Inferential Statistics – Lending Club Data

The Jupyter notebook used to explore this data is available at the following URL:

https://github.com/mhardcastle0/Springboard/blob/master/lending_club_capstone/inferential_statistics.ipynb

Introduction

This analysis seeks to evaluate the impact of macroeconomic factors at the time that a loan is originated on the likelihood that the borrower will default at any point during the term of the loan. The null hypothesis for this analysis is that macroeconomic variables at the time of origination have no impact on loan default likelihood; the alternative hypothesis is that there is a non-zero impact.

The economic variables evaluated are the US unemployment rate during the month of origination, the number of bankruptcies filed nationally during the month of origination, and the change in the S&P 500 stock index from the beginning of the month of origination to the end of the month of origination.

Methods

To estimate the impact of each macroeconomic variable on the loan default rate, univariate OLS regression was used. To calculate confidence intervals, bootstrap replication was used to re-sample the data 10,000 times for each subgroup.

When Lending Club originates a loan, it is assigned a grade, from A through G, where A is expected to be the highest-risk and G the lowest-risk. Grade G loans are excluded from this analysis, as very few are originated. Loans can have either 36- or 60-month terms.

This analysis treats each grade-term combination as its own subgroup, evaluating, for example, grade A 36-month loans separately from grade B 36-month loans. This is done for two reasons. Firstly, changes in portfolio composition over time might change portfolio-wide default rates; if there are more F-grade loans in later periods than earlier periods, then portfolio-wide default rates would likely increase over time regardless of macroeconomic impacts. Secondly, macroeconomic impacts might vary by risk tier—higher-risk borrowers may be more impacted by the unemployment rate than lower-risk borrowers, for example.

Results

The table in the appendix shows the confidence interval and sample size for each specification.

The large majority of results are not statistically significant. Of those that are significant, all results that correlate the unemployment rate with charge-off rates have the *opposite* sign than expected - a lower unemployment rate at the time of origination is correlated with a higher default likelihood. Similarly for the bankruptcy rate - lower monthly national bankruptcies at origination are associated with higher default rates.

There are two significant results correlating the monthly change in the S&P 500 with default rates, but one term is positive and the other negative, suggesting low generalizability of any results.

Discussion

As mentioned, the signs on nearly all statistically significant coefficients are the opposite of what might be expected. There may be a variety of reasons that this may be the case, including:

1. Improving Economic Conditions - Both the unemployment rate and the monthly bankruptcy rates have decreased significantly in the time period analyzed (2010 to 2015). It is possible that what matters for loan default rates is the *improvement* in economic conditions rather than their absolute state. As high unemployment and bankruptcy rates for the time period analyzed are followed by rapid decreases in their respective rates, poor economic conditions may predict lower default rates.
   If this is the cause of the observed results, then evaluation of loan performance over larger periods would be required, as larger periods would include more macroeconomic fluctuations - the period evaluated is characterized by constant and relatively rapid macroeconomic improvement, which is not generally expected.
2. Independent Variable Selection - If changes in economic conditions are more important than baseline economic conditions, then it is possible that independent variables with low explanatory power were used for this analysis. For example, using the *change in the unemployment rate* over the prior 12 months as the independent variable may be more informative than using the point-in-time unemployment rate at the time of origination.
   It is also possible that there are macroeconomic variables that impact loan default rates that were not analyzed, including average wages or GDP per capita, stock indices other than the S&P 500, or other proxies for macroeconomic performance.
3. Lending Club Risk Modeling - It could be that Lending Club already accounts for macroeconomic factors in their loan originations and reject borrowers with high risk profiles when the economy is doing poorly. If Lending Club does so, and overestimated the impact of macroeconomic factors on borrower default probability, then the observed results are expected.

Appendix – Confidence Intervals for Each Subset

The table below shows the confidence interval for the slope term for each variable of interest, broken out into each subgroup tested. Terms that are significant at the .05 level are in bold.

Unemployment Rate %

Term	Grade	95% CI Lower Bound	95% CI Upper Bound	Sample Size
36 months	A	-0.00135949	0.001641	60795
36 months	B	**-0.00468227**	**-0.00099431**	90624
36 months	C	**-0.01494143**	**-0.00967629**	68178
36 months	D	**-0.02550619**	**-0.01771471**	34891
36 months	E	**-0.04220207**	**-0.02550099**	9765
36 months	F	**-0.07110762**	**-0.02906948**	2113
60 months	A	-0.02670311	0.02487238	624
60 months	B	-0.00647652	0.02579703	4049
60 months	C	-0.00800087	0.02261887	6081
60 months	D	-0.02629441	0.012933	4112
60 months	E	**-0.04510762**	**-0.00650016**	4500
60 months	F	**-0.0741704**	**-0.01809013**	2380

National Monthly Bankruptcies (per 1,000)

Term	Grade	95% CI Lower Bound	95% CI Upper Bound	Sample Size
36 months	A	-7.85E-05	1.30E-04	60795
36 months	B	-2.40E-04	3.26E-05	90624
36 months	C	**-0.00095931**	**-0.00058006**	68178
36 months	D	**-0.00161385**	**-0.0010405**	34891
36 months	E	**-0.00260701**	**-0.00144237**	9765
36 months	F	**-0.00435358**	**-0.00160429**	2113
60 months	A	-0.00123861	0.00151179	624
60 months	B	-0.00044933	0.00099869	4049
60 months	C	-0.00084185	0.00061737	6081
60 months	D	-0.00137703	0.00043214	4112
60 months	E	**-0.00211259**	**-0.00030859**	4500
60 months	F	**-2.79E-03**	**-9.35E-05**	2380

Monthly Change in S&P 500 (EOM value minus BOM value)

Term	Grade	95% CI Lower Bound	95% CI Upper Bound	Sample Size
36 months	A	-8.49E-05	9.35E-06	60795
36 months	B	**-1.23E-04**	**-1.88E-05**	90624
36 months	C	-1.12E-04	3.08E-05	68178
36 months	D	-1.45E-04	7.54E-05	34891
36 months	E	-8.70E-05	3.80E-04	9765
36 months	F	-5.19E-04	0.00059952	2113
60 months	A	-4.22E-04	0.00090613	624
60 months	B	-3.11E-04	0.00028922	4049
60 months	C	-1.94E-05	5.79E-04	6081
60 months	D	-3.90E-04	0.000366	4112
60 months	E	-4.04E-04	0.00034348	4500
60 months	F	**4.41E-05**	**1.11E-03**	2380