프로젝트 기획서

1. 주제

가계 부채와 금리와 상관관계 및 은행의 대출 리스크 분석

은행의 대출 리스크-> 신용위험

금리 변화에 따른 은행의 신용위험 변화

신용위험을 연체율로 보겠다.

연체율에 영향을 주는 요인: 금리, 물가, 경기선행지수, 시간당 소득, 최저임금

경기 국면 변화에 따른 은행 신용 위험 분석

1. 개요(모티베이션)

최근 코로나의 영향력이 감소하고 예상치 못한 국제 정세 등의 변화, 심각한 기후 위기 등으로 전세계적인 물가 상승 추세가 나타나고 있다. 이런 현상에 대해 각국의 중앙은행들은 기준금리를 물가상승 추세에 대응하여 올리고 있다. 즉, 사실상 금융위기 이후 이어져온 양적 완화의 시대에서 긴축의 시대의 전환이 일어나고 있는 것이다. 이런 상황에서 주요 언론에서 가계 대출의 위험성에 대한 경고 기사를 쏟아내고 있다. 이에 대해서 직접 가계부채와 금리와의 상관관계를 분석하고 향후 은행의 리스크가 어떻게 변화해 나갈 것인지 확인하고자 한다.

경기에 따라 금리 외의 다양한 경제 지표에 의해서 연체율이 올라갈 것이로 생각하기 때문에 다양한 요인을 같이 고려하고자 한다.

1. 목적

가계의 부채와 상관관계가 높은 금리를 선정하여 향후 금융위기 이전 수준 이상으로 금리가 상승할 경우 가계 부채의 연체율이 어떻게 변할 것인지 확인한다.

경제 펀더멘털 중 상관관계가 높은 요소를 선택하여

* 다른 지표도 표현하면서 최종적으로 금리의 방향으로 간다면 우리 가설과 맞는 방향이긴 할 것이다.

1. 팀원 및 역할 분담

* 팀장: 박준배
* 경제 도메인: 조해원
* PPT제작: 박희연
* 기록: 김민성
* 코드 구현: 박준배

1. 조사 기간

* 교수님 질문 연계 예정

분석하고자 하는 분석 방법에 대한 이해를 깊이 해야한다.

분석 방법이 무엇이든 기본 추정이 있다.

상관분석의 기본 추정- 선형성, 정규성 (피어슨 상관분석) 아닌 경우 켄달이나 스피어만

피어슨 제시 경우 히트맵, scatplot -> 시각적 전달

회귀 -> 선형성, 정규성 + 독립성, 등분산성

최소 10년 이상

원하는 결과가 나오는 기간을 설정하는 것도 좋다.

기간에 대해서 여러 구간을 나눠서 진행하되(ex:기본 10년 데이터 연구 진행시 -> 5년, 3년 등) 시간이 부족하다면 추후 연구로 제시함

특정 검증을 하지 못한 경우 못했다고 고백하면 된다.

(,ex: 로버스트 검증)

1. 프로젝트 구성

* 프로젝트 주제 설정 및 전개
* 가설 설정 및 검증
* 분석 대상 정의 및 조사 기간 설정
* 데이터 수집
* 데이터 전처리 및 분석
* 모델 설정 및 결과 도출

1. 주요 가설
2. 은행의 전체 대출 대비 가계 대출의 비중이 크기 때문에 해당 대출의 리스크를 무시할 수 없을 정도일 것이다.

* 실제 한국은행 자료를 통해 은행 전체 대출 대비 가계의 부동산 담보 대출 비중이 37.4%에 이르는 것을 확인

1. 금리는 장기적으로 상승할 것이다.
2. 기존: 코로나 이전의 금리수준의 회복을 가정-> 이미 회복

수정: 세계 금융위기 이전의 금리 수준의 회복을 가정

가설(귀무가설): 원인과 결과의 상관관계가 없다.

가정(대립가설): 원인과 결과의 상관관계가 있다.

다 지워라.

부득이한 경우 검증이 불가능한 경우 가정을 사용한다.

1. 모델링
2. 피처선정: 독립변수 설정

방법: 전진, 후진, stepwise, lasso

1. 분석 방법
2. 상관분석: scatterplot, heatmap, 상관계수, -> insight 제시
3. 시차 상관분석:
4. 회귀분석: 다중회귀분석,

경제지표들은 시계열 데이터이기 때문에 시계열 데이터를 쓸때는 자기 상관을 검증해야 한다.

자기상관이 높다면 staionary해야한다-> 분산 일정(차분)

1. 모델링 평가
2. 회귀 rmse,
3. 분류 f1\_score
4. 분석 대상 및 모델
5. 분석 대상: 가계 대출, 금리, 연체율
6. 가계 대출

* 부동산 담보대출로 한정

-크게 담보대출과 신용대출로 대출 종류가 구분되지만 신용 대출의 경우 개인의 신용 스프레드의 영향력이 담보 대출보다 크기 때문에 실제 분석에 어려움이 있어 상대적으로 신용 스프레드 영향력이 적은 부동산 담보 대출로 한정한다.

* 여신 취급 방식에 따른 대출을 구분하여 분석 가능할까?

- 실제로 이 자료자체가 찾기 힘든 상황이다.

- 현행 가계대출 상품 중에서 고정금리, 변동금리 상품의 비중 정도만 알 수 있다.

- 교수님 질문사항 연계.

-> 찾아보니까 2020년 주담대에서 사실상 97%정도가 변동금리다.

1. 금리

* 주요 시중은행 조사 결과 COFIX 신규물, 금융채 5년물, CD 91일물 수익률을 주택담보 대출의 기준금리로 사용하는 것을 확인
* 금융채: 무보증 은행채 AAA 5년물의 평균 금리로 활용
* 관련 피드백
* 기본적으로 금리의 영향력은 다음과 같이 발생한다.

미국기준금리(가능한 경우 조사 대상에 추가하자….)->

RP금리(한국은행 기준금리) -> CD금리 -> COFIX금리 -> 금융채 금리

1. 연체율

* 한국은행에 나와있는 가계 대출 연체율로 분석하기에는 적절성이 조금 떨어지는 것 같다.
* 교수님 질문사항 연계

1. 분석 방법
2. 기본적으로 상관관계 분석

* 기획안 발표 이후 시계열 상관관계 분석으로 전환

- 교수님 질문사항 연계

1. 프로젝트 진행 예상
2. 1주일 동안 가계부문 프로젝트 코드화 및 결과 도출
3. 남은 인원은 기업부문 프로젝트 구체화 및 자료 조사
4. 남은 기간 동안 기업 부문 프로젝트 마무리
5. 교수님 질문사항 정리

* 조사 기간 관련 질문사항

1. **금융 위기 이전의 금리수준으로 올라갔을 때, 은행의 리스크를 확인한다고 했을 때, 조사 기간을 어떻게 설정 하는 것이 적절한 것인가?**

**(금융위기 이전~조사 가능한 오래된 시기)**

1. **조사 기간을 특정 기간을 구분하는 것이 좋다는 조언을 해 주셨는데 그러면 특정 기간별 은행의 리스크를 파악하는 것이 맞는가?(가령 5년 간격으로 은행의 리스크를 파악한다고 할 때, 그 리스크의 변동이 어떻게 나타나는지를 확인?)**

* 금리 관련:

1. **한국은행 조사 자료에서 가계의 부동산 담보 대출의 총액 정도는 알 수 있지만 고정금리나 변동금리인 대출의 비중은 알 수 없었다.**

**-> 찾아보니까 2020년 주담대에서 사실상 97%정도가 변동금리다.(50%정도 변동, 나머지가 혼합)**

**현행 가계대출 상품 중에서 고정금리, 변동금리 상품의 수 정도만 알 수 있다.**

**교수님 면담까지 고정, 변동 금리 상품의 비중을 알 수 없다면 프로젝트 진행 방향에 대한 조언을 구한다.**

1. **RP금리를 알 수 없어서 한국은행 기준금리로 조사했는데 이게 괜찮을까요?**

* 연체율 설정 관련 질문사항

1. **한국은행 가계대출 연체율은 알 수 있지만 가계의 부동산담보 대출의 연체율은 알 수 없었다. 이런 상황에서 연체율을 그대로 사용해도 되나요?**

* **전체 연체율을 한번에 고려하는 것이 좋긴 하다. 여기서 대출별로 연체율을 구분하는 것이 좋다.**
* **여기서 가정을 통해서 연체율의 변화율을 위험으로 생각한다.**

1. **전체 가계대출 대비 부동산 담보 대출 비중은 알 수 있다. 그렇다면 몇 % 수준이 되면 가계대출 연체율을 부동산 담보 대출 연체율 대신에 사용이 가능한가요?**

**(2021 전체 가계대출: 6.5조, 주택담보 대출: 5.7조)**

1. **혹시나 하고 여쭙는데 가계의 부동산 담보대출 연체율 자료를 갖고 계신가요?**
2. **위의 연체율 데이터가 적절하지 못하고 부동산 담보 대출의 연체율을 알 수 없는 경우 리스크에 대한 정의를 총 가계대출 중 부동산 담보 대출의 비중으로 하려고 하는데 이것이 적절한가요?**
3. **이것이 적절하지 않다면 리스크에 대한 정의에 대해서 도움을 받고 싶습니다.**

* 분석 방식에 대한 질문사항

1. **시계열 상관관계 분석에 대한 학습자료 요청**
2. **현재 수준에서 가능한지(?) 질문**
3. **파이썬 구현 방법????**

* 이거 끝나고 바로 기업 분석 가능할까요?????
* 피처: 물가(cpi), 경기심리지표(원계열 데이터), 최저임금, 환율(달러환율), 국제유가(사우디), gni(국민총소득의 측면), gdp(총생산 측면에서 연체율과의 상관관계)
* 물가: 물가와 연체율의 상관관계
* 경기심리지표(원데이터): 심리적 요인과 연체율의 상관관계
* 최저임금: 소득과 연체율의 상관관계
* 환율(달러): 환율과 연체율의 상관관계
* 국제유가(두바이유):
* Gni: 국민총소득과 연체율의 상관관계
* Gdp: 국내총생산과 연체율의 상관관계
* 유동성(m2 )

변수의 파급 효과를 고려해서 골라야 되는가? 아니면 그냥 피처 대상을 다 분석하면 되는가?