

List of Data Sources

“Investment versus Output Subsidies: Implications of Alternative Incentives for Wind Energy”

This document provides instructions for obtaining the data used in Aldy, Geraden and Sweeney (2022). Additional information on the contents of these data and how they are cleaned is provided in the text and appendix of the paper.

The paper relies on both publically available and proprietary data sources, which are discussed in turn. Although we are not permitted to post the raw proprietary data, intermediate datasets, sufficient for reproducing all of the tables and figures in the text, are provided on [GitHub](#).

Public Data Sources

Energy Information Administration (EIA) Data:

860, 923, and 861M data-sets below are used in “eia_prep.do”

EIA 860 –collects generator-level specific information about existing and planned generators and associated environmental equipment at electric power plants with 1 megawatt or greater of combined nameplate capacity

URL: <https://www.eia.gov/electricity/data/eia860/>

EIA 923 - collects detailed electric power data -- monthly and annually -- on electricity generation, fuel consumption, fossil fuel stocks, and receipts at the power plant and prime mover level. Specific survey information provided

URL: <https://www.eia.gov/electricity/data/eia923/>

EIA 861M (formerly 826) - collects sales of electricity and associated revenue, each month, from a statistically chosen sample of electric utilities in the United States.

URL: <https://www.eia.gov/electricity/data/eia861m/index.html>

EIA 861- include information such as peak load, generation, electric purchases, sales, revenues, customer counts and demand-side management programs, green pricing and net metering programs, and distributed generation capacity. We use these data to obtain retail sales of electricity in the industrial sector.

Used in “rec_prep.do”

URL: <https://www.eia.gov/electricity/data/eia861/>

Powercurves from PelaFlow Consulting

Used in “prep_power_curve.do”

Collection of power curves that are provided for a wide variety of turbines. It's updated every few months, but our code uses the August 2015 version. This is one of three sources for power curve.

URL: <http://www.wind-power-program.com/>

Department of Energy (DOE) REC-Tracking:

REC-Tracking-data-viewer.xlsx - Contains information for how RECS are claimed by other States in order to meet RPS requirements. In particular, it gives REC shares for all pairwise combination of states.

Used in "out_of_state_rec.do"

URL: <https://energy.gov/eere/analysis/downloads/cross-state-renewable-portfolio-standard-compliance>

Lawrence Berkeley National Labs (LBL) RPS Compliance:

RPS Compliance Data.xlsx - Spreadsheet that tracks annual RPS obligations for states across the U.S.

Used in "rec_prep.do"

URL: <https://emp.lbl.gov/projects/renewables-portfolio>

Google Cost Searches

We searched for press releases and regulatory filings for PTC claimants that entered between 2009 and 2012. The cost figures obtained and their sources are provided in USplants_Costs_GoogleSearch.csv. These are used in "clean_bnef_cost_data.do"

Holland, Mansur, Muller, and Yates (AER, 2016)

Replication data available at: <https://www.aeaweb.org/articles?id=10.1257/aer.20150897>

Locational Marginal Pricing Data

We gathered locational marginal pricing for each of the major regional transmission organizations across the United States. For all markets, we used real time LMP pricing as opposed to day ahead.

These data are publically available, but the raw data files obtained are large (~500 GB). In lieu of posting these on dataverse, we have posted minimally processed Stata .dta files which read this raw data in and save separate datasets for each ISO-year.

For those interested in obtaining the raw data:

California ISO (CAISO) – Scraped the "PRC_INTVL_LMP" data from California OASIS API. This API request provides LMP prices every 5 minutes. NOTE: The script takes weeks to run as the API only allows you to pull one hour at a time.

Used in “caiso_imp_clean.do”

URL: <http://oasis.caiso.com/mrioasis/logon.do>

FullNetworkModel_NodeMapping- This excel file was downloaded from the URL below (Select the sheet that says “Full Network Model Pricing Node Mapping (based on Full Network Model Release DBYYYYQX”). It provided a way to link LMP raw data with node type as we were interested only in generator nodes across the data

Used in “match_nodes_to_lat_long.do”

URL: <http://www.caiso.com/market/Pages/NetworkandResourceModeling/Default.aspx>

Electric Reliability Council of Texas (ERCOT) – Requested “LMPs by Resource Nodes, Load Zones and Trading Hubs” from ERCOT website. These data go back to December 2010

Used in “ercot_imp_clean.do”

URL: <http://www.ercot.com/about/contact/inforequest>

MISO – Provides Real-time LMP prices. CSVs were downloaded from the MISO website.

Used in “miso_imp_clean.do”

URL: <https://www.misoenergy.org/markets-and-operations/real-time--market-data/market-reports/#t=10&p=0&s=MarketReportPublished&sd=desc>

New England ISO (NEISO) - Downloaded Final Real Time Hourly LMPs from NEISO website.

Used in “neiso_imp_clean.do”

URL: <https://www.iso-ne.com/isoexpress/web/reports/pricing/-/tree/lmps-rt-hourly-final>

New York ISO (NYISO) – Used a python script to scrape data from NYISO website. Scraped both zonal and generator nodes but only used generator node pricing in final analysis. The data are provided at 5 minute increments.

Used in “nyiso_imp_clean.do”

URL: <http://mis.nyiso.com/public/P-24Alist.htm>

PJM Interconnection – Used a python script to scrape data from PJM website. The data were provided at the hourly level in PJM.

Used in “pjm_imp_clean.do”

URL: <http://www.pjm.com/markets-and-operations/energy/real-time/lmp.aspx>

Proprietary Data Sources

American Wind Energy Association (AWEA):

AWEA makes several databases available to its members (for an annual fee). The Project database includes more than 60 fields of data including: project location, owner and developer, project size, turbine model and manufacturer, project offtake, power purchase agreement data, hub heights, Congressional districts, and more.

The Turbine database allows users to query for turbine-level data including turbine model, hub height, and the latitude and longitude coordinates.

Both data-sets are used in “awea_prep.do”

URL: <https://cleanpower.org/cleanpower-iq/>

Bloomberg New Energy Finance (BNEF)

Used in “clean_bnef_cost_data.do”

Proprietary information on plant investment costs and PPAs. Once a subscription to BNEF is obtained, you need to download three datasets:

- “2020-02-03 - U.S. Power Plant Stack Raw Data and User Guide.xlsx”
- “2018-12-19 - U.S. Renewable PPA Prices Hit Record Lows in 2018 “
- All of the wind farms in the United States in the “renewable assets” tool

Powercurves

Both of these data-sets are processed in “prep_power_curve.do”

Power curves for different wind turbines provided by Joern Hunteler. Available from the authors upon request.

SNL

SNL Energy provided us with location information on which ISO each wind farm operates in and location information (latitude and longitude) for every ISO node. These data are available to SNL Energy subscribers.

Used in “snl_prep.do” and “match_nodes_to_lat_long.do”

URL: <https://marketintelligence.spglobal.com/client-solutions/sectors/energy>

We also use cost data from the SNL “Power Plant Screener” tool in “clean_bnef_cost_data.do”. This data can be downloaded by selecting all US power plant units, selecting all fields, and exporting to excel.

RECS Market Data

Market Data Price Report - Contains bid and offer prices for various different REC/energy programs. It was obtained by purchasing the “full REC database” from Marex Spectron by contacting them via email.

Used in “rec_prep.do”

URL: <http://www.marexspectron.com/>

1603 Grant Information from Treasury:

Used in “treasury_1603_prep.do”

Data on the size and award amount of each 1603 grant recipient provided to the authors by the US Treasury Department.

3Tier Wind Data

Hourly wind speed estimates for all US wind farms were provided to us by Gabe Chan and Joern Huenteler, who obtained them from [3Tier](#), a commercial wind vendor. We ran each hourly estimate through the wind farms’ power curve to get an estimate of output, and then summed these at the year-month level. We have provided these monthly summaries, which are sufficient to replicate all results in the paper, since the EIA output data is monthly. Those who have access to 3Tier can obtain the raw data by entering the latitude and longitude for each plant from EIA into 3Tiers Global Wind Prospecting Tool.