

[3주차] 방학 스터디 최종 보고서

작성일: 2026년 2월 X일

팀 명: 핀셋(fin-set)

팀 원: 김예정 김한별 이유준 이지은

1. 주요 결과물 요약

1.1. 데이터 개요 및 분석 인사이트

본 프로젝트는 2010년 12월부터 2025년 12월까지 약 15년간의 한국과 미국 주요 거시경제 지표를 수집하여 원달러 환율 변동의 구조적 원인을 심층 분석함. 수집된 핵심 지표는 양국의 소비자물가지수, 광의통화, 산업생산지수, 기준금리 및 10년물 국채금리 등임. 거시경제 데이터 특성상 발생하는 변수 간의 강한 상관관계를 통제하고, 특정 시점에 발생하는 구조적 변화를 포착하는 데 주안점을 둠.

전체 분석 기간을 거시경제의 안정적 추세가 유지되던 정상 구간과 환율이 급등한 이상 구간으로 분할하여 분석을 수행함. 정상 구간인 2024년 10월 이전까지 약 14년간 환율의 장기적 대세 흐름을 결정지은 펀더멘털은 양국의 금리차가 아닌 미국의 물가와 한국의 통화량이었음. 이 시기 한미 금리차는 환율과 뚜렷한 음의 상관관계를 보이며 한국 금리가 상대적으로 높을 때 환율이 하락하는 교과서적인 환율 방어 기제로 정상 작동함.

그러나 2024년 11월부터 2025년 12월까지의 이상 구간에서는 기존의 금리 방어력이 완전히 무력화되는 디커플링 현상이 확인됨. 한미 금리차와 환율의 상관관계가 양으로 역전되어, 한국이 금리 우위를 점하더라도 환율이 동반 상승하는 비정상적인 장세가 전개됨. 이에 머신러닝 분석을 도입하여 원인을 추적한 결과, 시중에 급격히 풀린 한국의 초단기 부동자금이 환율 급등을 주도한 핵심 원인임을 규명함. 1년여 만에 173조 원이 넘는 신규 유동성이 공급되었고, 이 중 60% 이상이 언제든 투기적 자산이나 달러 매수로 이동할 수 있는 수시입출식 예금 및 요구불예금 형태였음이 확인됨.

1.2. 활용 모델 및 고도화 계획

전통적인 선형 회귀 모형은 거시경제 지표 간의 다중공선성 문제를 해결하기 어렵고 변수 간의 복잡한 상호작용을 포착하지 못하는 한계가 있음. 이를 극복하기 위해 의사결정나무 기반의 앙상블 기법인 랜덤 포레스트 모델을 도입함. 비선형적인 패턴을 학습하는 데 탁월한 성능을 보이며, 이상 구간에서 발생한 경제 지표 간의 비정상적 관계를 포착하는 데 적합함.

또한 트리 기반 블랙박스 모델의 예측 결과를 논리적으로 해석하기 위해 SHAP 기법을 활용함. 각 변수가 환율 예측에 미친 한계 기여도를 개별 데이터 포인트 단위로 산출함으로써, 왜 금리가 아닌 통화량이 이상 구간의 1위 상승 동력으로 지목되었는지 명확한 근거를 제시함. 차후 본 학기에는 짧은 기간 내에 폭발적으로 유입된 유동성과 환율 사이의 비선형적 임계점 가설을 검증하기 위해 동적 상관분석과 시계열 특화 딥러닝 알고리즘을 추가 도입할 계획임.

2. 이론적 배경 및 심화 학습

2.1. 전통적 환율 결정 모형의 한계 고찰

외환시장의 메커니즘을 설명하는 가장 대표적인 이론인 이자율 평가설은 자본 이동이 자유로운 시장에서 양국 간의 이자율 차이가 환율의 기대 변동률과 일치해야 한다는 가정을 전제로 함. 즉, 한국의 금리가 미국보다 높다면 더 높은 수익률을 좋아 외국 자본이 유입되고, 이는 원화 수요 증가로 이어져 환율이 하락해야 함. 이론을 검증하기 위해 장기 데이터로 상관관계 분석을 돌린 결과, 아래와 같은 히트맵 결과가 도출되었고 이는 이론이 실제로 유효함을 검증함.

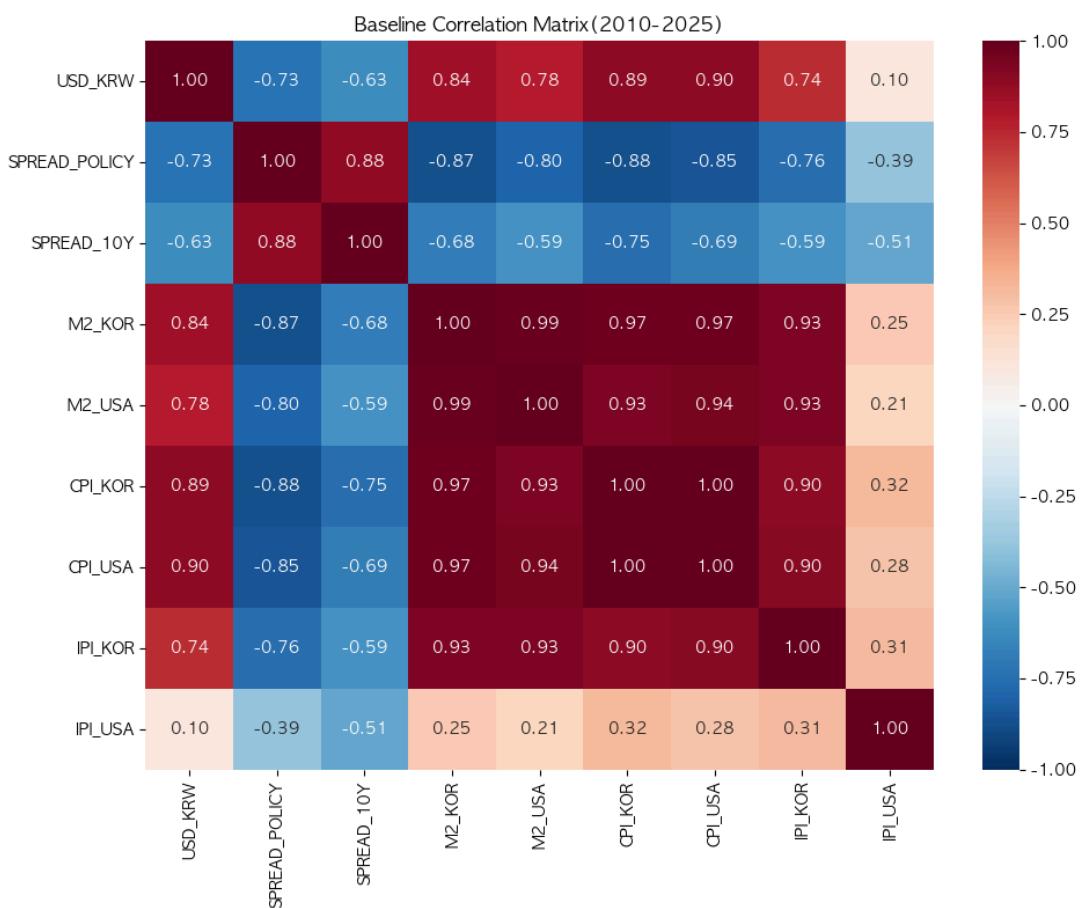


그림 1: 환율 및 주요 변수 상관관계 히트맵

그러나 최근 1년여간의 이상 구간에서는 금리를 인상하거나 높은 수준으로 유지했음에도 환율이 하락하지 않는 현상이 발생함. 이는 시장의 관심이 단순한 이자 수익률의 격차에서 벗어났음을 시사함. 글로벌 경제의 불확실성이 극도로 커지면서, 자본은 이자율이 아닌 안전자산 선호 심리에 따라 이동하기 시작함. 이 과정에서 금리 조절을 통한 전통적인 통화정책의 환율 방어 메커니즘이 훼손된 것임.

이를 보완하기 위해 화폐수량설에 기반한 유동성 관점을 심화 학습함. 시장에 공급된 통화량이 경제의 실물 산출량 대비 과도하게 많아질 경우 자산 가격이 왜곡되며 자국 통화의 가치가 절하됨. 본 분석에서는 한국 통화량의 증가가 원화 가치 하락, 즉 원달러 환율 상승의 직접적 원인으로 작용했음을 경제학적 관점에서 재해석함. 특히 장기 투자처를 찾지 못한 잉여 현금이 단기 예금에 머물다 외환시장으로 유입되는 투기적 경로를 주요 작용 기제로 상정함.

2.2. 랜덤 포레스트 모형의 수학적 원리

거시경제 변수들은 본질적으로 서로 강하게 연관되어 있어 단일 선형 회귀식에 투입할 경우 분산이 과도하게 커지는 문제가 발생함. 이를 통제하기 위해 의사결정나무를 여러 개 생성하여 그 결과를 평균 내는 배깅 기법의 일종인 랜덤 포레스트를 채택함. 랜덤 포레스트는 데이터를 샘플링할 때 중복을 허용할 뿐만 아니라, 노드를 분할할 때 사용할 설명 변수의 집합도 무작위로 선택함. 이를 통해 개별 나무 간의 상관관계를 낮추고 모델 전체의 분산을 줄여 예측의 안정성을 확보함.

분할 기준으로는 지니 불순도나 평균 제곱 오차의 감소량을 사용함. 부모 노드에서 자식 노드로 분할될 때 환율의 분산이 가장 크게 감소하는 지표를 찾아 데이터를 나눔. 이 과정을 통해 금리차 지표가 예측 오차를 줄이는 데 기여하는 정도와 통화량 지표가 기여하는 정도를 정량적으로 비교할 수 있으며, 전체 모델에서 각 변수의 전역적 중요도를 산출하는 기반이 됨.

2.3. SHAP을 통한 모델 해석론

랜덤 포레스트는 높은 예측력을 보이지만 내부에 수백 개의 트리가 존재하여 어떤 이유로 특정 환율 값이 도출되었는지 직관적으로 알기 어려움. 이를 해결하기 위해 협조적 게임 이론에 기반한 샐플리 값을 모델 해석에 도입함. 샐플리 값은 전체 변수들이 함께 환율을 예측하는 과정에서, 특정 변수가 추가되었을 때 발생하는 예측값의 변화량을 모든 가능한 변수 조합에 대해 평균 낸 값임.

본 스터디에서는 TreeSHAP 알고리즘을 적용하여 연산 효율을 높였음. 기존의 변수 중요도 지표가 단순히 모델 학습 시 불순도를 얼마나 줄였는지만을 보여준다면, SHAP은 특정 시점의 데이터에 대해 해당 변수가 환율을 상승 시켰는지 하락시켰는지 방향성과 크기를 동시에 제공함. 특히 정상 구간과 이상 구간의 SHAP 값을 분리하여 비교함으로써, 과거에는 환율 상승에 기여하지 않던 한국 통화량 변수가 최근 들어 강력한 상승 압력으로 돌변한 과정을 수학적으로 증명할 수 있었음.

3. 수행 내용 상세

3.1. 데이터 전처리 및 파생 변수 생성

한국은행 경제통계시스템과 미국 연방준비제도 경제 데이터 API를 활용하여 한미 양국의 경제 지표를 월별 단위로 수집함. 수집 항목은 원달러 환율, 양국의 소비자물가지수, 광의통화, 산업생산지수, 기준금리, 10년물 국채금리임. 각 지표는 발표 시차와 집계 기준이 다르므로 공통 날짜를 기준으로 내부 조인하여 병합하였으며, 결측치가 포함된 행은 제거하여 분석 데이터셋을 구성함. 랜덤 포레스트 알고리즘의 특성상 변수의 스케일 차이가 트리의 분할 결과에 영향을 주지 않으므로 별도의 스케일링은 생략함.

양국의 금리 격차가 시장에 미치는 심리적 요인을 반영하기 위해 한국 금리에서 미국 금리를 뺀 파생 변수를 두 가지 형태로 생성함. 하나는 단기 정책 방향성을 대변하는 기준금리 격차이며, 다른 하나는 장기 국채 시장의 자본 이동을 대변하는 10년물 국채금리 격차임.

3.2. 시계열 분할에 따른 상관관계 구조 변화 확인

데이터 전처리 후 가장 먼저 피어슨 상관계수를 활용한 탐색적 데이터 분석을 진행함. 전체 데이터를 정상 구간과 이상 구간으로 분리하여 변수 간의 상관관계 변화를 추적함. 정상 구간 히트맵 분석 결과, 기준금리 격차와 환율의 상관관계수는 음수로 나타나 한국의 금리가 상대적으로 높을 때 환율이 안정화되는 패턴이 확인됨. 반면 양국의 물가와 유동성 지표는 환율과 강한 양의 상관관계를 보임.

그러나 최근 14개월의 이상 구간만을 따로 분리하여 히트맵을 도출한 결과 구조적 붕괴가 관찰됨. 정책 금리 격차와 환율의 상관계수가 극단적인 양수로 역전되었음. 이는 더 이상 금리를 통해 환율을 통제할 수 없음을 시각적으로 방증하는 데이터임. 흥미로운 점은 이 짧은 이상 구간 내에서 한국 통화량과 환율의 단순 선형 상관관계수는 0에 가깝게 나타났다는 사실임. 짧은 기간의 심한 변동성으로 인해 선형적 관계성이 수치상으로 소거된 것으로 추정되며, 이는 후속 머신러닝 분석을 필요로 하는 핵심 동기가 됨.

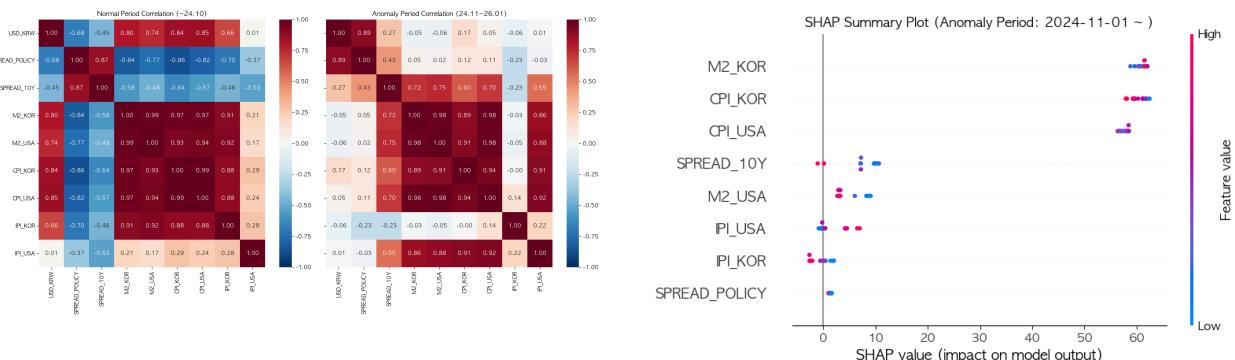


그림 2: 기간별 변수 상관관계 변화 및 이상 구간 SHAP 기여도 요약 플롯

3.3. 모델링 수행 및 이상 구간 변수 기여도 도출

전처리된 변수들을 투입하여 랜덤 포레스트 회귀 모델을 학습시킴. 시계열 데이터의 특성을 고려하여 2024년 10월 이전의 정상 구간 데이터만을 훈련 세트로 분리하여 모델을 학습시켰으며, 트리의 개수는 200개로 지정하여 앙상블 효과를 확보함. 정상 시장의 규칙을 학습한 이 모델에 2024년 11월 이후의 이상 구간 데이터를 테스트 세트로 투입하여 SHAP 요약 플롯을 추출함.

분석 결과 단순 선형 상관관계에서는 유의미성이 보이지 않던 한국 통화량이 SHAP 분석에서는 환율을 비정상적으로 밀어 올린 압도적 1위 원인으로 도출됨. 플롯 상에서 한국 통화량 값이 커질수록 환율 상승에 미치는 기여도가 급격히 붉은색 군집으로 나타남을 확인함. 이는 통화량과 환율 사이에 단순한 직선의 비례 관계가 아니라 임계점을

넘어가면 폭발적으로 영향을 미치는 비선형적 관계가 성립하고 있음을 모델이 자체적으로 학습해 냈음을 의미함. 금리 지표들은 영향력 순위에서 하위권으로 밀려남.

3.4. 유동성 세부 구성 항목 심층 분석

도출된 머신러닝 분석 결과를 경제적 실체와 연결하기 위해, 환율 급등의 주원인으로 지목된 한국 광의통화의 세부 구성 항목을 분해하여 심층 분석함. 이상 구간 동안 증가한 통화량은 총 173조 원에 달함. 이를 상품별로 바 차트로 시각화한 결과 증가분의 성격이 명확히 드러남.

전체 증가분의 29%를 차지하는 1위 항목은 수시입출식 저축성예금이었으며, 이자가 거의 발생하지 않는 요구불 예금과 초단기 자산 운용 펀드인 머니마켓펀드 지분이 그 뒤를 이음. 이 세 가지 항목의 공통점은 언제든 즉시 현금화가 가능한 초단기 부동자금이라는 점임. 경제 불확실성이 커짐에 따라 가계와 기업의 여유 자금이 장기 정기예금이나 채권 시장 등 생산적 투자처로 흘러가지 못하고 대기성 자금으로 쌓임. 이렇게 팽창된 단기 유동성이 달러 가치 상승 기대감과 맞물려 외환시장으로 대거 유입되면서 환율 방어선을 무너뜨린 핵심 동력으로 작용했음을 증명함.

3.5. 팀원별 기여

일련의 분석 과정에서 팀원별 세부 기여사항은 다음과 같음.

TBD

4. 차후 본 학기 계획

방학 스터디를 통해 진행한 1차 분석은 최근 고환율 장세의 원인이 표면적인 한미 금리 격차가 아니라 시중에 누적된 초단기 잉여 유동성에 있음을 정량적으로 밝혀냄. 그러나 본 연구는 분석 과정에서 도출된 중요한 수학적 딜레마를 남겨두고 있으며 이를 완전히 해석하지 못한 한계가 존재함. 따라서 본 학기에는 이 한계를 극복하고 분석의 깊이를 더하기 위해 다음의 세 가지 계획을 수립하여 최종 발표까지 발전시킬 예정임.

4.1. 1. 선형 상관관계 고리 해석 및 비선형 임계점 가설 검증

이상 구간에서 한국 통화량과 환율의 단순 피어슨 상관계수는 사실상 무상관에 가깝게 나타났으나 머신러닝 모델은 통화량을 환율 급등의 가장 강력한 원인으로 지목함. 이러한 고리는 유동성과 환율의 관계가 선형적이지 않기 때문에 추정함. 즉 유동성이 일정 수준까지는 시장에 흡수되지만 특정 규모를 초과하는 순간 외환시장으로 범람하여 발작을 일으키는 비선형적 임계점이 존재한다는 가설을 수립함.

이를 검증하기 위해 본 학기에는 SHAP 의존성 플롯 분석을 본격적으로 도입할 계획임. X축을 통화량의 증가 규모로 두고 Y축을 환율에 미치는 SHAP 영향력으로 설정하여 산점도를 그림. 이를 통해 통화량이 증가함에 따라 영향력이 점진적으로 오르는 것이 아니라 특정 금액 구간에서 수직으로 상승하는 계단식 패턴을 탐색함. 이를 통해 향후 통화 당국이 유의해야 할 구체적인 과잉 유동성 경계선 규모를 정량적으로 제시하는 것을 목표로 함.

4.2. 2. 동적 상관분석 도입을 통한 단기 변동성 캡처

14개월이라는 짧은 이상 구간 전체의 데이터를 하나의 상관계수 수치로 요약하는 과정에서 단기적인 노이즈로 인해 실제 발생했던 변수 간의 동조 현상이 뭉개졌을 가능성이 높음. 이러한 한계를 보완하기 위해 3개월에서 6개월 단위로 윈도우 크기를 설정하고 1개월씩 이동해가며 상관계수의 추이를 시계열로 그리는 동적 상관분석 기법을 도입함.

단순 평균으로는 상관성이 낮아 보이지만 룰링 상관분석을 적용하면 초단기 부동자금이 머니마켓펀드 등에 집중적으로 쏟아져 들어오던 특정 몇 개월의 시점에 유동성과 환율 간의 양의 상관관계가 순간적으로 극단적 수치까지 치솟는 지점들을 캡처할 수 있음. 이를 통해 유동성 유입 속도와 환율 시장의 반응 지연 시간에 대한 추가적인 인사이트를 발굴할 계획임.

4.3. 3. 시계열 특화 예측 모형 추가 및 성능 비교

현재 적용된 랜덤 포레스트 모델은 각 월별 데이터를 독립적인 관측치로 취급하므로 과거 달의 경제 지표가 다음 달에 미치는 순차적 흐름과 시차 효과를 모델 내에 완벽하게 내재화하지 못하는 구조적 한계가 있음. 거시경제 지표는 관성에 의해 움직이는 경향이 크므로 시계열 데이터의 과거 패턴을 기억하고 미래를 예측하는 딥러닝 기반의 순환 신경망 모델 도입을 검토함.

장기 의존성 문제를 해결한 LSTM이나 연산 효율을 높인 GRU 알고리즘을 구축하여 랜덤 포레스트의 예측 성능과 비교 분석함. 이와 동시에 경제학 분야에서 다변량 시계열을 다룰 때 널리 쓰이는 벡터 자기회귀 모형을 통계적 비교군으로 설정함. 기계학습 기반의 알고리즘과 전통적인 통계 및 계량경제학 모형이 이상 구간의 환율 상승 원인을 동일하게 통화량으로 지목하는지 교차 검증함으로써 최종 결론의 신뢰성과 객관성을 극대화할 계획임.