Bayesian nonparametric methods in Econometrics

Maria Kalli

Bayesian Nonparametrics Networking Workshop 2023
Monash Business School

- Econometrics is the application of statistical methods to quantify economic and financial market phenomena.
- The goal is to build useful models/tools that can inform economic policy making and manage financial risk.
- So what is the challenge ???

- Economic and Financial data are typically collected over time. Often at different frequencies:
  - low (monthly, quarterly, annually)
  - high (daily)
  - ultra-high (intra-day)

- "Stylised facts", see R.Cont (2001), of such data are notoriously difficult to capture because :
  - differ according to frequency,
  - nature of time dependence.

And this is just in the univariate case!

- Moving from univariate to multivariate modelling, we need to consider:
  - the form of the joint dynamic relationship between such variables, and
  - how to adequately model the transition mechanism.
- These relationships are rarely well described by parametric models.

- To avoid making strong distributional assumptions, econometricians and financial economists have “embraced” classical nonparametric methods.
- However, there has been less work in Bayesian nonparametrics. This is due to:
  - lack of familiarity, and
  - fear of complex models which also come with computational complexity.

In the last ten years Econometricians have slowly ventured into the wonderful world of Bayesian nonparametrics.

- I believe that Bayesian nonparametric (BayesNP) methods can play an important role in developing econometric models with excellent forecasting performance.

  Why ?

- Let's view models as information channels from current observable data to future predictions.
- A finite number of parameters acts as a bottleneck in the information channel.
- BayesNP remove the bottleneck by placing a prior on an infinite dimensional parameter space that adapts its complexity to the data.
- Outcome is a richer class of models which can lead to superior out-of-sample performance relative to competitive models, see Norets and Pati (2017).

- Most of the initial (pre 2010) BayesNP work in economics is reviewed in Griffin et al. (2011).
- I am going to focus on more recent developments which concentrate on:
  - density estimation within a volatility model
  - long memory models, and
  - multivariate macroeconomic time series analysis.

- Measuring risk is central to financial management.
- It all starts with modelling the distribution of asset returns, y_t for $t = 1, \ldots, n$.
- A popular choice 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 and η_t's are i.i.d. $N(0, 1)$ and $N(0, \sigma_\eta^2)$ random variables respectively.

The SV-DPM model of Jensen and Maheu (2010) uses the stick breaking representation of the Dirichlet Process Mixture model (DPM), see Lo (1984) and Sethuraman (1994), to model the unconditional distribution of returns as,

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

where,

- w_j are the stick-breaking weights,
- μ_j and $\lambda_j^{-2} e^{h_t}$ are the mean and variance associated with the j^{th}component,
- with $\mu_j, \lambda_j^{-2} \stackrel{iid}{\sim} N(m, (\tau\lambda^2)^{-1}) - Ga(\nu_0/2, s_0/2)$.

The conditional volatility $h_t | h_{t-1} \stackrel{iid}{\sim} N(\phi h_{t-1}, \sigma_\eta^2)$

- Delatola and Griffin (2011) is another example where the DPM is used within an SV model to capture the distribution of y_t.
- They use the linear state space representation of an 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.
- 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 the DPM.

- The SV models of Jensen and Maheu (2010) and Delatola and Griffin (2011) capture the heavy-tails and asymmetry of the y_t distribution.
- Kalli et al. (2013) model the conditional distribution of y_t using an infinite scaled mixture of uniform distributions, where the mixing measure is a general stick-breaking process (SBP).
- The motivation is to capture the heavy-tails and asymmetry of y_t, the time varying volatility, as well as the 'leverage-effect' (the negative correlation between returns and volatility).

- The conditional return distribution is:

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

where,

- w_j is the stick-breaking weight, such as $w_1 = v_1, w_j = v_j \prod_{\ell<j}(1 - v_\ell)$ and $v_j \sim Be(a_j, b_j)$
- $\theta_j \sim$ standard exponential,
- λ is the skewness parameter, and
- the volatility σ_t^2 is modelled using the GARCH(1,1) (Bollerslev, 1987), the GJR-GARCH(1,1) (Glosten et al., 1993), and the EGARCH(1,1) (Nelson, 1991)

Delatola and Griffin (2013) and Jensen and Maheu (2014) consider alternative approaches for extending their nonparametric SV models to account for the leverage effect.

- All aforementioned volatility models are suitable for high-frequency financial data (daily returns).

- So, what about ultra-high frequency (intra-day returns)?

- Griffin et al. (2021) use DPM to model *ex post variance*.
  - *ex-post variance* is a forward-looking measure of risk.
  - Used to determine an investor's maximum amount of loss over a specific period of time within a certain degree of probability.

- We can use high-frequency data to estimate the *ex post variance* using intra-period returns.

- O.E.Barndorff-Nielsen and Shephard (2002) and Andersen et al. (2003) formalised the idea of using high-frequency data to measure the volatility of lower frequency returns.

- Consider the following stochastic process for logarithmic price $p(t)$:

$$dp(t) = \mu(t, p(t))dt + \sigma(t, p(t))dB(t)$$

  where,
  - $\mu(t, p(t))$ is a drift term,
  - $\sigma(t, p(t))$ is the instantaneous volatility, and
  - $B(t)$ is a Brownian montion

- V_τ is the true variance measure of the return over period $(\tau - 1, \tau)$ defined as

$$V_\tau = \int_{\tau-1}^{\tau} \sigma^2(\Delta\tau)d\tau$$

- O.E.Barndorff-Nielsen and Shephard (2002) and Andersen et al. (2003) show that realised volatility, defined as

$$RV = \sum_{i=1}^{n_t} y_{t,i}^2$$

where $y_{t,1}, \ldots, y_{t,n_t}$ are the intra-day returns, is a consistent estimator of V_t.

- Griffin et al. (2021) model the intra-day returns $y_{t,i}$ as

$$
\begin{aligned}
y_{t,i} &\sim N(\mu_t, \sigma^2_{t,i}) \\
\sigma^2_{t,i} &\sim G_t \\
G_t &\sim DP(\alpha_t, G_{0,t}) \\
G_{0,t} &= IG(\nu_{0,t}, s_{0,t})
\end{aligned}
$$

(Hierarchical representation of a DPM)

- This allows partial polling of the $\sigma^2_{t,i}$, leading to better estimation of RV.

- Jin and Maheu (2016) use the Infinite Hidden Markov Model (IHMM) of M.Beal et al. (2002) to capture the time dependence of realised covariance (RCOV) matrices.
- IHMM can be thought of as a first-order Markov switching model with a countably infinite number of states.
- The unbounded nature of the transition matrix allows for both recurring states from the past as well as new states to capture structure change.

- The prior for the infinite transition matrix is based on the Hierarchical Dirichlet Process (HDP) of Teh et al. (2006).
- The HDP is a distribution over multiple correlated probability measures, $G_1, \ldots, G_r$, sharing the same atom locations.

$$\begin{aligned} G_j &\sim DP(M, G_0) \text{ for } j = 1, 2, \ldots \\ G_0 &\sim DP(M_0, H), \end{aligned}$$

- Each row of the transition matrix is centred around a common draw from a top level Dirichlet process.

- Jin and Maheu (2016) follow Fox et al. (2011) and give base measure H (of the bottom level DP) an HDP prior.
- The resulting model is referred to as the "sticky" version of the IHMM because it favours self transitions between states.

- Other applications of IHMM are found in:

  Y.Song (2014) for real interest rate, and inflation modelling,

  L.Bauwens et al. (2017) for Autoregressive Moving Average (ARMA) models, and

  Maheu and Yang (2016) for modelling the term structure of short term interest rates.

- In the models of Jensen and Maheu (2010), Delatola and Griffin (2011), Kalli et al. (2013) the focus was estimation of the conditional and/or unconditional distribution of returns in volatility models.
  - The unknown innovation distribution via a DP prior, or a general SB prior.
- In the work of Jin and Maheu (2016) capturing the time dependence of the RCOV matrix relied on the HDP, and
- Griffin et al. (2021) modelled the intra-day returns distribution via a DPM.

- The reviewed so far, relied on infinite mixtures, and the stick-breaking representation of the DP, SBP, and HDP.
- Moving away from infinite mixtures, we can look at how the DP can be used in other models, such as aggregation models.
- Good example in financial econometrics is the aggregation of AR(1) processes introduced by Robinson (1978); Granger (1980); Zaffaroni (2004).

- Aggregation of such processes leads to a class of models that can exhibit long-range dependence.
- The autoregressive parameter ϕ plays a big role on whether or not such dependence exists.
- The shape of the distribution, F_ϕ is pivotal to this.

- Kalli and Griffin (2015) model the unknown F_ϕ using a Dirichlet process prior, within a stochastic volatility (SV) framework.

- Because the DP generates discrete probability measures, it allows for the decomposition of the aggregate AR(1) process into processes with different levels of dependence.

- Kalli and Griffin (2015) refer to their model as stochastic volatility with infinite cross sectional aggregation (SV-ICA).
- The motivation behind it is :
  - to capture the effect of uneven information flows on volatility, and
  - see if this is linked to the differences in effects of dissimilar types of information.

- Macroeconomists are interested in explaining:
  - what goes on in the aggregate economy, and
  - what informs changes to the business cycle.

- Macroeconomic analysis involves multivariate time series models.
  - Dynamic Factor Models (DFM)
  - Vector Autoregressive Models (VAR)

- Sims (1980) introduced the 'vanilla' VAR model

$$y_t = B_1 y_{t-1} + \ldots + B_L y_{t-L} + e_t \tag{1}$$

where,

- $y_t = (y_{1,t}, \ldots, y_{p,t})'$ is a p-dimensional vector of macroeconomic variables,
- $\{B_l\}_{l=1}^{L}$ are $(p \times p)$-dimensional matrices of unknown coefficients, with L the number of lags, and
- $e_t = (e_{1,t}, \ldots, e_{p,t})'$ is a p-dimensional innovation vector with distribution $\mathrm{N}(0, \Sigma)$

The limitations of the VAR models are now well understood and many people have considered either regime-switching VAR models or time varying parameter VAR models.

- Bassetti et al. (2014) is one of the early papers that consider an infinite mixture representation for a panel VAR (PVAR) model.

- The idea is to understand interdependencies between different economies. Their PVAR model is:

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

where, $y_{i,t}$ is a vector of observations for the i-th economy at time t, and X_t is a row vector containing lags of $y_{i,t}$ for all economies.

- Bassetti et al. (2014) introduce a dependent BayesNP prior, the beta-product dependent Pitman-Yor prior.
- They use this prior to model the $w_{i,k}$, the mixture weights.
- This allows for different time-invariant weights for multiple economies.

- Kalli and Griffin (2018) construct a Bayesian nonparametric mixture of VAR models.
- They focus on the transition density of y_t and y_t^L, which is an infinite mixture, with weights depending on observed lagged values of y_t.
- This allows for different component transition densities to be favoured at different periods - for example, in expansionary and contractionary periods.

- Huber and Rossini (2022) divert from infinite mixtures and consider BART approach of Chipman et al. (2010).

- BART assumes that $y = (y_1, \ldots, y_T)$ - T-dimensional response vector of depends on $X = (x_1, x_2, \ldots, x_T)$ - $T \times K$ matrix of variables - through a nonlinear function f,

$$y = f(X) + \epsilon, \;\; \epsilon \sim N(0_T, \sigma^2 I_T)$$

where $f(X)$ is approximated by summing over N regression trees:

$$f(X) \approx \sum_{j=1}^{N} g(X|\tau_j, m_j)$$

where τ_j is the tree structure associated with the jth binary tree and $m_j = (\mu_{j,1}, \ldots, \mu_{j,b_j})'$ is the vector of terminal node parameters associated with τ_j and b_j - the leaves of the jth tree.

- The binary trees are constructed by considering splitting rules of the form :

$$\{X \in A_{j,k}\} \text{ or } \{X \notin A_{j,k}\}$$

with $A_{j,K}$ being the partition set.

- These rules typically depend on selected columns of X, denoted as $X_{\cdot j}$ for $(j = 1, ..., K)$, and a threshold c.

- The set $A_{j,k}$ is then defined by splitting the predictor space according to

$$\{X_{\cdot j} \leq c\} \text{ or } \{X_{\cdot j} > c\}$$

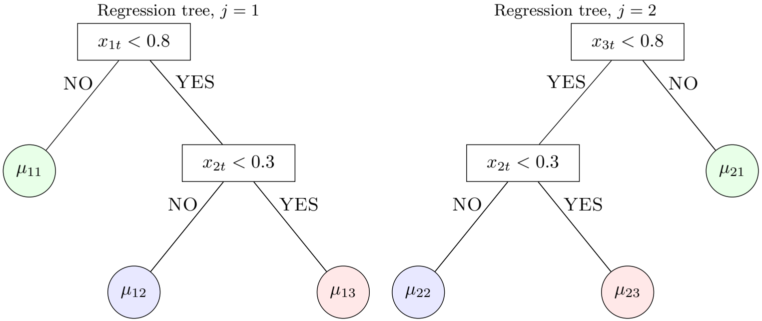
- The step function g is constant over the elements of

$$A_{j,k} : g(X|\tau_j, m_j) = \mu_{j,k}, \text{ if } X \in A_{j,k} \text{ for, } k = 1, ..., b_j$$

  Hence, the set $A_{j,k}$ defines a tree-specific unique partition of the covariate space s.t. the function g returns a specific value $\mu_{j,k}$ for specific values of x_t.

- Huber and Rossini (2022) extend BART to multivariate setting and call their model BAVART.

Sum of regression trees, $g(x|\tau_j, m_j)$, with internal nodes labeled by their splitting rules and leaf nodes labeled with the corresponding parameters $\mu_{j,k}$, where $j = 1, 2$ and $k = 1, 2, 3$.



- A drawback of BAVART (and BART) is that the ϵ's are assumed to be Gaussian
- Empirical evidence shows that shocks in key macroeconomic variables, such as inflation and GDP, are not Gaussian.
- So, Clark et al. (forthcoming) return to the infinite mixture approach and model the distribution of the ϵ's using a DPM.

- Bayesian nonparametric priors can be defined by using normalised random measures with independent increments (NRMII).

- NRMII are defined by normalising a completely random measure (CRM).

- CRM's date back to Kingman (1967) and can be represented as follows,

$$\tilde{\mu}^c = \sum_{i=1}^{\infty} J_i \delta_{X_i}$$

where both the positive jump heights J_i and the jump locations X_i are random.

- (J_i, X_i) have Lévy intensity ν
- Usually we take $\nu_J \times h_X$ where h_X is a probability density.
- For example the Gamma process can be represented as the limit

$$\lim_{K \to \infty} \sum_{i=1}^{K} J_i \delta_{X_i}$$

where J_i is distributed $Gamma(M/K, 1)$

- Griffin and Leisen (2017) define a compound random measure (CoRM) 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

- M.Beraha (2023) use CoRM as prior to derive a factor model for random measures.
- They apply their model to 265 areas of California to understand the variation of the personal income distribution.

  For more on this please see my student Ziyou Wang's poster.

- The behaviour of financial markets, and the aggregate economy can change abruptly.
- The time period when this happens is referred to as a change point.
- Product Partition Models (PPMs) date back to Barry and Hartigan (1992), and can be used in change point modelling.

- We can use PPMs to define a prior on the number and locations of change points.
- We assume that the data can be partitioned into non-overlapping homogeneous segments, that are independent.
- My student Cason McKee has a poster where PPMs are used to identify change points is in macroeconomic time series.
- Other work on PPMs goes back to Quintana and Iglesias (2003) and more resent work can be seen in S.Paganin et al. (2023) and Mueller et al. (2023).

References I

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

Barry, D. and Hartigan, J. (1992). Product partition models for change point problems, *Annals of Statistics* **20**: 260–279.

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.

Chipman, H., George, E. and MvCulloch, R. (2010). Bart: Bayesian additive regression BART : Bayesian additive regression trees., *Annals of Applied Statistics* **4**(1): 266–298.

Clark, T., Huber, F., Koop, G. and Marcellino, M. (forthcoming). Forecasting US inflation using Bayesian nonparametric models, *Annals of Applied Probability* .

References II

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.

Fox, E., Sudderth, E., M.Jordan and Willsky, A. (2011). A sticky HDP-HMM with application to speaker diarization., *Annals of Applied Statistics* **5**: 1020–1056.

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., J.Liu and Maheu, J. (2021). Bayesian Nonparametric Estimation of Ex Post Variance, *Journal of Financial Econometrics* **19**(5): 823–859.

References III

Griffin, J. E. and Leisen, F. (2017). Compound random measures and their use in Bayesian non-parametrics, *JRSS 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.

Huber, F. and Rossini, L. (2022). Inference in Bayesian Additive Vector Autoregressive Tree models, *Annals of Applied Statistics* **16**: 104–123.

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

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.

References IV

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. (2018). Bayesian Nonparametric Vector Autoregressive Models, *JoE* **203**: 267–282.

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 Comparisons with ARCH Models, *RES* **65**: 361–393.

Kingman, J. (1967). Completely random measures, *Pacific Journal of Mathematics* **21**: 59–78.

L.Bauwens, Carpantier, J. and Dufays, A. (2017). Autoregressive Moving Average Autoregressive infinite Markov-switching models, *Journal of Business and Economic Statistics* **35**(2): 162–182.

References V

Lo, A. Y. (1984). On a Class of Bayesian Nonparametric Estimates: I. Density Estimates, *Annals of Statistics* **12**: 351–357.

Maheu, J. and Yang, Q. (2016). An infinite hidden markov model for short-term interest rates, *Journal of Empirical Finance* **38**: 202–220.

M.Beal, Z.Gharamani and C.Rasmussen (2002). The Infinite Hidden Markov Model, *In Advances in Neural Information Processing Systems* pp. 577–584.

M.Beraha, a. (2023). Normalised Latent Measure Factor Models, *Journal of the Royal Statistical Society, Series B* **85**(4): 1247–1270.

Mueller, P., Quintana, F. A. and Page, G. (2023). Regression with variable dimension covariates, *Sankhya A* pp. 1–14.

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

Norets, A. and Pati, D. (2017). Adaptive Bayesian Estimation of Conditional Densities, *Econometric Theory* **33**: 980–1012.

References VI

O.E.Barndorff-Nielsen and Shephard, N. (2002). Estimating quadratic variation using realized, *Estimating Quadratic Variation Using Realized Estimating Quadratic Variation Using Realized Variance* **17**: 457–477.

Quintana, F. A. and Iglesias, P. (2003). Bayesian clustering and product partition models, *Journal of the Royal Statistical Society, Series B* **65**: 557–574.

R.Cont (2001). Empirical properties of asset returns: stylized facts and statistical issues, *Quantitative Finance* **1**(2): 223–236.

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

Sethuraman, J. (1994). A constructive definition of Dirichlet priors., *Statistica Sinica* **4**(2): 639–650.

Sims, C. (1980). Macroeconomics and Reality, *Econometrica* **48**: 1–48.

S.Paganin, Page, G. and Quitana, F. (2023). Informed random partition models with temporal dependence, arXiv preprint arXiv:2311.14502.

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

References VII

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.

Y.Song (2014). Modeling Regime Switching and Structural Breaks with an Infinite Hidden Markov Switching Model, *Journal of Applied Econometrics* **29**: 825–842.

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