

# Natural Gas Data Summary

Jason Charles Ross

September 13, 2022

## Contents

<b>1</b>	<b>Natural Gas Transactions Data</b>	<b>1</b>
<b>2</b>	<b>Tariff Summary Data</b>	<b>2</b>
<b>3</b>	<b>Rate Calculator Data (in progress)</b>	<b>3</b>
3.1	Rate Data . . . . .	3
3.2	List of Points . . . . .	4

## 1 Natural Gas Transactions Data

We have data on 193,578,719 natural gas transactions from January 2017 through August 2022 across 119 different pipelines.<sup>1</sup> The cleaned data is stored on midway2 at

`/project2/hortacsu/naturalgas/ngasdata.fixed.csv`

The original data is stored at the same directory with the file name `ngasdata.csv`.<sup>2</sup> The data separated by pipeline is stored at

`/project2/hortacsu/naturalgas/ngasdata.by_pipeline/`

The data separated by year and the 11 pipelines in the MISO South region and the 108 not in the MISO South region is stored at

`/project2/hortacsu/naturalgas/ngasdata.by_miso/`

For the latter data, for example, one file contains all natural gas transactions in the MISO South region in 2019.

For each transaction, we observe the following 12 characteristics, where items in parenthesis denote either the set of possible entries when there are few possibilities or a brief explanation of the data in the column:

---

<sup>1</sup>These pipelines can be found on midway2 at `/project2/hortacsu/naturalgas/pipeline_names.csv`

<sup>2</sup>The cleaned data has human-readable dates and corrects a formatting issue across POINT\_NAME observations than hindered reading into R or Stata.

1. POINT\_NAME
2. OPERATIONAL\_CAPACITY\_KEY
3. GAS\_DAY (transaction date)
4. RECEIPT\_DLTV (Delivery, Mainline, Receipt, Bi-Directional, Segment)
5. SCH\_CAPACITY
6. PIPELINE\_NAME
7. POINT\_TYPE (Delivery to an LDC, Mainline, Gas Processing Plant, Delivery to End User, Compressor, Interconnect, Storage Quantity, Gathering, Receipt by LDC, Exchange, Unknown, Pool, Segment, Power Plant, LNG, Storage Withdrawal, Stand Alone Meter, Wellhead, Storage Injection)
8. FLOW\_INDICATOR (Delivery, Line Capacity, Receipt, Injection, Withdrawal, Bi-Directional, Forwardhaul, Backhaul, South, West, East, North)
9. CYCLE (Timely, Evening, Intraday 1, Intraday 2, Intraday 3)
10. AVAIL\_CAPACITY
11. FREQ (Daily, Monthly)
12. MAX\_CAPACITY

## 2 Tariff Summary Data

We have data on 6,818,135 tariff observations from January 2018 to December 2021 across the 11 MISO South pipelines. The data is stored on midway2 at

`/project2/hortacsu/naturalgas/tariff.csv`

Each tariff observation has 10 characteristics:

1. tariff\_rate\_type (Electric Power Costs, ACA, Commodity, Fuel, Overrun, Reservation, Injection, Storage Demand, Storage Quantity, Withdrawal, Injection Fuel, Injection Overrun, Withdrawal Fuel, Withdrawal Overrun, Loaning, Parking, Inventory Cycling, Sourcing)
2. tariff\_date\_begun
3. tariff\_rate\_end
4. receipt\_zone

5. delivery\_zone
6. tariff\_rate
7. magnitude (\$/Dth%, \$/Dth/day, \$/Dth/mth, ¢/Dth/day, ¢/Dth,)
8. pipeline
9. effective\_date (date selected by web scraper to make download)
10. rate\_schedule

The set of available rate schedules varies across different pipelines (e.g. some pipelines may have one rate schedules, others may have 15). Tariffs are either defined between two zones (points are grouped by zones) or throughout an entire system.

### 3 Rate Calculator Data (in progress)

#### 3.1 Rate Data

The rate calculator data download is in progress. We will have observations from January 2018 through December 2021 across the 10 MISO South pipelines. There are 10 characteristics for each rate observation:

1. tariff\_rate\_type
2. tariff\_rate
3. magnitude
4. tariff\_rate\_structure
5. rate\_zone
6. date (we download all data at the first day of every month)
7. pipeline
8. rate\_schedule
9. receipt\_zone
10. delivery\_zone

The set of available rate schedules varies across different pipelines. The sets of delivery zones and receipt zones vary across different pipelines. Rates are either defined between two zones or throughout an entire system. Interzonal rates vary with rate schedules. Delivery and receipt points are grouped by different zones.

As an example, we select a pipeline  $i$ . Then, we select rate schedule 1 in the list of available rate schedules for pipeline  $i$ . Next, we download the interzonal rates for each pair of receipt and delivery zones. We then move on to rate schedule 2 and download the interzonal rates for each pair of receipt and delivery zones. We then move onto rate schedule 3, etc.

We download all rates data on the first date of each month. In a random sample, we did not observe any rates that changed within a month (e.g., different rates between the beginning and end of each month; same rate between the beginning and end of each month, but a change and change-back in the middle of the month, etc.). We download data in this manner, as scraping data for each day of the four year period would take a prohibitively long time (at least 30 days by my estimate).

## **3.2 List of Points**

We will have a list of the points included in each delivery and receipt zone for each pipeline. This list will be downloaded at the beginning of each month from January 2018 through December 2021. We do not anticipate that this list will change between months.