Inflation Expectations and Readiness to Spend: (XG)Boosting Previous Results

STAT 27420—Intro to Causality with Machine Learning

Wesley Janson
wrjanson@uchicago.edu

Drew Keller
drewkeller@uchicago.edu

December 9, 2022

Abstract

Inflation expectations have become a pivotal tool in gauging the direction in which future inflation is headed. Policymakers follow consumer inflation expectations closely due to the belief that consumer inflation expectations affect behaviors that in turn affect future inflation, such as purchasing decisions. However, while economic theory supports the idea of such a causal relationship, there is a debate over the degree to which consumer inflation expectations *actually* affect behavior, with important policy implications. Bachmann et al. (2015) explore the relationship between 1-year expected inflation and consumer spending attitudes – more specifically, attitudes towards durable good purchasing. To address the same question while relaxing parametric assumptions those authors rely on, we propose an alternative method implementing the augmented inverse probability of treatment weighting (AIPTW) estimator with gradient-boosted random forests. We find similar results to Bachmann et al: a generally small but significant average negative effect of consumer inflation expectations on attitudes towards spending on durable goods.

1 Introduction

Inflation expectations have remained on the forefront of policymakers' minds since their importance first became evident in the 1970s. In the midst of the highest modern inflationary episode, central bankers finally began to build monetary policy around not only monetary aggregates and realized inflation, but inflation expectations. Theory around the very idea of inflation expectations became (and continues to be) a topic widely studied by economists.

Inflation expectations are captured in two primary forms: surveys and financial market expectations, as measured through Treasury inflation-protected securities (TIPS). This paper will focus on the former. Survey measures of inflation expectations come from a wide range of target audiences. Expectations of professional forecasters are widely used to examine whether central bank inflation targets adequately anchor inflation beliefs (Binder et al., 2022). Surveys of consumers are important for different reasons: knowing consumers' outlooks and sentiment towards the economy can indicate trends in consumption. The University of Michigan's Survey of Consumers has been continuously maintained since 1946 (though more reliably since the 1970s). This data was used by Bachmann et al. 2015 to estimate the impact of inflation expectations on consumers' attitude towards purchasing durable goods.

We have chosen to revisit the methodology and data used in the Bachmann paper, but with alterations. We will move beyond their ordered-probit estimation to use a more flexible, non-parametric machine learning model. This will allow us to assess the same question without relying on strong functional form assumptions like previous authors.

2 Data & Relevant Literature

The University of Michigan's Survey of Consumers (MSC) is a widely-used public dataset that each month asks a representative sample of households a comprehensive set of questions regarding their outlook on the economy and their own personal finances[1]. It has become a popular source of data for a wide range of topics from displaying the partisan divide in inflation expectations (Gillitzer & Robinson, 2021) to household responses to stimulus payments (Sahm et al., 2010), along with providing real-time feedback of how the average American views the outlook of our economy. Bachmann et al. (2015), is the most relevant paper for us. The authors examine the relationship between expected inflation and spending attitudes using MSC microdata. Using an ordered-probit regression they estimate how one-year-ahead expected inflation affects attitudes towards spending on durable goods. In their baseline specification, they find a one percentage point increase in expected inflation during the zero lower bound period of 2008-2014 reduced households' probability of having a positive attitude towards spending by about 0.5 percentage points.

We update the sample used by Bachmann et al. to include the years since the paper's publishing, as well as second interviews with some respondents. This leaves us 249,549 observations from January 1984-October 2022, compared to Bachmann et al's 68,000 (each observation corresponds to a household completing a survey at a given point in time). We have access to all of the same variables as Bachmann, except that for some unknown reason it appears MSC no longer publishes race microdata. We augment this data with the federal funds rate, unemployment rate, and headline and durables CPI from the Federal Reserve Economic Data (FRED) service. We remove 31,054 observations in which the respondent refused to answer the durable purchase or predicted price change questions described below, giving us a final dataset of 218,495 observations.

The two most pivotal attributes are our treatment and outcome. The latter, $durable\_purchases$, is a categorical variable describing the respondent's self-professed perception of the favorability of buying durable goods in the next year: "Negative", "Neutral", or "Positive". Our treatment variable is derived from the MSC's $price\_change\_amt\_next\_yr$ variable, which is the respondents' prediction of the percent change in prices for the next year (denoted as π). We assign this continuous

[1] An exhaustive list of variables used can be found in the appendix

quantity to 5 categories:

$$treatment\_bins = \begin{cases} 0 & \text{if } \pi < 5 \\ 1 & \text{if } 5 \leq \pi < 10 \\ 2 & \text{if } 10 \leq \pi < 15 \\ 3 & \text{if } 15 \leq \pi < 20 \\ 4 & \text{if } \pi \geq 20 \end{cases} \tag{1}$$

Treatment bin 0 acts as our "control" group, and the remaining bins are our "treatments". We will index the treatment categories with $t = \{0, 1, 2, 3, 4\}$.

3 Models

As described previously, our main contribution to the work done by Bachmann et al. is using a more flexible non-parametric modeling process. Specifically, we use XGBoost's classifier model to implement augmented inverse probability of treatment weight (AIPTW) estimation.

We aim to estimate the average treatment effect (ATE) of inflation expectation category on categorical attitude towards durable good buying. Concretely – let A be the former and Y be the latter, with indicator variables for each category (for example, $A_1 \in \{0, 1\}$ and $Y_{Good} \in \{0, 1\}$). This results in 12 causal estimands in total, one for each combination of non-control treatment and outcome; overall $ATE_{t,Y} = E[Y|do(A_t = 1)] - E[Y|do(A_t = 0)]$, so for example our estimand for effect of treatment level 2 on "Good" purchasing attitude is $ATE_{2,Good} = E[Y_{Good}|do(A_2 = 1)] - E[Y_{Good}|do(A_2 = 0)]$.

In order to carry out causal identification, we produced a causal directed acyclic graph (cDAG) describing our a priori beliefs about the causal structure of the variables we have available. We include an abbreviated version of the cDAG below for space reasons. X represent the set of observed confounders, U represent the set of unobserved confounders, Z represent the set of instruments, and Q represent the set of factors affecting the outcome but not treatment. Based on a priori reasoning and knowledge of other existing research, we sorted each of the variables that we had available in the survey or that we could obtain from economic indicators into these sets. See Appendix for a list of variables that we sorted into X, which generally fall into three categories: macroeconomic indicators, respondent attitude on other personal and general economic conditions, and some demographics.

We can identify our ATE estimands with $\tau_{t,Y} = E_X[E[Y|X, A_t = 1] - E[Y|X, A_t = 1]]$ using a similar assumption as Bachmann et al: that there are no unobserved confounders U. (We also need overlap so $P(A = 1|X = x) \in (0,1) \forall x \in$ sample space, but direct checks of the data confirm this requirement, so we do not need to assume it.) So, we include all of X in our adjustment set. While some of the variables in X arguably may not be confounders, we erred on the side of caution, since including a non-confounder instrument or cause of the outcome would only increase variance. We also took care that no possible colliders or mediators are in X.

While it is highly unlikely that there are *no* unobserved confounders, we believe that for the most part our identifying assumption likely holds. Other than actual inflation, inflation predictions, and policy levers meant to influence inflation, it seems unlikely that many factors would influence the typical respondents' expectations of inflation that would also influence attitudes towards buying durable goods, other than general views of the economy or general pessimism/optimism. We expect these latter factors to be likely highly causative of both inflation predictions and purchasing attitude, but we believe they should be reasonably controlled for by including other questions on general economic outlook in the adjustment set.

We carry out finite sample estimation of τ via the augmented inverse probability of treatment weight (AIPTW) estimator. This relies on fitting two separate models: one predicting treatment level probabilities $g(X)$ ("propensity scores") and another predicting outcome category probabilities $Q(A, X)$. The main benefits of this estimator are its "double robustness", meaning that only one of the models need to be properly specified for the ATE to remain consistent, and its efficiency; as discussed in class, the AIPTW is the best non-parametric estimator of τ. Finally, AIPTW does not require any particular functional form assumptions, allowing us to use purely nonparametric models and therefore rely only on structural assumptions.

Propensity Score Model:

$$g(x) = P(A = 1|X)$$

For each observation i, we estimate the conditional probability of treatment for each level: $\hat{g}_t^*(x_i) = P(A = t|X, A = t \cup 0) = \frac{\hat{g}_t(x_i)}{\hat{g}_t(x_i)+\hat{g}_0(x_i)}$.

We checked the distribution of $\hat{g}_t^*(x)$ to ensure that most propensity scores do not fall near 0 or 1, which is equivalent to checking the overlap assumption as mentioned above.

We could estimate τ using only the propensity scores (and did so as a sanity check for our final results) via:

$$\hat{\tau}_{t,Y}^{g} = \frac{1}{n}\sum_i A_{t,i}\frac{Y_i A_{t,i}}{\hat{g_t^*}(x_i)} + \frac{Y_i(1 - A_{t,i})}{1 - \hat{g_t^*}(x_i)}$$

Outcome Model:

$$Q(A, X) = E[Y|A, X]$$

For each observation we estimate the probability of each outcome level $P(Q(A_i, X_i)) = \{Good, Neutral, Bad\}$.

We could estimate τ using only the propensity scores (and did so as a sanity check for our final results) via:

$$\hat{\tau}^{Q} = \frac{1}{n}\sum_i \hat{Q}(1, X_i) - \hat{Q}(0, X_i)$$

AIPTW Estimator:

$$\widehat{ATE}_{t,Y} = \frac{1}{n}\sum_i \hat{Q}(A_t = 1, X_i) - \hat{Q}(A_t = 0, X_i)+$$

$$A_{t,i}\frac{Y_i - \hat{Q}(A_t = 1, X_i)}{\hat{g_t^*}(x_i)} + (1 - A_{t,i})\frac{Y_i - \hat{Q}(A_t = 0, X_i)}{1 - \hat{g_t^*}(x_i)}$$

We also estimate the variance of each ATE and use this to assess the confidence level in each effect estimate being different from 0.

We selected the XGBoost Classifier model for both $g(x)$ and $Q(A, X)$ models. XGBoost is a library that implements gradient-boosted random forest. XGBoost iteratively minimizes a regularized objective function, adding new trees that predict the errors of prior trees that are then combined with previous trees to make the final prediction. The gradient boosting portion of "XGBoost" reflects the use of a gradient descent algorithm to minimize the loss when adding new models.

We carried out hyperparameter tuning separately for the g and Q models, with final hyperparameters listed in the Appendix. We evaluated model performance based on several criteria with a focus on accuracy (defined as fraction of test observations correctly classified). We found it was difficult to improve much on baseline accuracy (that is, frequency of the most common class), which seems to be due to the nature of the dataset; that is, even conditional on the adjustment set it seems that "Good" is typically the most likely outcome. However, this limitation is somewhat mitigated by the fact that we used predicted probabilities rather than predicted class labels.

To verify our choice of model, we compared the models to linear and neural network approaches, with the results appearing encouraging for our chosen direction. We also experimented with using multiple models to incorporate the fact that there is a natural ordering to our treatment and outcome classes (a mild functional assumption implemented as in Frank and Hall, 2001), but did not find that this resulted in significant accuracy improvement so decided not to pursue this direction.

Finally, we reviewed feature importance in all of our models for reasonableness. We include top 10 most important features for the g and Q models in the Appendix.

4 Results

We fit estimates of the ATE for each outcome/treatment pair. The results in Table 1 shows that by and large, we find a significant treatment effect.

Table 1: XGBoost Results

Outcome	Treatment	ATE	StdErr	P-Value	95% Significant?
Negative	5-10% inflation	0.022	0.003	0.000	Yes
Negative	10-15% inflation	0.041	0.006	0.000	Yes
Negative	15-20% inflation	0.041	0.008	0.000	Yes
Negative	20+% inflation	0.023	0.006	0.000	Yes
Neutral	5-10% inflation	0.0032	0.0014	0.002	Yes
Neutral	10-15% inflation	0.000	0.002	0.9	No
Neutral	15-20% inflation	0.001	0.003	0.8	No
Neutral	20+% inflation	0.003	0.002	0.196	No
Positive	5-10% inflation	-0.025	0.003	0.000	Yes
Positive	10-15% inflation	-0.042	0.006	0.000	Yes
Positive	15-20% inflation	-0.042	0.008	0.000	Yes
Positive	20+% inflation	-0.027	0.006	0.000	Yes

Each ATE is in comparison to 0-5% inflation group.

We observe that the results are largely not significant for the "Neutral" outcome variable. This is an interesting finding which could not be demonstrated using Bachmann et al's methods. There could be a fairly intuitive story behind this result: Those who feel neutral about whether it is a good or bad time to buy durable goods may be agnostic to how other economic indicators (e.g. expected inflation) might affect these spending habits. These individuals may not change their spending habits regardless of economic conditions, or may not internalize their general economic outlook when planning purchases of durable goods.

We find that expecting higher levels of inflation in the coming year causes respondents to be less likely to believe it will be a good time to buy durable goods and more likely to believe it will be a bad time to buy durable goods. Though the differences in treatment groups are not monotonic, this could be attributed to differences in the nature of respondents that answer "20+% inflation" in the next year, a figure that one might argue is an unreasonable prediction for inflation in the U.S. Overall, these are interesting results in the context of theoretical debates over the direction and magnitude of the effect of inflation expectations on purchasing behavior.

5 Conclusion

Our results may have been somewhat surprising to us (who expected no significant effect), but are not all that dissimilar to those found in Bachmann et al. Those authors found that higher expected inflation reduces households' probability of having a positive attitude towards spending by about 0.5 percentage points while monetary policy was at the zero-lower bound.

We find reasonably strong results that show that one's inflation expectations have an affect on their attitudes towards spending on durable goods. In fact, we predominantly find that there is an even stronger negative relationship between inflation expectations and willingness to spend on durables than Bachmann et al. This could due to our estimation process, as we were able to observe non-linearities in effect of treatment on the outcome.

Of course, Bachmann et al were able to carry out many more robustness checks, which would be important next steps for our work if we were to carry it further. Other further steps would include continuing to attempt to improve the accuracy of the Q and g models, perhaps by experimenting with other model families, and exploring differences in treatment effect within different respondent groups. We would also want to continue to seek out other possible adjustment variables to better exclude remaining unobserved confounders. Finally, we might attempt to take advantage of the natural ordering of treatment and outcome classes via certain choices of loss functions or model setups, as we briefly experimented with.

Despite its limitations and obviously very exploratory scope, we believe that our approach demonstrates a direction for improvement upon Bachmann et al's empirical method to answer the causal question of inflation expectations' effect on attitudes towards durable good spending.

References

Bachmann, Rüdiger, Tim O. Berg, and Eric R. Sims. (2015). "Inflation expectations and readiness to spend: Cross-sectional evidence." American Economic Journal: Economic Policy, 7 (1): 1-35.

Binder, C., Janson, W., & Verbrugge, R. (2022). Out of bounds: Do SPF respondents have anchored inflation expectations?. Journal of Money, Credit and Banking. *forthcoming*.

Dräger, L., & Lamla, M. J. (2013). Anchoring of consumers' inflation expectations: Evidence from microdata.

Gillitzer, C., Prasad, N., & Robinson, T. (2021). Political attitudes and inflation expectations: Evidence and Implications. Journal of Money, Credit and Banking, 53(4), 605-634.

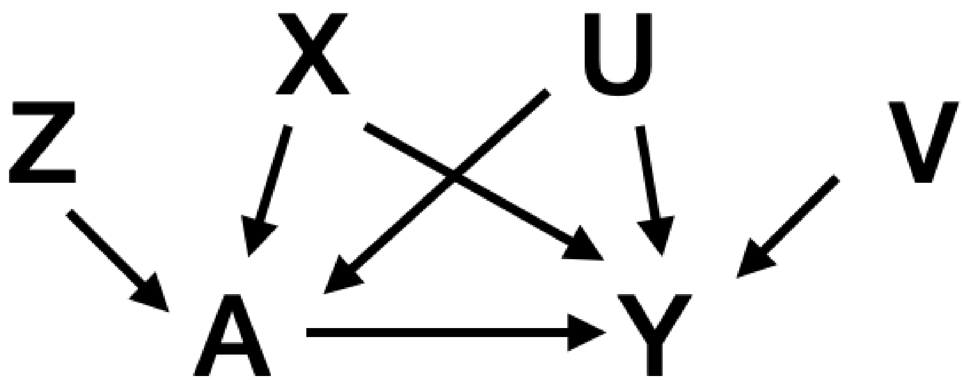
Frank, E., Hall, M. (2001). A Simple Approach to Ordinal Classification. In: De Raedt, L., Flach, P. (eds) Machine Learning: ECML 2001. ECML 2001. Lecture Notes in Computer Science(), vol 2167. Springer, Berlin, Heidelberg.

Sahm, C. R., Shapiro, M. D., & Slemrod, J. (2010). Household response to the 2008 tax rebate: Survey evidence and aggregate implications. Tax Policy and the Economy, 24(1), 69-110.

A Appendix

A.1 Causal DAG

Figure 1: Causal Structure of Problem



A.2 Feature Variables

Table 2: Observed Confounders X

Variable	Description
fed_funds_rate	Overnight federal funds rate, expressed in terms of percent.
$unemployment\_rate$	Official U3 unemployment rate for prime-age workers in U.S.
cpi_1mo_lag	Lagged CPI inflation, expressed in terms of percent.
$cpi\_durable\_1mo\_lag$	Lagged CPI inflation for durable goods, expressed in terms of percent.
$personal\_finances\_next\_yr$	Categorical variable on attitude of personal finances in the next year.
$income\_change\_amt\_next\_yr$	Categorical variable on attitude on change in personal income in the next year.
$conditions\_next\_yr$	Categorical variable on attitude regarding economic conditions in the next year.
$unemployment\_next\_yr$	Categorical variable on attitude of the unemployment rate in the next year.
$income\_quintile$	Categorical variable denoting income quantile respondent falls in.
age	Age of respondent.
sex	Sex of respondent.
$education$	Categorical variable denoting respondent's education level.
$household\_size$	Number of residents in household of respondent.
$price\_related\_yr\_ago$	Categorical response whether price changes are relevant to respondent's current financial situation.
zlb	Indicator variable denoting whether federal funds rate is at the zero-lower bound (>0.25%)

A.3 Model Hyperparameters

Table 3: XGBoost Results

Model	Loss Function	Learning Rate	Max Depth	N Estimators	Min. Child Weight
$g(x)$	Multiclass log cross-entropy	0.05	6	150	3
$Q(A, X)$	Multiclass log cross-entropy	0.07	5	120	1

Each model uses $k = 5$ folds.

A.4 Feature Importance

The following graphs display the 10 most important features in both the $Q(A, X)$ and $g(X)$ models.

Figure 2: Most important features of $Q(A, X)$ model

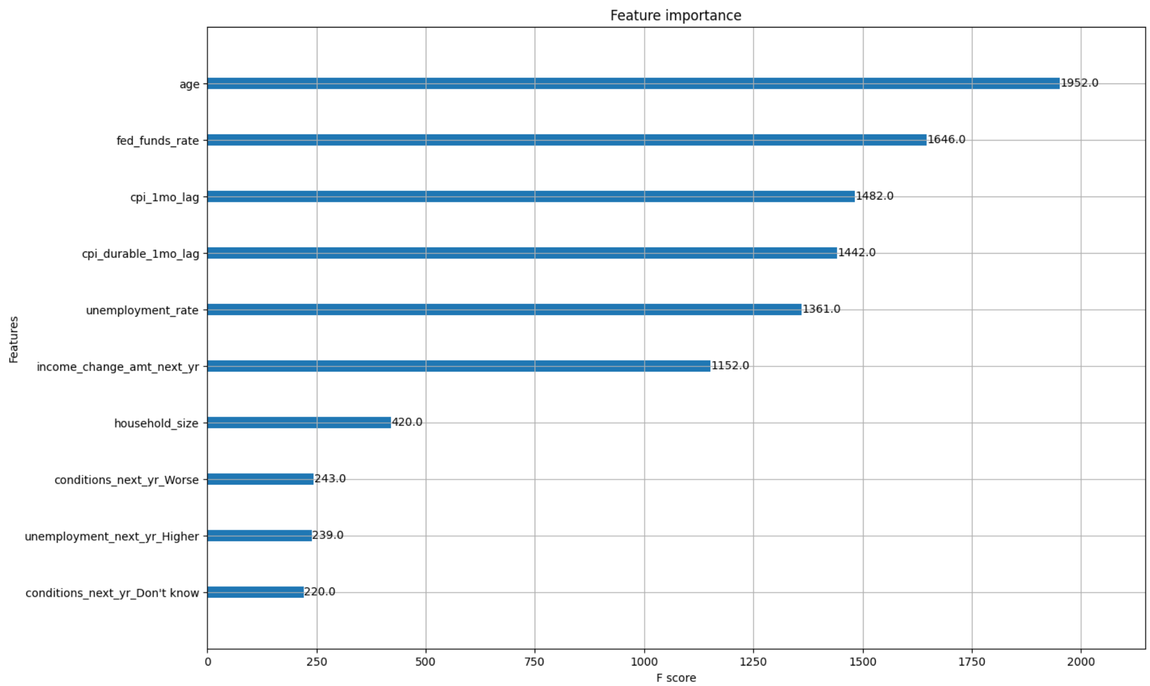


Figure 3: Most important features of $g(X)$ model