Forecasting Design Day Demand Using Extremal Quantile Regression

David J. Kaftan, Jarrett L. Smalley, George F. Corliss,
Ronald H. Brown, and Richard J. Povinelli
GasDay Project, Marquette University, Milwaukee, WI

Abstract

Extreme events occur rarely, making them difficult to predict. Extreme cold events strain natural gas systems to their limits. Natural Gas distribution companies need to be prepared to satisfy demand on any given day that is at or warmer than an extreme cold threshold. The hypothetical day with temperature at this threshold is called the Design Day. To guarantee Design Day demand is satisfied, distribution companies need to determine the demand that is unlikely to be exceeded on the Design Day.

We approach determining this demand as an extremal quantile regression problem. We review current methods for extremal quantile regression. We implement a quantile forecast to estimate the demand that has a minimal chance of being exceeded on the design day. We show extremal quantile regression to be more reliable than direct quantile estimation. We discuss the difficult task of evaluating a probabilistic forecast on rare events.

Probabilistic forecasting is a quickly growing research topic in the field of energy forecasting. Our project contributes to this field in three ways. First, we forecast quantiles during extreme cold events where data is sparse. Second, we forecast extremely high quantiles that have a very low probability of being exceeded. Finally, we provide a real world scenario on which to apply these techniques.

Index terms: Energy demand, natural gas demand, extremal quantile regression

1   Introduction

Natural gas Local Distribution Companies (LDCs) need to provide steady flow to their customers. It is important for these LDCs to be able to forecast future demand, as to be able to reserve and send out an appropriate amount of Natural Gas. With a large portion of Natural Gas usage being for spatial heating, Natural Gas Demand is highly weather dependent [7]. Extreme cold events stress the limits of the distribution system, as this is often when natural gas demand is at its maximum. The **Design Day** is the hypothetical day when

temperatures break a given extreme cold threshold. The Design Day condition is the temperature of an extreme event that only occurs once in N years. The Design Day demand is the natural gas demand that is forecasted to occur on a day with the Design Day condition. This is the demand that the LDC should be able to supply up to.

Probabilistic forecasting methods offer an appropriate model to forecast extreme events, since they are characterized as only happening once in n years. Probabilistic forecasts of extreme events are advantageous as they reflect the inherent uncertainty of extreme event forecasting, while giving forecast end users the information needed to make system-wide decisions [5].

2   Background

Quantile regression is a method for predicting the cumulative density function of a response variable conditional on a predictor variable [3]. For example, quantile regression can answer the question: what is the demand that has 10 percent chance of being exceeded given temperature is 20 °F? The relationship between the response variable and the predictor variable for a given quantile (i.e., 10 percent chance of being exceeded) can be found by minimizing the pinball loss function - typically through linear programming methods.

Problems arise when we choose an extremely high (or low) quantile. Consider the situation where we want to predict the demand that will be exceeded 0.01 percent of the time for a given temperature. Perhaps we only have 100 data points on which to fit this relationship. That means we are fitting a line above every data point in our data set. In this case, using traditional quantile regression will not give us reasonable results. Extremal Quantile Regression extends the concept of quantile regression to the extreme tails of a distribution [1]. One particularly useful method is introduced by Wang [8]. The conditional relationship is assumed to be the same across higher quantiles. For example, the conditional relationship between demand and temperature is the same for the 90th and 99th quantiles. The only difference between these two quantiles is constant with temperature (i.e., a bias term). The bias term is determined using extreme value theory.

3   Methods

We estimate the 99.38^{th} quantile of flow, given temperature. We start with a quantile regression model of demand and flow, making the following 4 adjustments. First, we address the non-linear relationship between temperature and flow. Second, we account for our desire for performance during extreme cold days. Third, we address to challenge of modeling an extremely high quantile. Fourth, we account for the long term out of sample uncertainty.
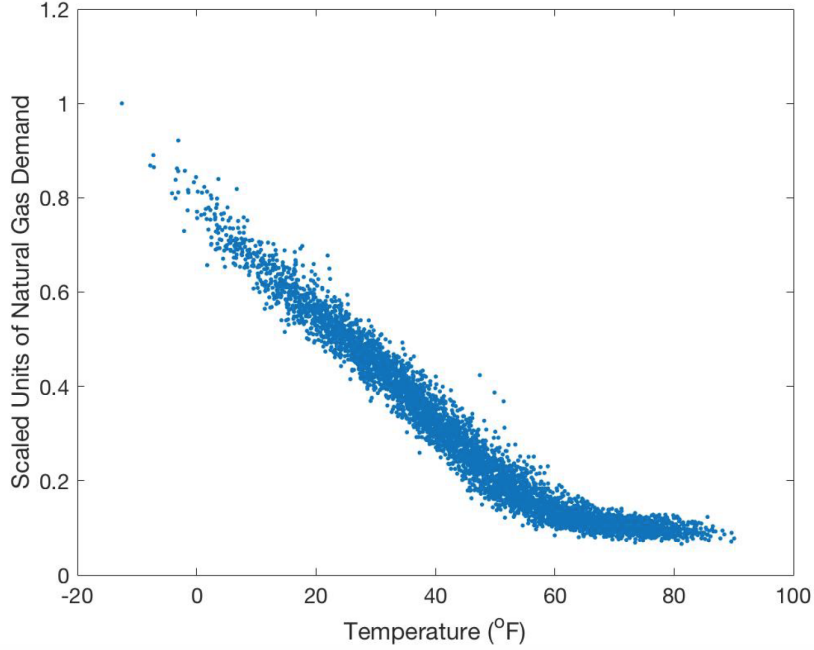
Figure 1: Demand is non-linear with temperature. However, when temperature is less than 65 °F it is approximately linear.

3.1 Non-Linear relationship between Temperature and Flow

As seen in Figure 3.1, there is a non-linear relationship between temperature and flow. Typically, this relationship is accounted for by transforming temperature into heating degree days (HDD), based on a reference Temperature. This transformation can be seen in Equation 1.

$$HDD = \max(0, T_{Reference} - T_{Current}) \tag{1}$$

Heating degree days allow us to assume temperature independence from flow at warm temperatures and linear temperature dependence at colder temperatures. However, it does not account for the bias shift in uncertainty around 65 degrees F. Rather than using an indicator variable to represent temperatures colder than 65 degrees F, we ignore all temperatures warmer than 50 degrees F. Conveniently, we are only concerned about the coldest temperatures, so removing the warm temperatures does not bother us.

3.2 Performing Well on Cold Days

After removing the warm days, we have significantly less data. We would like to model the demand on the coldest days, but we lack enough data to do so.
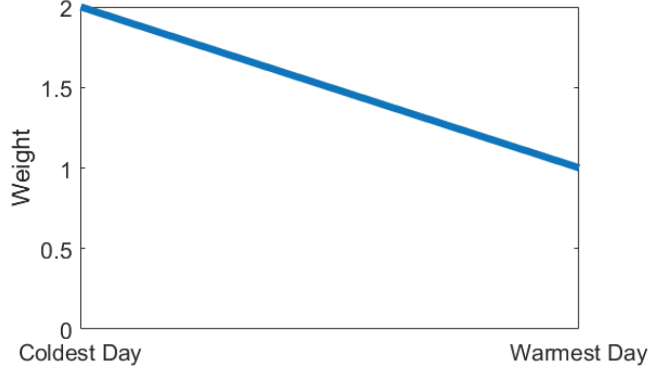
Figure 2: Weighting of pinball loss optimization

We assume the remaining data can give us information about demand on the coldest days, but we know that our coldest data is best. We therefore weigh the Pinball Loss by temperature in the quantile regression optimization. This is done by sorting pinball loss vector by temperature (the first index is the warmest temperature) and adjusting each element i in the vector of length N accordingly:

$$AdustedPinball_i = SortedPinball_i \times (1 + \frac{i}{N}).\qquad(2)$$

The adjusted pinball loss on the coldest day is now weighted twice as heavy as the warmest day. When minimizing the adjusted pinball loss a quantile will be fit better to the coldest days. These weights are visualized in Figure 3.2.

3.3   Estimating Extreme High Quantiles

When estimating in the extreme quantiles, you are like to get some wild results. At high enough quantiles with little data, quantile regression will fit a line through the two most extreme outliers (given a 2 parameter model). We have already begun to mitigate this by weighting the optimization of the coldest points rather than removing non-extreme cold data. We further account for this by using composite quantile regression (also known as weighted quantile regression, or Hogg's method [2]). Composite quantile regression allows us to determine the temperature parameter using less extreme quantiles, leaving only the bias parameter to be determined for the extreme high quantile.

3.4   Long Term Out Of Sample Uncertainty

It is well known that errors are often greater in magnitude for out-of-sample testing than in-sample testing. Since we are forecasting the quantile for the following year, we need to adjust based on out-of-sample uncertainty. We first

train (detrend data and perform quantile regression) our model on a single year of data. We then validate on a single year of data and re-fit the bias term based on the residuals. We record the change in the bias term. This process is repeated using the first two years as a training set and the third as a validation set, then again until the last year of training data is used as the validation set. We determine the mean absolute change of the bias term. Finally, we fit a model to all of the training data. We add the mean absolute change to the bias term in order to account for the out-of-sample uncertainty.

4 Results

The experiment is run on the 100 most temperature sensitive datasets GasDay forecasts. The experiment is run in two folds. First, we hold out the 2017 heating season for testing and we train on all previous years. Second, we hold out the 2016 heating season for testing and train on all previous years.

Our experiment reflects our goal; we need to estimate the quantile representing demand that is exceeded 0.62% of days in a heating season. We therefore count the number of exceedances of the quantile in the test set. In particular, we want to coldest days in winter. We therefore limit our analysis to the 10% coldest days of the test set.

Across the two test folds, we expect 44.70 exceedances. In total, our method incurs 38 exceedances: 85.0% of the expected exceedances. For the sake of evaluating the strategy for dealing with long-term out of sample uncertainty described in Section 3.4, we evaluated our experiment absent of said strategy. The result showed our bounds unreasonably tight with over 200% of the expected exceedances.

5 Conclusion

This work introduces an interesting problem to the forecasting community: forecasting the Design Day demand. Using methods common in forecasting literature - along with novel extensions - we are able to successfully predict the demand that is extremely unlikely to be exceeded during the Design Day.

References

[1] Victor Chernozhukov. Extremal quantile regression. *Annals of Statistics*, pages 806–839, 2005.

[2] Roger Koenker. Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, 91(1):74–89, 2004.

[3] Roger Koenker and Gilbert Bassett Jr. Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50, 1978.

[4] Samuel Kotz and Saralees Nadarajah. *Extreme value distributions: theory and applications.* World Scientific, 2000.

[5] Allan H Murphy. Probabilities, odds, and forecasts of rare events. *Weather and forecasting*, 6(2):302–307, 1991.

[6] Zoltan Toth, Olivier Talagrand, Guillem Candille, and Yuejian Zhu. Probability and ensemble forecasts, 2003.

[7] Steven Vitullo, Ronald H Brown, George F Corliss, and Brian M Marx. Mathematical models for natural gas forecasting. *Canadian applied mathematics quarterly*, 2009.

[8] Huixia Judy Wang, Deyuan Li, and Xuming He. Estimation of high conditional quantiles for heavy-tailed distributions. *Journal of the American Statistical Association*, 107(500):1453–1464, 2012.