# Preliminary Memo

# Data

ERCOT’s [2018 LTSA report](http://www.ercot.com/content/wcm/lists/144927/2018_LTSA_Report.pdf) was central in guiding my LMP predictions, specifically the data presented in its Figure I.1. Figure I.1 in the LSTA was generated by a survey of ERCOT stakeholders. They concluded that the key drivers for the ERCOT grid were Texas economic conditions, natural gas prices, capital costs for renewable energy, environmental regulations, and weather conditions, followed by several less significant drivers. Therefore, I selected natural gas prices, Texas GDP, and renewable energy capital costs as predictors in my model, along with year, load zone, season, and peak versus off peak timing. Table 1 summarizes the data I used in my model and their sources.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table 1: Data Sources | | | | |
| **Name** | **Source** | **Notes** | **Date Range** | **Time Resolution Downloaded** |
| Historical SPP | [ERCOT](http://mis.ercot.com/misapp/GetReports.do?reportTypeId=13061&reportTitle=Historical%20RTM%20Load%20Zone%20and%20Hub%20Prices&showHTMLView=&mimicKey) |  | 2011-2019 | 15 minutes |
| Historical Temperature | [NOAA](https://www.ncdc.noaa.gov/cdo-web/datatools/lcd) | Houston Airport | 2011-2020 | 1 hour |
| Temperature Forecasts | [USGS](http://regclim.coas.oregonstate.edu/visualization/rccv/states-counties/) | Texas-wide Used, Mean Model Used | 2020-2099 | 10 years |
| US Natural Gas Price Forecasts | [EIA](https://www.eia.gov/outlooks/aeo/data/browser/#/?id=13-AEO2018&cases=ref2018&sourcekey=0) | Industrial Used | 2016-2050 | 1 year |
| Historical US Natural Gas Price | [EIA](https://www.eia.gov/dnav/ng/hist/n3035us3A.htm) | Monthly available, Industrial Used | 1997-2019 | 1 year |
| Historical and Forecasted Peak and Annual Demand | [ERCOT](http://www.ercot.com/content/wcm/lists/114580/2017_Long-Term_Hourly_Peak_Demand_and_Energy_Forecast.pdf) | Extrapolate for 2027-2039, not yet included in model | 2007-2026 | 1 year |
| Historical Texas GDP | [US Bureau of Economic Development](https://fred.stlouisfed.org/series/TXNGSP) |  | 1997-2019 | 1 year |
| Texas GDP Forecasts | [ERCOT](http://www.ercot.com/content/wcm/lists/144927/2018_LTSA_Report.pdf) | Growth rate applied to 2019 GDP | 1.4% growth (normal), 2.2% (high) | 1 year |
| Capacity-Weighted Average Renewable Costs | [EIA](https://www.eia.gov/electricity/generatorcosts/) | Extrapolate exponentially for years outside range, past and future | 2013-2018 | 1 year |

# Scenarios

ERCOT’s [2018 LTSA report](http://www.ercot.com/content/wcm/lists/144927/2018_LTSA_Report.pdf) also provided guidance on scenarios likely to impact the ERCOT grid and thus future prices. Two of the scenarios it discusses are (1) High Economic Growth and (2) High Renewable Penetration. My third scenario was (3) the Base Case. Forecasts for the next 20 years for all my predictors were available with the exception of renewables capital costs, which I extrapolated by fitting an exponential model to the available data.

(1) the Base Case was predicted using the available forecasts data from Table 1. For (2) the High Economic Growth scenario, EIA provides its own predictions of natural gas prices under high economic growth, and ERCOT describes an annual Texas GDP increase of 2.2% rather than the 1.4% in the base case. Again, future LMPs were predicted using forecasts as predictors, this time with 2.2% GDP growth and High Economic Growth conditions gas prices in place of the base case numbers. For (3) the High Renewable Penetration, it was assumed that this adoption was driven by lower renewable energy costs, so forecasted renewable energy capital costs were multiplied by a discount factor (assumed to be 0.9). Future LMPs were once more predicted using the forecast data from Table 1, with the substitution of the discounted renewable energy capital cost forecasts.

# Aggregation

Historical LMPs were aggregated by season (“summer” as April through October and “non-summer” as the rest of the year) and peak (hours 8 though 23 in the summer and 1 and 24 in the non-summer) and off-peak the rest of the time. This aggregation was done both because the financial models that will use the LMP predictions as inputs will likely not need a finer temporal resolution and because the forecast data available was generally only at 1 year temporal resolution, making finer prediction difficult and likely inaccurate. Historical LMPs in each of the time-categories (summer peak, summer off-peak, non-summer peak, and non-summer off-peak) were averaged by median value in that year, though other functions such as mean or max could be computed, depending on the requirements of downstream models. Therefore, my model predicts the median LMP in each load zone, each year from 2020-2039, in each of the four time-categories.

# Model

I used linear regression to fit a model to LMPs from 2011 to 2019. I then applied this model to 2020 to 2039 to predict future LMPs. My model is of the following form:

Since Category and LoadZone are both categorical variables, they were each broken up into N-1 dummy variables, where N is the number of unique values in that category. The resulting model is summarized below.

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 3.103e+03 3.747e+02 8.281 1.20e-15 \*\*\*

Year -1.550e+00 1.866e-01 -8.308 9.79e-16 \*\*\*

categorynon-summer peak -2.368e+00 1.632e-01 -14.510 < 2e-16 \*\*\*

categorysummer off-peak -1.913e+00 1.632e-01 -11.723 < 2e-16 \*\*\*

categorysummer peak 5.843e+00 1.632e-01 35.808 < 2e-16 \*\*\*

ZoneHB\_HOUSTON 5.061e-01 3.065e-01 1.651 0.099325 .

ZoneHB\_HUBAVG -5.944e-02 3.065e-01 -0.194 0.846296

ZoneHB\_NORTH 4.819e-02 3.065e-01 0.157 0.875117

ZoneHB\_PAN -2.310e+00 6.970e-01 -3.314 0.000988 \*\*\*

ZoneHB\_SOUTH 2.040e-01 3.065e-01 0.666 0.505928

ZoneHB\_WEST -4.525e-01 3.065e-01 -1.476 0.140490

ZoneLZ\_AEN 2.593e-01 3.065e-01 0.846 0.397947

ZoneLZ\_CPS 4.303e-01 3.065e-01 1.404 0.160996

ZoneLZ\_HOUSTON 5.972e-01 3.065e-01 1.949 0.051923 .

ZoneLZ\_LCRA 2.594e-01 3.065e-01 0.846 0.397694

ZoneLZ\_NORTH 1.503e-01 3.065e-01 0.490 0.624133

ZoneLZ\_RAYBN 2.969e-01 3.065e-01 0.969 0.333104

ZoneLZ\_SOUTH 3.749e-01 3.065e-01 1.223 0.221898

ZoneLZ\_WEST 5.083e-01 3.065e-01 1.659 0.097853 .

NG\_Price 4.700e+00 1.332e-01 35.297 < 2e-16 \*\*\*

GDP 1.360e-05 2.199e-06 6.187 1.31e-09 \*\*\*

Solar\_PV\_Cost -1.857e-03 3.694e-04 -5.028 6.99e-07 \*\*\*

Onshore\_Wind\_Cost 3.491e-03 1.456e-03 2.397 0.016901 \*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

# Future Work

Future work for my LMP prediction model includes additional data gathering, fine-turning aggregation, and backtesting. First, more data describing the ERCOT grid could be included such as planned projects and capacity/ transmission expansion, reserve margins, annual and peak demands. Second, while I did download temperature data, both historical and forecasted, it was not included in this model because I struggled to find temperature forecasts; The temperature forecasts in Table 1 are at a very coarse time resolution. I would also like to explore finer temporal aggregation, if it could be of use to downstream models, as well as different aggregation functions than “median.” Finally, I would like to assess the accuracy of my model by backtesting it on past years and comparing model predictions to actual LMPs.

# Code

My code and some of my data are available on GitHub at <https://github.com/rmoglen/EnviroDev_Policy>.