Causal Estimation of the Korea-USA Interest Rate Inversion on Samsung Electronics Stock Price





- Introduction
 - 1. Reason for Topic
 - 2. Analysis Direction
 - 3. Modeling Process

- **Modeling**
 - 1. Balancing Method
 - 2. DAG 1
 - 3. DAG 2
 - 4. DAG 3
 - 5. DAG 4

- Data & Assumption
 - 1. Data Description
 - 2. Data Assumption: DAG 1
 - 3. Theoretical Assumption
 - 4. Additional Assumption : DAG 2
 - 5. Mediation Assumption: DAG 3 & DAG 4

- Conclusion
- 1. Comparison
- 2. Summary
- 3. Follow-up Study

I. Introduction

- 1. Reason for Topic
- 2. Analysis Direction
- 3. Modeling Process

Reason for Topic

> Reason for Topic

- 도메인 지식이 중요한 인과 추론에 있어 **비교적 접근이 쉬운 국내 주식을 분석**
- 다양한 교란 변수(Confounder)들과 종속 변수(Response variable), 처치 변수(Treatment variable)간의 병합이 중요한데, 주식 데이터의 경우 일자(Date)를 기준으로 병합이 용이
- 실제 뉴스에서만 접하던 **한미 금리 역전이 우리에게 어떤 영향이 있는지에 대한 지적 호기심**
- 다수에게 친숙하고 유명한 삼성전자 주가를 통해 분석 동기 부여(14개월만에 7만전자..!!)



▲ 접근이 쉬운 국내주식



▲ 변수 간 병합 용이



▲ 한미 금리역전 영향 분석

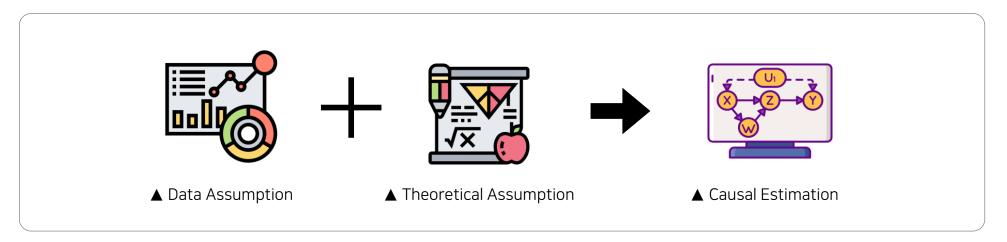


▲ 다수에게 유명한 삼성전자

Analysis Direction

> Analysis Direction

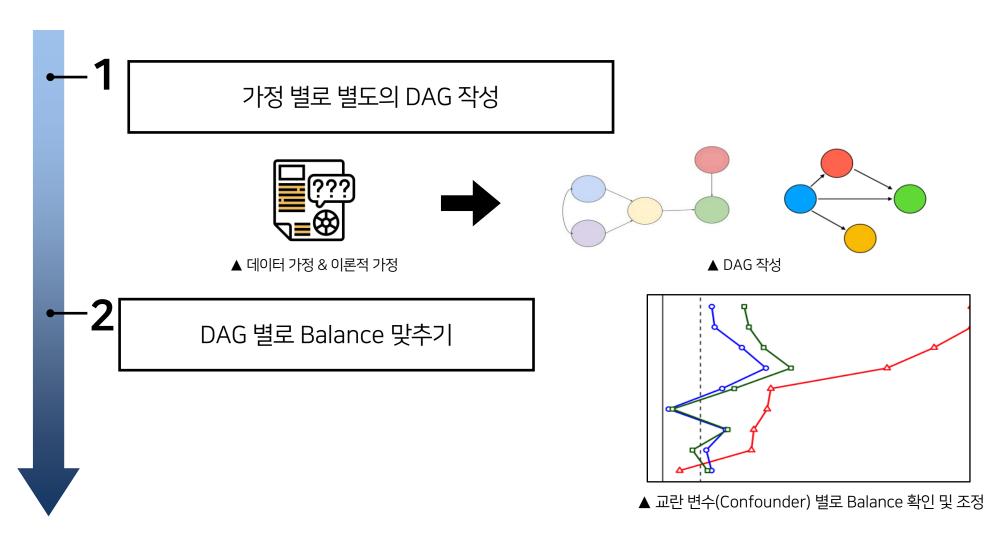
- 가정 설정: 수집한 데이터를 토대로 데이터 간의 구조 파악을 위한 가정(데이터 가정)과 인과 추론을 위한 가정(이론적 가정) 설정
- 설정한 가정들을 체크하고 각 설정된 가정 세팅 별로 모델링을 통해 인과 효과를 추정 및 비교



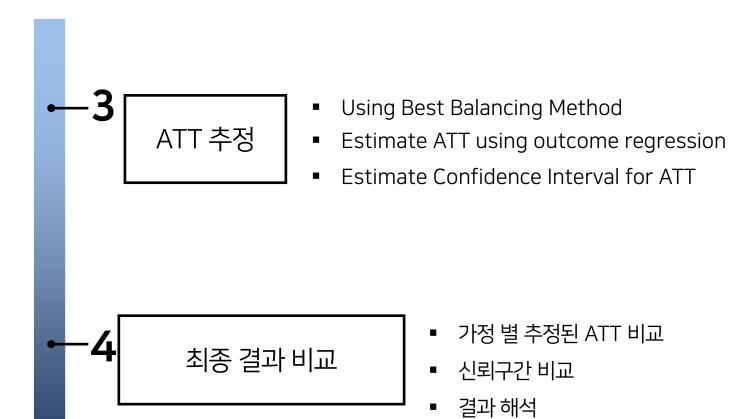
➤ Key Point

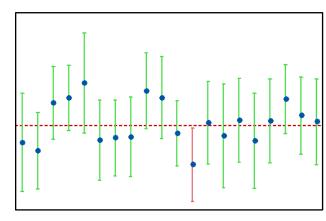
- 각 가정별로 추정된 인과 효과의 유의성과 그 정도가 달라지는지?
- 설정한 가정들이 잘 만족하는 경우와 그렇지 않은 경우 결과가 어떻게 달라지는지?

➤ Modeling Process



Modeling Process





II. Data & Assumption

- 1. Data Description
- 2. Data Assumption: DAG 1
- 3. Theoretical Assumption
- 4. Additional Assumption: DAG 2
- 5. Mediation Assumption : DAG 3 & DAG 4

Data Description

- > Response variable & Treatment variable
 - 2018년 1월 1일 ~ 2023년 4월 30일: 한국 영업일 기준으로 1,550일
 - **종속 변수(Response variable)**: 한국 영업일 기준 삼성전자 일 별 종가
 - **처치 변수(Treatment Variable)**: 한국 기준 금리가 미국 기준 금리보다 높은 지 여부(높거나 같으면 treatment group , 아니면 control group)

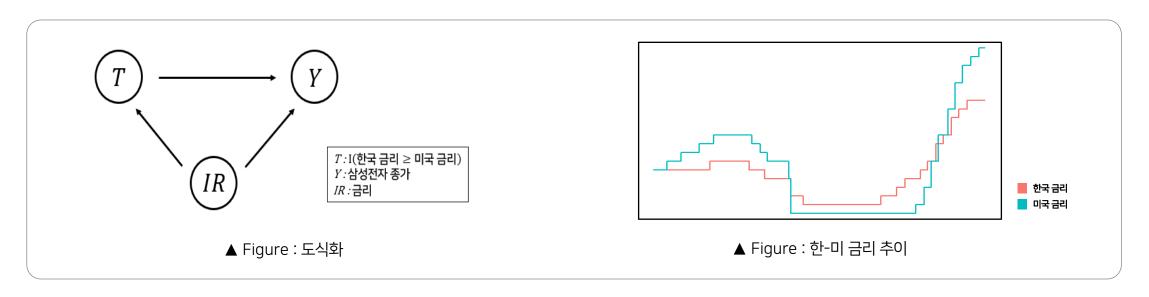
> Candidates for Confounder



Candidates for Confounder ①: IR

Interest Rate

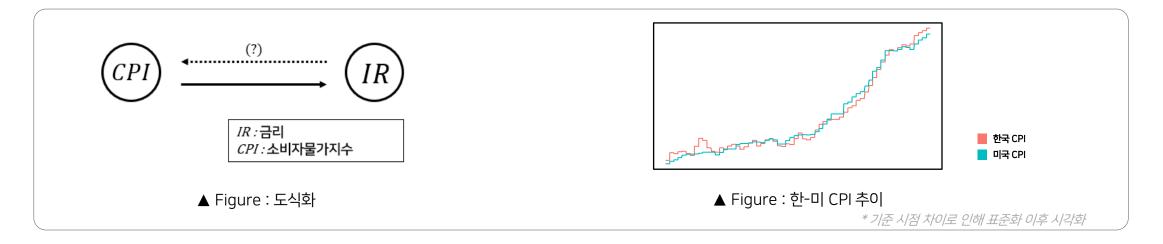
- 한국 기준금리: 한국은행 통화정책위원회에서 결정 *출처: 한국은행
- 미국 연준금리: 연방공개 시장위원회(FOMC)에서 결정 *출처: Federal Reserve
- 교란 변수(Confounder)로서 평가
 - : 금리는 통화 가치를 평가하는 1차적 요소이기 때문에 주식 시장(종속 변수)에 끼치는 인과 영향이 있다고 볼 수 있음 정책 등으로 인해 특정 국가의 금리가 너무 높거나 낮음으로 인해, 한미 금리 수준 여부(처치 변수)가 결정될 수 있으므로 처치 변수에 대한 인과성 존재



Consumer Price Index(CPI)

- 한국 소비자물가지수 : 통계청에서 2020년을 기준(=100)으로 매월 산출 *출처 : 한국 통계청
- 미국 소비자물가지수: 미국 노동부에서 1982~1984년을 기준(=100)으로 매월 산출 *출처: U.S. BUREAU OF LABOR STATISTICS
- 교란 변수(Confounder)로서 평가

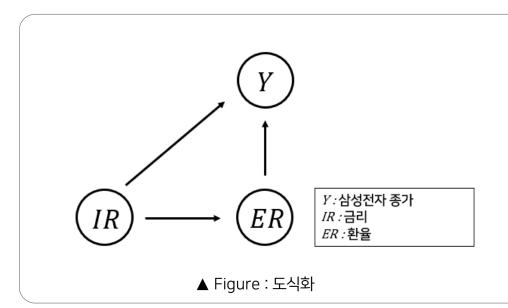
: 소비자물가지수는 그 변동률을 통해 인플레이션을 측정할 수 있고, 인플레이션 정도를 조절하기 위해 정부가 금리를 조정할 수 있음 직접적으로 종속 변수나 처치 변수에 영향을 주지는 않지만, 금리를 조정하는 데 주요한 인과 관계를 가지고 있음. 금리가 소비자물가지수에 끼치는 인과 효과에 대해서는 향후 가정을 통해 완화할 예정

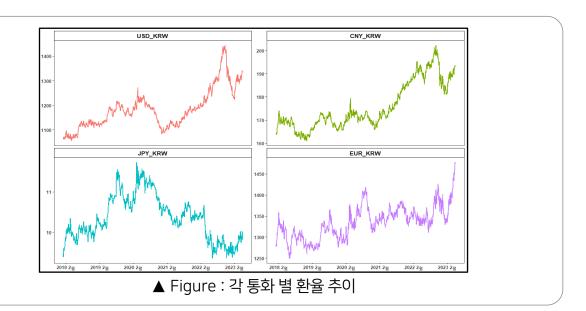


Candidates for Confounder ③: Exchange Rate

> Exchange Rate

- 선정 환율 : 미달러 원 환율(USD KRW), 중국위안 원 환율(CNY KRW), 일본엔 환율(JPY KRW), 유로 원 환율(EUR KRW) * 출처 : 인베스팅닷컴
- 선정 배경: 재무제표 공시 시 국가 별 매출을 공개하는 나라 중 상위 4개국으로 선정
- 교란 변수(Confounder)로서 평가
 - : 환율로 인해 외화로 벌어들이는 수익 정도가 달라지기 때문에, 종속 변수에 대한 인과 관계를 유추할 수 있음 금리 정도가 높아지면 해당 국가의 투자 수익률이 높아지므로, 외국인 투자자들이 해당 통화를 많이 매수하게 되어 해당 통화 기준 환율이 상승할 수 있음

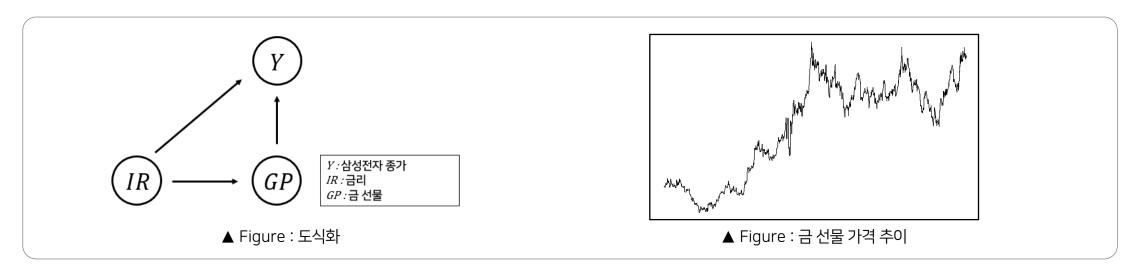




Candidates for Confounder 4: Futures

> Futures

- 선정 선물 : 금 선물 *출처 : 인베스팅닷컴
- 선정 배경: 농산물 종류의 상품 선물이나 금속 종류의 은 선물, 구리 선물 등도 존재하지만, 대중적으로 인지도와 관심도가 높은 금 선물만을 활용
- 교란 변수(Confounder)로서 평가
 - : 안전 자산으로 취급되는 금 가격은 금리 변화로 인한 시장의 변화에 영향을 받는다고 유추할 수 있음 금 가격 경향 추이를 보고, 주식 시장의 전망 정도를 판단할 수 있음(경기가 안 좋을 때에도 안전 자산으로서, 금이 대체 투자 자산 가격 상승 경향 있음) 투자자들이 헷징이나 포트폴리오 다변화를 위해 금을 많이 사용하는데, 금 가격 상승으로 인해 일부 주식 투자금이 금으로 이동할 수 있음.

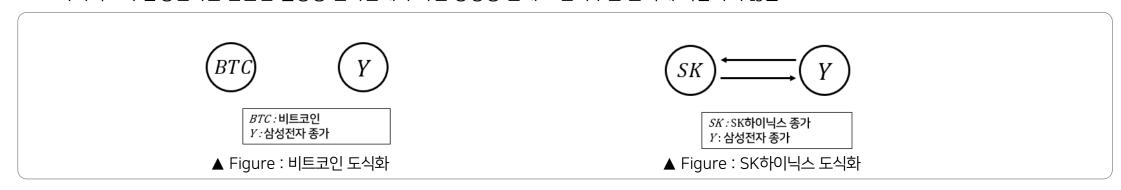


Cryptocurrency

- 선정 암호화폐 : 비트코인(BTC) *출처 : 인베스팅닷컴
- 선정 배경 : 암호화폐 중 대장 격인 비트코인을 활용. 이더리움 같은 다른 코인의 경우 비트코인과 상관 관계가 너무 높아 비트코인 하나만 고려
- 교란 변수(Confounder)로서 평가
 - : 암호화폐 시장이 주식시장과 어느 정도 연관성이 있다고 볼 수는 있지만, 그것이 명확히 인과 관계를 의미한다고 볼 수 없다는 연구 존재(*Bouri et al. (2017)*) 암호화폐 자체가 영향을 받는 다른 요소들이 너무 많기 때문에(ex. 머스크의 트위터...) 통제하기 어려움

Similar Stock Price

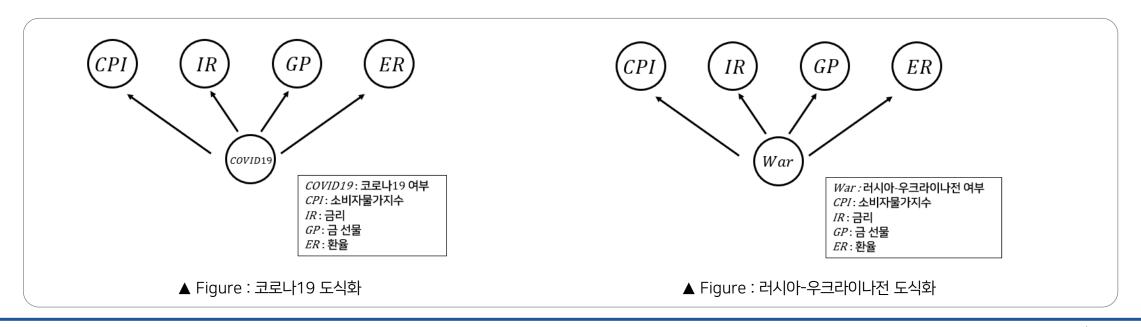
- 선정 주식 : SK하이닉스 * 출처 : 인베스팅닷컴
- 선정 배경: 반도체 테마 주 중, 삼성전자만큼 인지도가 높은 SK하이닉스 고려
- 교란 변수(Confounder)로서 평가
 - : SK하이닉스와 삼성전자는 단순한 일방향 인과관계가 아닌 양방향 관계로 인과추론 분석에 적합하지 않음



Candidates for Confounder 7: Global Impact

Global Impact

- 선정 요인: 코로나19 여부, 러시아-우크라이나전 여부 *출처: 검색을 통해 자체 제작
- 선정 배경: 전 세계의 경제 상황에 큰 영향을 미쳤던 감염병 코로나19와, 러시아-우크라이나 전을 고려하는 것이 분석에 도움이 될 것이라 판단
- 교란 변수(Confounder)로서 평가
 - : 코로나19가 직접적으로 삼성전자 주가나, 한미 금리 수준에 영향을 미쳤다고 보기는 어렵지만 다른 거시적 요인에 영향을 미쳤다고 볼 수 있음. 러시아-우크라이나전도 코로나19와 마찬가지로 직접적으로 종속 변수와 처치 변수에 영향을 미치기 보다는 다른 요인들에 영향을 미쳤다고 볼 수 있음



Final Analysis Setting

Variable Description

- 2018년 1월 1일 ~ 2023년 4월 30일 : 한국 영업일 기준으로 1,550일
- 종속 변수(Response variable): 한국 영업일 기준 삼성전자 일 별 종가
- 처치 변수(Treatment Variable): 한국 기준 금리가 미국 기준 금리보다 높은 지 여부(높거나 같으면 treatment group, 아니면 control group)
- 교란 변수(Confounder) : 금리(한국/미국), 소비자물가지수(한국/미국), 환율(미달러 원, 중국위안 원, 일본엔 원, 유로 원), 금 선물, 코로나19, 러시아-우크라이나 전

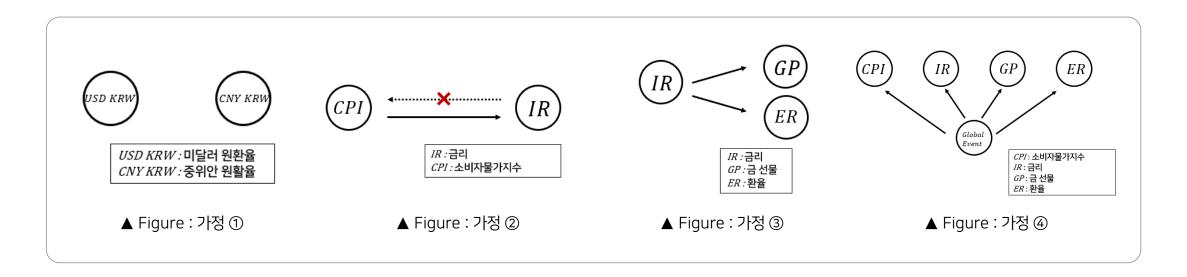
Causal Assumption & Target population

- Causal Assumption: 한미 금리 수준이 한달 후의 삼성전자 주가에 미치는 인과 영향
 - ➡ 금리 수준 차이(처치 변수)의 효과가 시장에 반영되기까지의 시기를 한달로 보고 한달 후의 주가(종속 변수)에 미치는 인과 영향을 분석
- **Target Population 종속 변수(Response variable)**: 2018년 2월 1일 ~ 2023년 4월 30일(한국 영업일 기준) **처치 변수(Treatment variable) & 교란 변수(Confounder)**: 2018년 1월 1일 ~ 2023년 3월 31일(한국 영업일 기준)

Data Assumption: DAG 1

Data Assumption

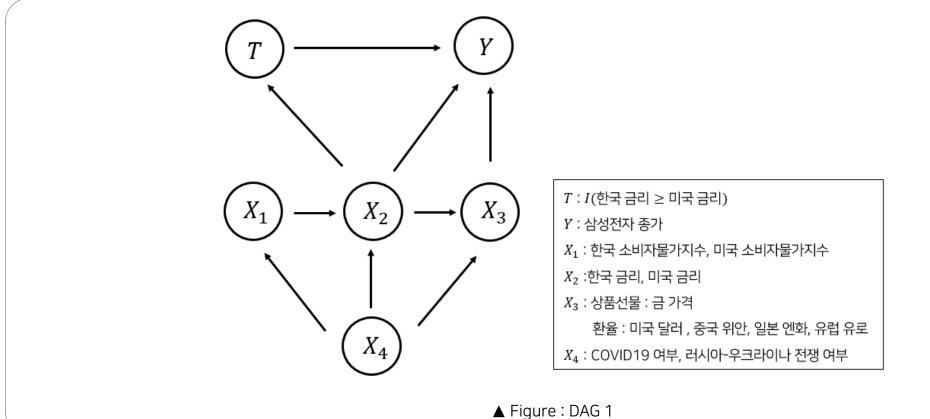
- 교란 변수간의 인과 관계를 명확하게 하기 위해 몇 가지 데이터적 가정을 설정
 - 데이터 가정 ①: 동등한 지위를 갖는 교란 변수(ex. 환율) 간에는 서로 독립적인 관계를 가진다. 즉 서로 상관 관계나 인과 관계가 없다
 - 데이터 가정②: 인플레이션(소비자물가지수)은 가장 외생적인 변수로 다른 거시적 요인에 영향을 받지 않는다.
 - 데이터 가정 ③: 금리가 환율과 상품선물에 인과 영향을 가지며, 환율과 상품선물은 서로 동등한 지위를 갖는다.
 - 데이터 가정 ④: 범세계적으로 영향을 미치는 이벤트적인 사건/사고(ex. 코로나19, 전쟁)는 모든 거시적 요인에 영향을 가진다.



Data Assumption: DAG 1

> DAG 1

앞선 데이터 가정들을 따르는 DAG를 그리고 해당 DAG를 앞으로 DAG 1으로 명명



Theoretical Assumption

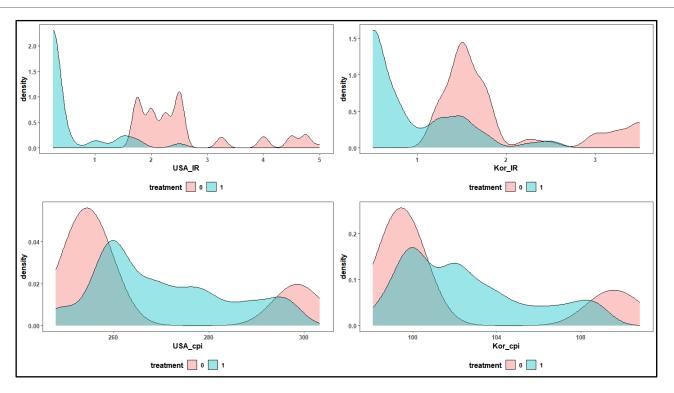
> Theoretical Assumptions

- 인과추론 분석을 위한 기본적인 이론적 가정 설정
 - Ignorability: $Pr(Z_i|Y_i(0),Y_i(1),X_i) = Pr(Z_i|X_i)$
 - Overlap Assumption: $0 < Pr(Z_i = 1 | X_i) < 1 \text{ for all } i$
 - SUTVA: $Y_i = Z_i Y_i(1) + (1 Z) Y_i(0)$
- 본 분석에서 설정한 추가적인 이론적 가정
 - 이론적 가정 ①: 주가 데이터가 시간에 따라 기록되었지만 각 시간 별 기록을 하나의 Obs로 보고 Cross-sectional 데이터로 간주한다.
 - 이론적 가정 ②: 본 분석에 영향을 미치는 더 이상의 교란 변수(Confounder)는 없다고 가정한다.
 - 이론적 가정 ③ : 본 분석의 처리 변수를 보면 준실험적 연구로 볼 수 있지만, 실험적 연구로 간주하고 실험적 연구 방법론을 적용하는 데 무리가 없다고 본다.
 - 이론적 가정 ④: 한국 금리가 더 높은 경우에만 관심을 가지고, Target Estimand는 ATT로 설정하여 위의 과정을 다음과 같이 약화한다.

Unconfoundedness for untreated: $Y_i(0) \perp Z_i | X_i$ Weak Overlap: $Pr(Z_i = 1 | X_i = x) < 1$ for any x

Checking Overlap of Confounders

Checking Overlap Assumption ①

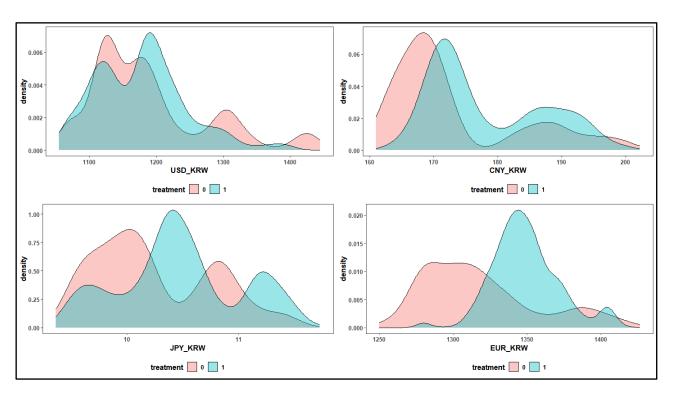


▲ Figure: Overlap plot for USA IR, Kor IR, USA CPI and Kor CPI

- 금리 변수의 경우, 양쪽 금리가 낮은 값에서 P(Z=1|X) = 1 값을 가짐 ➡️ Overlap Assumption 위배
- CPI 변수의 경우, 중간 지점에서 P(Z=1|X) = 1 값을 가짐 ➡ Overlap Assumption 위배

Checking Overlap of Confounders

Checking Overlap Assumption ②

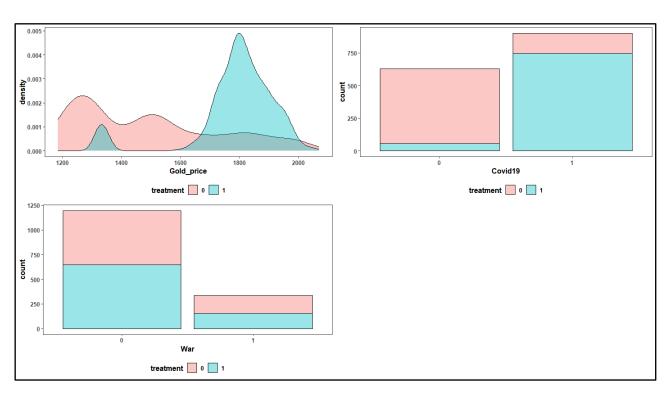


▲ Figure: Overlap plot for USD KRW, CNY KRW, JPY KRW, and EUR KRW

유로 원 환율(EUR KRW)의 경우 낮은 값에서, P(Z=1|X) = 0의 값을 가지지만 약화된 Overlap 가정을 위배하지 않음

Checking Overlap of Confounders

Checking Overlap Assumption ③



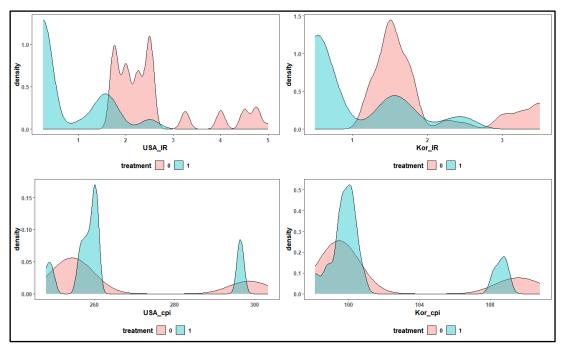
▲ Figure : Overlap plot for Gold price, Covid19, and War

■ 금 선물(Gold price)의 경우 특정 구간들에서, P(Z=1|X) = 0의 값을 가지지만 약화된 Overlap 가정을 위배하지 않음

Re-checking Overlap Assumption

➤ Re-checking Overlap Assumption ①

한/미 소비자물가지수에 대해 Overlap Assumption이 위배되던 구간을 삭제 후 다시 Overlap Assumption check



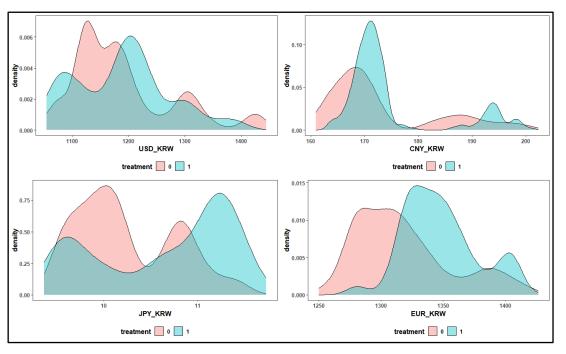
▲ Figure: Overlap plot for USA IR, Kor IR, USA CPI and Kor CPI

- 금리 변수의 경우, 양쪽 금리가 낮은 값에서 P(Z=1|X) = 1 값을 가짐 ➡ Overlap Assumption 위배
- CPI 변수의 경우, 특정 구간에서 P(Z=1|X) = 0의 값을 가지지만 약화된 Overlap 가정을 위배하지 않음

Re-checking Overlap Assumption

Re-checking Overlap Assumption ②

한/미 소비자물가지수에 대해 Overlap Assumption이 위배되던 구간을 삭제 후 다시 Overlap Assumption check



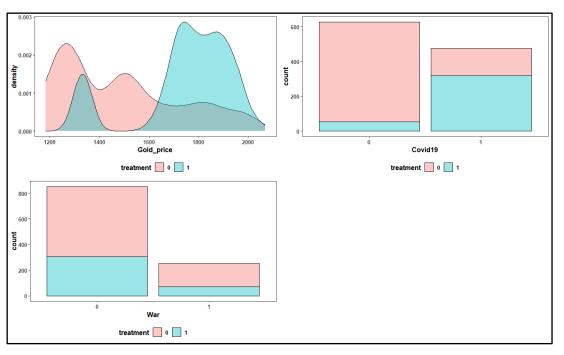
▲ Figure: Overlap plot for USD KRW, CNY KRW, JPY KRW, and EUR KRW

유로 원 환율(EUR KRW)과 중위안 원환율(CNY KRW) 경우 낮은 값에서, P(Z=1|X) = 0의 값을 가지지만 약화된 Overlap 가정을 위배하지 않음

Re-checking Overlap Assumption

➤ Re-checking Overlap Assumption ③

한/미 소비자물가지수에 대해 Overlap Assumption이 위배되던 구간을 삭제 후 다시 Overlap Assumption check



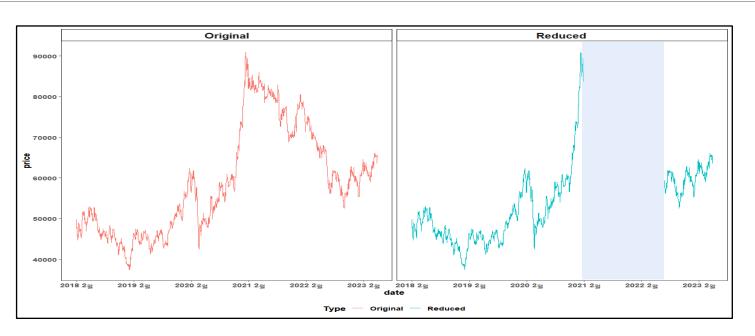
▲ Figure: Overlap plot for Gold price, Covid19, and War

금 선물(Gold price)의 경우 특정 구간들에서, P(Z=1|X) = 0의 값을 가지지만 약화된 Overlap 가정을 위배하지 않음

Re-checking Target Population

Re-checking Target Population

Overlap Assumption 만족을 위해 일정 구간의 데이터를 삭제함으로써 변화한 Target Population 재확인



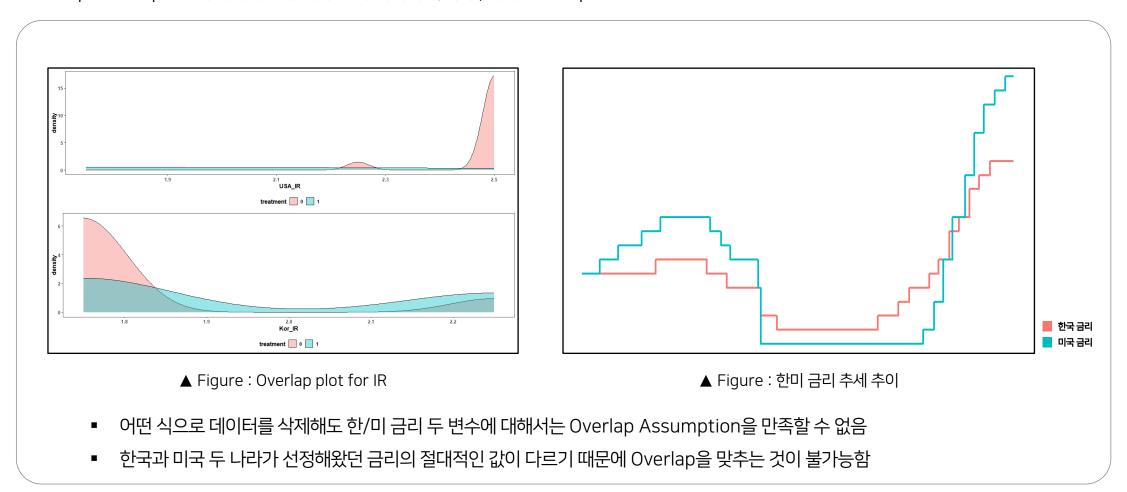
▲ Figure: Original Target Population vs Reduced Target Population

- 기존 Target Population에서 삭제된 부분은 한/미 모두 코로나19가 가장 심했던 시기로 볼 수 있다.
- 기존 Target Population: 2018년 2월 1일 ~ 2023년 4월 30일(한국 영업일 기준)
 - 변경된 Target Population: 2018년 2월 1일 ~ 2023년 4월 30일 중, 시장에 큰 충격이 없는(=코로나19가 없는) 주가

Overlap Assumption of Interest Rate

Overlap Assumption of Interest Rate

Overlap Assumption이 위배되는 금리 구간에 대해서 삭제 후, 다시 Overlap Check



Additional Assumption: DAG 2

Additional theoretical assumption for IR

기존 DAG 1에서 교란 변수(Confounder) 금리(IR)이 Overlap Assumption을 심하게 위배하기 때문에 DAG 1에 추가적인 이론적 가정을 더함

이론적 가정 ⑤-1:

한/미 금리 변수가 Overlap assumption을 다소 위배하지만, 해당 교란 변수의 중요도가 높다는 점에서 해당 변수를 포함했을 때 얻을 수 있는 모형적 이점이 Overlap Assumption이 위배됨으로 인해서 생길 수 있는 Bias보다 크다.

금리 변수가 다른 교란 변수(ex. CPI, War, COVID19)들에 의해 영향을 받으므로 다른 교랸 변수들로 인해 Bias가 어느 정도 상쇄될 수 있다.

하지만, 가정 위배가 가져오는 위험도가 클 수 있으므로 다음의 또 다른 가정을 설정하여 새로운 세팅의 DAG를 구성

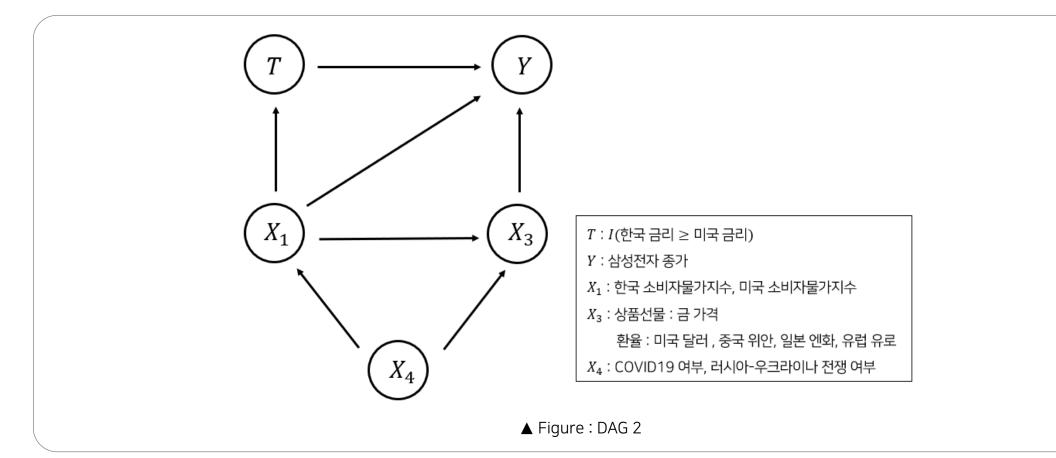
이론적 가정 ⑤-2:

한/미 금리 변수의 Overlap assumption 위배 정도는 무시할 수 없는 정도이다.

금리 변수가 다른 교란 변수(ex. CPI, War, COVID19)들에 의해 영향을 받지만, 그 Bias를 상쇄하는 것에는 어느 정도 한계가 있다.

> DAG 2

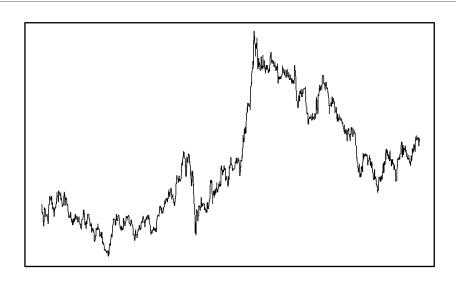
■ 앞서 추가로 가정한 금리 교란 변수를 제거한 상황에서의 DAG를 DAG 2라고 명명



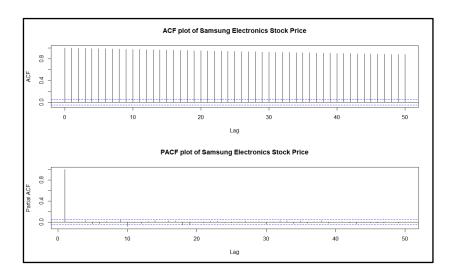
Causal Effect of Lagged price

Causal Effect of Lagged price

- Cross-sectional 분석을 가정하긴 했지만 주가는 대표적인 시계열 데이터로 이전 시점의 주가에 영향이 있음
- 과거 시점 주가들이 어느 정도 수준인지에 따라 현재 주가가 달라질 수 있다는 점은 보편 타당하게 생각될 수 있음



▲ Figure : 삼성전자 주가 TS plot



▲ Figure : 삼성전자 주가 ACF/PACF plot

삼성전자 주가 데이터는 명백한 시계열 데이터임과 동시에, ACF 플롯을 보면 주가 데이터에서 주로 나타나는 **장기 기억성 특성** (Long Memory property)가 나타나는 것을 볼 수 있음

Long Memory Property

Long Memory Property

- 장기 기억성 특성(Long Memory Property) : 시계열 데이터에서 시점 별 의존성(time-correlated)이 매우 긴 시차(lag)동안 지속되는 특성
- 장기 기억성 특성을 지닌 시계열 데이터를 분석하는 방법으로 차분, 소수점 차분 등 다양한 방법이 있지만, 차분 방법의 경우 종속 변수에 대한 처리임으로 본 분석에서는 다른 방법을 고려
- Heterogeneous AutoRegressive Model(HAR Model) (Corsi et al. (2009))
 - HAR Model : 차분을 통해서 장기 기억성 시계열 데이터를 설명하는 것이 아닌, 하루 전, 일주일 전, 한달 전 값들을 활용해서 장기 기억성을 설명하는 모형

$$\begin{aligned} y_t^{daily} &= \phi_1 y_{t-1}^{daily} + \phi_2 y_{t-1}^{weekly} + \phi_1 y_{t-1}^{monthly} + \epsilon_t \\ y_{t-1}^{weekly} &= \frac{1}{5} \sum_{j=1}^{5} y_{t-j}^{daily}, \quad y_{t-1}^{monthly} = \frac{1}{22} \sum_{j=1}^{22} y_{t-j}^{daily} \\ &* \text{88CM of the property of the property$$

현 시점의 삼성전자 주가는 하루 전 주가, 과거 일주일 평균 주가, 과거 한달 평균 주가에 영향을 받는다는 식으로 모형을 세울 수 있음

Lagged Price as Mediator

➤ Lagged Price as Mediator

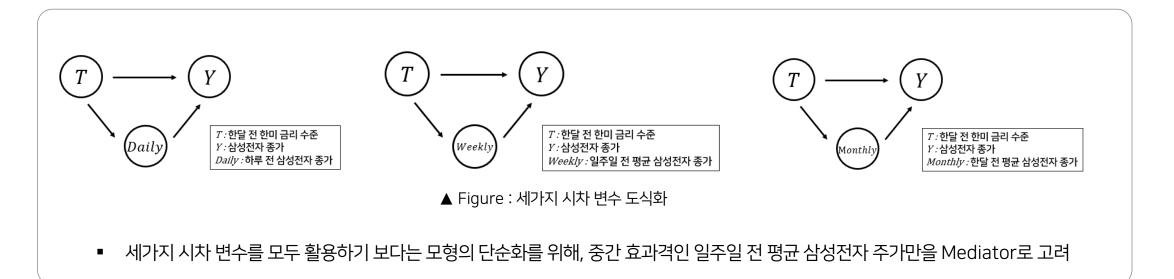
Data Assumption (5):

현재 시점의 주가는 하루 전 주가, 일주일 전 평균 주가, 한달 전 평균 주가에 영향을 받는다. 또한 과거 주가들도 주가라는 점에서 마찬가지로 처치 변수에 영향을 받는다

Remark:

본 분석에서의 Causal Assumption = 한미 금리 수준이 한달 후의 삼성전자 주가에 미치는 인과 영향

처치 변수와 종속 변수 간 한달 간의 시차가 있어, 하루 전 주가, 일주일 전 평균 주가, 한달 전 평균 주가를 활용하는 데 있어 시점적 오류는 없음



Mediation Assumption

Mediation Assumption

Sequential Ignorability

$$\{Y_i(t',m),M_i(t)\}\perp T_i|X_i=x$$

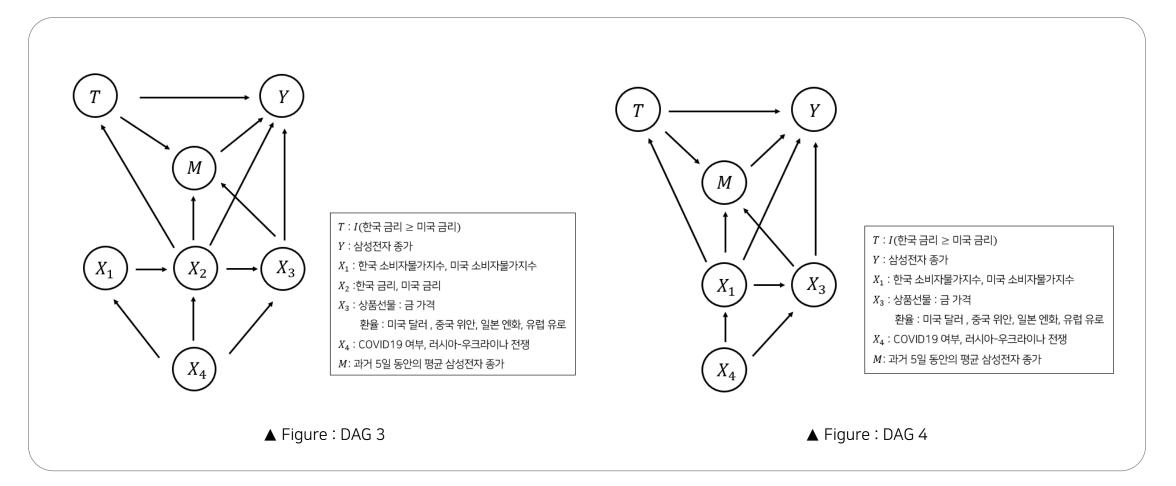
$$Y_i(t',m)\perp M_i(t)|T_i=t, X_i=x$$

Effect Decomposition

- Treatment Effect(per unit) : $au_i = Y_i(1, M_i(1)) Y_i(0, M_i(0))$
- Direct Effect : $\zeta(t) \equiv Y_i(1, M_i(1)) Y_i(0, M_i(0))$
- Indirect Effect : $\delta_i(t) \equiv Y_i(t, M_i(1)) Y_i(t, M_i(0))$
- Sum of the direct effect and indirect effect : $au_i = \delta_i(t) + \zeta_i(1-t)$

> DAG 3 & DAG 4

앞선 DAG 1, DAG 2에 각각 일주일 전 평균 종가를 Mediator로 추가한 DAG



III. Modeling

1. Balancing Method

2. DAG 1

3. DAG 2

4. DAG 3 5. DAG 4

Balancing Method

Balancing Method

- 앞서 가정한 이론적인 가정들을 만족하기 위해 처치 변수(Treatment variable)와 교란 변수(Confounder) 간의 Balance가 맞도록 처리
- 일반적으로 Standardized Mean Difference(SMD)가 -0.1과 0.1 사이에 있으면 Balance가 되었다고 간주

$$d = \frac{(\overline{X}_1 - \overline{X}_0)}{\sqrt{\frac{s_1^2 + s_0^2}{2}}}$$

▲ Figure : SMD for continuous confounder

$$d = \frac{(\hat{p}_1 - \hat{p}_0)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)+\hat{p}_0(1-\hat{p}_0)}{2}}}$$

▲ Figure: SMD for binary confounder(using dummy encoding)

* SMD is calculated using "cobalt" R package,

- 다음의 3가지 방법을 이용해서 Balancing 진행
 - ① Matching using Propensity Score
- ② Stratification using Propensity Score
- ③ Inverse Probability Weighting(IPW)

Propensity Score & Mean SMD

Propensity Score using BART

- Propensity Score를 추정하기 위해 BART 모형을 사용
- Why BART??
 - ① 본 분석에서 사용하는 교란 변수(Confounder)들 간의 다중공선성(Multicollinearity)로 인해 GLM 알고리즘이 수렴하지 않음
 - ② 분석에서 사용한 R 함수들에서 BART 모형이 사용 가능함
 - ③ 다른 Bagging 모형들(ex. RandomForest, Boosting)을 사용하기에는 데이터 수가 충분하지 않음

Mean SMD

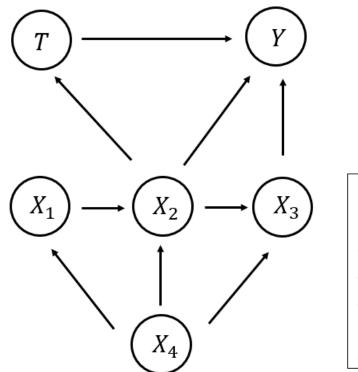
SMD가 모든 변수에 대해서 -0.1과 0.1 사이에 있으면 문제가 되지 않지만, 실제 데이터에서는 어렵다는 점에서 모든 변수에 대한 Balance 정도를 고려하기 위해 새로운 평가 지표로서 도입

$$Mean SMD = \frac{1}{\# of \ confounders} \sum_{i=1}^{\# of \ confounders} (|SMD_i| - 0.1) \times I(|SMD_i| > 0.1)$$

SMD의 절대값이 0.1보다 큰 교란 변수(Confounder)들에 대해서 평균적으로 얼마나 0.1과 멀리 떨어져 있는지 산출하는 지표

DAG 1: Remark

Remark DAG 1



 $T: I(한국 금리 \ge 미국 금리)$

Y : 삼성전자 종가

 X_1 : 한국 소비자물가지수, 미국 소비자물가지수

 X_2 :한국 금리, 미국 금리

 X_3 : 상품선물: 금 가격

환율: 미국 달러, 중국 위안, 일본 엔화, 유럽 유로

 X_4 : COVID19 여부, 러시아-우크라이나 전쟁 여부

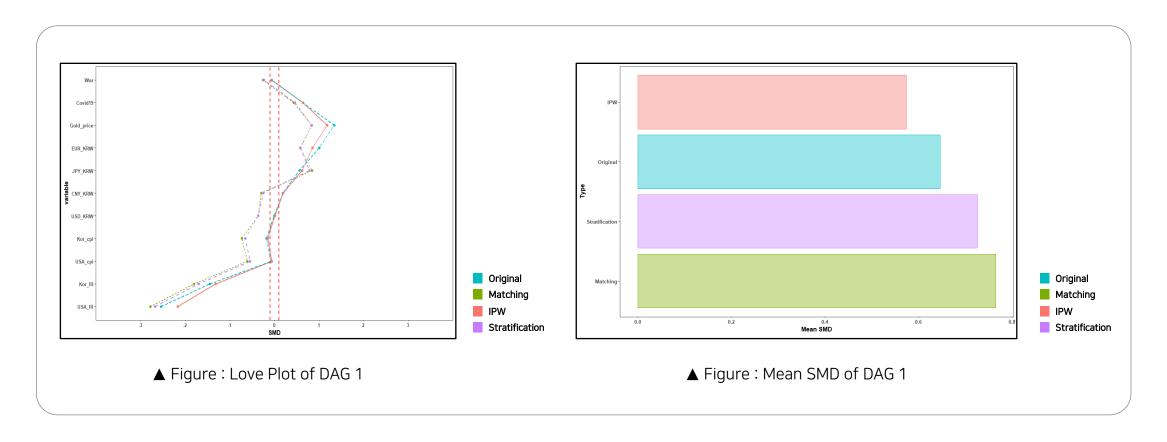
▲ Figure: DAG 1

■ DAG 1: 한/미 금리 변수(Interest Rate)가 포함되고 Mediator(시차 변수)는 없는 모형

DAG 1: Balancing

> Balancing

- Overlap Assumption이 위배되는 IR 변수들의 경우 Balancing이 모든 방법에서 크게 개선되지 않음
- 세가지 방법들 중 IPW를 활용한 방법이 가장 작은 Mean SMD 값을 가지므로, IPW 방법을 통해 Balancing 진행 후 ATT 추정



DAG 1: Estimating ATT

Outcome Regression

$$log(Price) \sim Treatment * (USA IR + Kor IR + USA CPI + Kor CPI + Gold Price + USD KRW + CNY KRW + IPY KRW + EUR KRW + Covid19 + War)$$

- 처치 변수(Treatment)와 교란 변수(Confounder) 간의 교호 작용(Interaction)도 고려
- 종속 변수(Response variable)이 0이상의 값만을 가지는 가격이라는 점에서 Gaussian 분포에 log link 적용

Estimated ATT

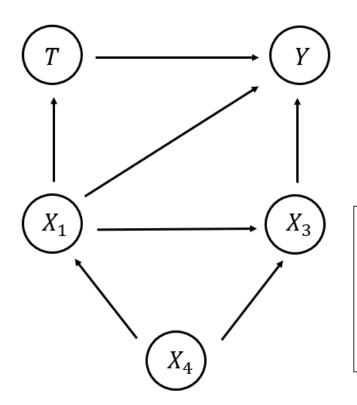
Estimated ATT

$$\widehat{ au}_{ATT} = E\{Y_i(1) - Y_i(0)|Z=1\}$$
 = -11,334(원)

- **95 % Confidence Interval** : [-13,464, -9,205]
 - → 95% 신뢰구간에 0이 포함되지 않으므로 유의수준 0.05하에서 유의미한 효과가 있다고 볼 수 있음

DAG 2: Remark

Remark DAG 2



T: I(한국 금리 \geq 미국 금리)

Y : 삼성전자 종가

 X_1 : 한국 소비자물가지수, 미국 소비자물가지수

 X_3 : 상품선물: 금 가격

환율 : 미국 달러 , 중국 위안, 일본 엔화, 유럽 유로

 X_4 : COVID19 여부, 러시아-우크라이나 전쟁 여부

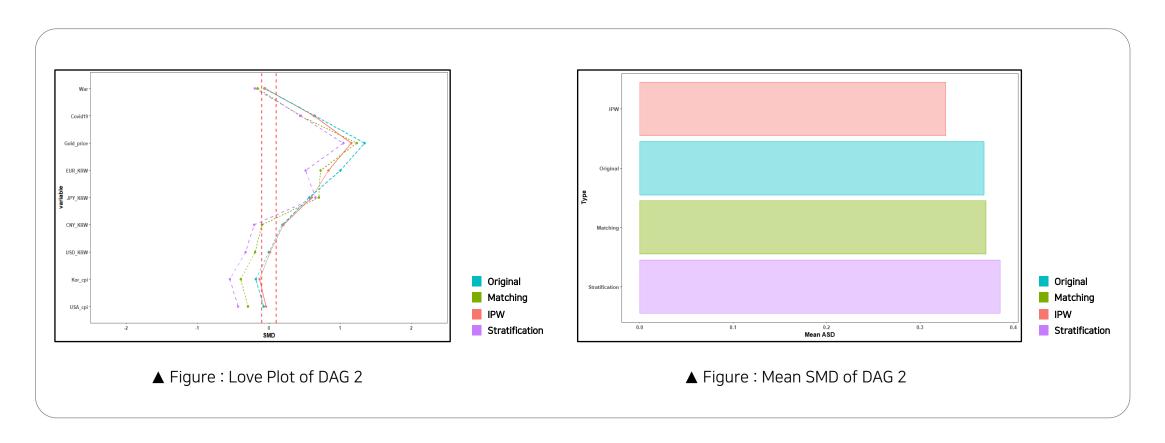
▲ Figure : DAG 2

■ DAG 2: 한/미 금리 변수(Interest Rate)와 Mediator(시차 변수) 모두 없는 모형

DAG 2: Balancing

> Balancing

- 유로 환율(EUR KRW), 금 선물(Gold price)와 같이 초기 Balance가 심하게 위배된 변수들의 경우 Balance가 크게 개선되지 않음
- 세가지 방법들 중 IPW를 활용한 방법이 가장 작은 Mean SMD 값을 가지므로, IPW 방법을 통해 Balancing 진행 후 ATT 추정



DAG 2: Estimating ATT

Outcome Regression

$$log(Price) \sim Treatment * (USA CPI + Kor CPI + Gold Price + USD KRW + CNY KRW + JPY KRW + EUR KRW + Covid19 + War)$$

- 처치 변수(Treatment)와 교란 변수(Confounder) 간의 교호 작용(Interaction)도 고려
- 종속 변수(Response variable)이 0이상의 값만을 가지는 가격이라는 점에서 Gaussian 분포에 log link 적용

Estimated ATT

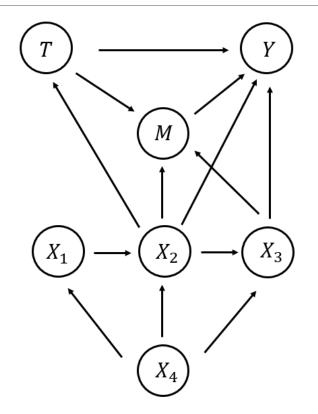
Estimated ATT

$$\widehat{ au}_{ATT} = E\{Y_i(1) - Y_i(0) | Z = 1\}$$
 = -197(원)

- 95 % Confidence Interval : [-1,516, 1,121]
 - 95% 신뢰구간에 0이 포함하므로 유의수준 0.05하에서 유의미한 효과가 있다고 볼 수 없음

DAG 3: Remark

Remark DAG 3



T : I(한국 금리 ≥ 미국 금리)

Y : 삼성전자 종가

 X_1 : 한국 소비자물가지수, 미국 소비자물가지수

 X_2 :한국 금리, 미국 금리

 X_3 : 상품선물: 금 가격

환율: 미국 달러, 중국 위안, 일본 엔화, 유럽 유로

 X_4 : COVID19 여부, 러시아-우크라이나 전쟁

M: 과거 5일 동안의 평균 삼성전자 종가

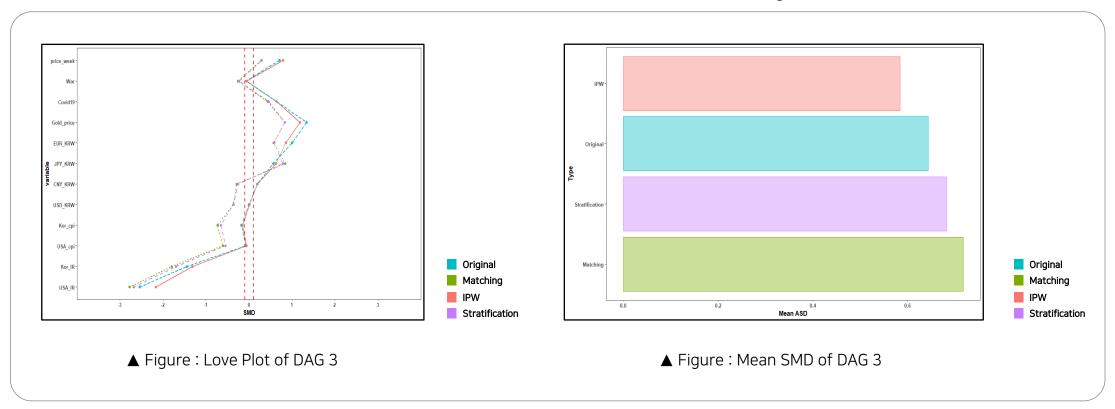
▲ Figure : DAG 3

■ DAG 3: 한/미 금리 변수(Interest Rate)와 Mediator(시차 변수) 모두 고려한 모형

DAG 3: Balancing

> Balancing

- 교란 변수(Confounder) 외에도 Mediator 간의 Balance 여부도 체크. 매칭이나 층화에서는 크게 개선되지만 IPW에서는 크게 개선 되지는 않음
- Overlap Assumption이 위배되는 IR 변수들의 경우 Balancing이 모든 방법에서 크게 개선되지 않음
- 세가지 방법들 중 IPW를 활용한 방법이 가장 작은 Mean SMD 값을 가지므로, IPW 방법을 통해 Balancing 진행 후 ATT 추정



DAG 3: Outcome Regression Formula

Outcome Regression

```
M_i(t): \log(Price\_week) \sim Treatment*(USAIR + KorIR + USACPI + KorCPI + Gold Price + USD KRW + CNY KRW + JPY KRW + EUR KRW + Covid19 + War)
Y_i(m,t): \log(Price) \sim Treatment*(USAIR + KorIR + USACPI + KorCPI + Gold Price + USD KRW + CNY KRW + JPY KRW + EUR KRW + Covid19 + War) + Price week + Price week * Treatment
```

- Mediator Model: 처치 변수(Treatment)와 교란 변수(Confounder)간의 교호 작용(Interaction) 고려
- Full Model: 처치 변수(Treatment)와 교란 변수(Confounder) 간의 교호 작용(Interaction) 뿐 아니라 Mediator와의 교호 작용도 고려
- 종속 변수와 Mediator 모두 0이상의 값만을 가지는 가격이라는 점에서 Gaussian 분포에 log link 적용
- 부트 스트랩(Bootstrap)을 통해 ATT, Direct Effect, Indirect Effect 신뢰구간 추정 * 1000 번 시행

DAG 3: Estimated ATT and Direct/Indirect Effect

Estimated ATT

Estimated ATT

$$\widehat{ au}_{ATT} = E\{Y_i(1) - Y_i(0)|Z=1\}$$
 = -10,500(원)

- 95 % Confidence Interval : [-12,800, -8,517]
 - ➡ 95% 신뢰구간에 0이 포함되지 않으므로 유의수준 0.05하에서 유의미한 효과가 있다고 볼 수 있음

Direct Effect

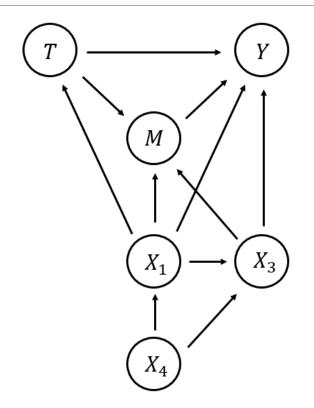
- Estimated Direct Effect = 855(원)
- 95 % Confidence Interval : [-373, 2,026]
 - 95% 신뢰구간에 0이 포함하므로 유의수준 0.05하에서 유의미한 효과가 있다고 볼 수 없음

> Indirect Effect

- Estimated Indirect Effect = -11,400(원)
- 95 % Confidence Interval : [-13,200, -9,810]
 - ➡ 95% 신뢰구간에 0이 포함되지 않으므로 유의수준 0.05하에서 유의미한 효과가 있다고 볼 수 있음

DAG 4: Remark

Remark DAG 4



T: I(한국 금리 \geq 미국 금리)

Y : 삼성전자 종가

 X_1 : 한국 소비자물가지수, 미국 소비자물가지수

 X_3 : 상품선물: 금 가격

환율: 미국 달러, 중국 위안, 일본 엔화, 유럽 유로

 X_4 : COVID19 여부, 러시아-우크라이나 전쟁

M: 과거 5일 동안의 평균 삼성전자 종가

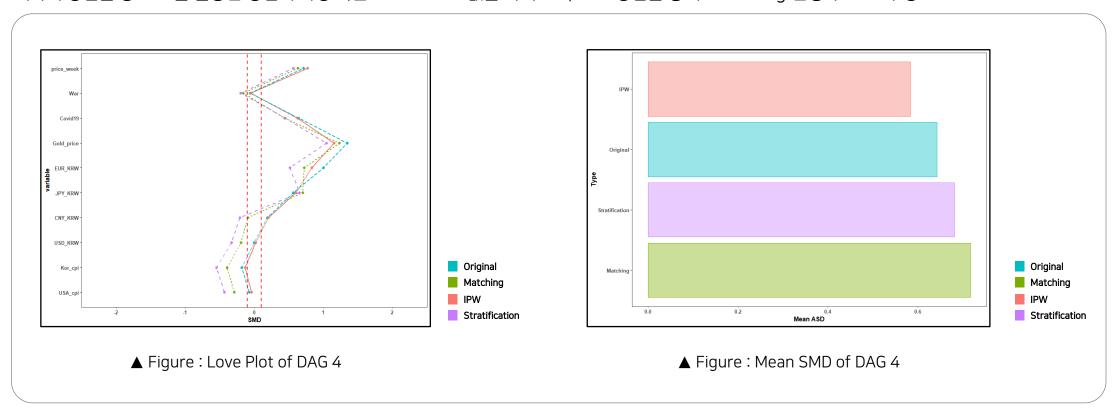
▲ Figure : DAG 4

■ DAG 4 : 한/미 금리 변수(Interest Rate)는 제외하고 Mediator(시차 변수)만 고려한 모형

DAG 4: Balancing

> Balancing

- 교란 변수(Confounder) 외에도 Mediator 간의 Balance 여부도 체크. 모든 방법에서 크게 개선되지는 않음
- 유로 환율(EUR KRW), 금 선물(Gold Price) 같이 초기 Balance가 심하게 위배되었던 변수들은 Balancing후에도 크게 개선되지는 않음
- 세가지 방법들 중 IPW를 활용한 방법이 가장 작은 Mean SMD 값을 가지므로, IPW 방법을 통해 Balancing 진행 후 ATT 추정



DAG 4: Outcome Regression Formula

Outcome Regression

```
M_i(t): \log(Price\_week) \sim Treatment * (USA CPI + Kor CPI + Gold Price +
                                      USD\ KRW + CNY\ KRW + IPY\ KRW + EUR\ KRW + Covid19 + War)
Y_i(m,t): \log(Price) \sim Treatment * (USA CPI + Kor CPI + Gold Price +
                                     USD\ KRW + CNY\ KRW + JPY\ KRW + EUR\ KRW + Covid19 + War) +
                        Price week + Price week * Treatment
```

- Mediator Model: 처치 변수(Treatment)와 교란 변수(Confounder)간의 교호 작용(Interaction) 고려
- Full Model: 처치 변수(Treatment)와 교란 변수(Confounder) 간의 교호 작용(Interaction) 뿐 아니라 Mediator와의 교호 작용도 고려
- 종속 변수와 Mediator 모두 0이상의 값만을 가지는 가격이라는 점에서 Gaussian 분포에 log link 적용
- 부트 스트랩(Bootstrap)을 통해 ATT, Direct Effect, Indirect Effect 신뢰구간 추정 * 1000 번 시행

DAG 4: Estimated ATT and Direct/Indirect Effect

Estimated ATT

Estimated ATT

$$\widehat{ au}_{ATT} = E\{Y_i(1) - Y_i(0)|Z=1\}$$
 = -5,729(원)

- **95 % Confidence Interval** : [-7,451, -4,073]
 - 95% 신뢰구간에 0이 포함되지 않으므로 유의수준 0.05하에서 유의미한 효과가 있다고 볼 수 있음

Direct Effect

- Estimated Direct Effect = 1,206(원)
- **95 % Confidence Interval**: [132, 2,234]
 - ➡ 95% 신뢰구간에 0이 포함되지 않으므로 유의수준 0.05하에서 유의미한 효과가 있다고 볼 수 있음

> Indirect Effect

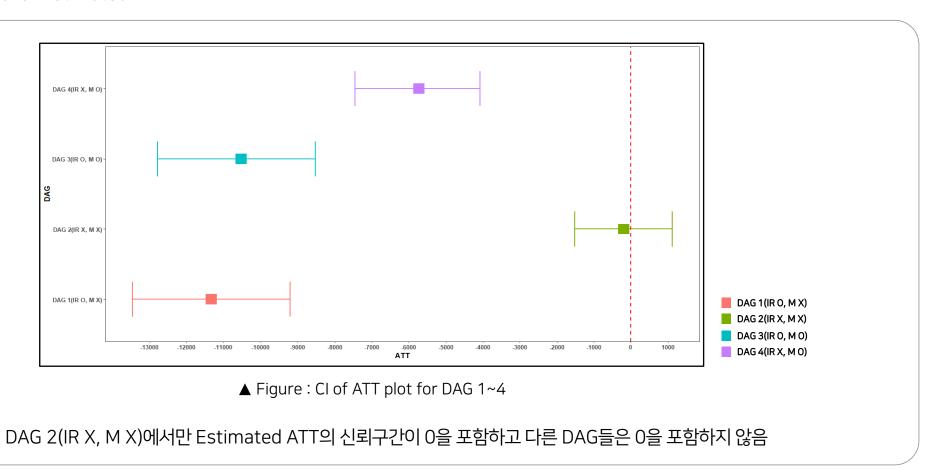
- Estimated Indirect Effect = -6,935(원)
- **95 % Confidence Interval** : [-8,547, -5,287]
 - ➡ 95% 신뢰구간에 0이 포함되지 않으므로 유의수준 0.05하에서 유의미한 효과가 있다고 볼 수 있음

IV. Conclusion2. Summary3. Follow-up Study

Comparison: 1) Confidence Interval of Estimated ATT

Comparison

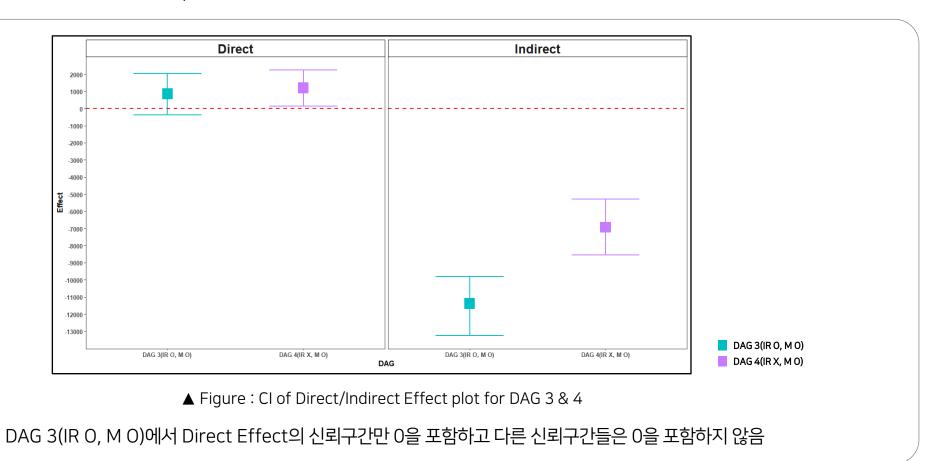
① Confidence Interval of Estimated ATT



Comparison: 2 Confidence Interval of Estimated Direct/Indirect Effect

Comparison

② Confidence Interval of Estimated Direct/Indirect Effect



Summary for ATT

> Summary

① Summary for Estimated ATT

	Contain IR as Confounder	Contain weekly price as Mediator	Balancing Method	Estimated ATT	Lower Bound	Upper Bound	Contain zero
DAG 1	0	X	IPW	-11,334	-13,464	-9204	X
DAG 2	X	X	IPW	-197	-1515	1,120	0
DAG 3	0	0	IPW	-10,522	-12,786	-8,517	X
DAG 4	Х	0	IPW	-5,729	-7,451	-4,073	Х

▲ Table: Summary table for Estimated ATT of DAG 1~4

- 비록 DAG 2의 경우 신뢰구간의 0이 포함되지만, 모든 Estimated ATT가 음의 ATT 값을 줌
- 한/미 금리와 Mediator 모두 고려하지 않은 DAG 2의 경우 유의수준 하에서 유의미한 Treatment effect를 추정하지 못함
- 한/미 금리를 교란 변수(Confounder)로 고려하는 경우, 그렇지 않은 경우보다 더 큰 음의 ATT 값을 줌
- 결과론적으로 해석하면, 한국 금리가 미국 금리보다 높은 경우 삼성전자의 종가가 더 낮아진다는 결과로 해석할 수 있음
- 높은 금리는 경기 약화 신호로, 특히 한국 금리가 미국 금리보다 높은 경우, 약화된 수익 전망으로 주식을 매도하는 경향을 기대할 수 있음

Summary for Direct/Indirect Effect

> Summary

② Summary for Estimated Direct & Indirect Effect

	Contain IR as Confounder	Contain weekly price as Mediator	Туре	Estimated Effect	Lower Bound	Upper Bound	Contain zero
DAG 3	0	0	Direct	855	-372	2,026	0
DAG 3	0	0	Indirect	-11,377	-13,239	-9,810	Х
DAG 4	Х	0	Direct	1,206	132	2,234	Х
DAG 4	X	0	Indirect	-6,935	-8,547	-5,287	Х

▲ Table: Summary table for Estimated Direct & Indirect Effect of DAG 3~4

- 앞선 Estimated ATT의 경우 모두 음의 값을 추정했지만, 추정된 Direct Effect를 보면 비록 DAG 3가 0을 포함하지만 모두 양의 값을 추정함
- 반면에 모든 추정된 Indirect Effect는 Direct Effect에 비해 비교적 큰 유의미한 음의 추정값을 줌
- 결과론적으로 해석하자면, 현재 주가에 대한 영향 정도는 금리 같은 외생적 요인 보다는 해당 주가의 과거 행태가 더 중요하다고 볼 수 있음
- 앞선 결과와 다르게 Direct Effect가 양의 값을 주는 것에 대해선, 금리 인상으로 인한 외국인 투자자들의 유입으로 주가가 상승하는 경향으로 유추 가능함

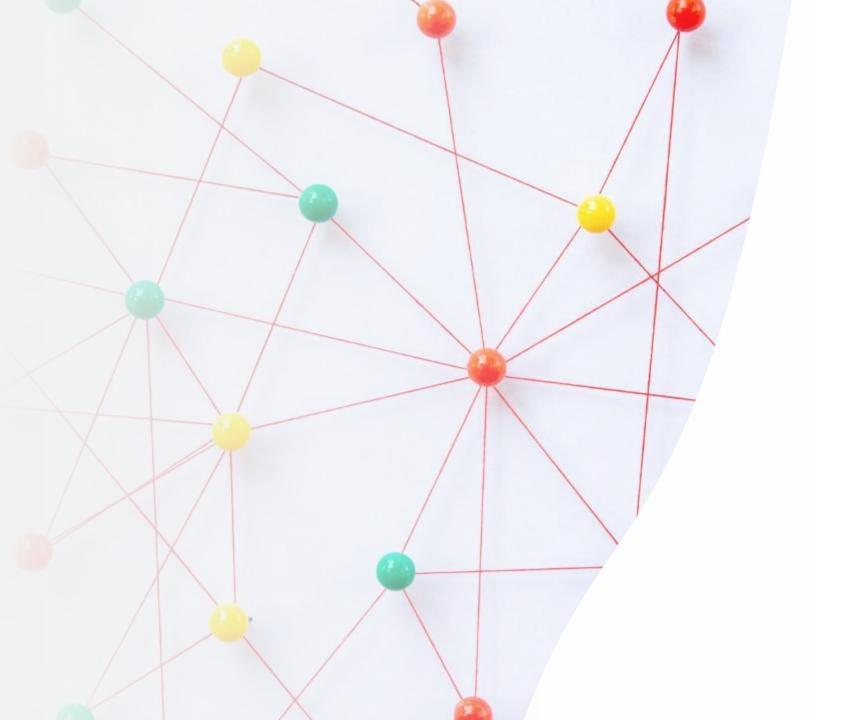
Follow-up Study

> Follow-up Study

- 본 분석에서 고려한 교란 변수(Confounder)들 외에 추가적인 변수들을 고려할 수 있음(ex. 반도체 시장 유망 정도를 지수화)
- 전반적인 데이터 가정(ex. CPI가 가장 외생적이다)들을 변경함으로써 다른 DAG 구조로 모형을 적합할 수 있음
- Cross-sectional 데이터로 가정하고 분석을 진행했지만, Longitudinal 분석 기법을 적용할 수 있음
- 준실험적(Quasi-experimental) 데이터이지만 실험적(Experimental) 데이터로 간주해서 분석을 진행했지만, 준실험적 데이터 분석 기법도 활용 가능함
- Mediator를 하나로만 보고 분석을 진행했지만 여러 Mediator를 고려한 모형을 통해 여러 개의 Path별로 Indirect Effect를 추정할 수 있음
- 종속 변수(Response variable)과 상호 연관성이 있는 변수들에 대해서 상호 연관성을 제거하고 분석에 포함시키는 방법을 고민할 수 있음

➤ Take Home Message(Just a joke)

- 아이작 뉴턴(Isaac Newton, 1643~1727) : *나는 천체의 움직임은 계산할 수 있지만 인간의 광기는 계산할 수 없다.* "I can calculate the motion of heavenly bodies, but not the madness of people." (Issac Newton)
- 조지 박스(George E. P. Box, 1919~2013) : 모든 모형은 틀렸다. 다만 몇몇 모형들은 유용하다. "All models are wrong, but some are useful." (George E. P. Box)



Thank you

V. Appendix

1. Reference

Reference

Reference

- Gil-Alana, Luis Alberiko, Emmanuel Joel Aikins Abakah, and María Fátima Romero Rojo. "Cryptocurrencies and stock market indices. Are they related?." Research in International Business and Finance 51 (2020): 101063.
- Corsi, Fulvio. "A simple approximate long-memory model of realized volatility." *Journal of Financial Econometrics* 7.2 (2009): 174-196.
- Greifer, Noah. "Covariate balance tables and plots: a guide to the cobalt package." Accessed March 10 (2020): 2020.
- Ho, D., Imai, K., King, G., Stuart, E., & Whitworth, A. (2018). Package 'Matchlt'.
- Greifer, Noah, and Maintainer Noah Greifer. "Package 'Weightlt'." (2019).
- Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). Mediation: R package for causal mediation analysis.