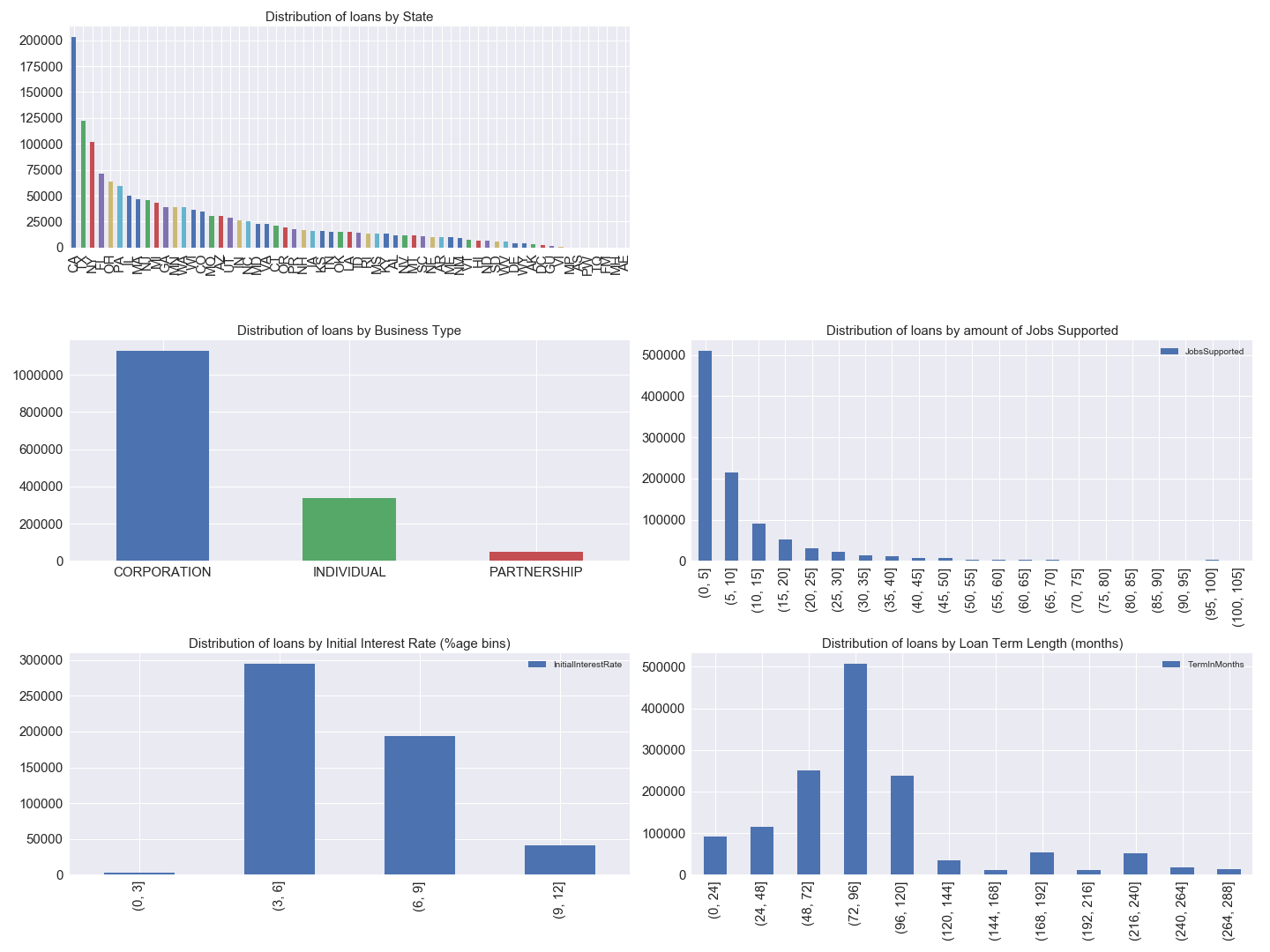
Before proceeding, lets get a few assumptions made out of the way:

1. Pool Construction: We assume loans start entering our holding portfolio from 1990-10-01 (since SBA data is available from 1990-10 forward). As loans amortize, the amount of balance of loans in our pool also decreases
2. Definition of Loss Rate: We implement an open pool loss rate calculation. The open pool approach is a non-cohort approach in which charge-offs are tracked over a period of time, with no tracking of activity or balances of specific loans. Essentially we just calculate the rolling 12 month loss rate
3. Loss Rate Difference:
4. Loans are considered “Seasoned”: when they cross the half way point of their term length
5. We will attempt to provide an estimate or trend of where the loss rate increases are going to be for another period of time that some what mimics a recession.

We have loan level data that dates back from 1990-10 forward. Let us first take a look at the distribution of loans originated and the charge-offs experienced.

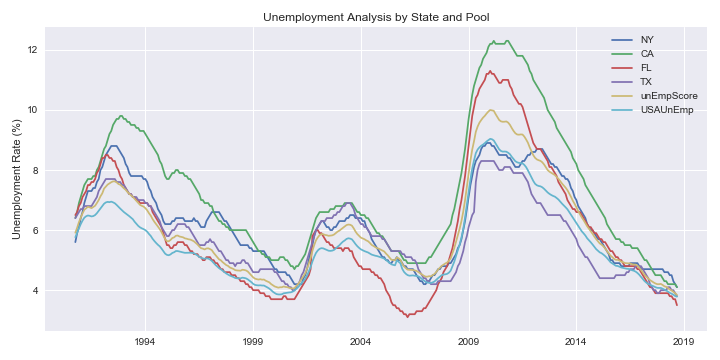


Some of the key takeaways are:

1. There are a few states (CA,NY,FL,TX) that originate the most loans and charge-offs (obviously overlapping). This gives us intuition that some measure of how the state itself is doing may impact the loan pool loss rate depending on how diversified by state the pool is.
2. We notice that most of the loans charged off are “Seasoned”.
3. Another way to dissect the loan pool is by business type, i.e corporation vs individual.

**How to use the State Feature of Loans:**

We need an economic variable that we can use to judge the performance of the State itself. This is as if a pool of loans is heavily based out of a State that has failing macroeconomic outlook in comparison to other states, the loss rate of the pool will be higher than otherwise better geographically distributed loan pool. We use State level *Unemployment* *Rates* at a monthly level. We approximate how much of our portfolio of loans alive at every month are distributed by state (proportion). We then cross multiply loan proportion by state with the unemployment score by state every month to receive a unemployment score for our pool of loans every month (unEmpScore in the graph below).



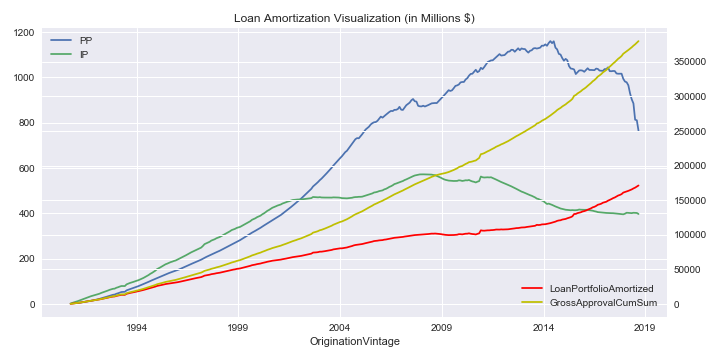
We now look at the loan level and charge off data by on a monthly vintage basis. We summarize our data in a dataframe called “summaryDF”. We discuss what it contains and give a visual representation of some of the important items it contains.

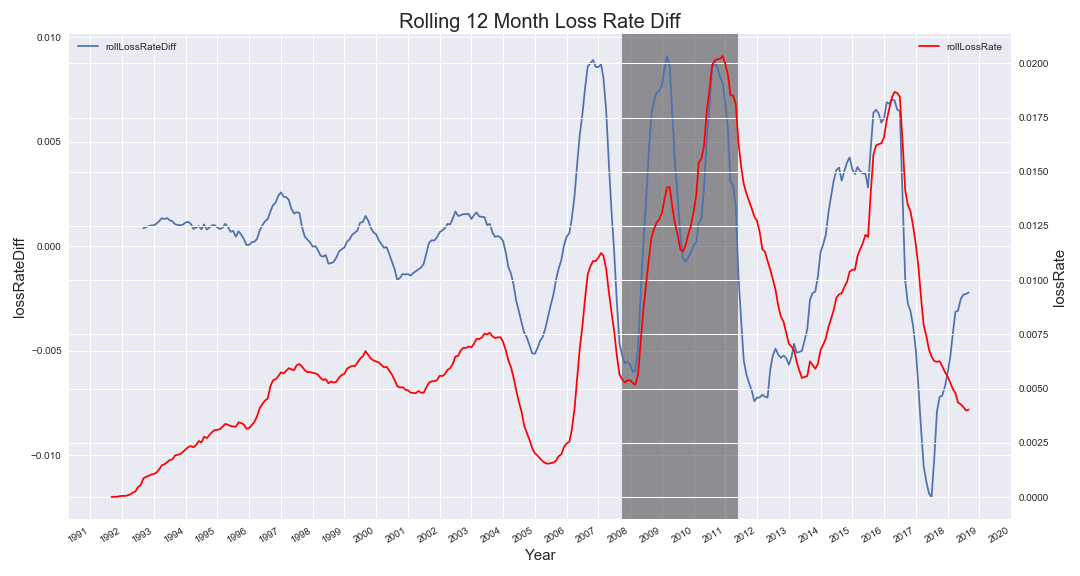
|  |  |
| --- | --- |
| **Column** | **Description** |
| GrossApproval | Sum of loan principal originated by monthly vintage |
| SBAGuaranteedApproval | Sum of loan principal originated by monthly vintage guaranteed by SBA |
| CORPORATION | Count of loans originated with business Type as CORPORATION that vintage |
| INDIVIDUAL | Count of loans originated with business Type as INDIVIDUAL that vintage |
| PARTNERSHIP | Count of loans originated with business Type as PARTNERSHIP that vintage |
| TermInMonths | Weighted Average of Term Length of loans originated (weighted by GrossApproval Amount) |
| InitialInterestRate | Weighted Average of InitialInterestRate of loans originated (weighted by GrossApproval Amount) |
| JobsSupported | Weighted Average of JobsSupported of loans originated (weighted by GrossApproval Amount) |
| seasonedCount | Count of loans seasoning every vintage (i.e how many loans have crossed the half way point of their term length) |
| MPRIME | Bank Prime Rate for that month as taken from the FED (seperate data source) |
| PP | Amount of Principal of loans being paid back that month |
| IP | Amount of Interest of loans being paid back that month |
| CORPORATIONCount | Count of loans in the pool that are active (i.e not fully amortized) that are of CORPORATION Type |
| INDIVIDUALCount | Count of loans in the pool that are active (i.e not fully amortized) that are of INDIVIDUAL Type |
| PARTNERSHIPCount | Count of loans in the pool that are active (i.e not fully amortized) that are of PARTNERSHIP Type |
| LoanPortfolioAmortized | The amortized value of the loan pool (taking into account the principal repaid till date) |
| GrossApprovalChargedOff | Sum of loan principal (Gross Approval) of loans by month (not neccesarrily equal to the amount of charge off) |
| SBAGuaranteedApproval ChargedOff | Sum of loan principal charged off by month guaranteed by SBA |
| GrossChargeOffAmount | Amount charged off every month |
| CORPORATIONChargedOff | Count of loans charged off with business Type as CORPORATION that month |
| INDIVIDUALChargedOff | Count of loans charged off with business Type as INDIVIDUALthat month |
| PARTNERSHIPChargedOff | Count of loans charged off with business Type as PARTNERSHIPthat month |
| TermInMonthsChargedOff | Weighted Average of Term Length of loans charged off (weighted by GrossChargeOffAmount Amount) |
| InitialInterestRateChargedOff | Weighted Average of InitialInterestRate of loans charged off (weighted by GrossChargeOffAmount Amount) |
| seasonedTime | Weighted Average of time the loans were into their lives at the time of charge off(weighted by GrossChargeOffAmount) |
| JobsSupportedChargedOff | Weighted Average of JobsSupported of loans charged off (weighted by GrossChargeOffAmount Amount) |
| rollSumChargeOff | Rolling 12 month sum of amount charged off |
| rollAvgPortfolioAmortized | Rolling 12 month mean of LoanPortfolioAmortized |
| rollLossRate | rollSumChargeOff/rollAvgPortfolioAmortized |
| rollLossRateDiff | YoY (Year on Year) difference of rollLossRate ~ THIS IS WHAT WE ARE INTERESTED IN |
| unEmpScore | unemployment score of the loan pool that is still active every month |
| unEmpScoreDiff | change of unEmpScore MoM (Month on Month) |
| seasonedCount | Count of Loans that are seasoned |
| LoansInPool | Total number of loans alive in pool |
| seasonedRatioPool | seasonedCount/LoansInPool |
| CorporationRatio | Loans in Pool by Corporation type/LoansInPool |
| IndividualRatio | Loans in Pool by Individual type/LoansInPool |
| PartnerShipRatio | Loans in Pool by PartnerShip type/LoansInPool |

\*\*\*We build our own amortization schedule of the loan pool assuming payments are done at the end of month. This derives our PP,IP and LoanPortfolioAmortized Values. LoanPortfolioAmortized is essentially the PV of the Principal, Interest and Charged off Amounts that are pending at the end of every month.

\*\*\*From 1991-2008 no InitialInterestRate Data was available, so instead we just took the Bank Prime Rate from the FEDs Website

Below I give a representation of how the amortized loan value of the portfolio changes with time, along with principal and interest payments received. This shows how the amortized value of the loan pool grows against cumulative sum of loan principal originated (GrossApprovalCumSum) We also look at PP and IP (principal and interest payments received from pool every month). I also show how our Rolling Loss Rate and Rolling Loss Rate YoY Difference looks (Shaded region is 2007-10 to 2011-06 ~ Recession+post Recession)



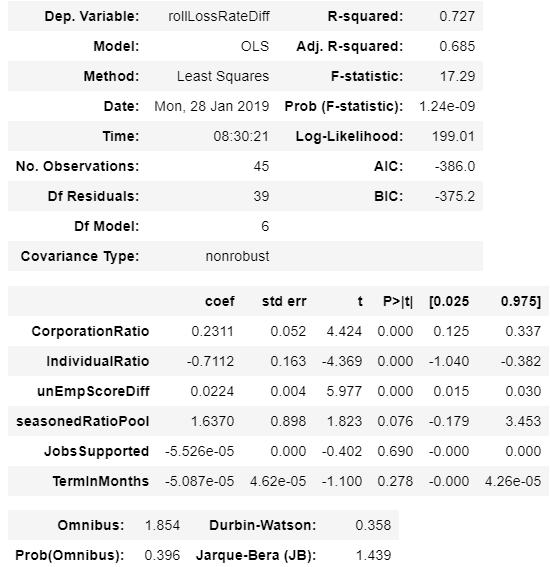


**Loss Rate Estimation**

We train an OLS model to predict the rollLossRateDiff on the recession period where the features are the ones identified below:

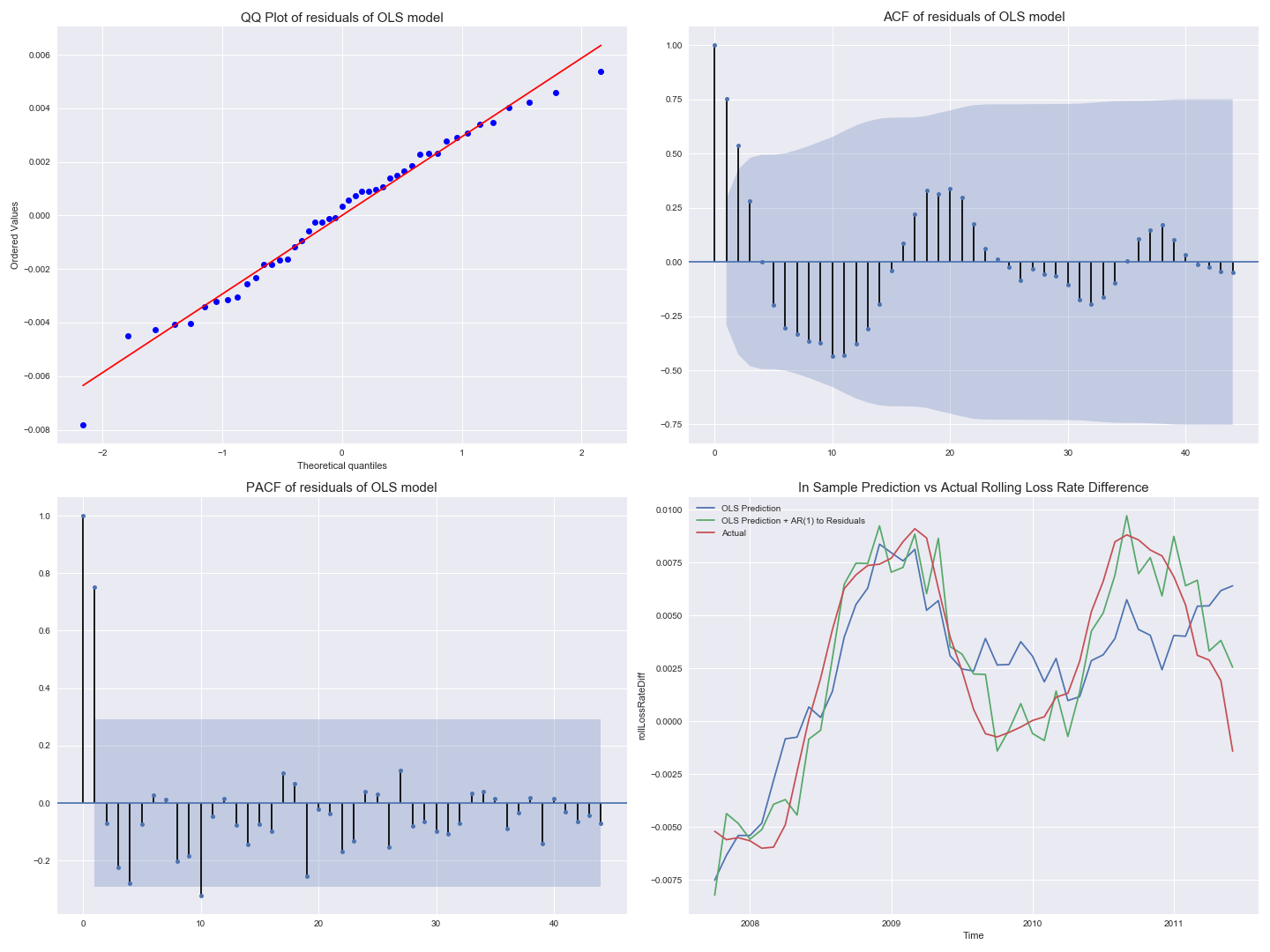
1. seasonedRatioPool: seasonedCount/LoansInPool
2. CorporationRatio: Loans in Pool by Corporation type/LoansInPool
3. IndividualRatio: Loans in Pool by Individualtype/LoansInPool
4. unEmpScoreDiff: MoM change of the unemployment score of the loan pool
5. JobsSupported: Average of Jobs Supported by loans in pool active now
6. TermInMonths: Average of TermInMonths of loans in pool active now

The OLS results are summarized below:

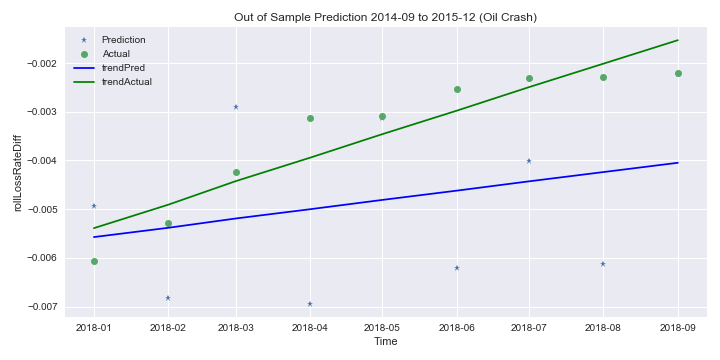
We do a residual Diagnostic test as well (QQ plot) along with Shapiro Wilk Test that yields a p-val of .67, hence we do not reject the assumption that the errors are normally distributed.

Since the Durbin-Watson stat =.32, we are pretty sure the errors have positive autocorrelation. Hence we also look at the PACF to identify the AR model we can fit into or residuals. We clearly otice AR(1) would fit best.

After fitting an AR(1) model to the residuals, we look at our predicted loss rate changes with the OLS estimates only and the ones with AR(1) added to correct for error autocorrelation.



We have essentially trained a model on small but for sure a period that suffered a recession, and not just any recession, one which was credit related. We can try to use this OLS on another Out Sample Time Series data that we think mimics recession time. The closest guess is 2014-09 to 2015-12, a period of time when the oil price took a huge hit and we actually do see a spike in rollLossRateDiff and rollLossRate during that period. We however should note that the Oil Crash of 2015 still doesn't mimic the recession as it did not come with a spike in credit risk as the 2008 Great Recession did. Nevertheless, the results of our prediction and actual values of rollLossRateDiff are shown below as well.



We look at the correlations of these features with rolling loss rate change during a recission period (2007-10 to 2011-06) and a non recession period (2002-01 to 2007-06). (These are shown in the python notebook attached).