Dear LDS,

To test the hypothesis that churn is driven by the customers' price sensitivities, I think a binomial regression model with a logit link function would be appropriate, and our goal is the churn rate of a client at a specific time in this case. To build a model and do some further prediction, we may need the following data as independent variables:

1. The type of the client. Are they corporate, SME or residential customers, as different groups might have different sensitivities to the prices.

2. The type of product they are using, gas or electricity.

3. The unit price that a client pays.

4. The time a client has been paying for our products.

5. We can also take into account the total price that a client pays and we can classify this into multiple groups (rather than using the specific value) in order to have sufficient data for each group.

First, we need to divide all the data we got into 10 folds, and randomly merge them into two groups, one is for training (e.g. 9 folds, around 90%) and the other is for testing (10%). Then we set various time interval (1 month, 2 months, …) and count the number of churning customers after a specific time interval for different client type, product type, and unit price.

Next, we can fit a linear logistic regression model using the above data. For each coefficient we should look at its p-value and decide if it should be retained in the model. Also, we need to consider multicollinearity, for example, if the size of a client and its unit price and total price are strictly linearly correlated then we may only need to use one of the three variables. If the dataset is relatively small, we can use cross validation to select a "best" model (10 folds have been set as before). Then by checking the estimate and the corresponding p-value, we can see if the clients are price-sensitive. If the estimate is far away from zero with a significant p-value under 0.05, we can conclude they are price-sensitive.

Then, by reducing 20% of unit price of a group of clients in SME (and others) with a high probability to churn and predicting the resulting churn rate and comparing with the previous rate, we can see if a discounting strategy is efficient.

Best Regards,
Lingyu