* CO2 and local emissions
* Hour of the day and seasons
* Save point estimates – really small table
  + One table that has hour of the day, season, region, co2, local pollutants
  + 24 HOD, 2 seasons, 13 regions
* Statistical model
  + Remove week fixed effects
* Standard errors
* BELOW IS AT FACILITY LEVEL BY FUEL TYPE – 2 COAL GENS AT ONE PLANT AND 3 NAT GAS = 2 GENERATORS
  + 82 File: Uses python command to map bhats into b tildes
    - Bhats are obtained from generator specific regressions
      * All gens from CEMS files, regress generation, include controls (sunshine, wind blow, week effects)
      * Measuring each generators response to increases in electricity consumption by region
      * Brute force regression
    - Btildes are regularized regression coeffs
      * 2 restrictions
        + If electricity consumption is going up in any region, no generator will systematically decrease its output

B tilde is always positive

* + - * + Some generators will pick up excess demand

Sum of increase in generation across all generators will equal increase in consumption

Sum b tildes is always 1

Weakly positive

* + - * Lots of corner solutions
        + Many b tildes will be 0
        + Reveal a
      * One vector of b tildes for each region
        + Some for NE will be positive, estimator is revealing that these facilities are selected into the model
        + B tilde = 0 = not selected into model
* Standard error
  + Sample splitting
    - When you use data to select a model and then do inference, there will be sample selection bias
    - Mistakes in model selection tend to correlate to mistakes in inference estimation
    - Split the sample into 2 subsamples
      * Randomize at the week level to split
      * Use subsample 1 to select the model
        + 50% of the weeks
      * Go to subsample 2 and recompute the bhats **but** using the selected sample from subsample 1
      * Subsample 1 shows activation of generators
        + Subsample 2 goes through all generators one by one
        + Only want to regress on active regions, but in addition to same controls (wind, sun, week fixed effects)
        + Subsample 2 running provides **beta hat 2**
      * **Add loop to select twice if possible**
    - Hard part
      * 9 regions, 1100 generators
      * Zeros are never selected
      * Constraint minimization procedure (done in 82)
* SE and Debiasing
  + Only do inference after debiasing estimator
  + Lasso regression – very rich class of models, which model should I use?
    - Implicitly always doing model selection which needs debiasing
  + Point estimator before inference (called delta tilde)
  + Separate W of I for co2 and local pollution
  + W of I =
  + More precision at the expense of model selection bias
  + Debias estimator for inference