Residential Commercial Modeling of Natural Gas

Nathaniel Rogalskyj

January 2026

1 Introduction

Residential and Commercial usage of Natural Gas is vitally important to the modeling of the Supply and Demand Balance for Natural Gas historically. At a number of hedge funds, due to the requirement to produce quickly, many Residential and Commercial models are not properly bench-marked. Likewise, a significant amount of theory is glossed over or not looked at. In the codes that are developed, a number of written code and algorithms have things that upon second glance and examination should be redone and rethought. In this paper, Spectral Technologies examines a Residential and Commercial Natural Gas Consumption Model Benchmarks and Theory. Spectral Technologies also examines the creation of a theory for modeling Residential/Commercial Natural Gas Consumption, a model built off this theory which is exposed in an API and how said model performs against the historical benchmark.

2 Design of a Benchmark

The goal of a benchmark is to be a practical method that other methods/models can be compared against. The benchmark aims to show how another method can beat or even lose to the said benchmark. The benchmark must be something that can be easily implemented, otherwise it is not a benchmark but instead a well-constructed method. Hence in the design of the benchmark, one needs (1) an implementable method, (2) that is not too complicated.

3 Historical Benchmark

Spectral Technologies proposes a historical benchmark for the prediction of unseen EIA Natural Gas Consumption Reports for various states. Assign to each year/month tuple the values (Y, M). As an example, the month February 2023 will be assigned the tuple $Y = 2023$ and $M = 2$. Likewise January 2025 will be assigned the values $Y = 2025$ and $M = 1$. The data that the Energy Information provides a set of consumption information $X_{Y,M}$ indexed by (Y, M) where

$Y = 1980...C$, where C is the current year, and months where $M = 1...12$. To predict $(\bar{Y}, \bar{M})$, we calculate

$$E_H(\bar{Y}, \bar{M}) = \frac{1}{N} \sum_{Y<\bar{Y}, M=\bar{M}} X_{Y,M}$$

where $N = |Y < \bar{Y}, M = \bar{M}|$, where $||$ is a measure of cardinality. E_H is the historical benchmark that we will benchmark against. The benchmark does quite a good job at estimating the realized consumption. The accuracy of this estimator can be measured as:

$$e(E_H, \bar{Y}, \bar{M}) = E_H(\bar{Y}, \bar{M}) - X_{\bar{Y},\bar{M}}$$

Upon averaging over all months and years, the percentage error for a Historical Benchmark is around 9%.

4 Definition of Skill

Skill is how much can one beat the benchmark. The benchmark has a 9% error rate. The skill for PredictGas API is 5% since, the percentage error of PredictGas is 4%.

5 Leveraging Daily Weather Values

A core issue that is of interest is when does theory state that daily weather values matter in the prediction of natural gas consumption versus monthly weather aggregated values? This was the core reason I decided to develop this API. I saw that only the summation of weather values, reduces information by 97%. The conclusion to this problem is that one needs a nonlinear daily relationship to make the daily weather values useful over aggregated values over the appropiate period.

Suppose that there are two functions that could stand for the daily relationship between Weather and Natural Gas Consumption. f_1 is a linear relationship and f_2 represents a quadratic relationship.

$$f_1(x) = \alpha x + \beta$$
$$f_2(x) = \theta x^2 + \alpha x + \beta$$

The representation of the monthly values can be simply represented as:

$$\sum_{i=1}^{n} f(x_i)$$

We can apply this directly to the linear and quadratic functions: f_1 and f_2 respectively.

$$\sum_{i=1}^{n} f_1(x_i) = \alpha \sum_{i=1}^{n} x_i + n\beta$$

$$\sum_{i=1}^{n} f_2(x_i) = \theta \sum_{i=1}^{n} x_i^2 + \alpha \sum_{i=1}^{n} x_i + n\beta = \theta \sum_{i=1}^{n} x_i^2$$

Clearly, $\sum_{i=1}^{n} f_1(x_i)$ is really just dependent on $\sum_{i=1}^{n} x_i$, while f_2 is not just dependent on the sum. For the more nonlinear f is, the more important knowledge of x_i are. If $f(x)$ is linear, then the monthly weather values are sufficient. Nonlinearity and high levels of nonlinearity in the weather-to-consumption impulse function are required to make full usage of the value of the daily weather values.

5.1 Generality

The goal is to now the problem in its generality. This can be done by analyzing a function $f(x)$ in its parametric form provided by:

$$f(x) = \sum_{j=0}^{\infty} \frac{f^j(0)}{j!} x^j$$

It is now important to consider:

$$\sum_{i=1}^{n} f(x_i) = \sum_{i=1}^{n} \sum_{j=0}^{\infty} \frac{f^j(0)}{j!} x_i^j$$

Rearranging terms, one can see that:

$$\sum_{j=0}^{\infty} \frac{f^j(0)}{j!} \sum_{i=1}^{n} x_i^j$$

It is clear that $\sum_{i=1}^{n} f(x_i)$ that relevant information extracted from the weather data can be expressed in the quantities provided via:

$$\sum_{i=1}^{n} x_i^j$$

What can be concluded is that the high levels of nonlinearity is required to make usage of the weather values. One can conclude that higher orders of the Taylor's Series of f are required to make good usage of full weather dataset that is provided by Weather Vendors.