Bayesian Nonparametric functional forecasting with locally-autoregressive particle systems: application to virtual gas markets

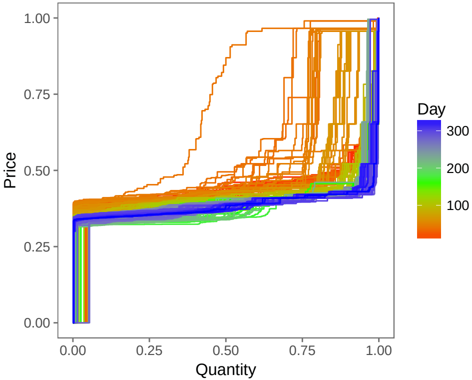
Guillaume KON KAM KING[1,2]
Matteo RUGGIERO[1,2]
Antonio CANALE[2,3]

[1]University of Torino, Italy
[2]Collegio Carlo Alberto, Moncalieri, Italy
[3]University of Padova, Italy

Mathematical Methods of Modern Statistics, 11/07/2017

Gas offer curve

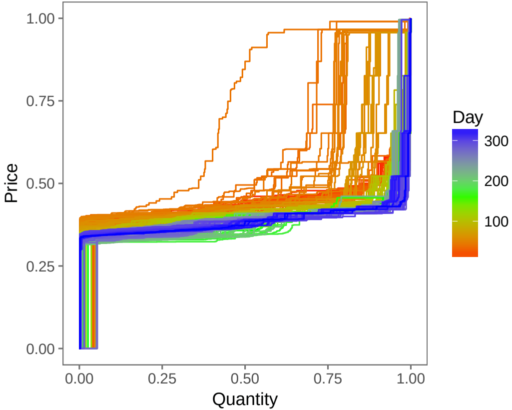
Liberalisation of the Italian energy market: different providers who trade gas to equilibrate their stocks.



- daily auctions
- offer and demand curve

Gas offer curve

Some comments on the data:



- bounded functions
- monotone functions
- time variability
- local trends
- noise-free data

Motivation for Bayesian forecast

Statistical challenge = h-steps-ahead functional forecasting

- Market traders want to predict the next days' curves (short term forecast)
- Full curve needed to design strategies
- Uncertainty needed

Some strengths of Bayesian Nonparametrics:

- great flexibility for the curves' irregular shapes
- propagation of uncertainty to forecast is simple

Description through a latent particle system

More options available for specifying a stochastic process for particle systems than curves.

- Monotone functional data can be represented by a **latent n-particle system** $\{X_i(t)\}_{t=1,\dots,T}$
- $X^{(n)}(t) = (X_1(t), \dots, X_n(t))$ is a vector of interacting $[0, 1]$ valued processes
- This latent particle system is related to the data by:

$$D_t(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(X_i(t) \leq x)$$

A population genetics model (Moran/Wright-Fisher model)[Canale and Ruggiero, 2016]

Given the state $X^{(n)}(t-1)$ the next state $X^{(n)}(t)$ is obtained as follows:

- delete M particles with:

$$M \sim \text{Binom}(n, p)$$

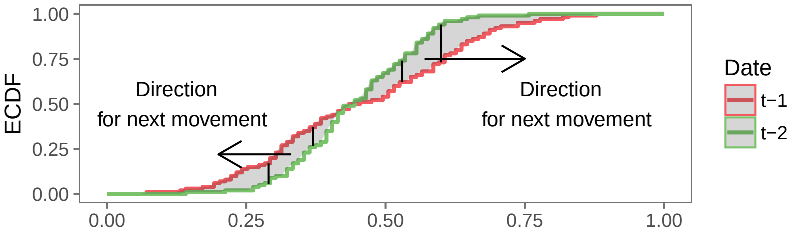
- and replace them **sequentially** by M particles sampled from a Blackwell-MacQueen Polya urn with total mass parameter θ and base measure $P_0(.) = \text{Beta}_{\alpha,\beta}(.)$ conditionally on the remaining $(n-M)$ particles.

$$X_{i_k} \sim \frac{1}{\theta + n - M + k} \left(\theta P_0(.) + \sum_{j \in \text{remaining particles}} \delta_{X_j}(.) \right)$$

This induces a **Dependent Dirichlet process** structure in time.
Markovian process $\rightarrow$ prediction **around the last observed value**.

Adding a locally auto-regressive rule (Trended model)

Inertia phenomenon

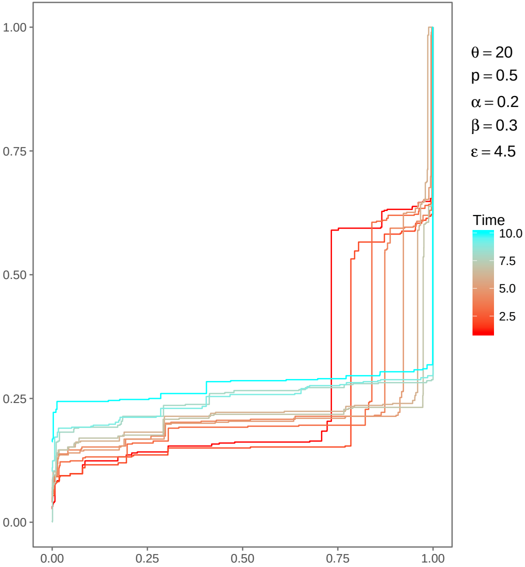


$$X_i(t) = X_i(t-1) - \epsilon_1 \int_{\mathscr{B}(X_i(t-1),h)\cap[0,1]} D_{t-1}(x) - D_{t-2}(x) - \dots$$

where:

$$\mathscr{B}(X,h) = [X - \frac{h}{2}, X + \frac{h}{2}]$$

Illustration of the new model



Likelihood is not available

Likelihood is unavailable:

- Data are fully observed curves because all trading activity is recorded.
- No obvious expression for the likelihood on this functional data on a functional space.

$\rightarrow$ We resort to a likelihood-free method, Approximate Bayesian Computing (ABC).

Another possible (frequentist) option: Functional Data Analysis approach[Canale and Vantini, 2016]

General principle behind ABC

Simplest ABC algorithm

1. Sample some θ_k from prior
2. Simulate $(D_t)_{1 \leq t \leq T} | \theta_k$ and compute S_k
3. Retain all θ_k s.t. $d(S_0, S_k) < \epsilon$ to form the posterior

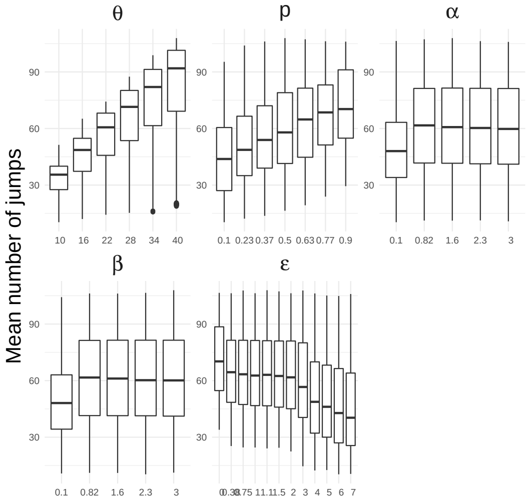
L dimensional summary statistic:

$$S((D_t)_{1 \leq t \leq T}) = (s_1, \ldots, s_L)$$

Distance function:

$$d(.,.) : \mathbb{R}^L \times \mathbb{R}^L \mapsto \mathbb{R}^+$$

Quality of summaries is paramount



Good summaries:

- have small dimension
- reduce sampling variability
- must capture information about the parameters

Sufficient summary statistics are best (no bias on posterior).

Semi-automatic summaries[Fearnhead and Prangle, 2010]

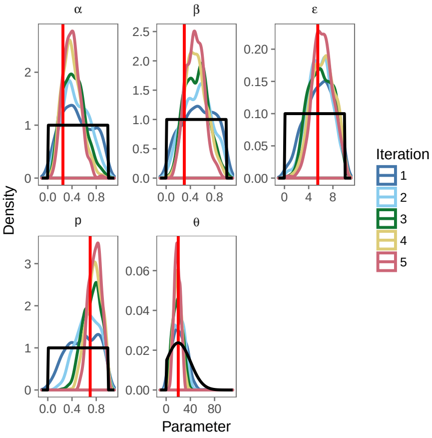
- Finding good summaries is **instructive but difficult.**
- Alternative solution: large number of good or bad summaries (potentially the whole data) and **selection of influential summaries.**

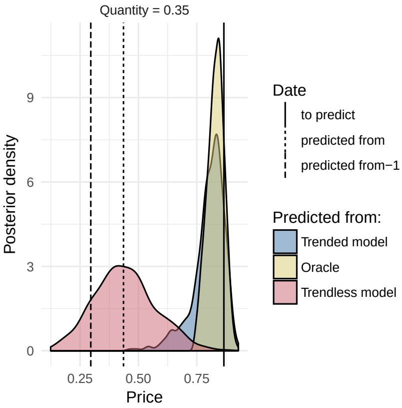
One solution: multivariate regression on a pilot run

$$\theta_k = m(S_k) + \epsilon_k$$

Several Implementations: Partial Least Squares, Lasso, neural networks ...
Accept-reject algorithm: Population ABC[Prangle, 2016]

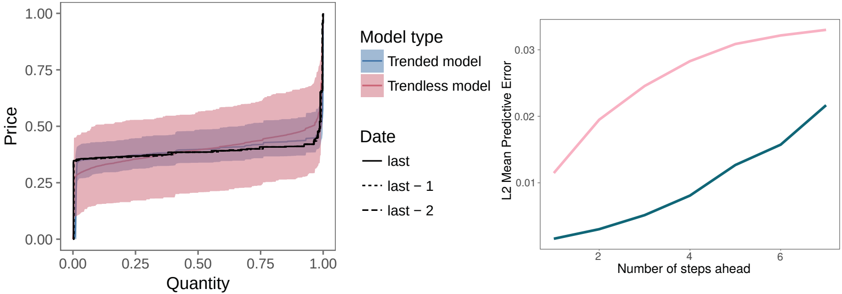
Check model identifiability on simulated data





Prediction of the trendless model around the **last observed value**. Trended model anticipates the movement.

Application to the Italian natural gas market data



- The trended model offers less forecast uncertainty
- Trendless model needs larger p to explain trend
- smaller average predictive error 7 days ahead

Concluding remarks

- Allowing for a trend mechanism can be necessary, else the prediction is stuck around the last value.
- Likelihood-free method $\rightarrow$ general framework for inference
- In particular, interesting developments could be:
  - Time-dependent trend
  - Discrete base measure (mortality data)
  - Two dependent samples

Thanks for your attention !

See also the poster if you want more details !

contact: guillaume.konkamking.work@gmail.com

Bibliography I

Blum, M. G. B., Nunes, M. A., Prangle, D., and Sisson, S. A. (2013).
A Comparative Review of Dimension Reduction Methods in Approximate Bayesian Computation.
Statistical Science, 28(2):189–208.

Canale, A. and Ruggiero, M. (2016).
Bayesian nonparametric forecasting of monotonic functional time series.
Electronic Journal of Statistics, 10(2):3265–3286.

Canale, A. and Vantini, S. (2016).
Constrained functional time series: Applications to the Italian gas market.
International Journal of Forecasting, 32(4):1340–1351.

Bibliography II

- Fearnhead, P. and Prangle, D. (2010).
  Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation.
  Journal of the Royal Statistical Society - Series B: Statistical Methodology, 74(3):419–474.

- Marin, J. M., Pudlo, P., Robert, C. P., and Ryder, R. J. (2012).
  Approximate Bayesian computational methods.
  Statistics and Computing, 22(6):1167–1180.

- Prangle, D. (2016).
  Adapting the ABC distance function.
  Bayesian Analysis, pages 1–21.