# 1
# Introduction

Recent years have witnessed a vast development of nonlinear time series techniques (58, 24). From a parametric point of view, the Smooth Transition (Auto-)Regression, ST(A)R, proposed by (13)[1] and further developed by (40) and (56), has found a number of successful applications; see (61) for a recent review. In the time series literature, the STAR model is a natural generalization of the Threshold Autoregressive (TAR) models pioneered by (57) and (59).

On the other hand, nonparametric models that do not make assumptions about the parametric form of the functional relationship between the variables to be modeled have become widely applicable due to computational advances. For some references on nonparametric time series models, see (29), (30), (26), and (21). Another class of models, the flexible functional forms, offers an alternative that leaves the functional form of the relationship partially unspecified. While these models do contain parameters, often a large number of them, the parameters are not globally identified. Identification, if achieved, is local at best without imposing restrictions on the parameters. Usually, the parameters are not interpretable as they often are in parametric models. In most cases, these models are interpreted as nonparametric sieve (or series) approximations (14).

The artificial neural network (ANN) model is a prominent example of such a flexible functional form. It has found applications in a number of fields, including economics, finance, energy, epidemiology, among others. Although the ANN model can be interpreted as a parametric alternative (37, 60, 46), its use in applied work is generally motivated by the mathematical result stating that, under mild regularity conditions, a relatively simple ANN model is capable of approximating any Borel-measurable function to any given degree of accuracy; see, for instance, (22), (17), Hornik, Stinchombe, and White (1989,1990), (63), (23), (15), and (25).

The above mentioned models aim to describe the conditional mean of the series. In terms of the conditional variance, Engle's (1982) Autoregressive Conditional Heteroskedasticity (ARCH) model, Bollerslev's (1986) Generalized ARCH (GARCH) specification, and the Stochastic Volatility (SV) model proposed by (55) are the most popular alternatives for capturing time-varying volatility in time series

[1](13) called the model Smooth Threshold Auto-regression.

data, and have motivated the development of a myriad of extensions (50, 44, 2, 3).

However, when the attempt is to model the entire conditional distribution, the mixture-of-experts (ME) proposed by (34) becomes a viable alternative. The core idea is to have a family of models, which is flexible enough to capture not only the nonlinearities in the conditional mean, but also to capture other complexities in the conditional distribution. The model is based on the ideas of (49), viewing competitive adaptation in unsupervised learning as an attempt to fit a mixture of simple probability distributions into a set of data points. (36) generalized the above ideas by proposing the hierarchical mixture-of-experts (HME). Both ME and HME have been applied with success in different areas. For example, (62) showed an application to financial time series forecasting. Good applications of HME in time series are also given by Huerta, Jiang, and Tanner (2001,2003). Recently, Carvalho and Tanner (2002a,b) proposed the mixture of generalized linear time series models and derived several asymptotic results. It would worth mentioning the Mixture Autoregressive model proposed by (66) and its generalization developed in (67).

In this paper we contribute to the literature by proposing a new class of mixture of models that is based on regression-trees with smooth splits. Our proposal has the advantage of being flexible but less complex than the HME specification, providing possible interpretation for the final estimated model. Furthermore, a simple model building strategy has been developed and Monte Carlo simulations show that the it works well in small samples. A quasi-maximum likelihood estimator (QMLE) is described and its asymptotic properties are carefully analyzed. The small-sample properties of the QMLE are also evaluated via a Monte Carlo experiment.

The paper proceeds as follows. In Section 2 a brief review of the literature on mixture of models for time series is presented. Our proposal is presented in Section 3. In Section 4, parameter estimation and the asymptotic theory are considered. The modeling cycle is described in Section 5. Simulations are shown in Section 6, and Section 7 presents some examples with actual data. Finally, Section 8 concludes. All technical proofs are relegated to the appendix.