DATA MERGING: MAPPING OVERNIGHT COSTS TO INSTALLED CAPACITIES

ANDREW BLOHM

1. Merging EIA overnight cost estimates with Form 860 generator level data

In this section we put forward the underlying assumptions on the mappings we use to merge data from Form EIA-860 with EIA overnight cost estimates. The result of this work is a dataset to be used in the multinomial logit.

1.1. **Overnight cost data.** The overnight cost data used in this analysis comes from the US Energy Information Administration (EIA) annual energy outlook (AEO) series, specifically the underlying assumptions on characteristics of new and existing power plants. Each annual energy outlook estimates the characteristics of new generating plants including the operational and cost assumptions. We use the overnight capital cost assumptions from each AEO to approximate the investment cost for each technology in each year of the analysis. The earliest AEO for which the overnight cost is available is the 1997 AEO; having been published in each subsequent year. However, as new technologies emerge and existing technologies evolve the categories within the overnight cost database have necessarily changed. Therefore, the database does not have consistent categories over time.

1.2. **EIA Form EIA-860 (plant level electricity sector data).** The EIA uses the survey form EIA-860 to collect generator level data on existing and planned power plants with an installed capacity of greater than 1 MW. Form EIA-860 compiles information on the utility, plant, and generator for each of the generators, including information on environmental regulation compliance, ownership structure, etc. Of particular interest to the study is the plant level dataset, which amongst other information contains capacity, installation date, operating status, primary fuel, and technology use at a disaggregated scale not found elsewhere. However, the technology categories available in the Form EIA-860 do not readily map onto the overnight cost technology categories. In the next section, we discuss the mappings that we employ and the reasoning behind those. We discuss the time dimension of the overnight costs in subsequent chapters.

The units of the installed summer capacity are not consistent across the years of the survey. For the EIA Form-860 years 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, and 2014 the summer capacity of each generator is in MW (and correctly identified as such). However, for EIA Form-860 years 1998, 1999, and 2000 the summer capacity is in kW despite the accompanying documentation stating the units as MW. Only 1992, 1993, 1994, 1995, 1996, 1997

Date: July 28, 2016.

has the summer capacity correctly listed as kW. The data also includes negative values in the summer capacity field.

1.3. **Mappings.** There are 37 prime movers (i.e., power production technologies) and 78 primary fuels in the Form EIA-860 database, which together form 190 unique, pair-wise combinations that need to be mapped to the overnight cost database. The issue is that the overnight cost database has between 15 and 20 overnight cost categories from which to choose; the exact number of categories fluctuates year-to-year. In this section, we identify the mapping selections that we made and expound on the assumptions underlying them. We work from the more obvious to the less clear mappings.

There are a few categories that either don't have enough information for assignment (or would require too much effort to do so). First, we ignore the prime mover designation CG as it is undefined by the survey (3 natural gas fired CG plants with an installed summer capacity of 0.33 TW over the period of analysis). Second, we ignore the prime mover designation other (OT) and primary fuel other (OTH) because the cost of determining how to allocate these resources is too high, as compared to the benefit (i.e., it would require reading the notes for approximately 575 generators with an installed capacity of 0.069 TW over the period of analysis). Further, the other category (OT) has an installed summer capacity of approximately 63,000 MW, which is a very small fraction of the total installed capacity (0.000009%) over the period of analysis. Third, we ignore any generator without an identified primary fuel, except in the case of hydro power technology, where the choice of fuel is obvious. There are 6 non-hydro generators with an installed summer capacity of 0.096 TW that do not presently have a primary fuel listed. Fourth, we ignore the primary fuel code WOC, which is not defined in any of the supporting literature. There are 12 generators with a total installed summer capacity of 0.30 TW with a primary fuel WOC. Finally, we ignore the multi-fueled generators (MF) given that the database doesn't indicate the set of fuels that they are using. There are 21 generators with a total installed summer capacity of 0.635 TW. In total we drop from the dataset 617 units, with an installed summer capacity of 1.43 TW or approximately 0.0001% of total installed summer capacity for all generators across the period of analysis.

1.3.1. *Renewables, nuclear, hydropower, biomass, and geothermal resources.* The overnight cost dataset includes one cost estimate for conventional hydropower. We assume that all of the production technologies with water (WAT) listed as the primary fuel can be mapped to the hydropower investment cost. These include prime mover designations hydraulic turbines - conventional (HC), hydraulic turbine - pipeline (HL), hydraulic turbine - reversible (HR), hydraulic turbine - conventional (HY), and hydraulic turbine reversible - pumped storage (PS). We also map these production technologies to the hydro power investment cost even in cases where the primary fuel column entry is missing. There are 90,163 hydropower generators in the dataset with a total installed summer capacity of approximately 1030.87 TW for the period of analysis.

The overnight cost dataset includes one cost estimate for geothermal power plants. We assign the geothermal overnight cost to any generator with geothermal (GEO) or geothermal steam (GST) as a primary fuel. This includes prime

mover designations: turbines used in binary cycles (BT), other (OT), steam turbine (geothermal) (GE), and steam turbine (boiler) (ST). This categorization moves one plant type from other (OT) to geothermal. There are 3,269 geothermal generators in the dataset with a total installed summer capacity of approximately 0.0033 TW over the period of analysis.

The overnight cost dataset includes one cost estimate for photovoltaic power plants. We assign the photovoltaic overnight cost to both prime mover designations: photovoltaic (PV) and photovoltaic (SP) in the Form EIA-860 dataset. There are 3,460 photovoltaic generators in the dataset with a total installed summer capacity of approximately 0.062 TW over the period of analysis.

The overnight cost dataset includes one cost estimate for solar thermal power plants. We assign the solar thermal overnight cost to any generator with sun (SUN) as the primary fuel and a prime mover designation that is not photovoltaic (PV or SP). This includes other (OT), steam turbine - solar (SS), and concentrated solar power energy storage (CP) generators. This categorization moves one plant type from prime mover other (OT) to solar thermal. There are 170 solar thermal generators in the dataset with a total installed summer capacity of approximately 0.018 TW over the period of analysis.

The overnight cost dataset has two cost estimates for wind (i.e., onshore and offshore). However, we do not have this level of specificity in the Form EIA-860 dataset but we do know that most of the installed capacity in the US has been in onshore wind capacity. Therefore, we assign the onshore wind overnight cost to the wind turbine (WT) generator. There are 7,732 wind generators in the dataset with a total installed summer capacity of approximately 0.52 TW over the period of analysis.

The overnight cost dataset includes one cost estimate for nuclear power plants. We assign the nuclear power overnight cost to any generator using nuclear (NUC), uranium (UR), plutonium (PT), or thorium (TH) as the primary fuel. This includes steam turbine - boiling water nuclear reactor (NB), steam turbine - graphite nuclear reactor (NG), steam turbine - high-temperature gas-cooled nuclear reactor (NH), and steam turbine - pressurized water nuclear reactor (NP). There are 2,450 nuclear plants in the dataset with a total installed summer capacity of approximately 1062 TW.

The overnight cost dataset has one cost estimate for fuel cells. We assign the overnight cost estimate for fuel cells to fuel cell (electrochemcial) (FC). In the Form EIA-860 dataset the fuel cell plants use landfill gas (LFG), natural gas (NG), and other biomass gases (OBG) as there primary fuel.

The overnight cost dataset includes one cost estimate for biomass power plants. The DOE assumes that woody residues, round wood and woody energy crops, municipal solid wastes, wet herbaceous residues, and dry herbaceous residues and energy crops are all biomass feedstocks Source: `http://www.energy.gov/eere/bioenergy/biomass-feedstocks`. However, given that MSW is its own category in the overnight cost we assign that to its own overnight cost. We assign all plants with the following primary fuels to the biomass overnight cost, biomass generic (BIO), other biomass solids (OBS); refuse, bagasse and all other non-wood waste (REF); wood and wood waste (WD), and wood waste solids (WDS). This categorization moves one plant from the internal combustion category to the biomass category and, two plant types from other to biomass (i.e., (OT, REF) and (OT, WDS)). Further,

we prevent the assignment of a combined cycle plant using wood and wood waste solids as a primary fuel to biomass and instead keep it assigned to the combined cycle overnight cost.

1.3.2. *Steam turbines.* Gas-oil steam turbines exist as a category for the 1997 to 2000 AEOs. We assign all steam turbine plants that use either, any form of natural gas or fuel oil to this overnight cost estimate. On the natural gas side this includes landfill gas (LFG), synthetic natural gas (SNG), other gas (OG), natural gas (NG), liquified propane gas (LPG), blast furnace gas (BFG), black liquor (BLQ) (gasified before use), and other biomass gases (OBG). We assume that some data entry mistakes were made in the process of inputing the survey in that BL should have been BLQ. This change affects 7 generators. On the oil side this includes black liquor (BLQ), distillate fuel oil (DFO), fuel oil 1-6 (FO1 - FO6), residual fuel oil (RFO), oil-other and waste oil (WO), coal synfuel (SC), residual fuel oil (RFO), and sludge waste (SLW).

1.3.3. *Combined cycle plants.* The EIA has overnight cost estimates for conventional and advanced gas-oil combined cycle. However, the EIA Form-860 dataset does not include the information necessary to delineate between the two alternatives.[1] The prime mover designations in the Form-860 dataset include the combined cycle - total unit (CC), combined cycle combustion turbine part (CT), combined cycle single shaft - gas turbine and steam turbine (CS), combined cycle steam turbine - waste heat boiler (CW), and the combined cycle steam turbine with supplemental firing (CA). The cost estimates in the overnight cost database are for an oil and gas unit however, some of the CT, CC, CA, use purchased steam (PUR), bituminous coal (BIT), or wood and wood waste solids (WDS) as their primary fuel. We don't have enough information to assign these plants to anything but the combined cycle overnight cost.

1.3.4. *Combustion turbines.* The EIA overnight cost estimates for combustion turbines once again include both an advanced and conventional option. One option is to use the conventional turbine overnight cost for all of the combustion turbine plants in the Form EIA-860 dataset. This includes combustion (gas) turbine (GT), internal combustion (IC), and jet engine (JE). There are 156,666 conventional or advanced combustion turbines (not sure of the distribution between them) in the dataset with a total installed summer capacity of 595.422 TW. For the time being we assume that all of them are conventional combustion turbines. At some future point (with more time) we might be able to use the average size of an advanced versus conventional combustion turbine to improve upon this assumption.

1.3.5. *Integrated coal gasification plants.* The EIA overnight cost dataset includes two overnight cost estimates for the integrated coal gasification plant; with and without carbon capture and storage (CCS). We assume that all IGCC plants built to date do not include the CCS option. Therefore, we assign the overnight cost for the IGCC without CCS to all IGCC plants in the Form EIA-860 dataset. There are 13 IGCC plants in the dataset with a total installed summer capacity of 2.028 TW.

[1]There might be an average size difference that could be used to assign the Form-860 combined cycle plants to the advanced and conventional categories.

1.3.6. *Coal fired power plants not IGCC.* The EIA overnight cost dataset includes a few coal power plant options. In the 1997 AEO there is one coal power plant option; pulverized coal power plants. In the 1999 AEO there is a scrubbed coal new option. In the 2001 AEO the coal option is the conventional pulverized coal power plant. Despite the changes in the name of the plant each of these plants has similar characteristics (i.e., cost, etc.). Therefore, we assign all coal fired plants bituminous coal (BIT), subbituminous coal (SUB), lignite (LIG), anthracite (ANT), coal generic (COL), refined coal (RC), and waste coal (WC), which are not IGCC (IG), combined cycle, or combustion turbine power plants to the coal power plant overnight cost (of varying name). Online resources suggest that petroleum coke is used primarily in steam turbine, coal-fired power plants so we assign the petroluem coke (PC) primary fuel plants to the coal plants as well. Online resources suggest that tire derived fuel (TDF) are primarily used as an additive in coal plants.

1.3.7. *Distributed generation.* There are two categories of overnight costs for distributed generation: peak and non peak.

We assign all megawatt hour (MWH) primary fuel plants to distributed generation (i.e,. battery energy storage (BA) and flywheel energy storage (FW)). Further, we assign compressed energy storage (CE). HC, HL, HR, HY, PS all use water as a fuel (AEO only has the conventional hydropower overnight cost estimate so we use it for all) (IC) Internal combustion plant: A plant in which the prime mover is an internal combustion engine. An internal combustion engine has one or more cylinders in which the process of combustion takes place, converting energy releasted from the rapid burning of a fuel-air mixture into mechanical energy. Diesel or gas-fired engines are the principal types used in electric plants. The plant is usually operated during periods of high demand for electricity. (BIO) (GEO) and (GST), all involve geothermal energy sources (MSW) is assigned to municipal solid waste