README for replication archive:

"Market Power in Coal Shipping and Implications for U.S. Climate Policy"

Louis Preonas

HOW TO USE THIS ARCHIVE: CODE

- The folder "code" contains all Stata and R files required to fully replicate all tables and figures in the main text and appendices, starting from the original raw datasets.

- The file "code/MASTER.do" linearizes all code into six modules:
    - 1. Build [B]
    - 2. Merge [M]
    - 3. Demand Estimation [D]
    - 4. Analysis [A]
    - 5. Counterfactuals [C]
    - 6. Outputs [O]

- To do a full replication, the modules must be run sequentially. The individual scripts within each module must also be run sequentially.

- In "code/MASTER.do", each script has a list of direct dependencies. These are the upstream scripts that last updated each dataset called by that script. Indirect dependencies may also exist upstream of any direct dependency.

- The vast majority of code runs in Stata. The script "code/paths.do" lets users define their local directory, and also defines all global macros needed across all Stata scripts to define folder paths within the replication archive. Each script calls "code/paths.do" at the top.

- The script "code/packages.do" installs all Stata packages necessary to execute a full replication.

- It would be extremely impractical to actually execute "code/MASTER.do", since individual programs take many hours to run. I strongly recommend using this master do file only to reference which script to run next.

- Within, "code/MASTER.do", I list any auxiliary scripts required to execute each main script. Most auxiliary scripts use R, and are not fully integrated into the main Stata script. Each main script situates any auxiliary R scripts linearly, indicating when/how to execute them (e.g., run Steps 1-3 in Stata, then as Step 4 open up the auxiliary script in R and run, then proceed to run Steps 5-7 in Stata).

- R scripts call "code/path.R" at the top, which lets users define their local directory.

- R scripts load required packages at the top, alerting replicators on which packages they need to install to run each script. (For individual scripts, I've erred on the side of loading too many packages.)

DATA AVAILABILITY AND PROVENANCE STATEMENT

- The raw data used in my analysis are almost all available for download online for free.

- The only exception are proprietary data from S&P Global Market Intelligence (formerly SNL), which require a paid subscription. I have omitted these raw datasets from the archive. *(See notes and download instructions under Scripts B22 and B23 below.)*

- I have also omitted two raw datasets that are exceedingly large:

    - Quarterly Census of Employment and Wages (CEW) data, which are downloadable from the Bureau of Labor Statistics website. *(See note and download instructions under Script B6 below.)*

    - EPA Continuous Emission Monitoring Systems (CEMS) data, which are downloadable from the EPA's website. *(See notes and download instructions under Scripts B16 and B17 below.)*

- All other raw datasets are included in the folder "data". I provide links and instructions for accessing these original datasets online (where still available), in my Online Appendix G and within their respective build files.

HOW TO USE THIS ARCHIVE: DATA

- The folder "data" contains all raw datasets necessary to complete a full replication **starting at Script B1**, except:

    - Quarterly Census of Employment and Wages (CEW) data. *(See note and download instructions under Script B6 below.)*

    - EPA Continuous Emission Monitoring Systems (CEMS) data. *(See notes and download instructions under Scripts B16 and B17 below.)*

    - Data from S&P Global Market Intelligence (formerly SNL. *(See notes and download instructions under Scripts B22 and B23 below.)*

- For users seeking to replicate my demand estimation procedure (i.e., Figure 5, Appendix A, Appendix B.4), the folder "data" includes all merged datasets necessary to complete a partial replication **starting at Script D4 thru Script D10, jumping to Script A3, and then jumping to Module O**.

- For users seeking to replicate my main regression analysis and pass-through estimates (i.e., Tables 2–3, Figures 6–8, Appendix C.2, Appendix E), the folder "data" includes all merged datasets necessary to complete a partial replication **starting at Script A4 and skipping Module C**.

- For users seeking to replicate my counterfactual analysis (i.e., Section 7.3, Appendix D.2), the folder "data" includes all merged datasets necessary to complete a partial replication of **Script C4 thru Script C7**.

- Finally, the folder "data" includes all datasets that are direct inputs into any summary tables or figures not covered by the above three "partial" replications (i.e., Figures 1–4, Table 1, Appendix D.1, Appendix F, Appendix G).

- This may go without saying: any replication that is upstream of a partial replication will overwrite merged datasets that are already included in the replication archive.

- While some build and merge files include data prior to 2002 and after 2015 (e.g. 2016 CEMS data), my analysis only uses a sample period of 2002–2015.

CONFIGURATION FOR EXECUTING CODE

- Replication requires only a Stata MP license, and a basic installation of R and RStudio.

- Stata scripts should be forward-compatible with the most recent version of Stata, with the following possible exceptions:

    - New versions of Stata may have slightly altered the syntax of commands that read in raw data files from .csv/.xls/.xlsx formats.

    - The syntax within reghdfe may have changed for saving fixed effects, saving residuals, and/or constructing predicted values.

- R scripts may become finnicky as R and RStudio evolves. Running R scripts using Version 3.4.3 (https://cran.r-project.org/bin/windows/base/old/) should avoid compatibility issues.

- Each of the steps described below should fully execute in less than 2–4 hours, except:

- Steps B6, B7, B8, B16, and B17, which each may take up to 12–24 hours.

- Steps M8 and M9, which involve long R sub-routines that may last 4–8 hours.

- Steps D4, D5, and D7, which iterate over coal units/plants and over demand estimation sensitivities. I recommend running these files across 20–30 concurrent Stata instances (to the extent that computing resources allow), where each instance loops over a different subset of plants in parallel (e.g., looping over units 1–50 in the first instance, units 51–100 in the second instance, etc.). Even with 20–30 Stata instances, these steps may take up to 2–4 days each.

- Step A6, which includes all bootstrapped regressions. Each regression takes around 3 hours to fully execute, but each outer loop can be run in parallel. Multiple concurrent Stata instances are highly recommended, especially for the sensitivities in Step A6d. If run concurrently, all bootstrapped regressions should execute in under 7 days.

BUILD FILES (Module B)

B1. "code/build/build_eia860.do" : builds EIA 860 datasets of electric generator/plant characteristics; raw files live in "data/eia860/raw"; script deposits built datasets in "data/eia860/dta_panel"; the auxiliary file "code/build/extract_dbf_eia860.R" converts some pesky .dbf files to a usable .csv format

B2. "code/build/build_msha.do" : builds MSHA datasets of coal mine production and characteristics; raw files live in "data/msha/raw"; script deposits built datasets in "data/msha/dta_panel"

B3. "code/build/build_eia7a.do" : builds EIA 7A datasets of coal mine production and ownership; raw files live in "data/eia7a/raw_csv"; script deposits built datasets in "data/eia7a"

B4. "code/build/build_eia423.do" : builds EIA 423 datasets of coal transaction data; raw files live in "data/eia423/raw"; script deposits built datasets in "data/eia423"

B5. "code/build/build_ctrd.do" : builds EIA coal transportation rate data; raw files live in "data/ctrd/raw"; script deposits built datasets in "data/ctrd"

B6. "code/build/build_cew.do" : builds BLS CEW data; download raw Quarterly Single File CSVs at https://www.bls.gov/cew/downloadable-data-files.htm, and put the zipped folders in "data/cew/raw_csvs/qcew"; script deposits built datasets in "data/cew/dta_panel"

Downloading these CEW data should be self-explanatory. There is a single (massive) file for each year, named "YYYY.q1-q4.singlefile.csv".

B7. "code/build/build_stb_waybill.do" : builds STB Public Use Waybill Sample data; raw files live in "data/cew/raw_csvs/qcew"; script deposits built datasets in "data/cew/dta_panel"; the auxiliary file "code/build/extract_stb_raw.R" converts the raw files into a usable .csv format

B8. "code/build/build_usgs.do" : builds USGS stratigraphic coal data; raw files live in "data/usgs/raw_coalqual" and "data/usgs/raw_ustrat"; script deposits built dataset in "data/usgs" and "data/gis"

B9. "code/build/build_eia906.do" : builds EIA 759/906/920/923 datasets of electric generator operations data; raw files live in "data/eia423/raw"; script deposits built datasets in "data/eia423"

B10. "code/build/build_eia767.do" : builds EIA 767/860/923 datasets of pollution control and boiler operations; raw files live in "data/eia767/raw", "data/eia860/raw", and "data/eia423/raw"; script deposits built datasets in "data/eia767"

B11. "code/build/build_eia767_unit.do" : builds crosswalk between generator ID and boiler ID; input files live in "data/eia767/raw", "data/eia423/raw" , and "data/eia860/dta_panel"; script deposits built datasets in "data/eia767"

B12. "code/build/build_epa_egrid.do" : builds EPA eGRID datasets of pollution control mechanisms and emissions; raw files live in "data/epa_egrid/raw"; script deposits built datasets in "data/epa_egrid"

B13. "code/build/build_epa_ampd.do" : builds EPA allowance program participation dataset; raw files live in "data/epa_ampd/raw"; script deposits built datasets in "data/epa_ampd"

B14. "code/build/build_ferc714.do" : builds list of plant from FERC Form 714 to assign plants to balancing authorities; raw files live in "data/ferc714/raw 2006-2015"; script deposits built dataset in "data/ferc714"

B15. "code/build/build_ferc1.do" : build FERC Form 1 datasets on plant operations and stated variable costs; raw files live in "data/ferc1/raw csvs"; script deposits built datasets in "data/ferc1"

B16. "code/build/build_cems.do" : build CEMS gross generation datasets at the daily, monthly, and annual levels; download unit-level daily emissions files https://campd.epa.gov/data/custom-data-download for all states for all quarters from 2000 to 2016, and put the zipped files in "data/cems/zipped_raw"; script deposits built datasets in "data/cems"

> *I downloaded these daily CEMS data from the EPA's FTP server, which may no longer be operational? Look for a bulk download option, which may be through an API. There is a separate csv at the unit-day level for each year-state-quarter, named "DLY_2002caQ1.csv" (i.e. for all CEMS units in California, for all days in Q3 of 2002).*

B17. "code/build/build_cems_hourly.do" : build CEMS gross generation datasets at the hourly level; download unit-level hourly emissions files https://campd.epa.gov/data/custom-data-download for all states for all months from 2000 to 2016, and put the zipped files in "data/cems/zipped_raw_hrly"; script deposits built datasets in "data/cems/hourly_dtas"

> *I also downloaded these hourly CEMS data from the EPA's FTP server. Look for a bulk download option, which may be through an API. There is a separate csv at the unit-hour level for each year-state-month, named "2002ca04.csv" (i.e. for all CEMS units in California, for all hours April 2002).*

B18. "code/build/build_eia_acr.do" : build EIA Annual Coal Report datasets of mine-mouth coal prices by state-year; raw files live in "data/eia_acr/raw_xlsx"; script deposits built dataset in "data/eia_acr"

B19. "code/build/build_eia_epm.do" : build EIA Electric Power Monthly datasets of delivered coal prices by state-month; raw files live in "data/eia_epm/epm xls tables"; script deposits built datasets in "data/eia_epm"

B20. "code/build/build_usgov_gis.do" : build various US government GIS datasets; raw files live in "data/usgov_gis/raw"; script deposits built datasets in "data/usgov_gis"

B21. "code/build/build_eia_ng.do" : build EIA natural gas price datasets; raw files live in "data/eia_ng/raw"; script deposits built datasets in "data/eia_ng"

B22. "code/build/build_snl_supply.do" : build SNL supply curve datasets (including assumed non-fuel variable costs); raw files are downloadable from S&P Global Market Intelligence (https://www.spglobal.com/marketintelligence, subscription required, search "Power > Generation Supply Curve"); place raw .xls downloads in "data/snl_supply/raw xls"; script deposits built datasets in "data/snl_supply"

> *I downloaded these "SNL Generation Supply Curves" at the region level. There is a separate document for each year in {2009, … ,2015}. I downloaded a separate file for each Balancing Authority (68), ISO (7), and NERC Region (14). I only downloaded "Summer Capacity", and excluded allowance costs. This yielded 534 year-region combinations, 494 of which made it into the built dataset in Step 2 of script B22. These proprietary data get used downstream in two ways. First, I use the SNL BA names (variable name "snl_ba") as a cross-check on my preferred BA/PCA identifier; this alternate identifier does not make it into the final analysis anywhere, but it persists in the code. Second, I use SNL's technology-specific assumptions for non-fuel variable costs per MWh (a component of these supply curve data) in a sensitivity analysis for my preferred demand estimation (see Appendix Figures A6 and E10).*

B23. "code/build/build_snl_prices.do" : build datasets of hub-specific natural gas prices, coal prices, and allowance prices; the core price time series data are downloadable from S&P Global Market Intelligence (https://www.spglobal.com/marketintelligence, subscription required, search "Market Prices > Advanced Search > Commodity"); place raw .xls downloads in "data/snl_prices/raw xls" and "data/snl_prices/raw xls future"; supplementary raw data are in "data/other_data"; script deposits built datasets in "data/snl_prices"

> *I downloaded five sets of SNL price time series: (i) daily prices at natural gas trading hubs; (ii) monthly prices at natural gas trading hub; (iii) monthly allowance prices; (iv)monthly coal prices; (v) Henry Hub futures prices. The file "data/snl_prices/dta build/snl_prices_index.dta" indexes all price time series used (i)–(iv), which I batch downloaded to "data/snl_prices/raw xls". Prices (i)–(iii) are inputs into my demand estimation, which leverages daily variation in natural gas prices, spatial variation across gas trading hubs, and monthly variation in SO_2/NO_x/CO_2 allowance prices. For (v), I downloaded monthly and yearly Henry Hub futures prices as of {1/1/2007, 1/1/2008, 1/1/2009, 1/1/2010}, plus monthly Henry Hub futures prices as of 12/1/2008; these 9 files lived in "data/snl_prices/raw xls future". The as-of-12/1/2008 prices are a key input for my counterfactual analysis.*

B24. "code/build/build_rcafs.do" : build rail cost adjustment factor (RCAF) and AAR fuel price index; raw files live in "data/rcaf" subfolders; script deposits built datasets in "data/rcaf"

MERGE FILES (Module M)

M1. "code/merge/merge_coal_mines.do" : merges together coal production datasets; Step 1 uses the auxiliary file "code/merge/gis_mine_coord.R" to validate mine coordinates; Step 2 uses the auxiliary file "code/merge/gis_raster_coal.R" to spatially interpolate seam depth and thickness; script deposits merged datasets in "data/merged datasets"

M2. "code/merge/merge_plants.do" : merges together plant-level datasets; Step 4 uses the auxiliary file "code/merge/gis_facil_coord.R" to validate plant coordinates; script deposits merged datasets in "data/merged datasets"

M3. "code/merge/merge_fuel_trans.do" : merges together coal shipping datasets; script deposits merged datasets in "data/merged datasets"

M4. "code/merge/merge_plants_units.do" : merges together unit-level datasets; script deposits merged datasets in "data/merged datasets"

M5. "code/merge/merge_cems_missing_gload.do" : corrects missings in gross generation in CEMS datasets; script updates built datasets in "data/cems", and updates merged datasets in "data/merged datasets"

M6. "code/merge/merge_cems_agg_load.do" : aggregates hourly/daily CEMS load across regions; script deposits aggregate datasets in "data/merged datasets"

M7. "code/merge/merge_cems_unit_mc.do" : constructs daily marginal costs for all CEMS units; script deposits datasets in "data/merged datasets"

M8. "code/merge/merge_coal_routes.do" : constructs coal-route-specific datasets; Step 3 uses the auxiliary file "code/merge/gis_rail_distance.R" to execute my rail graph algorithm; Step 12 uses the auxiliary file "code/merge/gis_water_distance.R" to calculate plants' distance to navigable waterways; script deposits datasets in "data/usgov_gis/rail_graph" and "data/merged datasets"

M9. "code/merge/merge_temperature.do" : constructs daily maximum temperature at each plant; Step 1 uses the auxiliary file "code/merge/gis_prism_daily_max.R" to download and process PRISM temperature data; script deposits dataset in "data/merged datasets"

DEMAND ESTIMATION FILES (Module D)

D1. "code/demand/dest_prep_aux_datasets.do" : prepares auxiliary files used in demand estimation; script deposits datasets in "data/demand estimation"

D2. "code/demand/dest_coal_units_hourly_cems.do" : constructs unit-specific CEMS time series; script deposits datasets in "data/demand estimation/coal units hourly gload" and "data/demand estimation/coal units daily gload"

D3. "code/demand/dest_gen_files_plant.do" : constructs plant-specific files that serve as inputs to the generation regressions in the next step; script deposits datasets in "data/demand estimation/gen_reg_input_plant"

D4. "code/demand/dest_gen_regs.do" : estimates dispatch model (a.k.a. generation regressions), separately for each coal unit time series; designed to be iterated over alternate specifications by changing local macros at the top of the script *before* proceeding to the next steps; script deposits datasets in "data/demand estimation/ptildes"

D5. "code/demand/dest_demand_curves_units.do" : converts distributions of P_tilde into unit-month demand curves; script deposits datasets in "data/demand estimation/demand_unit"

D6. "code/demand/dest_demand_curves_plants.do" : aggregates unit-month demand curves up to the plant, yielding plant-month demand curves; script deposits datasets in "data/demand estimation/demand_plant"

D7. "code/demand/dest_demand_curves_smoothing.do" : smooths plant-month demand curves and estimates lambda parameters; script deposits estimated demand curves in "data/demand estimation/demand_plant_smoothed", and deposits lambda parameter estimates in "data/demand estimation/lambda_estimates"

D8. "code/demand/dest_prep_lambdas.do" : consolidates lambda estimates and runs some diagnostics; script deposits finalized lambda estimates in "data/analysis"

D9. "code/demand/dest_sample_lambdas.do" : constructs simulated distributions of lambda parameter estimates, for constructing M and for bootstrapping generated regressors; script deposits outputs in "data/analysis"

D10. "code/demand/dest_monthly_cfs.do" : constructs a single dataset of plant-month capacity factors; script deposits dataset in "data/demand estimation"

ANALYSIS FILES (Module A)

A1. "code/analysis/anal_datasets.do" : merges together master datasets for main regressions; script deposits datasets in "data/analysis"

A2. "code/analysis/anal_overlap.do" : assesses overlap between captive vs. non-captive plants, then executes matching and constructs nearest-neighbor weights; script deposits dataset in "data/analysis"

A3. "code/analysis/anal_construct_M.do" : construct multiple versions of M, using demand estimates, captiveness, and barge-option variables; script deposits finalized M estimates in "data/analysis/M_estimates"

A4. "code/analysis/anal_regs_markup_levels.do" : estimates all "markup levels" regressions; script deposits results datasets in "data/results"

A5: estimate all "markup changes" DD regressions that don't require bootstrapping for generated regressors; these regressions are split across 3 scripts (which can be run concurrently):

- A5a. "code/analysis/anal_regs_markup_dd_captive.do" : TREAT = CAPTIVE
- A5b. "code/analysis/anal_regs_markup_dd_captive_nowater.do" : TREAT = CAPTIVE*(NO_WATER)

- A5c. "code/analysis/anal_regs_markup_dd_captive_nowater_event_study_wide.do" : TREAT = CAPTIVE*(NO_WATER), event-study model (wide)

the subscript "code/analysis/anal_dd_reg_globals.do" sets global macros common to all DD regressions; the subscript "code/analysis/anal_dd_reg_prep.do" loads data and preps variables for DD regressions; each regression script deposits a results dataset in "data/results"

A6: estimate all "markup changes" DD regressions that require bootstrapping for generated regressors; these regressions are split across 11 scripts (which can be run concurrently):

- A6a. "code/analysis/anal_regs_markup_dd_captive_L0hi.do" : TREAT = CAPTIVE*LAMBDA0_HI
- A6b. "code/analysis/anal_regs_markup_dd_captive_L0hi_nowater.do" : TREAT = CAPTIVE*LAMBDA0_HI*(NO_WATER)
- A6c. "code/analysis/anal_regs_markup_dd_Mhi.do" : TREAT = M_HI
- A6d. "code/analysis/anal_regs_markup_dd_M_linear .do" : TREAT = M
- A6e. "code/analysis/anal_regs_markup_dd_M_quintiles .do" : TREAT = quintiles of M's positive support
- A6f. "code/analysis/anal_regs_markup_dd_M_linear_event_study_narrow .do" : TREAT = M, event-study model (narrow)
- A6g. "code/analysis/anal_regs_markup_dd_M_linear_event_study_wide .do" : TREAT = M, event-study model (wide)
- A6h. "code/analysis/anal_regs_markup_dd_M_linear_1stdiff .do" : TREAT = M, first-difference model
- A6i. "code/analysis/anal_regs_markup_dd_M_linear_collusive .do" : TREAT = M_COLLUSIVE
- A6j. "code/analysis/anal_regs_markup_dd_M_linear_two_step .do" : TREAT = M, two-step estimator
- A6k. "code/analysis/anal_regs_markup_dd_M_quintiles_extrabin .do" : TREAT = quintiles of M's positive support, adding an upper bin

the subscript "code/analysis/anal_dd_reg_globals.do" sets global macros common to all DD regressions; the subscript "code/analysis/anal_dd_reg_prep.do" loads data and preps variables for DD regressions; the subscript "code/analysis/anal_dd_reg_prep_1stdiff.do" loads data and preps variables for DD first-difference regressions; the subscript "code/analysis/anal_bootstrap_function.do" loads a function to construct bootstrapped confidence intervals; each regression script deposits a results dataset in "data/results"

A7. "code/analysis/anal_passthrough.do" : uses regression results to estimate pass-through and incidence of a carbon tax; the subscript "code/analysis/anal_dd_reg_globals.do" sets global macros to align with DD regressions; the subscript "code/analysis/anal_dd_reg_prep.do" loads data and preps variables to align with DD regressions; script deposits results datasets in "data/results"

A8: extract residuals for DD pre-trends chart; this step is split across 3 scripts (which can be run concurrently):

- A8a. "code/analysis/anal_regs_markup_dd_captive_pretrends.do" : TREAT = CAPTIVE

- A8b. "code/analysis/anal_regs_markup_dd_Mhi_pretrends.do" : TREAT = M_HI
- A8c. "code/analysis/anal_regs_markup_dd_M_quintiles .do" : TREAT = quintiles of M's positive support

the subscript "code/analysis/anal_dd_reg_globals_pretrends.do" sets global macros common to all 3 scripts; the subscript "code/analysis/anal_dd_reg_prep_pretrends.do" loads data and preps variables for each script; each script deposits residuals in "data/results/pretrends_residuals"

A9. "code/analysis/anal_dataset_gas_markups.do" : merges together dataset for natural gas markup regressions; script deposits dataset in "data/analysis"

A10. "code/analysis/anal_regs_gas_markups.do" : estimates natural gas markup regressions; script deposits results dataset in "data/results"

COUNTERFACTUALS FILES (Module C)

C1. "code/counterfactuals/cf_prep_aux_datasets.do" : prepares auxiliary files for counterfactual analysis; script deposits datasets in "data/counterfactuals"

C2. "code/counterfactuals/cf_coal_prices.do" : predicts counterfactual coal prices absent the fracking boom; the subscript "code/analysis/anal_dd_reg_globals.do" sets global macros consistent with the main DD regressions; script deposits datasets in "data/counterfactuals"

C3. "code/counterfactuals/cf_gen_files_daily_plant_coal.do" : constructs plant-specific files that serve as inputs to the daily counterfactual regressions in Step C5; script deposits datasets in "data/counterfactuals/gen_reg_input_daily_plant_coal"

C4. "code/counterfactuals/cf_gen_files_hourly_plant_coal.do" : constructs plant-specific files that serve as inputs to the hourly counterfactual regressions in Step C6; script deposits datasets in "data/counterfactuals/gen_reg_input_hourly_plant_coal"

C5. "code/counterfactuals/cf_gen_regs_daily.do" : estimates daily counterfactual regressions separately for each coal unit time series; designed to be iterated over alternate specifications by changing local macros at the top of the script *before* proceeding to the next steps; script deposits results datasets in "data/counterfactuals/cf_results_daily"

C6. "code/counterfactuals/cf_gen_regs_hourly.do" : estimates hourly counterfactual regressions separately for each coal unit time series; script deposits results datasets in "data/counterfactuals/cf_results_hourly"

C7. "code/counterfactuals/cf_process_results.do" : processes results of counterfactual regressions; script deposits final dataset in "data/counterfactuals"

OUTPUT FILES (Module O)

O1. "code/outputs/outputs_figures.do" : creates all main text figures; the subscript "code/outputs/outputs_maps.R" makes maps for Figures 1 and 4; script deposits figures and accompanying .tex files in "paper/figures"

O2. "code/outputs/outputs_tables.do" : creates all main text tables, and performs in-line calculations for Section 7.3; script deposits .tex files in "paper/tables"

O3. "code/outputs/outputs_appx_a.do" : creates figures in Appendix A; the subscript "code/outputs/outputs_appx_a_map.R" makes the map in Figure A3; the subscript "code/analysis/anal_dd_reg_globals.do" sets global macros to align with DD regressions; script deposits figures and accompanying .tex files in "paper/figures"

O4. "code/outputs/outputs_appx_b.do" : creates figure in Appendix B, and performs in-line calculations for Appendix B.1; script deposits figure and accompanying .tex file in "paper/figures"

O5. "code/outputs/outputs_appx_c.do" : creates figures and table in Appendix C; script deposits figures and accompanying .tex files in "paper/figures", and deposits table .tex file in "paper/tables"

O6. "code/outputs/outputs_appx_d.do" : creates figure and table in Appendix D; script deposits figure and accompanying .tex file in "paper/figures", and deposits table .tex file in "paper/tables"

O7. "code/outputs/outputs_appx_e.do" : creates figures and tables in Appendix E; script deposits figures and accompanying .tex files in "paper/figures", and deposits tables .tex files in "paper/tables"

O8. "code/outputs/outputs_appx_f.do" : creates figures in Appendix F; script deposits figures and accompanying .tex files in "paper/figures"

O9. "code/outputs/outputs_appx_g.do" : creates figures in Appendix G; the subscript "code/outputs/outputs_appx_g_map.R" makes the maps in Figures G3 and G4; script deposits figures and accompanying .tex files in "paper/figures"

DATA CITATIONS

Association of American Railroads. Rail Cost Adjustment Factor. https://www.aar.org/rail-cost-indexes/; https://www.stb.gov/proceedings-actions/decisions/

Bureau of Labor Statistics. Quarterly Census of Employment and Wages. https://www.bls.gov/cew/downloadable-data-files.htm

Bureau of Transportation Statistics. GIS maps of U.S. rail network. https://data-usdot.opendata.arcgis.com/search?q=rail

Energy Information Administration. Annual Coal Report. https://www.eia.gov/coal/annual/

Energy Information Administration. Coal Transportation Rate Database. https://www.eia.gov/coal/transportationrates/archive/2008/ctrdb/database.html

Energy Information Administration. Electric Power Monthly. https://www.eia.gov/electricity/monthly/

Energy Information Administration. Form 423 data. http://www.eia.gov/electricity/data/eia423/

Energy Information Administration. Form 76 data. http://www.eia.gov/electricity/data/eia767/

Energy Information Administration. Form 7a data. http://www.eia.gov/coal/data.cfm#production

Energy Information Administration. Form 860 data. http://www.eia.gov/electricity/data/eia860/

Energy Information Administration. Form 906 data. https://www.eia.gov/electricity/data/eia923/eia906u.html

Energy Information Administration. Form 923 data. http://www.eia.gov/electricity/data/eia923/

Energy Information Administration. Fuel price data. http://www.eia.gov/dnav/ng/hist/rngwhhdd.htm; http://www.eia.gov/dnav/ng/ng_pri_sum_dcu_nus_m.htm

Energy Information Administration. Natural Gas Trading Hubs. https://atlas.eia.gov/datasets/eia::natural-gas-trading-hubs/explore

Energy Information Administration. Natural Gas Interstate and Intrastate Pipelines. https://atlas.eia.gov/datasets/eia::natural-gas-interstate-and-intrastate-pipelines/explore

Energy Information Administration. Rail Coal Transportation Rates to the Electric Power Sector. https://www.eia.gov/coal/transportationrates/archive/2008/index.php; https://www.eia.gov/coal/transportationrates/archive/2010/index.php

Energy Information Administration. USA Coal Fields. https://atlas.eia.gov/datasets/eia::usa-coal-fields/explore

Energy Information Administration. USA Federal Lands. https://atlas.eia.gov/datasets/esri::usa-federal-lands/explore

Environmental Protection Agency. Clean Air Markets Program Data. https://campd.epa.gov/

Environmental Protection Agency. Emissions & Generation Resource Integrated Database (eGRID). http://www.epa.gov/energy/egrid/

Federal Energy Regulatory Commission. Form 1 data. https://forms.ferc.gov/

Federal Energy Regulatory Commission. Form 714 data. https://www.ferc.gov/industries-data/electric/general-information/electric-industry-forms/form-no-714-annual-electric/data

Mine Safety and Health Administration. Production and mines datasets. https://arlweb.msha.gov/OpenGovernmentData/OGIMSHA.asp

Natural Earth. Coastlines dataset. https://www.naturalearthdata.com/http//www.naturalearthdata.com/download/10m/physical/ne_10m_coastline.zip

Natural Earth. Lakes dataset. https://www.naturalearthdata.com/http//www.naturalearthdata.com/download/10m/physical/ne_10m_lakes.zip

PRISM Climate Group. Temperature data. https://prism.oregonstate.edu/

S&P Global Market Intelligence. Generation Supply Curves. https://www.spglobal.com/marketintelligence

S&P Global Market Intelligence. Price Time Series. https://www.spglobal.com/marketintelligence

Surface Transportation Board. Public Use Waybill Sample. https://www.stb.gov/reports-data/waybill/#Public-Use

U.S. Army Corps of Engineers. River Mile Markers. https://geospatial-usace.opendata.arcgis.com/datasets/604cdc08fe7d43cb90a0584a0b198875_0/explore?location=35.425688%2C-100.496056%2C4.95

U.S. Census Bureau. U.S. state map. https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html

U.S. Geological Survey. National Coal Resources Data System, COALQUAL Database. https://ncrdspublic.er.usgs.gov/coalqual/

U.S. Geological Survey. National Coal Resources Data System, USTRAT Database. https://ncrdspublic.er.usgs.gov/ncrds_data/