DYNAMIC DENSITY ESTIMATION WITH FINANCIAL APPLICATIONS

ABEL RODRIGUEZ AND ENRIQUE TER HORST

ABSTRACT. Empirical distributions in finance and economics might show heavy tails, volatility clustering, varying mean returns and multimodality as part of their features. However, most statistical models available in the literature assume some kind of parametric form (clearly neglecting important characteristics of the data) or focus on modeling extreme events (therefore, providing no information about the rest of the distribution). In this paper we develop a Bayesian nonparametric prior for a collection of distributions evolving in discrete time that is dense on the space of absolutely continuous distributions, and therefore allows for the special features mentioned above. The prior is constructed by defining the distribution at any time point as a Dirichlet process mixture of Gaussian distributions, and inducing dependence through the atoms of their stick-breaking decomposition. A general construction, which allows for trends, periodicities and regressors is described, but special emphasis is placed on developing autoregressive processes (AR) for sequences of distributions. The resulting model, labeled Distribution Autoregressive process (DAR) are applied to the estimation of the option-implied risk neutral distribution of the S&P500 index.

1. INTRODUCTION

One of the main constraints associated with classical time series analysis is the parametric assumptions involved in the analysis. Even if the evolution process is modeled in a flexible or non-parametric way, observational and evolution noise are typically assumed to follow some parametric distribution. This means that inferences end up being restricted to the moments of the assumed distributions, and changes not captured by those moments are overlooked by the model. Besides, in many applications, the natural measurement object is the distribution itself, which can potentially present skewness and multimodality as part of its features. For example in genetic epidemiology studies, the main object of inference is the distribution of DNA damage across time. Also, in option pricing, interest lies on the estimation of the risk-neutral distribution underlying the observed process (Melick and Thomas, 1997; Panigirtzoglou and Skiadopoulos, 2004).

Non parametric methods have proven also useful in the valuation of options and derivatives. Since the seminal papers of Black and Scholes (1973) and Merton (1973), the option pricing literature has concerned itself with relaxing the key simplifying assumptions such as constant volatility, zero transactions costs and a flat yield curve, as well as with using stochastic processes flexible enough to handle returns exhibiting fat-tails and skewness. Those modelling relaxations include working with kernel methods to price options (Yacine, 1996; Yacine and Jefferson, 2003), as well as extracting implied probability densities of the S&P 500 (Panigirtzoglou and Skiadopoulos,

---

Key words and phrases. Dependent Dirichlet process; Nonparametric Bayes; Random probability measure; Option implied Risk Neutral Distribution.

This work was supported in part by the Intramural Research Program of the NIH, National Institute of Environmental Health Science. We wish to thank Lars Stentoft and *The Centre for Analytical Finance, University of Aarhus, Denmark*, for providing us this data; and to Alan Gelfand, David Dunson and Mike West for their helpful comments.

2004). In the analysis of credit rating scores, Huang et al. (2004) compared credit rating prediction performance between backpropagation neural network (BNN) and support vector machines (SVM), obtaining around 80% of accuracy for both methods in the context of corporate credit rating analysis. In the context of corporate bond credit rating, Chaveesuk et al. (1999) explore three of the most well known supervised neural network paradigms-backpropagation, radial basis function and learning vector quantization-for the task of rating US corporate bonds.

In this paper we develop statistical methods appropriate to estimate and predict densities that evolve in discrete time. Our models use countably infinite mixtures of Gaussian distributions to represent the unknown density at each time point. These methods can be conceived as an extension of the Dirichlet Process Mixture model (Antoniak, 1974; Escobar and West, 1995) to collections of distributions that evolve in discrete time. It has been shown that, under mild conditions, these infinites mixtures have full support, in the sense of being dense on the space of absolutely continuos distributions (Lo, 1984). Dependence is built into the mixing distribution by allowing the atoms to evolve dynamically as linear state-space models. Indeed, the models we present can also be regarded as an extension of the Gaussian Dynamic Linear Models (DLMs) of West and Harrison (1997), which are Bayesian versions of the popular Kalman Filter. The DLMs represent a very flexible class of models with well-known properties, and have been successfully applied on a number of different areas, including econometrics, engineering and climatology (Kim, 1994; Pesaran et al., 1995; West, 1995; Lamon et al., 1998; West et al., 1999).

Although physical and economic phenomenon typically occur in continuous time, discrete-time models provide good approximations as long as a fine enough scale is used. Besides, most real life phenomenon are actually observed in discrete and equispaced intervals, making discrete-time models a natural tool for empirical analysis. This paper focuses on financial applications, but the class of dynamic Dependent Dirichlet Processes we present in this paper has multiple applications in areas like engineering (dynamic spectra estimation), climatology (modeling rainfall distributions) and biostatistics/epidemiology (genetic epidemiology studies, dynamic random effect models).

There is an extensive literature on Bayesian density estimation and clustering based on discrete mixtures (either finite or infinite), particularly using Gaussian distributions (Escobar and West, 1995; Richardson and Green, 1997; Green and Richardson, 2001). The Dirichlet Process (DP) (Ferguson, 1973, 1974; Antoniak, 1974) has become the tool of choice in Bayesian nonparametric analysis due to their flexibility, elegance and simplicity. Some examples of recent applications of the Dirichlet Processes include finance, econometrics, epidemiology, genetics and survival analysis among others (Kacperczyk et al., 2003; Chib and Hamilton, 2002; Hirano, 2002; Dunson, 2005; Medvedovic and Sivaganesan, 2002; Dunson et al., 2005; Kottas et al., 2002; Bigelow and Dunson, 2005; Laws and O'Hagan, 2002).

Most of this clasical literature on the Dirichlet Process focuses on exchangeable samples. However, recent work has started to develop methods for dependent distributions, either by forming convex combinations of independent processes (Müller et al., 2004; Dunson et al., 2004; Griffin and Steel, 2006a; Dunson, 2006; Pennell and Dunson, 2006) or by introducing dependence in the elements of the stick-breaking representation of the distribution (MacEachern, 1999, 2000; DeIorio et al., 2004; Gelfand et al., 2005; Griffin and Steel, 2006b). Particularly relevant for this paper are the works of Griffin and Steel (2006b), who induce dependency through the weights of the mixture; Griffin and Steel (2006a), who construct mixing distribution at a new time point as a linear combination of the mixing distribution at the previous time plus an innovation; and Tang and Ghosal (2006), who are concerned with estimating the conditional distribution of a general autoregressive process. Our approach, while sharing some of the characteristics of Griffin and Steel (2006b), induces dependency in the location of the mixture components instead of their weights. This formulation provides a more intuitive interpretation in finance applications, simplifies the implementation of the model and casts it as a direct extension of widely used models for time series analysis.

The paper is organized as follows: Section 2 reviews the basic properties of the Dirichlet Process. Section 3 defines our dynamic Dependent Dirichlet Process and discusses its main properties, while section 4 describe efficient computational tools to fit the model. Section 5 disscusses a specific subclass of models that generalizes the concept of autoregressive processes to densities. Finally section 6 shows an application of our methodology which extracts the risk-neutral probability distribution from the S&P 500 options data, and section 7 contains a disscussion and concludes.

2. The Dirichlet Process

Let $(\mathcal{X}, \mathcal{B})$ be a complete and separable metric space (typically $\mathcal{X} = \mathbb{R}^n$ and $\mathcal{B}$ are the Borel sets on $\mathcal{X}$) with associated probability distribution K. A Dirichlet Process (Ferguson, 1973, 1974) with baseline measure K_0 and precision α defines a distribution on the space of probability measures $K \in \mathcal{K}$, such that $(K(B_1), \dots, K(B_L)) \sim \mathsf{Dir}(\alpha K_0(B_1), \dots, \alpha K_0(B_L))$ for any partition $B_1, \dots, B_L$ of $\mathcal{X}$. If $K \sim \mathsf{DP}(\alpha K_0)$ then for any set $B \in \mathcal{B}$

$$\mathbb{E}(K(B)) = K_0(B) \qquad\qquad \mathbb{V}(K(B)) = \frac{K_0(B)(1 - K_0(B))}{\alpha + 1}$$

which justifies the interpretation of K_0 and α as mean and precision parameters. Sethuraman (1994) proved that K admits a *stick breaking* representation:

$$K(\cdot) = \sum_{l=1}^{\infty} w_l^* \delta_{\boldsymbol{\eta}_l^*}(\cdot)$$

where $\{\boldsymbol{\eta}_l^*\}_{l=1}^{\infty}$ are iid draws from K_0 and $w_l^* = z_l^* \prod_{k=1}^{l-1}(1 - z_k^*)$ with $\{z_l^*\}_{l=1}^{\infty}$ iid samples from a $\mathsf{Beta}(1, \alpha)$. This readily shows that the Dirichlet Process places probability one on the subspace of discrete Distributions. Since this is typically not a reasonable observational model, the DP is used as the mixing measure over the parameters of some continuous distribution $H(\cdot|\boldsymbol{\eta})$, leading to the well known Dirichlet Process Mixture models (Escobar, 1994; Antoniak, 1974):

$$y \sim \int H(y|\boldsymbol{\eta})K(d\boldsymbol{\eta}) \qquad\qquad K \sim \mathsf{DP}(\alpha K_0)$$

A common choice is $H(\cdot|\boldsymbol{\eta}) = \mathsf{N}(\cdot|\boldsymbol{\eta} = (\mu, \sigma^2))$, yielding a model that, under mild conditions, is dense in the space of absolutely continuous distributions (Lo, 1984). Choosing a baseline measure K_0 that is conjugate to the mixed distribution H yields efficient computational algorithms that exploit the Polya Urn representation of the process (Blackwell and MacQueen, 1973). By integrating out the random distribution K it can be shown that the predictive rule for a new observation $\boldsymbol{\eta}_n$ given the previous observations $\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_{n-1}$ in

$$\boldsymbol{\eta}_n \sim \frac{\alpha}{\alpha + n - 1} K_0 + \sum_{l<n} \frac{1}{\alpha + n - 1} \delta_{\boldsymbol{\eta}_l}$$

Since observations are exchangeable, this predictive rule also yields the full conditional distributions necessary to implement efficient Gibbs samplers (Escobar and West, 1995; MacEachern, 1994; Bush and MacEachern, 1996; MacEachern and Müller, 1998).

3. Dependent Dirichlet Processes in discrete time

Given a set D, consider replacing the distribution underlying Sethuraman's stick-breaking construction with a stochastic process $\{\boldsymbol{\eta}(t)\ \forall\ t \in D\}$ and define

$$K_t(\cdot) = \sum_{l=1}^{\infty} w_l^*(t)\delta_{\boldsymbol{\eta}_l^*(t)}(\cdot) \tag{1}$$

where $\boldsymbol{\eta}_l^*(t) \sim \boldsymbol{\eta}(t)$ and $w_l^*(t) = z_l^*(t)\prod_{s=1}^{l-1}(1 - z_s^*(t))$ with $z^*(t) \sim \mathsf{Beta}(1, \alpha(t))\ \forall\ t \in D$. The resulting stochastic process is called a Dependent Dirichlet Process (DDP) (MacEachern, 2000) and defines a distribution on the collection of random distributions on the space D, such that every K_t is marginally a Dirichlet Process. In the sequel, we consider mixtures of Gaussian distributions by a discrete-time DDP. Therefore, we take $D = \mathbb{N}$ and assume our underlying stochastic process to be a general random walk. For computational reasons, we focus on "single p" DDP models where $z_l^*(t) = z_l^* \sim \mathsf{Beta}(1, \alpha)$, independently of t. More specifically, letting y_{it} be the i-th observation obtained at time $t = 1, \ldots, T$, our model reduces to:

$$\begin{aligned}
y_{it} &\sim \int \mathsf{N}\left(\mathbf{F}_{it}\boldsymbol{\theta}_t, \sigma^2\right) K_t(d\boldsymbol{\theta}_t, d\sigma^2) & K_t &= \sum_{l=1}^{\infty} w_l^*\delta_{\left(\boldsymbol{\theta}_{lt},\sigma_l^2\right)} \\
w_l^* &= z_l^*\prod_{s=1}^{l-1}(1 - z_s^*) & z_l^* &\sim \mathsf{Beta}(1, \alpha) \\
\boldsymbol{\theta}_{lt}|\boldsymbol{\theta}_{l,t-1}, \sigma_l^2 &\sim \mathsf{N}(\mathbf{G}_t\boldsymbol{\theta}_{l,t-1}, \sigma_l^2\mathbf{W}_t) & \boldsymbol{\theta}_{0l}|\sigma_l^2 &\sim \mathsf{N}(\mathbf{m}_0, \sigma_l^2\mathbf{C}_0) \\
\sigma_l^2 &\sim \mathsf{IG}(s_0, s_0S_0)
\end{aligned} \tag{2}$$

where $\mathsf{N}(a, b)$ denotes a gaussian distribution with mean a and variance b, while $\mathsf{IG}(a, b)$ denotes an inverse-gamma distribution with mean $s_0S_0/(s_0 - 1)$ and s_0 degrees of freedom.

Our model assumes that the distribution for any observation y_{it} can be written as a mixture of normal components with means $\mathbf{F}_{it}\boldsymbol{\theta}_{lt}$ and variance σ_l^2, for some known matrix $\mathbf{F}_{it}$. The parameters defining the means of these distributions are allowed to move, with the evolution matrix $\mathbf{G}_t$ and the innovation variance $\mathbf{W}_t$ controlling the direction and magnitude of these changes. This formulation is rather general, and by appropriately choosing the structural matrices $\mathbf{F}_t$, $\mathbf{G}_t$ and $\mathbf{W}_t$ our model can easily accommodate patterns like trends, periodicities and dynamic regressions, or even autoregressive or moving average models for densities. The variances of the mixture components are assumed to be constant in time, but allowed to be change across component; and the weights associated each distribution is also estimated from the data, and assumed to be constant in time. Therefore, the estimates of the model can be interpreted as kernel density estimates with adaptive bandwidths and time varying kernels.

The choice of a constant weights model not only simplifies computation (MacEachern, 2000), but it allows a natural interpretation of the model as a multiprocess type II model (West and Harrison, 1997). In this multiprocess interpretation, an infinite number of DLM models with the same evolutionary structure but a different set of state parameters are available at every point in time.

These components represent different evolutionary paths consistent with a common underlying regime described by $\mathbf{G}_t$ and $\mathbf{W}_t$, and where each observation is allocated to one of these components with constant probability. Therefore, predictions using expected posterior means can be interpreted in this model as weighted averages of the predictions obtained under an infinite number of DLMs.

As argued by (MacEachern, 2000), the model can be reexpressed as a Mixture of Dirichlet processes

$$y_{it} \sim \int \mathsf{N}\left(\mathbf{F}_{it}\boldsymbol{\theta}_t, \sigma^2\right) K\left(d\boldsymbol{\Theta}, d\sigma^2\right) \qquad\qquad K \sim DP(\alpha K_0)$$

where $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1', \ldots, \boldsymbol{\theta}_T')'$ and K_0 is the joint distribution of $(\boldsymbol{\Theta}, \sigma^2)$ induced by the evolution equations described above, which reduces to a multivariate normal-inverse gamma where

$$\begin{aligned}
\mathbb{E}(\boldsymbol{\theta}_t|\sigma^2) &= \left[\prod_{r=1}^{t} \mathbf{G}_{t-r+1}\right] \mathbf{m}_0 \\
\mathbb{V}(\boldsymbol{\theta}_t|\sigma^2) &= \sigma^2 \left[\prod_{r=1}^{t} \mathbf{G}_{t-r+1}\right] \mathbf{C}_0 \left[\prod_{r=1}^{t} \mathbf{G}_{t-r+1}\right]' + \\
&\qquad \sigma^2 \sum_{r=1}^{t-1} \left[\prod_{s=1}^{t-r} \mathbf{G}_{t-s+1}\right] \mathbf{W}_r \left[\prod_{s=1}^{t-r} \mathbf{G}_{t-s+1}\right]' + \sigma^2 \mathbf{W}_t \\
\mathbb{C}\mathrm{ov}(\boldsymbol{\theta}_{t+k}, \boldsymbol{\theta}_t|\sigma^2) &= \left[\prod_{r=1}^{k} \mathbf{G}_{t+k-r+1}\right] \mathbb{V}(\boldsymbol{\theta}_t) \\
\mathbb{E}(\sigma^2) &= \frac{s_0}{(s_0-1)} S_0 \\
\mathbb{V}(\sigma^2) &= \frac{s_0^2}{(s_0-1)^2(s_0-2)} S_0
\end{aligned}$$

This representation as a DP mixture will be exploited in section 4 to develop efficient and simple computational strategies to fit these models. Although the "single p" structure in 2 suggests a rather restrictive model where the same number of components is used to represent every distribution (in principle, at least the largest one needed at any time point), the model is indeed flexible and dense on the space of distributions on D, as argued in MacEachern (2000). Note that a good approximation to a lower number of components can be achieved at any time point by assigning different components similar values of their parameters. Indeed, if $(\boldsymbol{\theta}_{it}, \sigma_i) \approx (\boldsymbol{\theta}_{jt}, \sigma_j)$ then $w_{it}\delta_{(\boldsymbol{\theta}_{it}, \sigma_i)} + w_{jt}\delta_{(\boldsymbol{\theta}_{jt}, \sigma_j)} \approx (w_{it} + w_{jt})\delta_{(\boldsymbol{\theta}_{it}, \sigma_i)}$. Therefore, by having components with similar parameters but that are not allocated to any observation at certain time points, we can approximate variable weights. Therefore, the price to pay for the constant-weight assumption is, in general, a slightly larger number of atoms being used.

Note that our dynamic DDP encompasses a number of other models as limiting cases. On one side by letting $\mathbf{W}_t = \mathbf{0}\ \forall\ t$ we have $\boldsymbol{\theta}_t = \boldsymbol{\theta}_0\ \forall\ t$ and thus $K_t = K_r\ \forall\ t, r$, which is the set up of Escobar and West (1995). On the other hand, by letting $\alpha \to 0$, we revert to the class of parametric DLMs with replicates as discussed in West and Harrison (1997). Note, however, that although increasing the value of $\mathbf{W}_t$ reduces the dependence among distributions, letting $\mathbf{W}_t \to \infty\ \forall\ t$

does not yield independent density estimates at each time point, but an improper distribution for θ_t at all times $t \geq 1$. Indeed, it is known that obtaining independent distributions from a single p DDP is not possible (MacEachern, 2000).

It is straightforward to obtain the a priori covariance structure induced on the observations by the process (see appendix A),

$$\mathbb{C}\text{ov}(y_{i',t+k}, y_{i,t}) = \frac{\mathbf{F}_{t+k}\left[\prod_{s=1}^{k}\mathbf{G}_{t+k-s+1}\right]\mathbb{V}(\boldsymbol{\theta}_t)\mathbf{F}_t}{1+\alpha}\frac{s_0}{s_0-1}S_0 \ \ \forall \ \ i \neq i', s_0 > 1$$

Note the similarities with the replicated, gaussian DLM model, where the covariance reduces to $\mathbf{F}_{t+k}\left[\prod_{s=1}^{k}\mathbf{G}_{t+s}\right]\mathbb{C}\text{ov}(\boldsymbol{\theta}_t)\mathbf{F}_t s_0 S_0/(s_0-1)$. Therefore, the covariance under the discrete time DDP is strictly smaller, which is due to the added uncertainty in the model specification. Also, since this is the induced covariance after marginalizing over the unknown collection $\{K_1, \dots K_T\}$, the replicates at a given time point are not independent unless $\alpha = 0$ (which corresponds to a single DLM model).

4. Inference

Inferences on the dynamic DDP can be performed using the same computational techniques employed for DP mixtures models (see, for example, Bush and MacEachern (1996); MacEachern and Müller (1998); Neal (2000); Ishwaran and James (2001)). In what follows, we describe an application of the algorithm of MacEachern and Müller (1998). For this purpose, a reparametrization of the model is helpful: let L be the current number of components that have observations allocated to them, n^*_{lt} be the number of observations in time t assigned to group l, $n_l = \sum_t n_{lt}$, $\{\boldsymbol{\Theta}^*_1, \dots, \boldsymbol{\Theta}^*_L\}$ be the current estimated values for those paths and $\xi_{it} = l$ iif $\boldsymbol{\Theta}_{it} = \boldsymbol{\Theta}^*_{lt}$. Given values for the structural parameters $\mathbf{F}_{it}$, $\mathbf{G}_{it}$ and $\mathbf{W}_{it}$ and after initialization of the parameters, an MCMC sampler alternates through the following steps:

(1) Generate $\Theta^*_l, \sigma^{*2}_l | \{\mathbf{y} | \xi_{it} = k, \} \cdots$ using the following Forward-Filtering/Backward-Sampling algorithm (Carter and Kohn, 1994; Fruhwirth-Schnatter, 1994)

(a) Forward filter using the recursions

$$\begin{aligned}
\mathbf{m}_{lt} &= \begin{cases} \mathbf{a}_{lt} + \mathbf{A}_{lt} e_{lt} & \text{if } n_{lt} > 0 \\ \mathbf{a}_{lt} & \text{if } n_{lt} = 0 \end{cases} \\
\mathbf{C}_{lt} &= \begin{cases} \mathbf{R}_{lt} - \mathbf{A}_{lt}\mathbf{Q}_{lt}\mathbf{A}'_{lt} & \text{if } n_{lt} > 0 \\ \mathbf{R}_{lt} & \text{if } n_{lt} = 0 \end{cases} \\
s_{lt} &= s_{l,t-1} + n_{lt} \\
s_{lt} S_{lt} &= \begin{cases} s_{l,t-1} S_{l,t-1} + \mathbf{e}'_{lt}\mathbf{Q}_{lt}^{-1}\mathbf{e}_{lt} & \text{if } n_{lt} > 0 \\ s_{l,t-1} S_{l,t-1} & \text{if } n_{lt} = 0 \end{cases} \\
\mathbf{A}_{lt} &= \mathbf{R}_{lt}\mathbf{F}^*_{lt}\mathbf{Q}_{lt}^{-1} \\
\mathbf{e}_{lt} &= \mathbf{y}_{lt} - \mathbf{f}_{lt} \\
\mathbf{f}_{lt} &= \mathbf{F}^{*\prime}_{lt}\mathbf{a}_{lt} \\
\mathbf{Q}_{lt} &= \mathbf{F}^{*\prime}_{lt}\mathbf{R}_{lt}\mathbf{F}^*_{lt} + \mathbf{I} \\
\mathbf{a}_{lt} &= \mathbf{G}_{lt}\mathbf{m}_{l,t-1} \\
\mathbf{R}_{lt} &= \mathbf{G}_{lt}\mathbf{C}_{l,t-1}\mathbf{G}'_{lt} + \mathbf{W}_{lt}
\end{aligned}$$

where $\mathbf{y}_{lt}$ is made of all observations assigned to group l at time t, $\mathbf{F}^*_{lt}$ is a matrix whose columns are the corresponding $\mathbf{F}_{it}$ vectors and $\mathbf{I}$ is the identity matrix.

(b) Sample $\sigma_l^2 | \mathbf{y}_l, \cdots$ from $\mathsf{IG}(s_{lT}/2, s_{lT}S_{lT}/2)$.

(c) Sample $\boldsymbol{\theta}_{lT} | \sigma_l^2, \mathbf{y}_l, \cdots$ from $\mathsf{N}(\mathbf{m}_{lT}, \mathbf{C}_{lT})$. Then recursively sample $\boldsymbol{\theta}_{lt} | \boldsymbol{\theta}_{l,t+1}\sigma_l^2, \mathbf{y}_l, \cdots$ from $\mathsf{N}(\mathbf{d}_{lt}, \mathbf{D}_{lt})$ where

$$\begin{aligned}
\mathbf{d}_{lt} &= \mathbf{m}_{lt} + \mathbf{B}_{lt}\left(\boldsymbol{\theta}_{l,t+1} - \mathbf{a}_{l,t+1}\right) \\
\mathbf{D}_{lt} &= \mathbf{C}_{lt} - \mathbf{B}_{lt}\mathbf{R}_{l,t+1}\mathbf{B}'_{lt} \\
\mathbf{B}_{lt} &= \mathbf{C}_{lt}\mathbf{G}_{t+1}R^{-1}_{l,t+1}
\end{aligned}$$

(2) Sample $\xi_{it} | \mathbf{y}, \boldsymbol{\xi}^-, \cdots$ from a multinomial distribution with probabilities:

$$\begin{aligned}
q_l &= n_l^- p(y_{it} | \mathbf{y}^-, \boldsymbol{\xi}^-) \\
&= n_l^- \mathsf{T}_{s_{lT}^-}\left(y_{it} | \mathbf{F}'_{it}\mathbf{h}^-_{lt}, S^-_{lT}(1 + \mathbf{F}'_{it}\mathbf{H}^-_{lt}\mathbf{F}_{it})\right) \\
q_0 &= \alpha p(y_{it} | S_0) \\
&= \alpha \mathsf{T}_{s_0}\left(y_{it} | \mathbf{F}'_{it}\mathbf{h}_{t0}, S_0(1 + \mathbf{F}'_{it}\mathbf{H}_{0t}\mathbf{F}_{it})\right)
\end{aligned}$$

where the superindex indicates removal of observation (i, t) from the sample, q_l for $l = 1, \ldots, L^-$ is the probability of allocation observation (i, t) to cluster l, q_0 is the probability of allocating the observation to a new cluster, $\mathbf{h}_{lT} = \mathbf{m}_{lT}$, $\mathbf{H}_{lT} = \mathbf{C}_{lT}$ and

$$\begin{aligned}
\mathbf{h}_{lt} &= \mathbf{m}_{lt} + \mathbf{B}_{lt}\left(\mathbf{h}_{l,t+1} - \mathbf{a}_{l,t+1}\right) \\
\mathbf{H}_{lt} &= \mathbf{C}_{lt} - \mathbf{B}_{lt}(\mathbf{H}_{l,t+1} - \mathbf{R}_{l,t+1})\mathbf{B}'_{lt}
\end{aligned}$$

Both $\mathbf{h}_{0t}$ and $\mathbf{H}_{0t}$ can be calculated using the same recursions with $n_{t0} = 0 \; \forall \, t$.

The filtering and smoothing relations described above are generalizations of those found in Carter and Kohn (1994), Fruhwirth-Schnatter (1994) and West and Harrison (1997), intended to deal with missing data or multiple observations. Note that Step 2 can be computationally expensive since it implies running Forward-Filtering/Backward-Smoothing (FFBS) steps for each observation in the sample. A careful implementation requires at least one and at most two such steps for each observation: one to recalculate the parameters for the group to which the observation currently belongs and possibly another one to calculate those of the group were it is to be assigned.

Typically, the matrices $\mathbf{G}_t$ and $\mathbf{W}_t$ governing the evolution of the system might be unknown. However, since $\mathbf{W}_t$ and $\mathbf{G}_t$ define the moments of the baseline measure, inference on these quantities can be performed as discussed in Escobar and West (1998). Therefore, inferences on structural parameters like discount factors, periodic components or autoregressive coefficients can be accommodated very easily. Details on some specific examples are discussed in the following sections. The sampler can also be extended to obtain backward-smoothed and/or k-step-ahead density estimates by calculating the corresponding predictive distributions. These predictive distributions, obtained from the Polya urn scheme, take the form of mixtures of T densities,

5. AN EXAMPLE: DISTRIBUTION AUTOREGRESSIVE MODELS (DAR)

Autoregressive (AR) models are one of the most popular tools in finance and econometrics, both because of their flexibility and interpretability. For example, as noted by a number of authors (Box and Jenkins, 1974; West, 1997; Aguilar et al., 1999), high order stationary AR processes can be seen as approximations to Moving Average (MA) processes. Even more, Wold's decomposition ensures that high order AR processes are good approximations to any stationary process.

AR models have been used to model not only the mean structure of the time series, but also to understand other aspects of the distributions of interst. For example, ARCH models (Engle, 1982) use an AR process to model the volatility of the process. The goal of this section is to push the idea of AR processes beyond that of a model for *the moments* of distributions and get an equivalent formulation for the whole distribution. For simplicity, we start our discussion with the first-order, distribution autoregressive process (DAR(1)).

5.1. DAR(1) models.

In what follows, we wonsider a discrete-time DDP process of the form

$$y_{it} \sim \mathsf{N}(\mu_{it}, \sigma^2_{it}) \qquad\qquad (\mu_{it}, \sigma^2_{it}) \sim K_t$$

$$K_t = \sum_{l=1}^{\infty} w^*_l \delta_{(\mu^*_{lt}, \sigma^2_l)} \qquad\qquad \mu^*_{lt} \sim \mathsf{N}(\phi \mu^*_{l,t-1}, \sigma^2_l U)$$

The name DAR(1) comes from the fact that the stochastic process defining the location of the gaussian distributions used to represent the unknown density follow an autoregressive process with autocorrelation ϕ and variance $\sigma^2 U$. This is a special case of the general model described in section 3 where $\mathbf{F}_t = 1$, $\mathbf{G}_t = \phi$ and $\mathbf{W}_t = U$. Therefore, the correlation a priori induced on the observations is

$$\mathbb{C}\text{or}(y_{i,t}, y_{i',t+k}) = \frac{\phi^k}{1+\alpha}$$

This formulation extends the latent AR process models (West and Harrison, 1997) to infinite mixtures. As in the latent AR process, and unlike the typical Gaussian AR(1) process, $\phi = 0$ implies uncorrelated but dependent observations. Indeed, the case $\phi = 0$ generates identifiability issues since it is not possible to separate the noise of the underlying process from the observational noise.

The model is completed by placing priors on ϕ, U, μ_0 and α. For computational simplicity a conditionally conjugate distribution for the variance of the autoregressive process is used, $U \sim \mathsf{IG}(a_U, b_U)$. Also, in order to ensure stationarity, we set

$$\phi \sim \mathsf{N}\left(0, \tau^2\right) \mathbf{1}_{(-1,1)} \qquad\qquad \mu^*_{l0} \sim \mathsf{N}\left(0, \frac{\sigma^2 U}{1-\phi^2}\right) \;\; \forall \;\, l$$

where $\mathsf{N}(a,b)\mathbf{1}_{\Omega}$ denotes the normal distribution with mean a, variance b and restricted to the set Ω. Finally, the DP precision factor α is given a $\mathsf{G}(a_\alpha, b_\alpha)$, which is conditionally conjugate.

Implementation of this models is a straightforward extension of that in section 4. Conditional on ϕ, U and α, the model is a discrete-time DDP. On the other hand, conditional on the allocation indicators, the sample paths $\{(\boldsymbol{\mu}^*_l, \sigma^{*2}_l)\}_{l=1}^{L}$ are iid samples from the baseline measure. Therefore, samples from ϕ and U can be easily obtained using the following full-conditional distributions.

- The variance of the autoregressive process can be obtained by sampling U from its full conditional distribution,

$$U|\cdots \sim \mathsf{IG}\left(a_U + \frac{T}{2}, b_U + \sum_l \left[\frac{(1-\phi^2)\mu^*_{l0}}{\sigma^{*2}_l} + \sum_{t=1}^{T} \frac{\mu^*_{lt} - \mu^*_{l,t-1}}{\sigma^{*2}_l}\right]\right)$$

- The full conditional distribution for the autocorrelation coefficient takes the form

$$p(\phi|\cdots) \propto (1-\phi^2)^{-L/2} \exp\left\{-\frac{1-\phi^2}{2U}\sum_{l=1}^{L}\frac{\mu^{*2}_{l0}}{\sigma^2_l}\right\}$$
$$\exp\left\{-\frac{1}{2}\left[\frac{b}{U}+\frac{1}{\tau^2}\right]\left[\phi - d\left(\frac{b}{U}+\frac{1}{\tau^2}\right)^{-1}\right]\right\}$$

where

$$b = \sum_{t=1}^{T}\sum_{l=1}^{L}\frac{\mu^{*2}_{l,t-1}}{\sigma^{*2}_l} \qquad\qquad d = \sum_{t=1}^{T}\sum_{l=1}^{L}\frac{\mu^*_{l,t}\mu^*_{l,t-1}}{\sigma^{*2}_l}$$

Note that this expression does not correspond to any known distribution. However, we recognize the third term (which happens to contain most of the information provided by the observations) as a normal kernel. Therefore an efficient independent-proposal Metropolis step can be devised to sample from this full-conditional distribution. Given the current value of the autoregression parameter $\phi_{(c)}$ in the previous iteration, propose $\phi_{(p)} \sim \mathsf{N}\left(d\left(\frac{b}{U}+\frac{1}{\tau^2}\right)^{-1}, \left(\frac{b}{U}+\frac{1}{\tau^2}\right)^{-1}\right)\mathbf{1}_{(-1,1)}$. Then, accept this proposal with probability

$$\min\left\{1, \left(\frac{1-\phi^2_{(p)}}{1-\phi^2_{(c)}}\right)^{-L/2} \exp\left\{-\frac{\phi^2_{(c)} - \phi^2_{(p)}}{2U}\sum_{l=1}^{L}\frac{\mu^{*2}_{l0}}{\sigma^{2*}_l}\right\}\right\}$$

Otherwise retain the previous value $\phi_{(c)}$.

If a uniform prior is chosen for ϕ, the corresponding posterior distribution and associated Metropolis step can be obtained by taking $\tau^2 \to \infty$. An alternative algorithm can be produced using a random-walk Metropolis, but the independent proposal does not require any tuning of the parameters.

5.2. **General DAR(p) models.** Extending the previous ideas to a DAR(p) is straightforward. The model takes the form

$$y_{it} \sim \mathsf{N}(\mu_{it}, \sigma^2_{it}) \qquad (\mu_{it}, \sigma^2_{it}) \sim K_t$$

$$K_t = \sum_{l=1}^{\infty} w^*_l \delta_{(\mu^*_{lt}, \sigma^2_l)} \qquad \mu^*_{lt} \sim \mathsf{N}\left(\sum_{r=1}^{p} \phi_r \mu^*_{l,t-r}, \sigma^2_l U\right)$$

Again, this is a especial case of the general discrete-time DDP where $\mathbf{F} = (1, 0, \dots, 0)$,

$$\mathbf{G}_t = \begin{pmatrix} \phi_1 & \phi_2 & \phi_3 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \qquad \mathbf{W}_t = \begin{pmatrix} U & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

As with regular AR(p) models, enforcing stationarity of this process through priors on the autoregressive coefficients is not an easy task since the conditions on $(\phi_1, \dots, \phi_p)$ take a complex form. Therefore, we choose priors

$$(\phi_1, \dots, \phi_p) \sim \mathsf{N}(\mathbf{0}, \mathbf{\Sigma}_\phi) \qquad (\mu^*_{l0}, \dots, \mu^*_{l,-p+1}) \sim \mathsf{N}(\mathbf{0}, \mathbf{\Sigma}_\mu)$$

Although these priors do not enforce stationarity of the resulting AR process, they produce conditionally conjugate posterior distributions, simplifying inference in the model. A different approach relies on placing informative priors on the roots of the characteristic polynomial $q(s) = 1 - \sum_{r=1}^{p} \phi_r s^r$ associated with the AR(p) instead (Huerta and West, 1999). This choice induces in turn a distribution on $(\phi_1, \dots, \phi_p)$, which can be used not only to enforce the stationarity of the process, but also to perform inferences on p and to identify quasi-periodicities *in the probability distribution itself*, providing insights into the mechanisms driving the process being modelled.

In terms of inference, the DAR(p) requires a slight adaptation of the FFBS algorithm described in section 4 due to the fact that $\mathbf{W}_t$ is a singular matrix. This modification is described in West and Harrison (1997), Chapter 15.3.2 for the Gaussian AR(p) model.

6. Estimating implied risk-neutral distributions

6.1. **Option implied risk-neutral distribution (RNPD).** It is common knowledge that market prices of options contain information regarding market expectations. Important information can be extracted from the derivatives markets and used for several purposes, such as the probabilities of adverse movements in the market, as well as monetary authorities assessing market expectations. This is why cross sections of option prices have been investigated in order to retrieve the implied probability density distribution of the underlying S_t (stock, inflation, currency, interest rates), which represents market expectations. Retrieving the RNPD is a typical example of an *Inverse Problem* (Tikhonov, 1963), and the Bayesian methodology is a way to regularize it through the elicitation of a prior distribution which acts as a penalization function (Wolpert and Ickstadt, 2004).

In any option pricing model, looking for a suitable and realistic stochastic process to model the underlying stock price is essential. Nonparametric methods using maximum entropy techniques have been successfully used in the case of Lévy processes (Cont and Tankov, 2003). Such models may exhibit stylized features common in financial applications, such as skewness, volatility clustering, jumps, fat-tails (Cont and Tankov, 2003), as well as multimodality of the log-returns $\frac{dS_t}{S_t}$, but are typically awkward to implement. Other approaches (Melick and Thomas, 1997; Rebonato, 2004) have used finite mixtures of parametric distributions to fit the RNPD using number of mixtures as well, but to the cost of overfitting the observed RNPD, which usually leads to poor prediction. The novelty or our estimation relies on making no assumptions other than using the call-put parity equation to generate observations under the RNPD. The no arbitrage condition from equation (3) enables us to imply the existence (although not the uniqueness) of a RNPD (Delbaen and Schachermayer, 2006), whose posterior distribution is the focus of interest. Besides, our algorithm allows us to not only determine sequentially in time the optimal number of components, but also the dependence among the RNPD.

In what follows, we focus on *European options*. The holder of a European call option has the right, but not the obligation, to buy an underlying security at a specified date (*expiration date*) for a contractually specified amount (*strike price*), irrespective of the market value of the security on that date. The underlying securities of options can be stocks, indices such as the *Standard and Poor's 500*, interest rates, etc. At the expiration date T, the value of the option is $(S_T - K, 0)^+$, the maximum of $S_T - K$ and zero. Payoff is at later time T, so under constant discount rate r present value of the call option at time t will be $\exp(-r(T-t))\,(S_T - K)^+$, where K is the strike price. In what follows, let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\geq 0})$ be a probability space, equipped with a filtration satisfying the *usual hypotheses* (see Protter, 1990, p. 3, for definition). The classical framework of option pricing assumes that a call option C_t whose payoff $(S_T - K, 0)^+$ depends on our underlying S_T at the expiration date T, can be computed via the following integration:

$$\begin{aligned} C_t &= \exp(-r(T-t))\, E_{\mathbb{Q}}\{(S_T - K, 0)^+ \mid \mathcal{F}_t\} \\ &= \exp(-r(T-t)) \int_{\mathcal{S}} (S_T - K, 0)^+ d\mathbb{Q}(S_T \mid \mathcal{F}_t) \end{aligned}$$

where $\mathcal{S}$ is the sample space for the stock price S_T at terminal date T, K is the strike and integration is performed under the risk-neutral measure $\mathbb{Q}$, such that the discounted stock price $\exp(-rt)\,S_t$ is a $\mathbb{Q}$-martingale (Delbaen and Schachermayer, 2006), yielding

$$C_t = \exp(-r(T-t)) \int_{\mathcal{S}} \max(S_T - K, 0) d\mathbb{Q}(S_T \mid \mathcal{F}_t)$$

Similarly, for the put option

$$P_t = \exp(-r(T-t)) \int_{\mathcal{S}} \max(K - S_T, 0) d\mathbb{Q}(S_T \mid \mathcal{F}_t)$$

From the put-call parity price we obtain (Hull, 2005):

$$\begin{aligned} C_t - P_t &= S_t - K\exp(-r(T-t)) \\ \Rightarrow S_t &= C_t - P_t + K\exp(-r(T-t)) \end{aligned} \tag{3}$$

Given the call and put prices for each trade (which are typically available as part of market data), equation (3) can be used to obtain observations from the risk neutral probability distribution of S_t across different strikes K. These observations can then be used to estimate the RNPD of the underlying S_t at time each time t in a nonparametric fashion, for example through a discrete time DDP. Using a nonparametric method in this setup is attractive because it assumes the least amount of assumptions regarding the governing dynamics of the probabilistic structure of S_t.

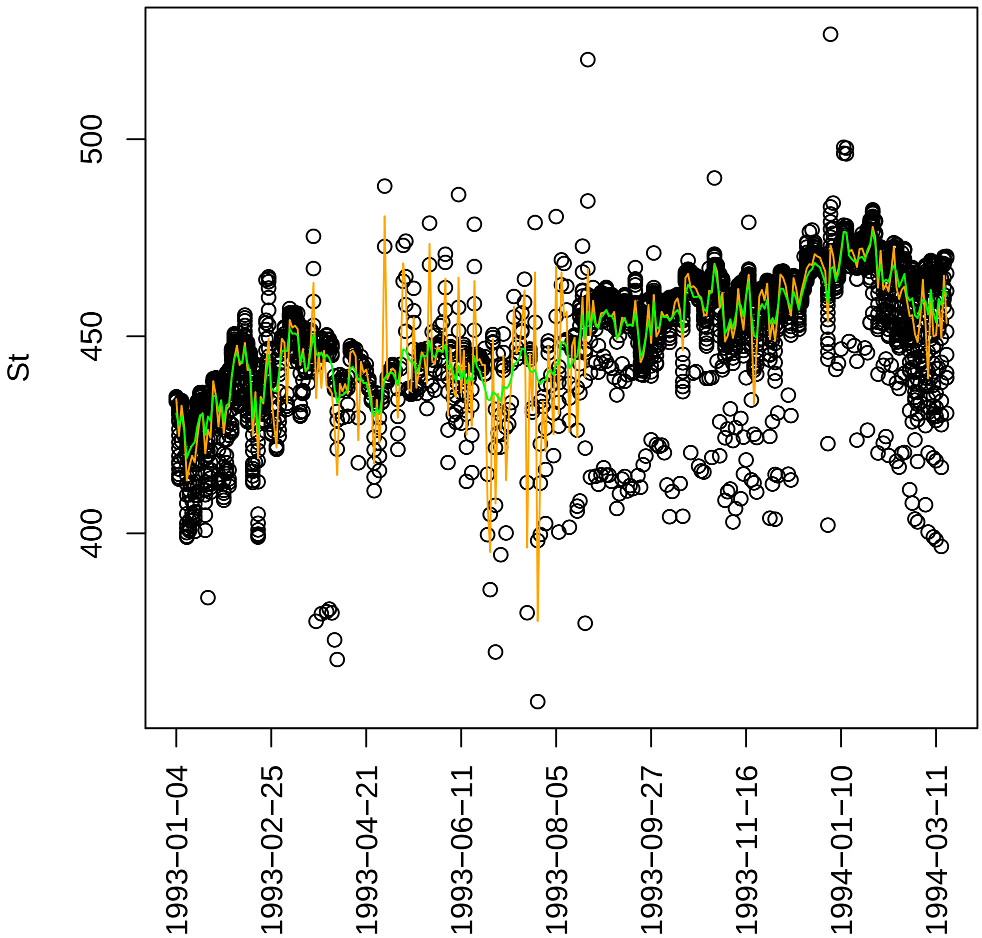

FIGURE 1. Underlying S&P500 prices. Dots correspond to the raw data, the orange line corresponds to the empirical mean of the observations at the corresponding time point and the green line to the smoothed mean under the DAR(1) model.

6.2. **RNPD in the S&P500 INDEX.** In what follows, we concentrate on options of the S&P500 index with three-month maturity times traded between January 4th 1993 and March 17th 1994, for a total of $n = 4385$ trades spread over $T = 293$ days, with sample sizes in any specific day varying between 2 and 26. The data was constructed by Yacine Ait-Sahalia and has been used in the empirical study of Duffie et al. (2000). A plot of the prices inferred from the put-call parity (3) using the libor as the interest rate (Panigirtzoglou and Skiadopoulos, 2004) is presented in Figure 1. Note that the distributions on any specific day may be highly skewed and may have very heavy tails. Means (plotted in orange) vary wildly, specially during the summer of 1993 when fewer trades occur and extreme values are highly influential. Even more, the kernel density estimates in Figure 2 show that the risk neutral distributions change dramatically across time and might actually be multimodal.

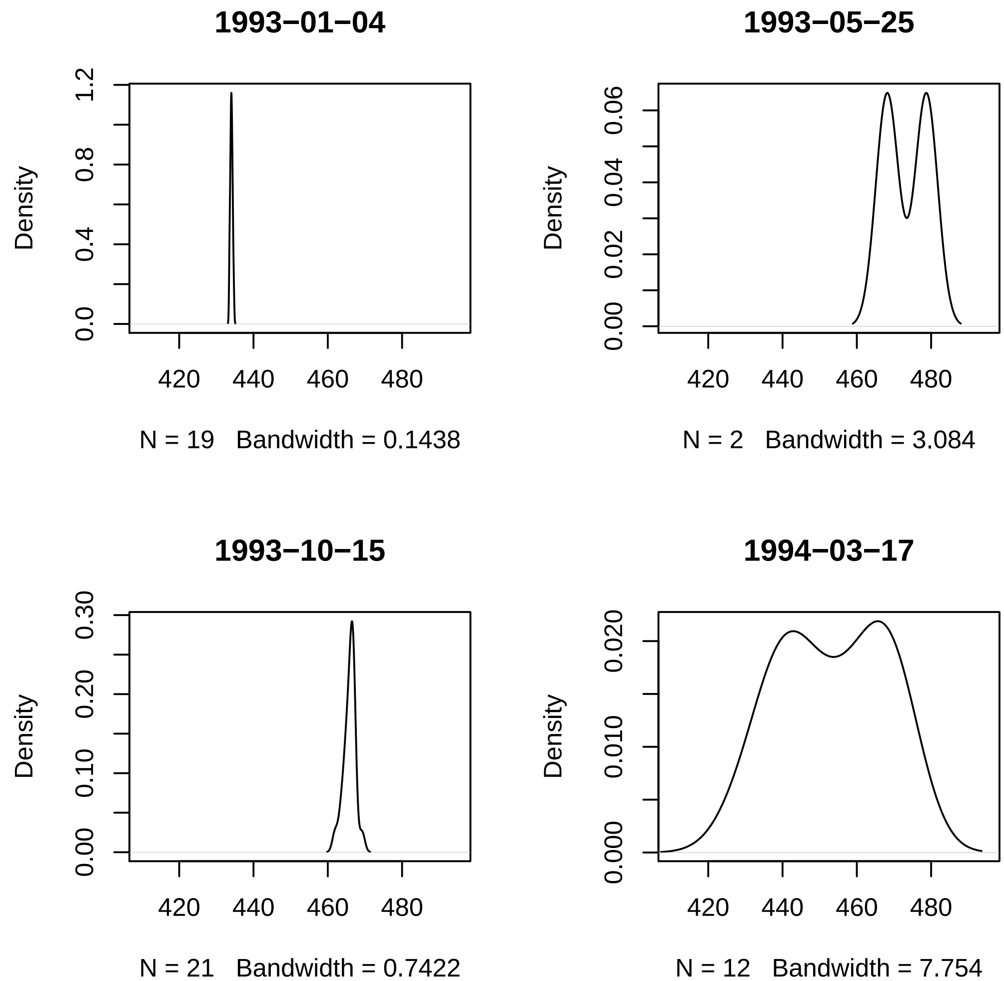

FIGURE 2. Kernel density estimates of S&P500 prices on Jan 4, 1993; May 25, 1993; Oct 15, 1993 and Mar 17, 1994. The number of observations N and the bandwidth estimated through cross-validation are shown bellow each plot.

We use the DAR(1) model described in section 5.1 to model the unknown distribution of S_t. In order to justify a zero mean DAR(1) process, the data has been recentered using the global mean. The goal is to obtain smoothed estimates for the mean and volatility of the process, as well as for the distribution itself, along with a predictive model that can be used to infer the distribution at subsequent dates.

Hyperparameters have been chosen as $\tau^2 = 1$, $s_0 = 1$ $S_0 = 40$ (about one sixth of the empirical observational variance on the whole sample) and $a_U = b_U = 1$. In order to increase the flexibility of the model, a hyperprior on the concentration parameter of the Dirichlet Process, $\alpha \sim \mathsf{G}(1, 1)$ was used. All results presented here are based on 25,000 samples of the posterior distribution obtained after a burn-in period of 5,000 iterations, and they seem to be robust to moderate changes in the prior distribution.

	Mean	Median	90% Prob Int
α	0.92	0.87	(0.36,1.76)
ϕ	0.974	0.974	(0.962,0.985)
U	13.12	13.05	(11.23,15.38)

TABLE 1. Posterior mean, median and symmetric 95% probability interval for some parameters in the DAR(1) model fitted to the S&P500 data

Table 1 shows posterior estimates for the concentration and baseline measure parameters in the model. As expected, the autocorrelation in the process is rather high. The evolution variance is also relatively large, being about 12 times larger than the average observational variance. Roughly speaking, the model uses between 7 and 10 mixture components to model the collection of 293 distributions, but three of those components capture around 90% of the trades (50%, 25% and 15% respectively), with the rest of the components with higher observational variance explaining the very extreme observed values. This observation has a potential explanation in the field of behavioral finance. It has been argued that individuals tend to over-weight small probability outcomes related to significant losses, which is therefore consistent with high-implied volatility on deep out of the money derivatives (Rasiel, 2003). Indeed, the observational variance of every component of the mixture decreases as the number of data points allocated to that group increases, which happens in the middle around the current market price. On the contrary, with extreme observations which might constitute a group (cluster) of their own, there is a much higher uncertainty and thus observational variance regarding the distribution. This analysis makes the implied volatility smile nothing but the consequence of human behavior which is to choose a direction of where the underlying might go, and then choose the probability distribution, explaining the fact that the RNPD is multimodal. This last feature of the RNPD also shows the proportion of different market participants (number of mixtures) and their bullish or bearish views regarding the future on the S&P 500.

The implied volatility is calculated using the original Black $\&$ Scholes formula that assumes a single lognormal with constant volatility. It then follows that we will have a volatility smile, where the implied volatility is an decreasing/increasing function of the strike price, because the higher probability of tail observations vis-a-vis the lognormal distribution will imply higher market prices for those deep out-of-the-money calls. Anything leading to fat tails vis-a-vis the lognormal will deliver the volatility smile, and mixture distributions are one such thing. However, our methodology gives the nice interpretation that fat tails are often a consequence of the market being divided into several broad groups, the bulls and the bears (together with their nuances), who make their bets accordingly in their private valuation of options.

Figure 3 shows smoothed density estimates $p(y_t|\mathbf{y}_1,\ldots,\mathbf{y}_T)$ and one-step-ahead predicted densities $p(y_t|\mathbf{y}_1,\ldots,\mathbf{y}_{t-1})$ for the last 8 days in the series. Note that the model seems to provide both a good fit to the observed data and sensible predictions of future observations, even for extreme observations. Both density estimates are effective in capturing the main characteristics of the data, namely high skewness and multimodality. However, the modes show up more sharply in the smoothed densities because of the additional noise being convoluted during the prediction.

Another interesting feature of the model is its robustness to small samples with extreme values, which is clearly hinted by behavior of the estimated mean during the summer of 1993 (green line

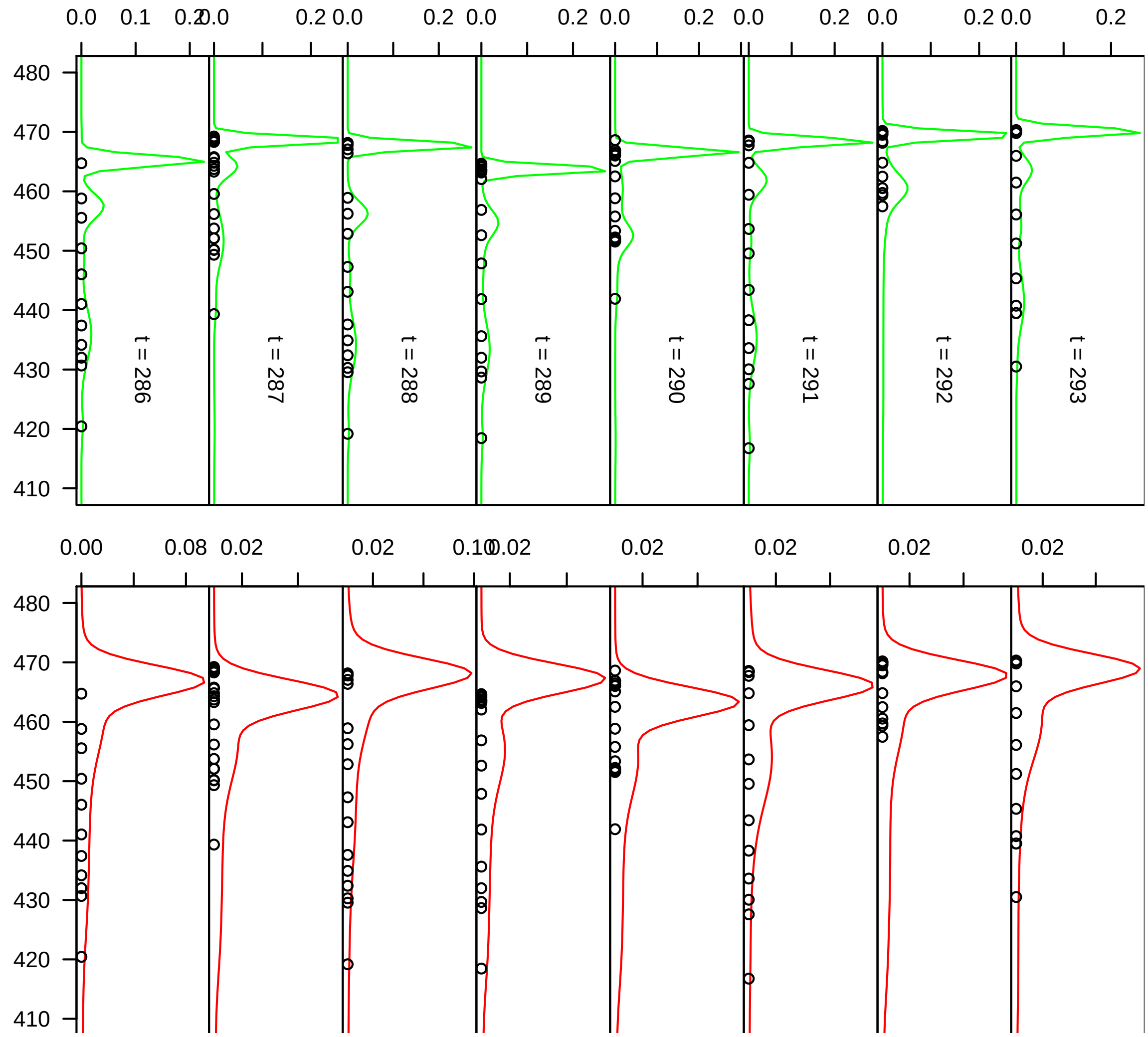

FIGURE 3. Smoothed (green) and one-step-ahead predicted densities (red) between March 8, 1994 ($t = 286$) and March 17, 1994 ($t = 293$). The dots correspond to the actual observations.

in Figure 1). For example, consider the market during May 25, 1993. In this date, only two trades occurred, and the raw density estimate was presented in the upper right panel of Figure 2 is highly skewed towards relatively high prices. Most parametric analysis would associate this with a large shift in the location of the distribution. The smoothed density on this date resulting from the DAR(1) model can be seen in figure 4. Since there is a bulk of information contained in the behavior of the market in adjacent days, the two observed market valuations are attributed by the model not to the mainstream investors (represented by the higher weight component), but to their more risk-prone counterparts. This is an interesting example of the regularization properties inherent to this class of models.

The model can also be used to obtain robust smoothed estimates for the mean and volatility underlying the process. The time varying aspect of the first and second moments can be of interest

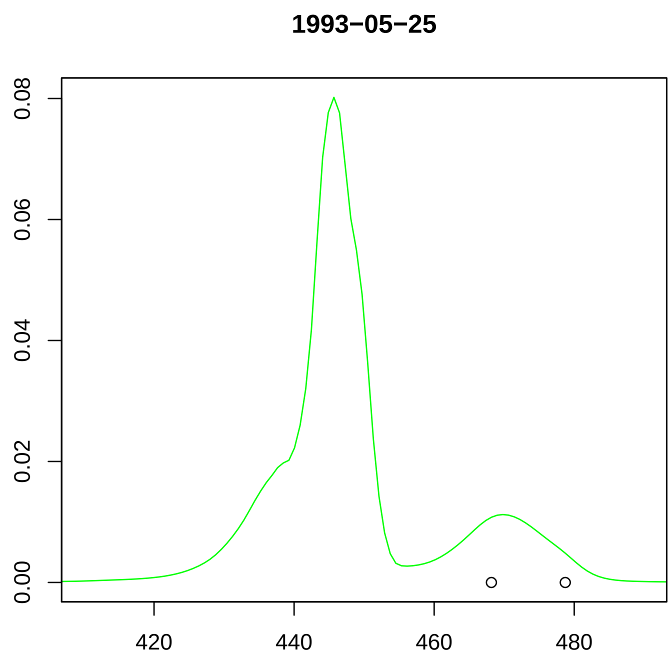

FIGURE 4. Smoothed density estimate for May 25, 1993 obtained through the DAR(1) model.

for financial institutions interested in determining and measuring the exposure in their portfolios to financial risk, and more precisely to market risk. By using the dynamic DDP model, the risk measures can then be computed under distributions not only exhibiting fat tails, volatility clustering, varying mean returns, multimodality, but estimation uncertainty as well. This last feature can therefore take into account model risk due to the use of nonparametric methods (Cont, 2006).

Figure 5 shows a plot of the interquartile range (IQR) associated with the smoothed densities generated by the model. In normal distributions, $IQR/1.349$ is a consistent estimator of the standard deviation of the process. For more general distributions, the IQR provides a robust alternative to the variance that can be used to compare the volatility at different points in time. From Figure 5 we can observe two high volatility periods: one in late June 1993 and another one by the end of the recorded series in March 1994. It is interesting to notice that the DDP extends Bollerslev (1986) to a nonparametric method exhibiting not only mean shifts in the distribution as well as volatility clustering, while working directly in levels. Working with data in levels is something that Garch models are not able to do, since they assume some long term variance around which the second order moments of the asset returns mean revert.

7. DISCUSSION

In this paper we discuss a class of models for dynamic density estimation in discrete time that allows us to borrow information across adjacent observations and obtain robust inferences, not only for the distribution itself, but also for other quantities of interest like the mean or the quantiles

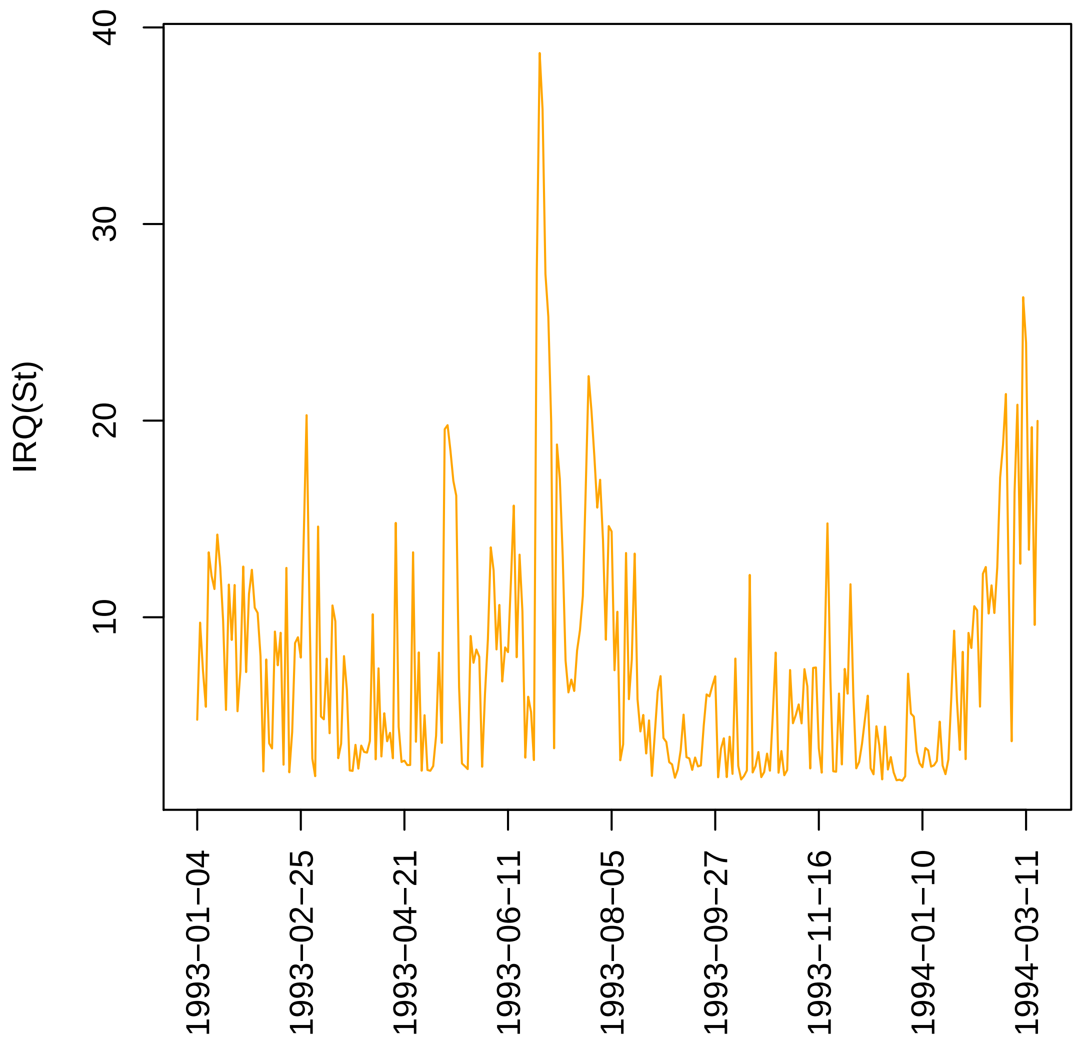

FIGURE 5. Estimated volatility in S&P500 series.

of the process. We provide details for the most general location-and-scale mixture model, with the development for location mixtures constituting a straightforward simplification.

The example in this paper focuses on distribution autoregressive models to estimate implied risk-neutral in options markets, which is in itself a novel contribution. It allows us to drop cumbersome parametric assumptions that are clearly not justified by the data but are part of the current state-of-the-art in the field. In spite of this, our formulation is not constrained to DAR models and is indeed much more general: trends, periodicities and even covariables can be easily included into the model. Also, applications extend not only to other areas in finance (stochastic volatility models, risk management, credit analysis and FX options) but also to other areas like epidemiology, climatology and engineering, some of which are the focus of current research. However, despite their generality, it is important to emphasize that the models described in this paper induce dependence in the distributions themselves and assumes that observations are exchangeable within every time point given that distribution. Therefore, our model is not suitable for the analysis of longitudinal studies where the same experimental unit is followed at different times points.

One of the key characteristics of these models is the simplicity of their implementation. By using common weights to define the collection of distributions, computational tools for standard Dirichlet Process can be employed. The sampler we have described in this section uses the Pólya urn trick to marginalize over the collection of unknown distributions and sample the paths of the different components in the mixture. However, there is an ample literature on MCMC techniques for Dirichlet processes that can be exploited to obtain alternative exact or approximate samplers. Some attractive options that can be readily implemented are truncation approximations or variational methods, with the latter being specially appealing for very large sample samples.

Another enormous advantage is its interpretability as a multiprocess type II model. The constant weight assumption allows us to think in term of alternative evolution paths for the process receiving different weights depending of the behavior of the process. These paths can have natural meaning in the context of a specific application, as in the option RNPD example. Also, the estimates of the parameters underlying $\mathbf{G}_t$ of $\mathbf{W}_t$ (if there are any) can be interpreted as averages over observed paths

There are two main drawbacks with our model formulation. As discussed by MacEachern (2000) and Griffin and Steel (2006b), discrete time DDPs do not posses the intuitive asymptotic behavior expected as $\mathbf{W}_t \to \infty$. Indeed, it is impossible to generate a collection of independent distributions within this framework. However, this is hardly a limitation (at least in the context of financial application) since in most cases the existence of dependence is not in question and it is rarely the case that dependence tests are required.

Identifiability issues also arise with single-observation time series. Indeed, it is rare in finance and econometric applications to have replicates at any given time. As noted by Griffin and Steel (2006b), it is hard for constant weight constructions to differentiate between variability in the baseline measure and multimodality unless replicates are available. The intuition in the context of discrete-time DDPs is straightforward: are observed changes in the process due to a unimodal process with fairly large evolution variance or to an almost constant multimodal distribution? We argue that the order in which new components are added as the number of observations grows contains valuable information, and therefore strong prior distributions enforcing small values of $\mathbf{W}_t$ will essentially solve the identifiability problem. This type of constrains, which might seem awkward in a spatial context like Gelfand et al. (2005), can be introduced in discrete time DDPs more naturally. This specific topic is a work in progress.

APPENDIX A. COVARIANCE STRUCTURE IN THE DISCRETE-TIME DDP

Note that our model implies that $y_{it}|\boldsymbol{\theta}_{it}, \sigma_i^2 = \mathbf{F}_t\boldsymbol{\theta}_{it} + \epsilon_{it}$ with errors $\epsilon_{it} \sim \mathsf{N}(0, \sigma_i^2)$ independent for every pair (i, t). Then

$$\begin{aligned}\mathbb{C}\text{ov}(y_{i,t}, y_{i',t+k}) &= \mathbb{C}\text{ov}(\mathbf{F}_t\boldsymbol{\theta}_{it}, \mathbf{F}_{t+k}\boldsymbol{\theta}_{i',t+k}) + \mathbb{C}\text{ov}(\epsilon_{it}, \mathbf{F}_{t+k}\boldsymbol{\theta}_{i',t+k}) \\ &\quad + \mathbb{C}\text{ov}(\mathbf{F}_t\boldsymbol{\theta}_{it}, \epsilon_{i',t+k}) + \mathbb{C}\text{ov}(\epsilon_{it}, \epsilon_{i',t+k})\end{aligned}$$

where the last three terms are zero as long as either $i \neq i'$ or $k \neq 0$. Therefore

$$\mathbb{C}\text{ov}(y_{i,t}, y_{i',t+k}) = \mathbf{F}_t\mathbb{C}\text{ov}(\boldsymbol{\theta}_{it}, \boldsymbol{\theta}_{i',t+k})\mathbf{F}'_{t+k}$$

Now,

$$\begin{aligned}\mathbb{E}(\boldsymbol{\theta}_{it}\boldsymbol{\theta}_{i',t+k}) &= \mathbb{E}\left(\sum_{l=1}^{\infty} w_l^*\delta_{(\boldsymbol{\theta}^*_{lt})} \sum_{l=1}^{\infty} w_l^*\delta_{(\boldsymbol{\theta}^*_{l,t+k})}\right) \\ &= \mathbb{E}\left(\sum_{l=1}^{\infty}\sum_{r=1}^{\infty} w_l^* w_r^*\delta_{(\boldsymbol{\theta}^*_{lt})}\delta_{(\boldsymbol{\theta}^*_{r,t+k})}\right) \\ &= \mathbb{E}\left(\sum_{l=1}^{\infty} w_l^{*2}\delta_{(\boldsymbol{\theta}^*_{lt})}\delta_{(\boldsymbol{\theta}^*_{l,t+k})}\right) + \mathbb{E}\left(\sum_{l=1}^{\infty}\sum_{r=1,r\neq l}^{\infty} w_l^* w_r^*\delta_{(\boldsymbol{\theta}^*_{lt})}\delta_{(\boldsymbol{\theta}^*_{r,t+k})}\right) \\ &= \frac{1}{1+\alpha}\mathbb{E}(\boldsymbol{\theta}^*_{1t}\boldsymbol{\theta}^*_{1,t+k}) + \frac{\alpha}{1+\alpha}\mathbb{E}(\boldsymbol{\theta}^*_{1t})\mathbb{E}(\boldsymbol{\theta}^*_{1,t+k})\end{aligned}$$

Therefore

$$\begin{aligned}\mathbb{V}(\boldsymbol{\theta}_{it}, \boldsymbol{\theta}_{i',t+k}) &= \mathbb{E}(\boldsymbol{\theta}_{it}\boldsymbol{\theta}_{i',t+k}) - \mathbb{E}(\boldsymbol{\theta}_{it})\mathbb{E}(\boldsymbol{\theta}_{i',t+k}) \\ &= \frac{1}{1+\alpha}\mathbb{E}(\boldsymbol{\theta}^*_{1t}\boldsymbol{\theta}^*_{1,t+k}) - \frac{1}{1+\alpha}\mathbb{E}(\boldsymbol{\theta}^*_{1t})\mathbb{E}(\boldsymbol{\theta}^*_{1,t+k}) \\ &= \frac{1}{1+\alpha}\mathbb{C}\text{ov}(\boldsymbol{\theta}^*_{1t}, \boldsymbol{\theta}^*_{1,t+k}) \\ &= \left[\prod_{r=1}^{k}\mathbf{G}_{t+k-r+1}\right]\mathbb{V}(\boldsymbol{\theta}^*_t)\end{aligned}$$

Since the pair $(\boldsymbol{\theta}^*_{1t}, \boldsymbol{\theta}^*_{1,t+k})$ is sampled from the baseline measure K_0.

REFERENCES

Aguilar, O., G. Huerta, R. Prado, and M. West (1999). Bayesian inference on latent structure in time series. In A. D. J.O. Berger, J.M. Bernardo and A. Smith (Eds.), *Bayesian Statistics 6*, pp. 3–26. Oxford University Press.

Antoniak, C. (1974). Mixtures of dirichlet processes with applications to bayesian nonparametric problems. *Annals of Statistics* **2**, 1152–1174.

Bigelow, J. L. and D. B. Dunson (2005). Semiparametric classification in hierarchical functional data analysis. Technical report, Institute of Statistics and Decision Sciences, Duke University.

Black, F. and M. Scholes (1973). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy 81*, 637–654.

Blackwell, D. and J. B. MacQueen (1973). Ferguson distribution via pólya urn schemes. *The Annals of Statistics* **1**, 353–355.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* **31**, 307–327.

Box, G. E. P. and G. M. Jenkins (1974). *Time Series Analysis: Forecasting and Control* (2nd ed. ed.). San Francisco: Holden-Day.

Bush, C. A. and S. N. MacEachern (1996). A semiparametric bayesian model for randomised block designs. *Biometrika* **83**, 275–285.

Carter, C. K. and R. Kohn (1994). On gibbs sampling for state space models. *Biometrika* **81**, 541–553.

Chaveesuk, R., C. Srivaree-Ratana, and A. Smith (1999). Alternative neural network approaches to corporate bond rating. *Journal of Engineering Valuation and Cost Analysis* **2**, 117–131.

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

Cont, R. (2006). Model uncertainty and its impact on the pricing of derivative instruments. *Mathematical Finance* **16**, 519–542.

Cont, R. and P. Tankov (2003). *Financial Modelling with Jump Processes*. Chapman & Hall.

DeIorio, M., P. Müller, G. L. Rosner, and S. N. MacEachern (2004). An anova model for dependent random measures. *Journal of the American Statistical Association* **99**, 205–215.

Delbaen, F. and W. Schachermayer (2006). *The Mathematics of Arbitrage*. Springer-Verlag.

Duffie, D., J. Pan, and K. Singleton (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica 68*, 1343–1376.

Dunson, D. (2005). Bayesian semiparametric isotonic regression for count data. *Journal of the American Statistical Association* **100**, 618–627.

Dunson, D. (2006). Bayesian dynamic modeling of latent trait distributions. *Biostatistics* **7**, 551–568. In press.

Dunson, D. B., A. H. Herring, and S. A. Mulheri-Engel (2005). Bayesian selection and clustering of polymorphisms in functionally-related genes. Technical report, Institute of Statistics and Decision Sciences, Duke University.

Dunson, D. B., N. Pillai, and J.-H. Park (2004). Bayesian density regression. Technical report, Institute of Statistics and Decision Sciences, Duke University.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of variance of united kingdom inflation. *Econometrica* **50**, 987–1008.

Escobar, M. D. (1994). Estimating normal means with a dirichlet process prior. *Journal of the American Statistical Association* **89**, 268–277.

Escobar, M. D. and M. West (1995). Bayesian density estimation and inference using mixtures. *Journal of American Statistical Association* **90**, 577–588.

Escobar, M. D. and M. West (1998). Computing nonparametric hierarchical models. In D. Dey, P. Müller, and D. Sinha (Eds.), *Practical nonparametric and semiparametric Bayesian statistics*, pp. 1–22. Springer - Verlag (Berlin, New York).

Ferguson, T. S. (1973). A bayesian analysis of some nonparametric problems. *Annals of Statistics* **1**, 209–230.

Ferguson, T. S. (1974). Prior distributions on spaces of probability measures. *Annals of Statistics* **2** , 615–629.

Fruhwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis* **15**, 183–202.

Gelfand, A. E., A. Kottas, and S. N. MacEachern (2005). Bayesian nonparametric spatial modeling with dirichlet process mixing. *Journal of the American Statistical Association* **100**, 1021–1035.

Green, P. and S. Richardson (2001). Modelling heterogeneity with and without the dirichlet process. *Scandinavian Journal of Statistics* **28**, 355–375.

Griffin, J. E. and M. Steel (2006a). Nonparametric inference in time series problems. In *Valencia Statistics 8*.

Griffin, J. E. and M. F. J. Steel (2006b). Order-based dependent dirichlet processes. *Journal of the American Statistical Association* **101**, 179–194.

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

Huang, Z., H. Chen, C. Hsu, and S. Chen (2004). Credit rating analysis with support vector machines and neural networks: a market comparative study. *Decision Support Systems* **37**, 543–558.

Huerta, G. and M. West (1999). Bayesian inference on periodicities and component spectral structure in time series. *Journal of Time Series Analysis 20*, 401–416.

Hull, J. (2005). *Options, Futures and Other Derivatives*. John Wiley & Sons.

Ishwaran, H. and L. F. James (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association* **96**, 161–173.

Kacperczyk, M., P. Damien, and S. G. Walker (2003). A new class of bayesian semiparametric models with applications to option pricing. Technical report, University of Michigan Bussiness School.

Kim, C. J. (1994). Dynamic linear models with markov-switching. *Journal of Econometrics* **60**, 1–22.

Kottas, A., M. D. Branco, and A. E. Gelfand (2002). A nonparametric bayesian modeling approach for cytogenetic dosimetry. *Biometrics* **58**, 593–600.

Lamon, E. I., S. Carpenter, and C. Stow (1998). Forecasting pcb concentrations in lake michigan salmonids: A dynamic linear model approach. *Ecological Applications 8*, 659–668.

Laws, D. J. and A. O'Hagan (2002). A hierarchical bayes model for multilocation auditing. *Journal of the Royal Statistical Society, Series D* **51**, 431–450.

Lo, A. (1984). On a class of bayesian nonparametric estimates: I. density estimates. *Annals of Statistics* **12**, 351–357.

MacEachern, S. N. (1994). Estimating normal means with a conjugate style dirichlet process prior. *Commnunications in Statistics, Part B - Simulation and Computation* **23**, 727–741.

MacEachern, S. N. (1999). Dependent nonparametric processes. In *ASA Proceedings of the Section on Bayesian Statistical Science*, pp. 50–55.

MacEachern, S. N. (2000). Dependent dirichlet processes. Technical report, Ohio State University, Department of Statistics.

MacEachern, S. N. and P. Müller (1998). Estimating mixture of dirichlet process models. *Journal of Computational and Graphical Statistics* **7**, 223–238.

Medvedovic, M. and S. Sivaganesan (2002). Bayesian infinite mixture model-based clustering of gene expression profiles. *Bioinformatics* **18**, 1194–1206.

Melick, W. and C. Thomas (1997). Recovering an asset's implied pdf from option prices: an application to crude oil during the gulf war. *Journal of Financial and Quantitative Analysis 32-1*, 91–115.

Merton, R. (1973). Theory of rational option pricing. *Bell Journal of Economics and Management Science 4*, 141–183.

Müller, P., F. Quintana, and G. Rosner (2004). Hierarchical meta-analysis over related non-parametric bayesian models. *Journal of Royal Statistical Society, Series B* **66**, 735–749.

Neal, R. M. (2000). Markov chain sampling methods for dirichlet process mixture models. *Journal of Computational and Graphical Statistics* **9**, 249–.

Panigirtzoglou, N. and G. Skiadopoulos (2004). A new approach to modeling the dynamics of implied distributions: Theory and evidence from the sp 500 options. *Journal of Banking and Finance 28*, 1499–1520.

Pennell, M. L. and D. B. Dunson (2006). Bayesian semiparametric dynamic frailty models for multiple event time data. *Biometrics In press*.

Pesaran, H., R. Smith, and K. Im (1995). Dynamic linear models for heterogeneous panels. Cambridge Working Papers in Economics 9503, Faculty of Economics (formerly DAE), University of Cambridge. available at http://ideas.repec.org/p/cam/camdae/9503.html.

Protter, P. (1990). *Stochastic Integration and Differential Equations*. Springer-Verlag.

Rasiel, E. (2003). *Hedging with a Smile: A Behavioral Explanation for the Volatility Put Skew*. Ph. D. thesis, Fuqua, Duke University.

Rebonato, R. (2004). *Volatility and Correlation: The Perfect Hedger and the Fox*. John Wiley & Sons.

Richardson, S. and P. Green (1997). On bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society, Series B* **59**(4), 731–792.

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

Tang, Y. and S. Ghosal (2006). A consistent nonparametric bayesian procedure for estimating autoregressive conditional densities. Technical report, Department of Statistics - North Carolina State University.

Tikhonov, A. N. (1963). Solution of incorrectly formulated problems and the regularization method. *Soviet Doklady 4*, 1035–1038.

West, M. (1995). Bayesian time series: Models and computations for the analysis of time series in the physical sciences. In *XV Workshop on Maximum Entropy and Bayesian Methods*.

West, M. (1997). Time series decomposition. *Biometrika* **84**, 489–494.

West, M. and J. Harrison (1997). *Bayesian Forecasting and Dynamic Models* (Second Edition ed.). Springer - Verlag, New York.

West, M., R. Prado, and A. Krystal (1999). Evaluation and comparison of eeg traces: Latent structure in non-stationary time series. *Journal of the American Statistical Association* **94**, 1083–1095.

Wolpert, R. L. and K. Ickstadt (2004). Reflecting uncertainty in inverse problems: A Bayesian solution using Lévy processes. *Inverse Problems 20*(6), 1759–1771.

Yacine, A.-S. (1996). Nonparametric pricing of interest rate derivative securities. *Econometrica* **64**, 527–560.

Yacine, A.-S. and D. Jefferson (2003). Nonparametric option pricing under shape restrictions. *Journal of Econometrics* **116**, 9–47.