

Modeling California Energy Prices Using Markov Chains

Andres Efren Rocha Jayasinha

1 Purpose of the Notebook

This notebook constructs a *discrete-time stochastic model* for electricity prices using observed market data for every single pricing node recorded in the California Independent System Operator's (CAISO) Open Access Same-Time Information System (OASIS) from January to March of 2025. The main goal is to get deeper insight into how pricing behaves in California and how this behavior overall fits into California having some of the highest energy prices in the nation.

The notebook can be found here: CAISO MARKOV

2 The Data Set

The data set consists of 35 millions rows of data containing hourly Locational Marginal Price data from January 1st all the way to March 31st, recording in \$/MW. A snapshot of the compiled data frame after merging all CSVs can be found in Table 1.

INTERVALSTARTTIME_GMT	NODE_ID	MW	OPR_HR	PriceType
2025-01-01 05:00:00+00:00	0096WD_7_N001	50.91579	22	MED
2025-01-01 03:00:00+00:00	0096WD_7_N001	51.23666	20	MED
2025-01-01 07:00:00+00:00	0096WD_7_N001	50.40439	24	MED
2025-01-01 06:00:00+00:00	0096WD_7_N001	50.57005	23	MED
2025-01-01 02:00:00+00:00	0096WD_7_N001	51.19425	19	MED

Table 1: Sample MW data for node 0096WD_7_N001

3 Creating Markov Chains

Rather than modeling prices as continuous values, the notebook partitions the real line into a finite number of *price regimes*. I defined the four intervals as:

$$\begin{aligned}\text{NEG} &:= [-20, 0), \\ \text{LOW} &:= [0, 40), \\ \text{MED} &:= [40, 60), \\ \text{HIGH} &:= [60, 80).\end{aligned}$$

4 Expected State Duration

After constructing transition matrices for each Price Node we can do the following. Suppose the process is currently in state i . At each time step, it:

- remains in i with probability P_{ii} ,
- exits i with probability $1 - P_{ii}$.

Thus, the number of consecutive time steps spent in state i follows a **geometric distribution** with success probability $1 - P_{ii}$.

The expected duration is therefore:

$$\mathbb{E}[\text{duration in } i] = \frac{1}{1 - P_{ii}}.$$

This quantity is computed for each price regime to quantify how long prices on average persist before changing regime.

5 Figures

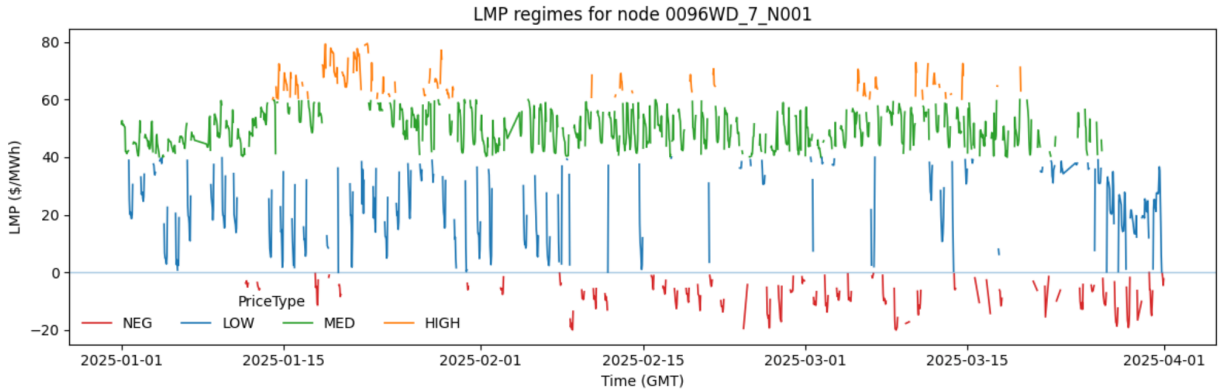


Figure 1: This is a time series for a node with ID 009WD.7_N001 broken up by PriceType. The expected state durations computed were computed to be 5.22, 3.48, **6.42**, 3.24 hours for the Negative, Low, **Medium**, and High price regimes respectively. Notice how compared to the other price categories, the medium price threshold has the longest expected duration and also tends to be less interrupted in the time series.

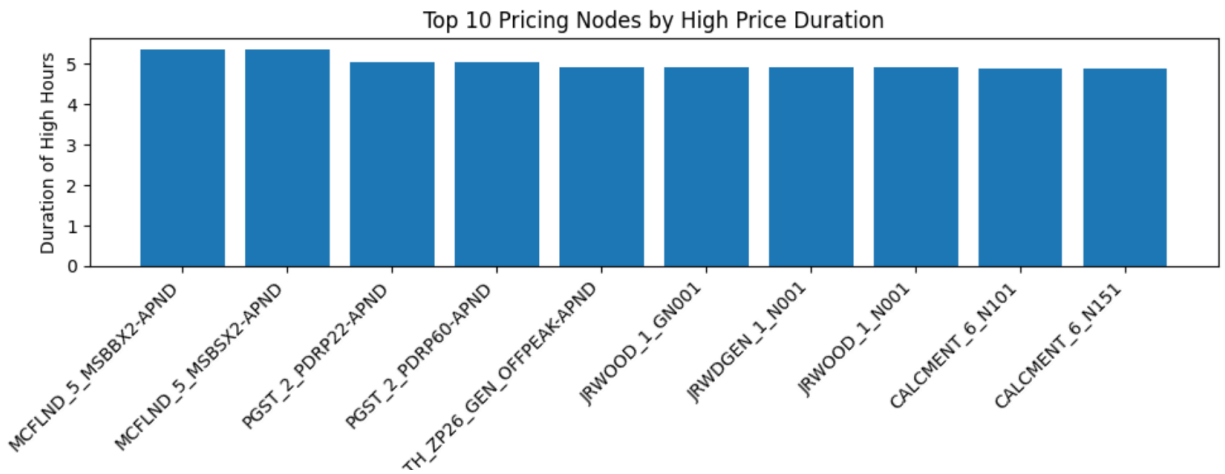


Figure 2: The expected state duration for high energy prices was computed. Pay special attention to the nodes with the tag “MCFLND” denoting McFarland California.

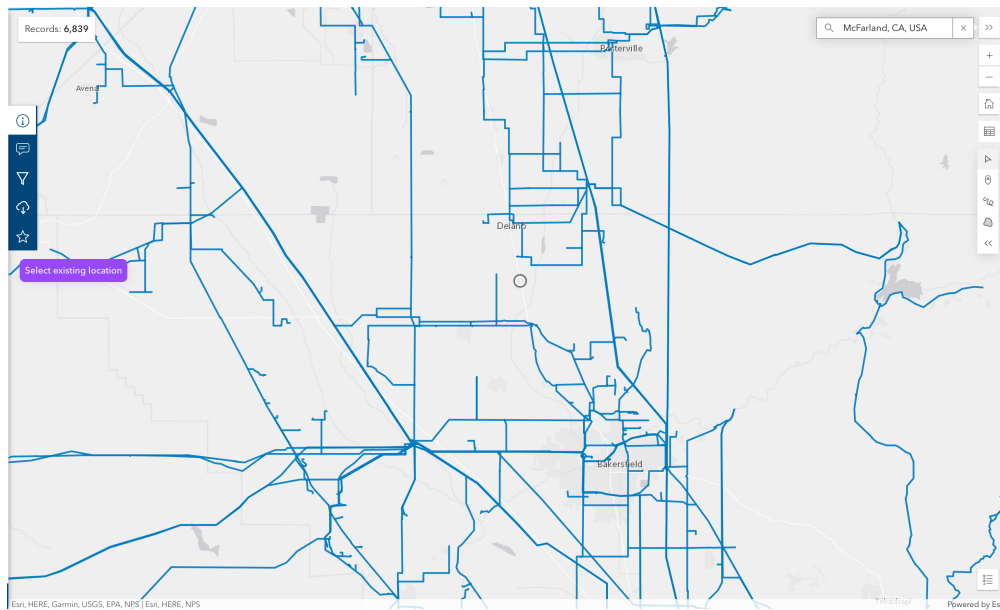


Figure 3: The circle with gray outline is McFarland California and the blue lines represent transmission lines, notice the lack of transmission lines in its vicinity.