Bayesian nonparametric methods in econometrics

by Jim Griffin, Maria Kalli and Mark Steel

The use of Bayesian nonparametrics has seen a rapid development since the work of Escobar and West (1995) showed that Markov chain Monte Carlo methods could be used to make inference in Bayesian nonparametric models. Subsequently, there has been the development of a huge number of models, priors and related computational methods. This has allowed Bayesian nonparametric methods to be applied to a range of statistical modelling problems. This paper provides an excellent introduction to the application of these methods to problems in areas such as clinical trial design, survival analysis and clustering of proteins. This motivates density estimation, nonparametric regression and nonparametric spatial modelling. The ability of Bayesian nonparametric methods to combine the flexibility of nonparametrics with the simplicity of the Bayesian modelling framework (such as hierarchical structure, automatic dimension penalisation and simple combination of different models) leads to attractive modelling approaches to applied problems. The authors concentrate on problems in medical statistics. Another applied area where Bayesian nonparametric approaches can play an important role is (financial) econometrics. In contrast to the data described in this paper, financial and economic data are typically collected over time at different frequencies which vary from intra-day (ultra high frequency), to daily (high frequency), monthly, quarterly, and annually, where the latter two mostly relate to business cycle data at the level of national or international economies.

The modelling challenge with financial and economic data is capturing their distributional characteristics (stylised facts), some of which differ depending on the frequency, as well as their time dependence, and this is just in the univariate case. Moving from univariate to multivariate, one needs to consider the dynamic relationships between such variables, and how to adequately model the transition mechanism. These relationships are rarely well described by simple parametric models. This has lead to a huge interest in classical nonparametric procedures which avoid making strong distributional assumptions. Traditionally, there has been less work in Bayesian nonparametric modelling, reflecting the relative lack of familiarity of econometricians and financial economists with Bayesian nonparametric methods and their computational complexity, but we note a steady increase in interest over the last ten years. We firmly believe that Bayesian nonparametric methods will play an important role in developing econometric models with excellent forecasting performance.

Much of the initial Bayesian nonparametric work in economics is reviewed in Griffin et al. (2011) and we focus on more recent developments which concentrate on density estimation within a volatility model or in the context of portfolio management, long memory models, models for the term structure of interest rates and vector autoregressive models.

1 Volatility modelling

Modelling of the distribution of financial time series, y_t, observed at regularly spaced times $t = 1\ldots, n$, is important for measuring risk. A starting point for many models is the stochastic volatility model of Taylor (1986)

$$y_t = e^{(h_t/2)}\epsilon_t,$$
$$h_t = \gamma + \phi h_{t-1} + \eta_t,$$

where the ϵ_t's (the innovations) and η_t's are i.i.d. $N(0,1)$ and $N(0,\sigma_\eta^2)$ random variables respectively. The SV model allows the log volatility to evolve, ensuring that the variance remains positive without need for further constraints. The latent process h_t can be interpreted as the random and uneven flow of information in the financial markets, and ϕ as the persistence in volatility.

Jensen and Maheu (2010) used the traditional DPM to model the unconditional distribution of returns by

$$p(y_t) = \sum_{j=1}^{\infty} w_j N(y_t|\mu_j, \lambda_j^{-2} e^{h_t})$$

where μ_j and $\lambda_j^{-2} e^{h_t}$ are the mean and variance associated with the j^{th}component, and w_j are the stick-breaking weights. The conditional volatility $h_t|h_{t-1}$ is generated from a normal distribution with mean ϕh_{t-1} and variance σ_η^2, the parameters $\mu_j, \lambda_j^{-2} \overset{iid}{\sim} G(\cdot)$ and $G(\cdot) \sim DP(G_0, M)$. The authors choose a normal-gamma as their base measure $G_0(\mu_j, \lambda_j^{-2})$ and refer to this model as SV-DPM.

They compared the SV-DPM to the SV with normal with Student-t innovations on a sample of daily asset returns over a period of 26 years, and showed that the predictive density of the SV-DPM model displays both negative skewness and high kurtosis, which are two of the 'stylised facts' of asset returns. The predictive densities of two parametric models did not capture these characteristics. In terms of out-of-sample predictive performance the log-predictive scores of the SV-DPM were better than those of the two parametric SV models.

Delatola and Griffin (2011), also use the stick-breaking representation of the DPM to model the distribution of asset returns using a stochastic volatility model. They use the Kim et al. (1998) linear state space representation of a stochastic volatility (SV) model where

$$y_t^\star = h_t + z_t \text{ for } t = 1, \ldots, n.$$

$y_t^\star = \log y_t^2$ (the log of the squared returns), $h_t = \phi h_{t-1} + \sigma_\eta \eta_t$, is the log-volatility at time t, and $z_t = \log(\epsilon_t^2)$. Both ϵ_t and η_t have zero mean and unit variance, and they are independent. In order to proceed to inference, Kim et al. (1998) suggest a mixture of normals to approximate the distribution of z_t whereas Delatola and Griffin (2011) consider a DPM. Their simulated examples show that this choice leads to narrower credible intervals when compared to the parametric method. In addition their model's out-of-sample predictive performance is superior to that of Kim et al. (1998).

Kalli et al. (2013) use a stick-breaking mixture to model the conditional distribution of asset returns. Their choice captures key 'stylised facts' of returns, specifically the heavy-tails and asymmetry of their distribution, the time varying volatility, and the 'leverage-effect' (the negative correlation between returns and volatility). They adopt a generalised autoregressive conditional heteroskedastic (GARCH) (Bollerslev, 1987) model for the conditional distribution. In contrast to a stochastic volatility model where h_t follows a latent stochastic process, GARCH models assume the h_t process can be modelled as a deterministic function of the asset returns (y_t's). Kalli et al. (2013) model the innovations using a general stick-breaking prior where $k(.|\theta)$ is a scaled uniform density, instead of the usual normal distribution. This allows for uni-modalilty, asymmetry and heavy tails of the innovation distribution.

Under their stick-breaking representation, the infinite mixture of uniforms for modelling the innovations' distribution has the following hierarchical setup:

$$\begin{aligned} y_t = \sigma_t \epsilon_t, \qquad & \epsilon_t \sim U(-\theta_{d_t} e^{-\lambda}, \theta_{d_t} e^{\lambda}) \qquad \text{for } t = 1, \ldots, n \\ Pr(d_t = j) = w_j, \qquad & \theta_j \sim G_0(\cdot) \qquad \text{for } j = 1, 2, \ldots \\ w_1 = v_1, w_j = v_j \prod_{\ell<j} (1 - v_\ell) \text{ and } & v_j \sim Be(a_j, b_j) \end{aligned}$$

where G_0 is a standard exponential distribution, and λ is the skewness parameter. The distribution of ϵ_t is therefore,

$$f_{v,\theta}(\epsilon_t) = \sum_{j=1}^{\infty} w_j \, U(-\theta_{d_t} e^{-\lambda}, \theta_{d_t} e^{\lambda})$$

and the conditional return distribution is:

$$f_{G,\lambda}(y_t|h_t) = \sum_{j=1}^{\infty} w_j \, U(y_t| - \theta_{d_t} \sigma_t \, e^{-\lambda}, \theta_{d_t} \sigma_t \, e^{\lambda})$$

where the volatility σ_t^2 is modelled using the GARCH(1,1), the GJR-GARCH(1,1) (Glosten et al., 1993) and the EGARCH(1,1) (Nelson, 1991). These choices of model allow the leverage effect to be modelled. Delatola and Griffin (2013) and Jensen and Maheu (2014) consider alternative approaches for extending their nonparametric SV models to account for the leverage effect.

Kalli et al. (2013) used a simulated GARCH(1,1) to compare the estimates of their modelling approach to those of the parametric and DPM alternatives and found it to be superior both in terms of intervals and mean integrated squared error. They compared their model with the three GARCH-type volatility specifications to the SV models of Jensen and Maheu (2010) and Jacquier et al. (2004) and showed that their out-of-sample performance was superior.

The recent work of Jin and Maheu (2016) focuses on estimating the conditional distribution of realised covariance (RCOV) matrices. These are daily volatility measures of the correlation between different assets which are estimated using ultra high frequency intraday data. The RCOVs theoretical underpinning is based on Andersen et al. (2003), and it is viewed as the quadratic variation of semi-martingale processes. Subsequently the econometric literature focused on improving this estimator in the presence of micro-structure noise and asynchronous trading. Jin and Maheu (2016) use the Hierarchical Dirichlet Process (HDP) of Teh et al. (2006) to capture the time dependence of the RCOV matrix. The HDP is a distribution over multiple correlated probability measures, $G_1, \ldots, G_r$, sharing the same atom locations. Each probability measure is generated from a DP with shared precision parameter and base measure. To ensure the sharing of atom locations, a second DP is used on the base measure, and hence the hierarchical set-up is:

$$G_0 \sim DP(M_0, H), \qquad G_j \sim DP(M, G_0) \text{ for } j = 1, 2, \ldots.$$

In Jin and Maheu (2016), the base measure H is also given a HDP prior, and the base measure of that

prior is an inverse Wishart distribution, a popular choice when modelling RCOVs. Their results show that density forecasts for both the conditional returns distribution and the distribution of the RCOV, based on the HDP are significantly better than any parametric method. This is because the parametric methods do not account for extreme observations in the RCOV matrices.

1.1 Portfolio allocation

The accurate modelling of the co-movement of asset returns, is key to risk management and portfolio allocation. Multivariate GARCH (MGARCH) type models, first introduced by Bollerslev et al. (1988), remain a popular choice when modelling the volatility of a portfolio as well as the volatility of the assets within it. For an extensive literature review see Silvennoinen and Terasvirta (2009). Similar to the univariate case these models should flexibly account for key "stylised facts" of asset returns, the asymmetry and heavy tails of both the conditional and unconditional distribution, the time-varying volatility and the negative correlation between volatility and returns. The GARCH-type models capture the time-varying volatility, and the distributional choice of the innovations' distribution serves in capturing the asymmetry and heavy-tails of the conditional and unconditional return distribution.

Jensen and Maheu (2013) depart from the popular choice of a Student-t or skewed Student-t distribution for the innovation vector. They choose to represent this distribution via DPM where the base measure G_0 is a multivariate normal-Wishart distribution. The vector of p asset returns, y_t, is assumed to be

$$y_t = H_t^{1/2}\epsilon_t \qquad \text{for } t = 1,\dots,T$$

where H_t is a $p \times p$ symmetric, scale matrix, and ϵ_t the $p \times 1$ innovation vector with an unknown distribution G. The matrix H_t is given by

$$H_t \;=\; \Gamma_0 + \Gamma_1 \odot y_{t-1}y_{t-1}^{'} + \Gamma_2 \odot H_{t-1},$$

where Γ_0 is a symmetric positive definite matrix, such that $\Gamma_0 = L_0 L_0'$ with L_0 a lower triangular matrix. $\Gamma_1 = \gamma_1\gamma_1'$ and $\Gamma_2 = \gamma_2\gamma_2'$ where γ_1 and γ_2 are $p \times 1$ vectors and the symbol $\odot$ denotes the Hadamard product. This is the construction of the MGARCH where each $h_{ij,t}$ element of H_t is only related to its lag $h_{ij,t-1}$ and to past returns. The hierarchical structure of this MGARCH-DPM is

$$\begin{aligned}
y_t|\mu_t,\Sigma_t,H_t &\sim N(H_t^{1/2}\mu_t, H_t^{1/2}\Sigma_t^{-1}(H_t^{1/2})') \\
H_t &= \Gamma_0 + \Gamma_1 \odot y_{t-1}y_{t-1}^{'} + \Gamma_2 \odot H_{t-1} \\
\mu_t,\Sigma_t|G &\overset{iid}{\sim} G \\
G|G_0,M &\sim DP(G_0,M) \\
G_0 &= N - \text{Wishart.}
\end{aligned}$$

The MGARCH-DPM model was compared to the MGARCH with Student-t innovations using three portfolios. Two equity portfolios (one with three assets and one with ten) and a foreign exchange portfolio with three assets. The weights of each portfolio were chosen by the authors. The out-of-sample

forecasting performance was compared using log-predictive scores, and the quality of the model fit using Bayes factors. The MGARCH-DPM was at least as good as the MGARCH with t-innovations for the foreign exchange portfolio, and performed significantly better in terms of density forecasts, particularly during the 2008 financial crisis, when the equity portfolios were considered.

Virbickaite et al. (2016) also consider portfolio risk, and just like Jensen and Maheu (2013) model the innovation distribution using the DPM with the base measure G_0 being the multivariate normal-Wishart distribution. They calculate the portfolio weights assuming that the investors' objective is to minimise the variance of the portfolio and model volatility using the asymmetric dynamic conditional correlation model (ADCC), of Cappiello et al. (2006) combined with a GJR-GARCH model. Their model is then applied to a portfolio with two assets, and its out-of-sample predictive performance is better than that of parametric ADCC models.

1.2 Long memory models

All the papers reviewed so far, have focused on estimating either the conditional or unconditional distribution of asset returns in volatility models. This was done by generating the unknown innovation distribution via a Dirichlet process prior, or a general stick breaking prior. Kalli and Griffin (2015) take a different approach to volatility modelling. They focus on the concept of long range dependence, i.e. the slow decay of the sample autocorrelation function, and model volatility h_t as the aggregate of AR(1) processes, see Robinson (1978); Granger (1980); Zaffaroni (2004). Aggregation of such processes leads to a class of models with long-range dependence, and the distributional choice for the autoregressive parameter, F_ϕ, has an effect on this dependence. Kalli and Griffin (2015) model the unknown F_ϕ using a Dirichlet process prior. The DP generates discrete probability measures, and this allows the decomposition of the aggregate process into processes with different levels of dependence. This models the effect of uneven information flows on volatility and can be linked to the differences in effects of different types of information (since some information may have a longer lasting effect on volatility than other pieces of information). Kalli and Griffin (2015) refer to their model as stochastic volatility with infinite cross sectional aggregation (SV-ICA). To construct it they first defined a suitable limiting process for cross-sectional aggregation as the number of elements tends to infinity and then used the Dirichlet process prior for F_ϕ, the distribution of the persistence parameter ϕ. For more details on this modelling approach refer to Kalli and Griffin (2015).

The authors applied the SV-ICA model to simulated data and found that a finite approximation to F_ϕ converged to the true F_ϕ. They also applied the SV-ICA to sampled returns from HSBC PLC and Apple Inc. They showed that the volatility dynamics can be decomposed into short-term, medium-term and long-term components, and that these dynamics depend on the sector in which a company is operating, with the banking sector exhibiting long range dependence when compared to the technology sector. The out-of-sample predictive performance of the SV-ICA was substantially better than that of the SV models of Jacquier et al. (2004) and Jensen and Maheu (2010).

1.3 Interest rate yields

The modelling of interest rate yields is an important problem in economics. The yield will depend on the length of time money is lent which is known as the maturity and yields at each maturity will also change over time. We assume that yields are observed at p maturities in discrete time and the yield of maturity m_i at the t-th time point is denoted by $y_{t,i}$. The Nelson-Siegel (Nelson and Siegel, 1987) model expresses the yields in terms of three latent factors. We use the parameterization of Diebold and Li (2006) who assume that

$$y_{t,i} = X_{t,i}\beta_t + \epsilon_{t,i} \tag{1}$$

where $\epsilon_{t,i} \sim \mathrm{N}(0, \sigma_t^2)$ and the latent factors are

$$X_{t,i} = \left(1, \frac{1}{\lambda_t m_i}(1 - e^{-\lambda_t m_i}), \frac{1}{\lambda_t m_i}(1 - e^{-\lambda_t m_i}) - e^{-\lambda_t m_i}\right).$$

These factors can be interpreted, respectively, as the yield level (which controls long term yields),the yield slope (which controls short-term behaviour) and the yield curvature (which controls medium term behaviour). The parameter λ_t in the second and third functions, controls the rate of decay with smaller values of λ_t leading to a slower decay.

A Bayesian nonparametric model can be constructed by allowing the conditional distribution of the parameters of the Nelson-Siegel three-component model to vary over time so that

$$(\beta_t, \lambda_t, \sigma_t^2) \sim G_t.$$

A time-varying nonparametric prior, the Dirichlet process autoregressive process (DPAR), introduced by Griffin and Steel (2011) defines

$$G_t(B) = \sum_{i=1}^{\infty} I(\tau_i < t) w_i(t) \delta_{\theta_i}(B) \tag{2}$$

where $\{\tau_i\}_{i=1}^{\infty}$ follow a Poisson process with intensity η and associates the marks (v_i, θ_i) with τ_i where $v_i \overset{i.i.d.}{\sim} \mathrm{Be}(1, M)$, $\theta_i \overset{i.i.d.}{\sim} H$ and

$$w_i(t) = v_i \prod_{\{j|\tau_i<\tau_j<t\}} (1 - v_j).$$

The structure implies that G_t follows a Dirichlet process with parameters M and H *a priori* with the dependence between G_t and G_{t+s} controlled by M and η. We will define this prior as $G_t \sim \mathrm{DPAR}(M, H; \eta)$. Suppose that τ_i^+ is a time slightly larger than τ_i then

$$G_{\tau_i^+}(B) = v_i \delta_{\theta_i}(B) + (1 - v_i) G_{\tau_i}(B).$$

and so the random probability measure evolves by jumping at arrival times $\{\tau_i\}_{i=1}^{\infty}$ with the addition of a new atom which is given weight V_i and all other atoms being down weighted by the factor $1 - V_i$. The correlation between $G_t(B)$ and $G_{t+s}(B)$ is given by $\rho(s) = e^{-[\eta/(M+1)]s}$.

We model zero coupon yields obtained from the Center for Research in Security Prices (CRSP) unsmoothed Fama and Bliss (1987) forward rates. The data set spanned January 1970 to December 2009 with maturities of 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108 and 120 months.

M	3.56 *(2.77, 4.65)*
η	0.089 *(0.010, 0.176)*

Table 1: Model 2: Parameter estimates as posterior median followed by 95% credible interval

The estimated parameters of the DPAR prior are shown in Table 1. The posterior median of M is relatively large (3.56) indicating that there are several distinct clusters at any given time point. The parameter η has a posterior median of 0.089. This parameter controls the arrival times of new clusters which would have a mean of 11.24 for this value of η. The combined effect of the two parameters can be understood through the autocorrelation function which, using the posterior medians, would be $\rho(s) = e^{-0.0195s}$. These results indicate that the underlying random measure is relatively stable over the course of the data period.

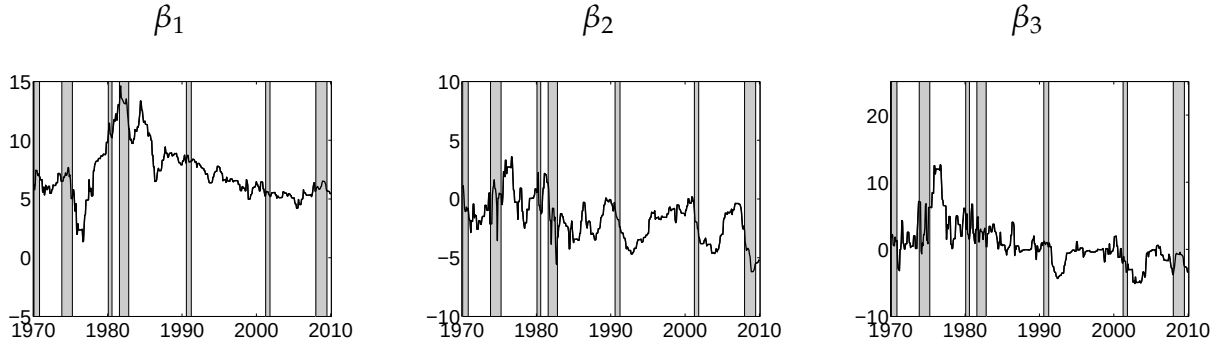


Figure 1: The smoothed posterior mean of $\mathrm{E}_{F_t}[\beta_t]$ (solid line). NBER periods of recession are shown with a grey band

Estimates of the mean of the time-varying factor weights are shown in Figure 1. The level parameter β_1 shows rapid changes in the 1970s with a rapid drop and recovery in the mid-1970s. The 1980s onwards show a largely decreasing value. The slope parameter β_2 shows a more interesting pattern with a rapid increase in the mid-1970s followed by a rapid decrease. The values become much more stable from the late 1970s onwards. The pattern is cyclical with decreases in the slope associated with periods of recession followed by increases in the next growth periods. The curvature parameter β_3 also shows a period of rapid change in the mid-1970s followed by more stable behaviour from the late 1970s onwards.

Figure 2 shows estimates of the mean of the time-varying factor loadings parameter λ_t and σ_t^2. These show fairly variable estimates for both λ_t and σ_t^2. The parameter λ_t shows a period of rapid change in the 1970s. Periods of recession seem to be associated with rapid drops in the value of λ_t. The mean of the volatility σ_t^2 has some peaks in the early period up to the mid-1980s but becomes much smaller during the late 1980s and 1990s (the period of the Great Moderation). The volatility increases again in the 2001 recession and in the late 2000s during the financial crisis.

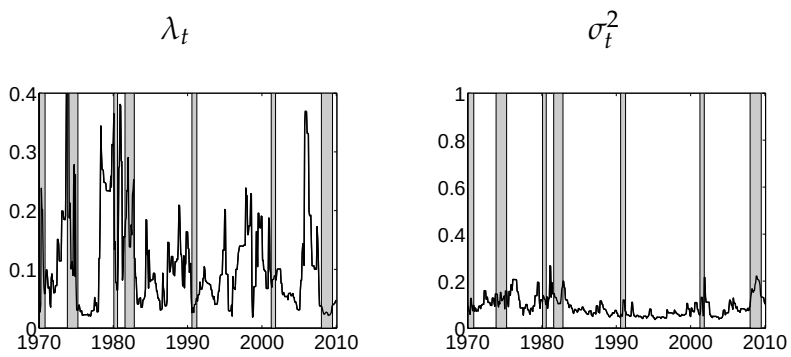


Figure 2: The smoothed posterior mean of $\mathrm{E}_{F_t}[\lambda_t]$ and $\mathrm{E}_{F_t}[\sigma_t^2]$ (solid lines). NBER periods of recession are shown with a grey band

1.4 Multivariate macroeconomic time series models

Multivariate time series models are widely used in macroeconomic modelling to understand the dynamic relationship between different economic variables (such as unemployment, inflation or output) in a particular economy or across different economies. The vector autoregressive (VAR) model has proved to be an important tool for analysing these types of time series. Suppose that y_t is a $(q \times 1)$-dimensional vector of economic variables measured at time t then the simplest VAR(1) model assumes that

$$y_t = \mu + \Phi(y_{t-1} - \mu) + \epsilon_t$$

where μ is a $(q \times 1)$-dimensional vector, Φ is a $(q \times q)$-dimensional matrix and ϵ_t is a $(q \times 1)$-dimensional multivariate normally distributed random vector with mean zero and covariance matrix Σ. The limitation of these models are now well-understood and many people have considered regime-switching models (which can be seen as a form of a dynamic mixture model) to more accurately capture the structure of the data and to provide better forecasting performance. Bassetti et al. (2014) propose a Bayesian nonparametric prior for a panel VAR model. They assume that there are r different economies and that $y_{i,t}$ is a vector of observations for the i-th economy at time t. They define

$$f_{it}(y_{i,t}) = \sum_k w_{i,k} \mathrm{N}((I_q \otimes X_t)\varphi_k, \Sigma_k)$$

where X_t is a row vector containing lags of $y_{i,t}$ for all economies. The framework allows flexibility in the conditional distribution of $y_{i,t}$ and allows different component weights for each economy. The introduction of indicator variables for the components of the mixture model allows inference to be made about the regime of economy i at time t. The authors introduce a novel prior, the beta-product dependent Pitman-Yor process, for $\{w_{i,k}\}_{i=1,k=1}^{r,\infty}$, by defining

$$w_{i,k} = v_{i,k} \prod_{l<k} (1 - v_{i,l})$$

where $v_{1,k}, \ldots, v_{r,k}$ are dependent and have beta marginal distributions. They consider two schemes for constructing such random variables. They apply their model to data from the US and EU economies to better understand the dynamics of the business cycle.

The model of Bassetti et al. (2014) uses time-invariant weights for multiple economies. Kalli and Griffin (2016) construct a Bayesian nonparametric model mixture of vector autoregressive models which has time-varying weights. This extends the univariate model of Antoniano-Villalobos and Walker (2016) and assumes that for multivariate observations y_t are modelled by a mixture of VAR's

$$f(y_t) = \sum_{k=1}^{\infty} w_k(y_{t-1}) \mathrm{N}(\mu_k + \Phi_k(y_{t-1} - \mu_k), \Sigma_k)$$

where

$$w_k(y) = \frac{\pi_k \mathrm{N}(y|\mu_k^\star, \Sigma_k^\star)}{\sum_{k=1}^{\infty} \pi_k \mathrm{N}(y|\mu_k^\star, \Sigma_k^\star)},$$

and $\pi_1, \pi_2, \ldots$ are given a stick-breaking prior, $\mu_k^\star$ and $\Sigma_k^\star$ are the stationary mean and covariance matrix for the VAR in the k-th component respectively. This implies that the stationary distribution and the transition density of y_t are both nonparametric mixtures of normal distributions. The model can be estimated using the adaptive truncation method of Griffin (2016). The authors demonstrate that this model has much better predictive performance than a VAR model and better predictive performance than the popular non-stationary time-varying parameter VAR model (Primiceri, 2005).

2 Normalised Random Measures with Independent Increments (NRMI models)

Most work on dependent random measures has concentrated on stick-breaking constructions. The authors discuss NRMI models in Section 2.3 and we believe that these offer an attractive framework for such measures. Griffin and Leisen (2017) define a compound random measure for related random probability measures $G_1(B), \ldots, G_r(B)$ by

$$G_j(B) = \frac{\sum_{k=1}^{\infty} m_{j,k} J_k \delta_{\theta_k}}{\sum_{k=1}^{\infty} m_{j,k} J_k}$$

where $m_{j,k} \overset{i.i.d.}{\sim} H$ are marks and $\tilde{\eta} = \sum_{i=1}^{\infty} J_i \delta_{\theta_i}$ is a realisation of a directing Lévy process with intensity ν which will often be taken to be a CRM (as described in Section 2.3). Many previously defined dependent NRMI priors fall within this class and it offers a rich framework for extension. For example, a regression model could be defined by assuming that G is indexed by a covariate x and defining

$$G_x(B) = \frac{\sum_{k=1}^{\infty} m_k(x) J_k \delta_{\theta_k}}{\sum_{k=1}^{\infty} m_k(x) J_k}$$

where $m_k(x)$ are independent stochastic processes. This approach is attractive from a modelling perspective since the dependence between measures is modelled through a sequence of parametric models (such as a Gaussian process for regression or an ARMA model for time series).

References

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2003). Modelling and forecasting realised volatility, *Econometrica* **71**: 579–625.

Antoniano-Villalobos, I. and Walker, S. G. (2016). A nonparametric model for stationary time series, *Journal of Time Series Analysis* **37**: 126–142.

Bassetti, F., Casarin, R. and Leisen, F. (2014). Beta-product dependent Pitman-Yor processes for Bayesian inference, *Journal of Econometrics* **180**: 49–72.

Bollerslev, T. (1987). A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return, *Review of Economics and Statistics* **69**: 542–547.

Bollerslev, T., Engle, R. and Wooldridge, J. (1988). A capital asset pricing model with time varying covariates, *Journal of Political Economy* **96**: 116–131.

Cappiello, L., Engle, R. and Sheppard, K. (2006). Asymmetric dynamics in the correlations of global equity and bond returns, *Journal of Financial Econometrics* **4**: 537–572.

Delatola, E. I. and Griffin, J. E. (2011). Bayesian nonparametric modelling of the return distribution with stochastic volatility, *Bayesian Analysis* **6**: 901–926.

Delatola, E. I. and Griffin, J. E. (2013). A Bayesian semiparametric model for volatility modelling with leverage effect, *Computational Statistics and Data Analysis* **60**: 97–110.

Diebold, F. X. and Li, C. (2006). Forecasting the term structure of government bond yields, *Journal of Econometrics* **130**: 337–364.

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

Fama, E. F. and Bliss, R. R. (1987). The information in long-maturity forward rates, *American Economic Review* **77**: 680–692.

Glosten, L., Jagannathan, R. and Runkle, D. (1993). On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks, *The Journal of Finance* **48**(5): 1779–1801.

Granger, C. (1980). Long memory relationships and the autoregression of dynamic models, *Journal of Econometrics* **14**: 227–238.

Griffin, J. E. (2016). An adaptive truncation method for inference in Bayesian nonparametric models, *Statistics and Computing* **26**: 423–441.

Griffin, J. E. and Leisen, F. (2017). Compound random measures and their use in Bayesian nonparametric models, *Journal of the Royal Statistical Society, Series B* **79**: 525–545.

Griffin, J. E., Quintana, F. A. and Steel, M. F. J. (2011). Flexible and nonparametric methods, *in* G. Koop, H. K. van Dijk and J. Geweke (eds), *Handbook of Bayesian Econometrics*, Oxford University Press.

Griffin, J. E. and Steel, M. F. J. (2011). Stick breaking autoregressive processes, *Journal of Econometrics* **162**: 383–396.

Jacquier, E., Polson, N. G. and Rossi, P. E. (2004). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors, *Journal of Econometrics* **122**: 185–212.

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

Jensen, M. and Maheu, J. M. (2013). Bayesian semiparametric multivariate GARCH modelling, *Journal of Econometrics* **176**: 3–17.

Jensen, M. and Maheu, J. M. (2014). Estimating a semiparametric asymmetric stochastic volatility model with a Dirichlet process mixture, *Journal of Econometrics* **178**: 523–538.

Jin, X. and Maheu, J. M. (2016). Bayesian semiparametric modelling of realized covariance matrices, *Journal of Econometrics* **192**: 19–39.

Kalli, M. and Griffin, J. E. (2015). Flexible modelling of dependence in volatility processes, *Journal of Business and Economic Statistics* **33**: 102–113.

Kalli, M. and Griffin, J. E. (2016). Bayesian nonparametric vector autoregressive models, *Technical report*.

Kalli, M., Walker, S. G. and Damien, P. (2013). Modelling the conditional distribution of daily stock index returns: an alternative Bayesian semiparametric model, *Journal of Business and Economic Statistics* **31**: 371–383.

Kim, S., Shephard, N. and Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models, *Review of Economic Studies* **65**: 361–393.

Nelson, C. and Siegel, A. (1987). Parsimonious modelling of yield curves, *Journal of Business* **60**: 473–489.

Nelson, D. B. (1991). Conditional Heteroscedasticity in Asset returns: A New Approach, *Econometrica* **59**: 347–370.

Primiceri, G. (2005). Time varying structural vector autoregressions and monetary policy, *Review of Economic Studies* **72**: 821—852.

Robinson, P. (1978). Statistical inference for a random coefficient autoregressive model, *Scandinavian Journal of Statistics* **5**: 163–168.

Silvennoinen, A. and Terasvirta, T. (2009). *Handbook of Financial Time Series*, Springer-Verlag, chapter Multivariate GARCH models, pp. 201–226.

Taylor, S. J. (1986). *Modelling financial time series*, John Wiley.

Teh, Y. W., Jordan, M. I., Beal, M. J. and Blei, D. M. (2006). Hierarchical Dirichlet Processes, *Journal of the American Statistical Association* **101**: 1566–1581.

Virbickaite, A., Ausin, C. M. and Galeano, P. (2016). A Bayesian nonparametric approach to asymmetric dynamic conditional correlation model with application to portfolio selection, *Computational Statistics and Data Analysis* **100**: 814–829.

Zaffaroni, P. (2004). Contemporaneous aggregation of linear dynamic models in large economies, *Journal of Econometrics* **120**: 75–102.