resulting model in a nonparametric copula dependence structure on a unit hypercube and further combine it with nonparametric marginals for general density estimation, establishing conditions for posterior consistency. The sparse posterior requires only a fraction of the run time and memory size relative to its non-sparse counterpart, which makes it suitable for implementation in higher dimensions. We illustrate its practical usefulness in an application forecasting the Value at Risk and Expected Shortfall of a financial portfolio in a scenario where implementation of the non-sparse counterpart would be infeasible.

Appendix

Implementation Algorithm

1. Initialize, for $s = 1, ..., s_{max}$:

   (a) Obtain the indices of *active* nodes $(s, \mathbf{h})$

   (b) Draw $\widetilde{\pi}_{s,\mathbf{h}}$ from the SBP prior

   (c) Evaluate $\prod_{j=1}^{d} Be(y_{ij}; h_j, 2^s - h_j + 1)$

2. For each $i = 1, ..., N$, using values from 1(c) and 1(d), update the latent allocation variable s_i^* by drawing from a multinomial over $s_i = 1, ..., s_{max}$, where

$$P(s_i = s|\cdot) \propto \widetilde{\pi}_{s,\mathbf{h}} \prod_{j=1}^{d} Be(y_{ij}; h_j, 2^s - h_j + 1)$$

   for *active* nodes $(s, \mathbf{h})$ only

3. For all *active* nodes $(s, \mathbf{h})$ update:

   (a) $n_{s,\mathbf{h}}$, $v_{s,\mathbf{h}}$

   (b) $S_{s,\mathbf{h}} \sim Be(1 + n_{s,\mathbf{h}}, a + v_{s,\mathbf{h}} - n_{s,\mathbf{h}})$, $S_{s_{max},\cdot} = 1$

   (c) $\mathbf{h}_{s,\mathbf{h}} \sim Dir(b + v_{s+1,\mathbf{h}_1}, \ldots, b + v_{s+1,\mathbf{h}_K})$ for the K active nodes at level $s + 1$

   (d) $\widetilde{\pi}_{s,\mathbf{h}} = S_{s,\mathbf{h}} \prod_{r=0}^{s-1} \left(1 - S_{r,\boldsymbol{g}_{s\mathbf{h}r}}\right) H_{r,\boldsymbol{g}_{s\mathbf{h}_r}}$

4. Loop over 2 and 3.

5. If applicable, output new density estimate over a grid, only including non-zero $\widetilde{\pi}_{s,\mathbf{h}}$ terms for active nodes $(s, \mathbf{h})$ in the summation (19).

Data Series Plots

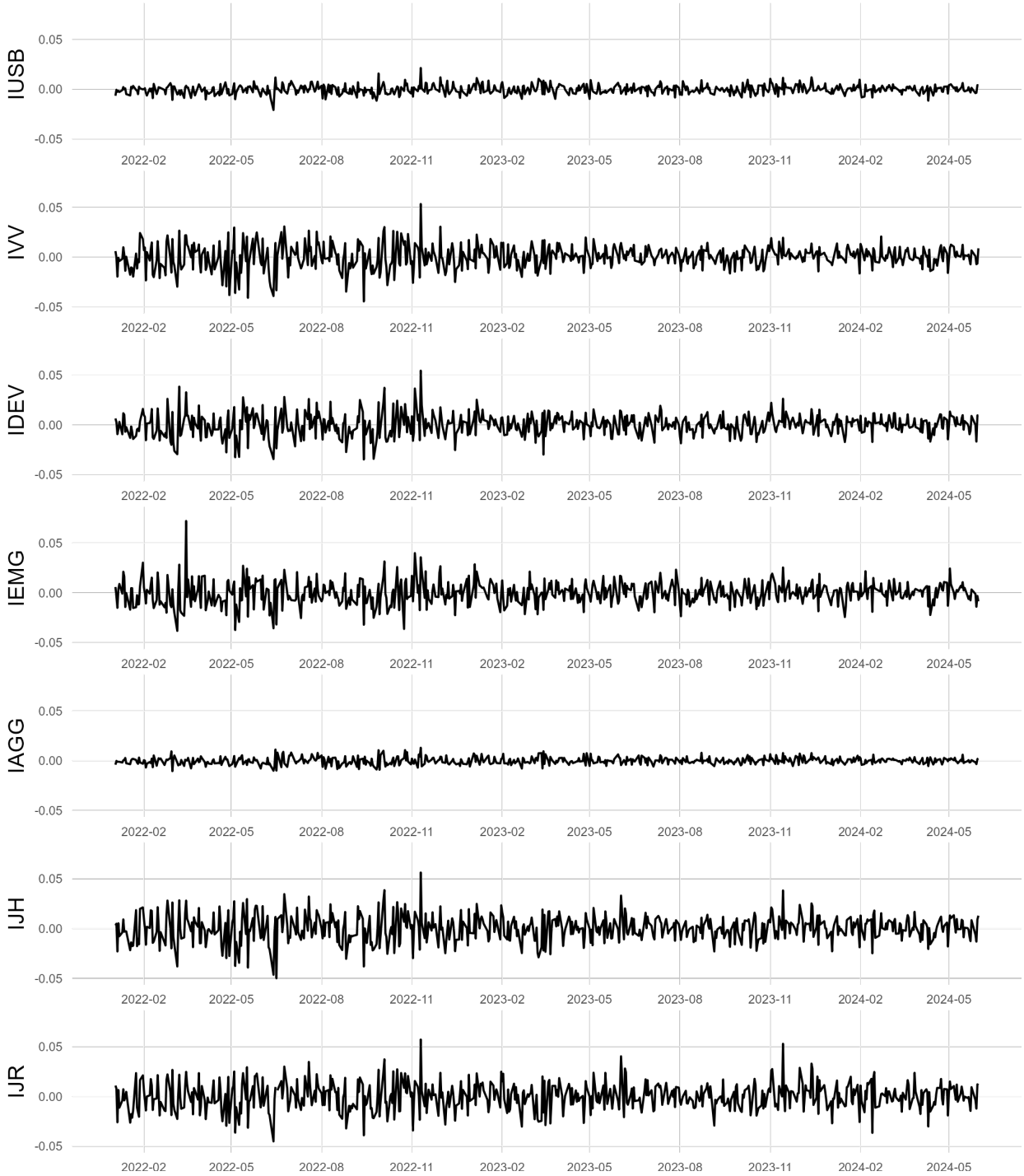

Figure 9: Data Series

References

Ausín, M. C., P. Galeano, and P. Ghosh (2014). A semiparametric Bayesian approach to the analysis of financial time series with applications to value at risk estimation. *European Journal of Operational Research 232*(2), 350–358.

Bai, R., V. Ročková, and E. I. George (2022). Spike-and-slab meets LASSO: A review of the spike-and-slab LASSO. In M. Tadesse and M. Vannucci (Eds.), *Handbook of Bayesian variable selection*, Chapter 4, pp. 28. Boca Raton: CRC Press.

Burda, M., M. C. Harding, and J. A. Hausman (2008). A Bayesian mixed logit-probit model for multinomial choice. *Journal of Econometrics 147*(2), 232–246.

Burda, M. and A. Prokhorov (2014). Copula based factorization in Bayesian multivariate infinite mixture models. *Journal of Multivariate Analysis 127*, 200–213.

Burda, M. and A. Prokhorov (2024). Multiscale Bernstein copula with Bayesian model averaging. Working paper.

Canale, A. and D. B. Dunson (2016). Multiscale Bernstein polynomials for densities. *Statistica Sinica 26*(3), 1175–1195.

Chib, S. and B. Hamilton (2002). Semiparametric Bayes analysis of longitudinal data treatment models. *Journal of Econometrics 110*, 67–89.

Conley, T., C. Hansen, R. McCulloch, and P. Rossi (2008). A semi-parametric Bayesian approach to the instrumental variable problem. *Journal of Econometrics 144*, 276–305.

Fiebig, D. G., R. Kohn, and D. S. Leslie (2009). Nonparametric estimation of the distribution function in contingent valuation models. *Bayesian Analysis 4*(3), 573–597.

Friedberg, B. and M. Adams (2024). 7 best balanced ETFs of April 2024. https://www.forbes.com/advisor/investing/best-balanced-etfs/.

Ghosal, S., J. K. Ghosh, and R. V. Ramamoorthi (1999). Posterior consistency of Dirichlet mixtures in density estimation. *The Annals of Statistics 27*(1), 143–158.

Ghosal, S. and A. W. van der Vaart (2017). *Fundamentals of Bayesian Nonparametric Inference*. Cambridge, UK: Cambridge University Press.

Hirano, K. (2002). Semiparametric Bayesian inference in autoregressive panel data models. *Econometrica 70*, 781–799.

Ishwaran, H. and J. S. Rao (2005). Spike and slab variable selection: Frequentist and Bayesian strategies. *The Annals of Statistics 33*(2), 730–773.

Jensen, M. and J. M. Maheu (2010). Bayesian semiparametric stochastic volatility modeling. *Journal of Econometrics 157*(2), 306–316.

Kim, J. G., U. Menzefricke, and F. Feinberg (2004). Assessing heterogeneity in discrete choice models using a Dirichlet process prior. *Review of Marketing Science 2*(1), 1–39.

Lempers, F. B. (1971). *Posterior Probabilities of Alternative Linear Models*. Rotterdam University Press.

Mitchell, T. J. and J. J. Beauchamp (1988). Bayesian variable selection in linear regression. *Journal of the American Statistical Association 83*(404), 1023–1032.

Scheipl, F., L. Fahrmeir, and T. Kneib (2012). Spike-and-slab priors for function selection in structured additive regression models. *Journal of the American Statistical Association 107*(500), 1518–1532.

Schwartz, L. (1965). On Bayesian procedures. *Probability Theory and Related Fields (Z. Wahrscheinlichkeitstheorie) 4*, 10–26.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica 4*, 639–650.

Wu, Y. and S. Ghosal (2008). Kullback Leibler property of kernel mixture priors in bayesian density estimation. *Electronic Journal of Statistics 2*, 298–331.

Wu, Y. and S. Ghosal (2009). Correction to: "Kullback Leibler property of kernel mixture priors in Bayesian density estimation". *Electronic Journal of Statistics 3*, 316 – 317.