2
Mixture of Models: A Brief Review of the Literature

In this section we present the class of models considered in this paper.

Definição 2.1 *The conditional probability density function (p.d.f.), $f(y_t|\mathbf{x}_t;\boldsymbol{\theta})$, of a random variable y_t is a Mixture of Models if*

$$f(y_t|\mathbf{x}_t;\boldsymbol{\theta}) = \sum_{i=1}^{K} g_i(\mathbf{x}_t;\boldsymbol{\theta}_i)\pi_i(y_t|\mathbf{x}_t;\boldsymbol{\psi}_i), \tag{2-1}$$

where $\mathbf{x}_t \in \mathbb{R}^q$ is a vector of covariates and $\boldsymbol{\theta} = [\boldsymbol{\theta}'_1,\ldots,\boldsymbol{\theta}'_K,\boldsymbol{\psi}'_1,\ldots,\boldsymbol{\psi}'_K]'$ is the conditional p.d.f. parameter vector, $\pi_i(y_t|\mathbf{x}_t;\boldsymbol{\psi}_i)$ is some known parametric family of distributions (basis distributions), indexed by the vector of parameters $\boldsymbol{\psi}_i$, and $g_i(\mathbf{x}_t;\boldsymbol{\theta}_i) \in [0,1]$ is the weight function. K is the number of basis distributions.

Suppose that y_t is distributed as in (2-1). The conditional expected value and variance are given by

$$\mathbb{E}[y_t|\mathbf{x}_t] = \sum_{i=1}^{K} g_i(\mathbf{x}_t;\boldsymbol{\theta}_i)\mathbb{E}_{\pi_i}[y_t|\mathbf{x}_t;\boldsymbol{\psi}_i] \tag{2-2}$$

$$\mathbb{V}[y_t|\mathbf{x}_t] = \sum_{i=1}^{K} g_i^2(\mathbf{x}_t;\boldsymbol{\theta}_i)\mathbb{V}_{\pi_i}[y_t|\mathbf{x}_t;\boldsymbol{\psi}_i], \tag{2-3}$$

where $\mathbb{E}_{\pi_i}$ and $\mathbb{V}_{\pi_i}$ are the expected value and the variance, taken with respect to the distribution π_i, respectively.

The simplest model belonging to this class is the TAR model, where a threshold variable controls the switching between different local Gaussian linear models. An indicator variable defines which local model is active and only one model is active each time. The conditional p.d.f. remains Gaussian and the conditional moments do not depend on the covariates. Many models have been proposed to overcome these limitations. The mixture autoregressive model proposed by (66) solves this problem using a mixture of Gaussian distributions with static weights. However, this model is still very limited because the weights do not vary across time (or with the covariate vector), so the authors suggest a generalization called a generalized mixture of autoregressive model (GMARX) (65). This generalization considers only two Gaussian local models and the weights are given by a logistic equation.

This model is more flexible then its predecessor but has a limited number of local models.

Many other alternatives have been derived in the neural networks literature. The mixture of experts model, by (34), describes the conditional distribution using gated neural networks to switch between local nonlinear models, such as neural networks models. This specification though very flexible, has a high number of parameters and is very hard to interpret. The Hierarchical Mixture of Experts (HME) is a tree-structured mixture of generalized linear models, where the weights are given by a product of multinomial logit functions. Each node of the tree can have any number of splits (branching factor), hence the specification and estimation of the model are very demanding. Furthermore, for the most general model there are no results that guarantee consistency of the estimators. Finally, the model is not completely interpretable once the subdivisions of the space are done by hyperplanes (functions of the covariates) which, in turn, are not necessarily interpretable.

To overcome some of the drawbacks caused by a profligate parametrization, (69) proposed the mixture autoregressive (MixAR) and (9) considered the mixture of generalized experts, which are simplifications of the HME model. In both cases the weights are given by a multinomial logit function. Probabilistic properties and approximation results were proved for both models; see (69), Carvalho and Tanner (2002a,b) and (7).

The model proposed in this paper combines the simplicity of the decision trees with the flexibility of the mixture of models. The main differences between our model and the previous ones are that our model is simpler, has a fewer number of parameters, is more easily interpretable and the model building strategy is well defined. The tree-structured mixture of models has a binary tree as the decision structure and the decision frontier is not a linear combination of the covariates, just one of the covariates each time.