Summary of Polson and Sokolov 2018

Deep Learning for Energy Markets

David Prentiss

OR750-004

November 12, 2018

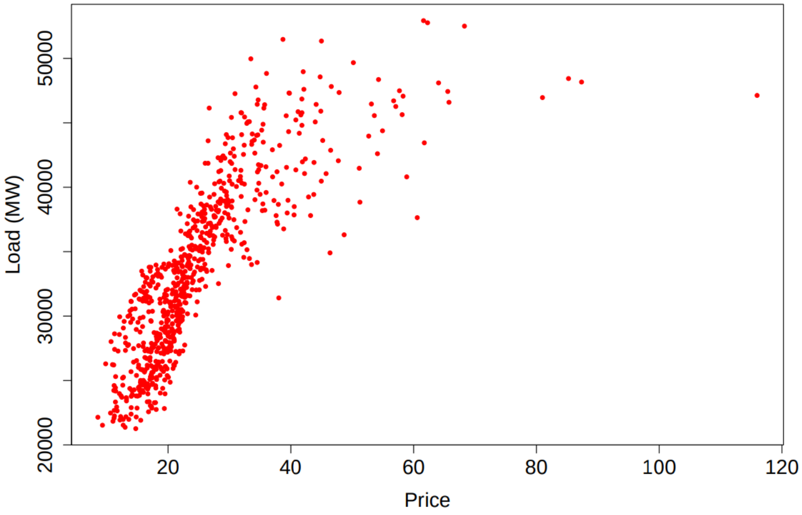
The PJM Interconnection

- The Pennsylvania–New Jersey–Maryland Interconnection (PTO) is a regional transmission organization (RTO).
- It implements a wholesale electricity market for a network of producers and consumers in the Mid-Atlantic.
- It's primary purpose is to prevent outages or otherwise un-met demand.
- Obligations are exchanged in bilateral contracts, the day-ahead market, and the real-time market.

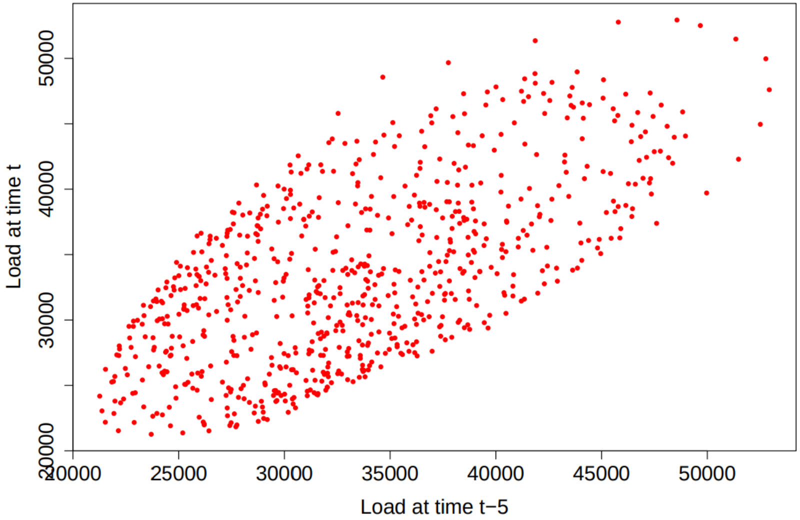
Local marginal price data

- Local Marginal Prices (LMP) are price data aggregated for prices in various locations and interconnection services is the network.
- They reflect the cost of producing and transmitting electricity in the network.
- Prices are non-linear because electricity.
- This paper proposes a NN to model price extremes.

Load vs. price



Load vs. previous load



RNN vs. long short-term memory

Vanilla RNN

$$h_t = \tanh\left(W\begin{pmatrix}h_{t-1}\\ x_t\end{pmatrix}\right)$$

LSTM

$$\begin{pmatrix}i\\ f\\ o\\ k\end{pmatrix} = \begin{pmatrix}\sigma\\ \sigma\\ \sigma\\ \tanh\end{pmatrix} \circ W\begin{pmatrix}h_{t-1}\\ x_t\end{pmatrix}$$

$$c_t = f \odot c_{t-1} + i \odot k$$

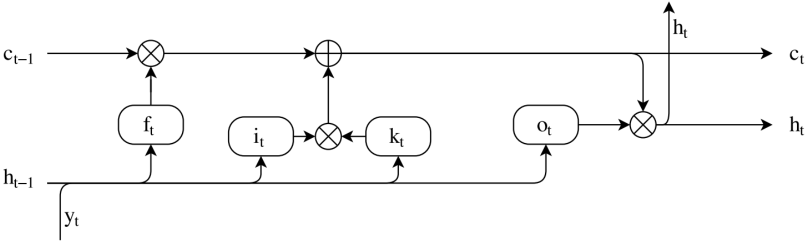
$$h_t = o \odot \tanh\left(c_t\right)$$

LTSM model

$$\begin{pmatrix} i \\ f \\ o \\ k \end{pmatrix} = \begin{pmatrix} \sigma \\ \sigma \\ \sigma \\ \tanh \end{pmatrix} \circ W \begin{pmatrix} h_{t-1} \\ x_t \end{pmatrix}$$

$$c_t = f \odot c_{t-1} + i \odot k$$

$$h_t = o \odot \tanh(c_t)$$



Extreme value theory

- Extreme value analysis begins by filtering the data to select "extreme" values.
- Extreme values are selected by one of two methods.
  - Block maxima: Select the peak values after dividing the series into periods.
  - Peak over threshold: Select values larger than some threshold.
- Peak over threshold used in this paper.

Peak over threshold

- Pickands–Balkema–de Hann (1974 and 1975) theorem characterizes the asymptotic tail distribution of an unknown distribution.
- Distribution of events that exceed a threshold are approximated with the generalized Pareto distribution.
- Low threshold increases bias.
- High threshold increases variance.

Generalized Pareto distribution

- CDF

$$H(y \mid \sigma, \xi) = 1 - \left(1 + \xi \frac{y-u}{\sigma}\right)_+^{\frac{-1}{\xi}}$$

- PDF

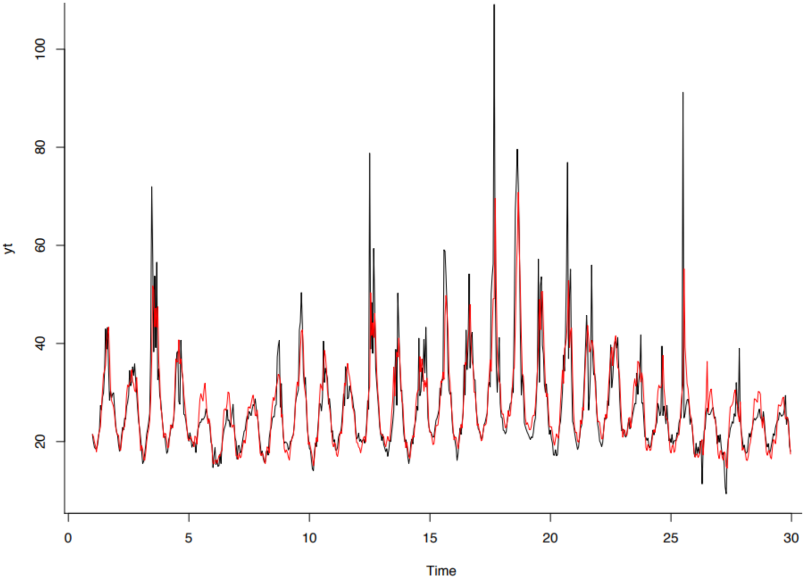
$$h(y \mid \sigma, \xi) = 1 - \frac{1}{\sigma}\left(1 + \xi \frac{y-u}{\sigma}\right)^{\frac{-1}{\xi}-1}$$

Parameters

$$h(y \mid \sigma, \xi) = 1 - \frac{1}{\sigma}\left(1 + \xi\frac{y-u}{\sigma}\right)^{\frac{-1}{\xi}-1}$$

- Location, u, is the threshold
- Scale, σ, is our learned parameter
- Shape, $\xi = f(u, \sigma)$? $\text{EX}[y] = \sigma + u \implies \xi = 0$?

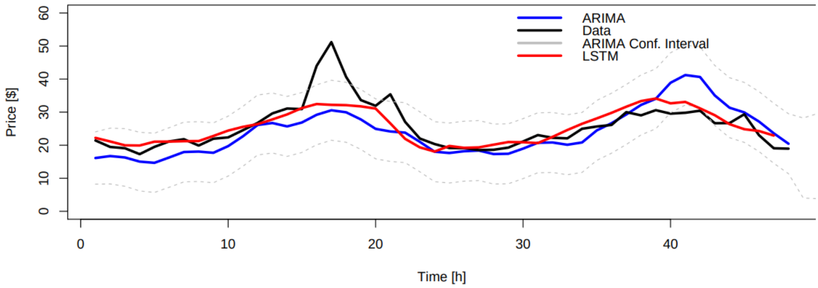
Fourier (ARIMA) model



Fourier (ARIMA) model vs DL

	mse	mrse	mae	mape
Fourier	26.6	5.1	4	0.19
LSTM	16.8	4.1	2.4	0.09

Table 2: Out-of-sample performace of DL and Fourier models



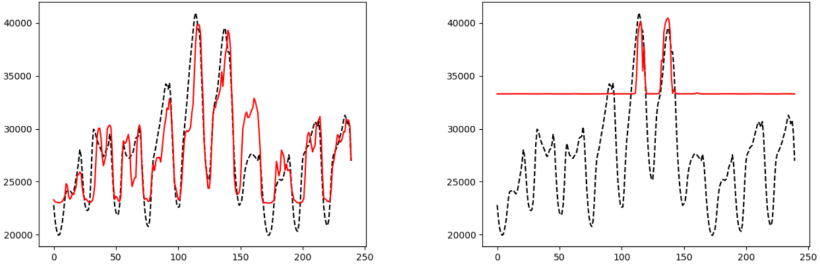
Demand forcasting DL–EVT

- DL–EVT Architecture

$$X \to \tanh\left(W^{(1)}X + b^{(1)}\right) \to Z^{(1)} \to \exp\left(\tanh\left(Z^{(1)}\right)\right) \to \sigma(X)$$

- $W^{(1)} \in \mathbb{R}^{p\times 3}$, $x \in \mathbb{R}^{p}$, $p = 24$ (one day)
- Threshold, $u = 31,000$

Vanilla DL vs. DL-EVT



(a) DL (MSE Loss)

(b) DL-EVT (GP Loss)