

Exploring Causal Inference in Galaxy Evolution via Double Machine Learning

Is machine learning good or bad for causal inference?

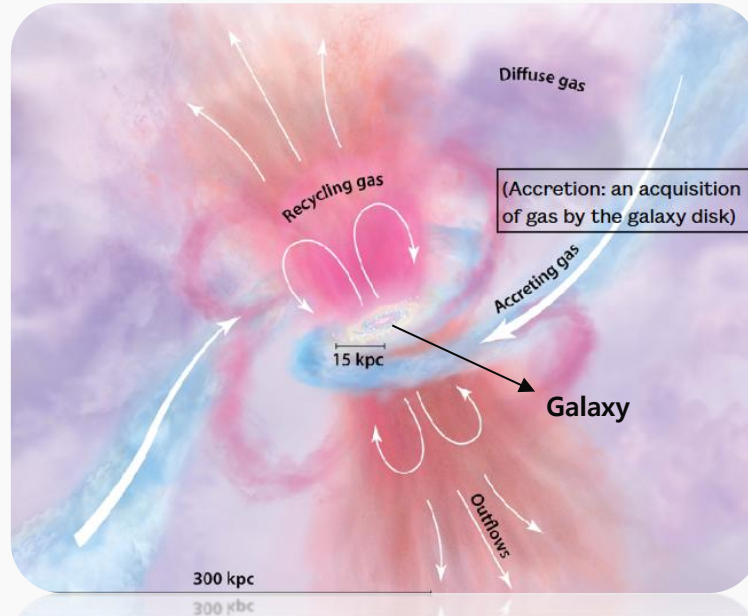
Minjae Kim (minjaeKim0827@github.io)

Dept. of Physics, KAIST

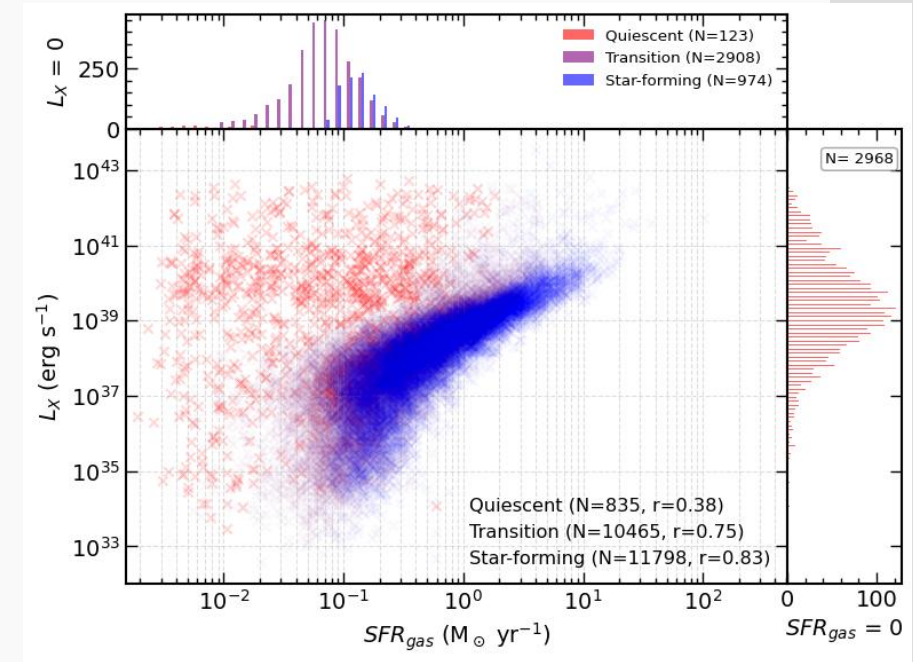


Galaxy and Circumgalactic Medium (CGM)

Circumgalactic Medium (CGM)



IllustrisTNG100-2

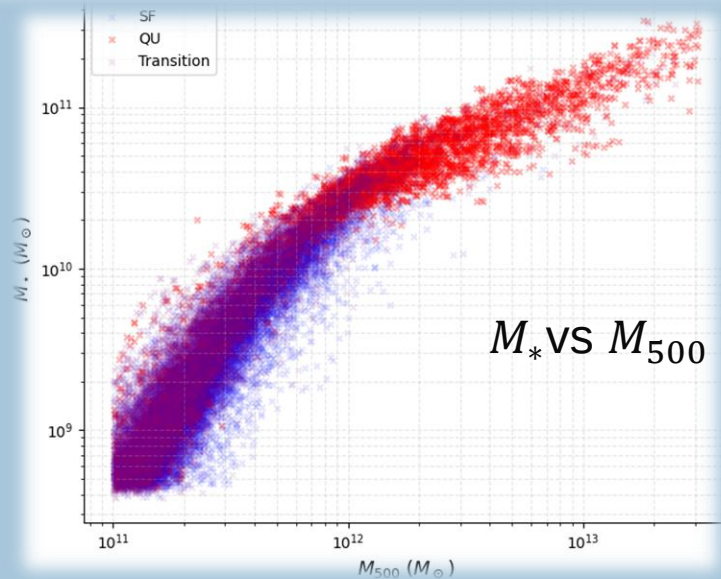
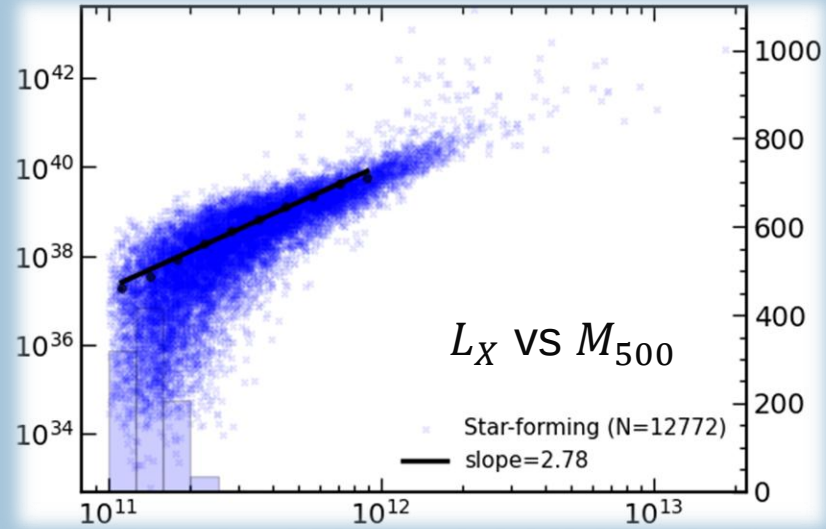


L_X vs SFR

Soft X-ray (0.25 – 2 keV) Luminosity (L_X) of CGM vs Star Forming Rate

- Strong Correlation ($r=0.83$) in **star forming galaxies**
- Does this imply $SFR \rightarrow L_X$? **Maybe ... But maybe not**

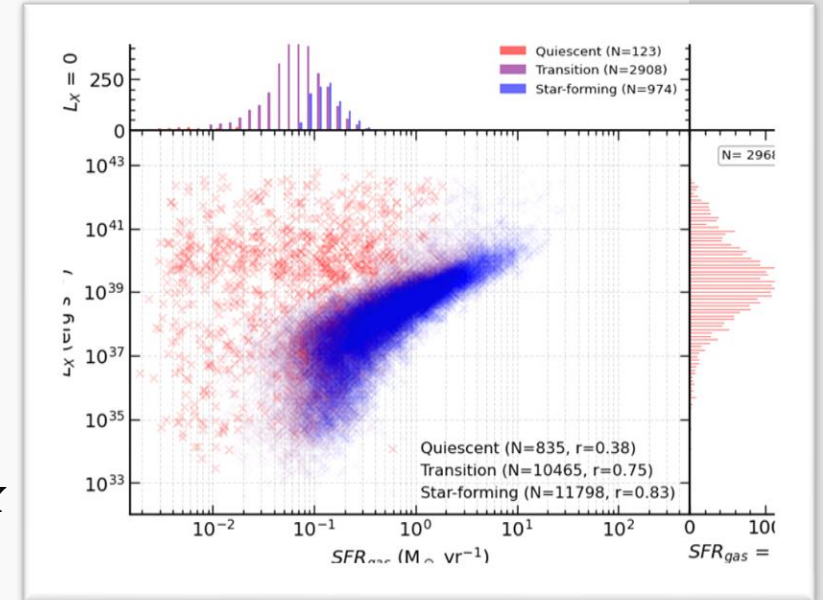
Realistic Example in Science : Galaxy



$$f_t(M_{500}, u_t) \rightarrow T = \text{SFR}$$

$$f_x(M_{500}, \text{SFR}, u_x) \rightarrow M_*$$

$$f_y(M_{500}, M_*, \text{SFR}, u_y) \rightarrow L_X$$

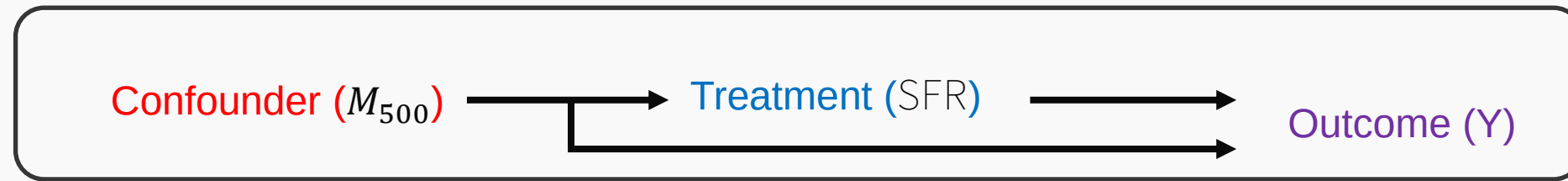


L_X vs SFR

Halo Mass (M_{500}) also has correlation with L_X & M_*
 Q. How to decompose the “contribution” of SFR vs Mass ?

Causal Inference with ML : 2 step

Science to infer **causal** relationships from correlation and to understand when and why they differ.



1) Identification (식별) : Getting causal estimation from observed data

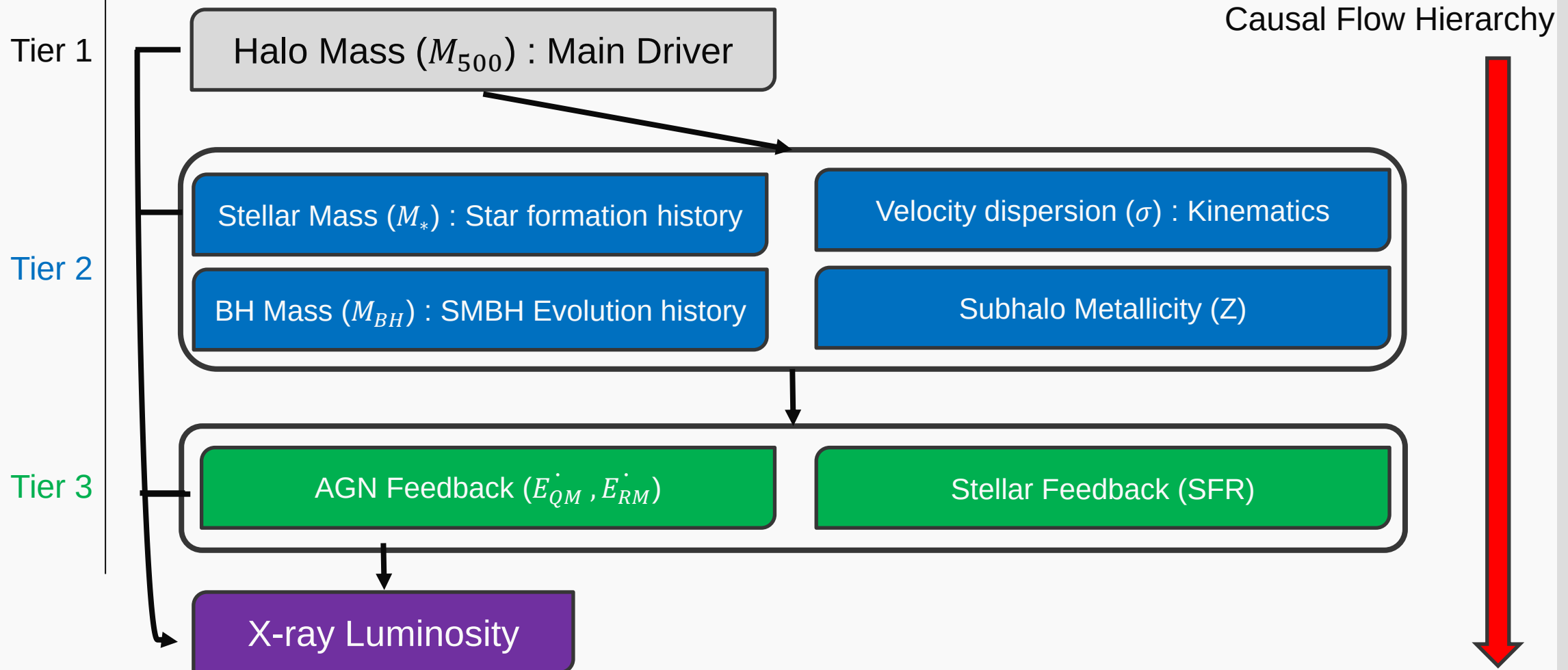
2) Estimation (추정) : **Double Machine Learning**

- Semi-parametric estimation method with Neyman orthogonality

Neyman orthogonality

Causal estimation is robust to the small error in ML model

Identification : Directed Acyclic Graph (DAG)



QM & RM

Each refers to the **Quasar Mode** and **Radio Mode**. In TNG, there are 2 modes for AGN feedback.

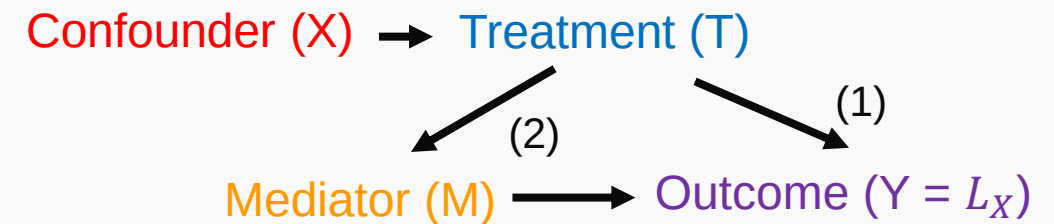
Estimation : Double Machine Learning

Redshift = 0, 0.1 & 0.2

Log scaled Sample : $[N_{sample}, N_{feature}] = [29665, 7]$

Double Machine Learning

Simple DAG

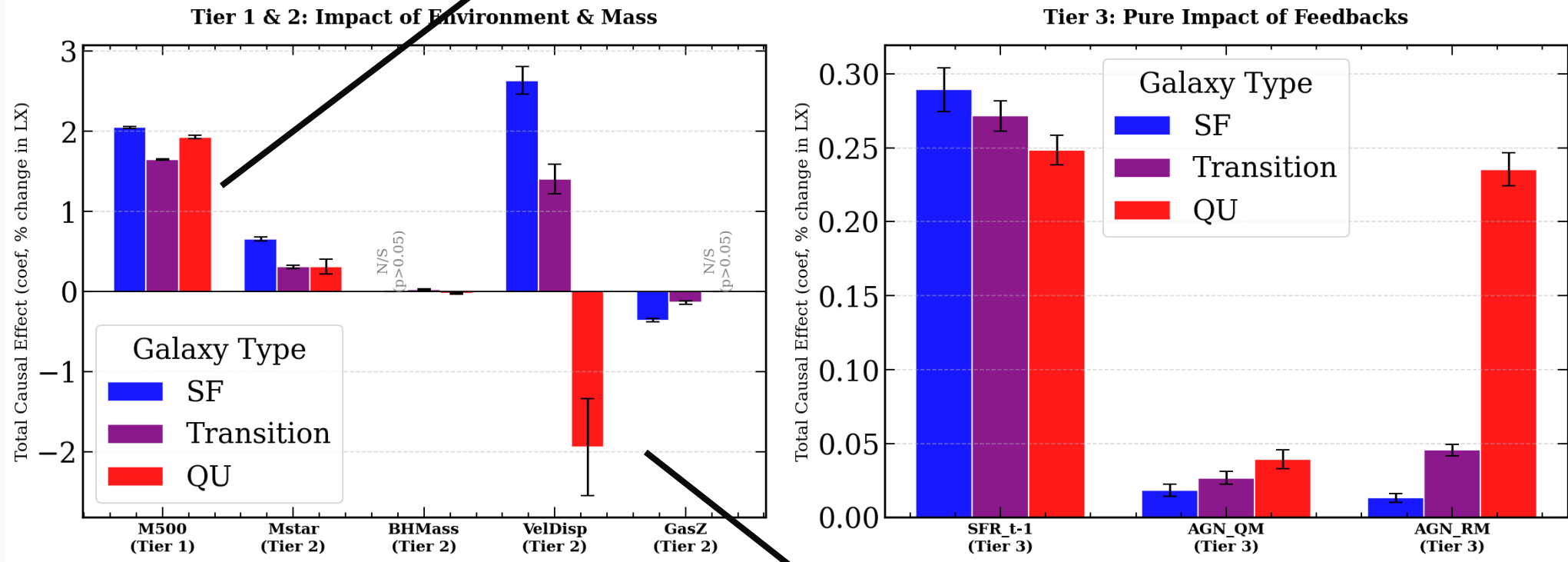


(1) : Direct route
(2) : Indirect (mediated) route

θ_T = If we fix **X**, how much **Y** changes because of the change of **T**?

Result

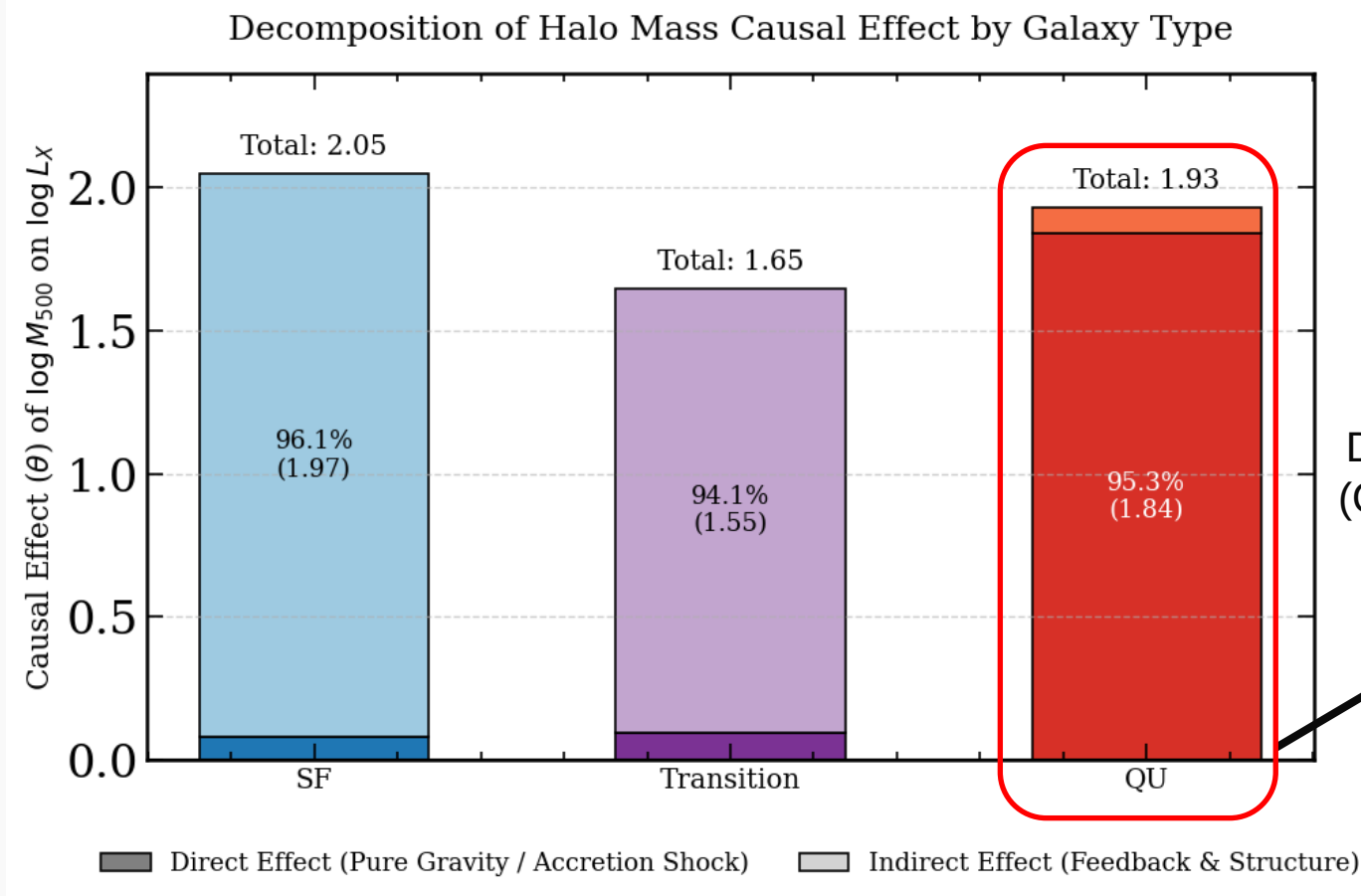
$\theta_{Total} \sim 2.0$ which is consistent with our initial fitting



In tier 2, velocity dispersion is dominant

SF = Star Forming, QU = Quiescent

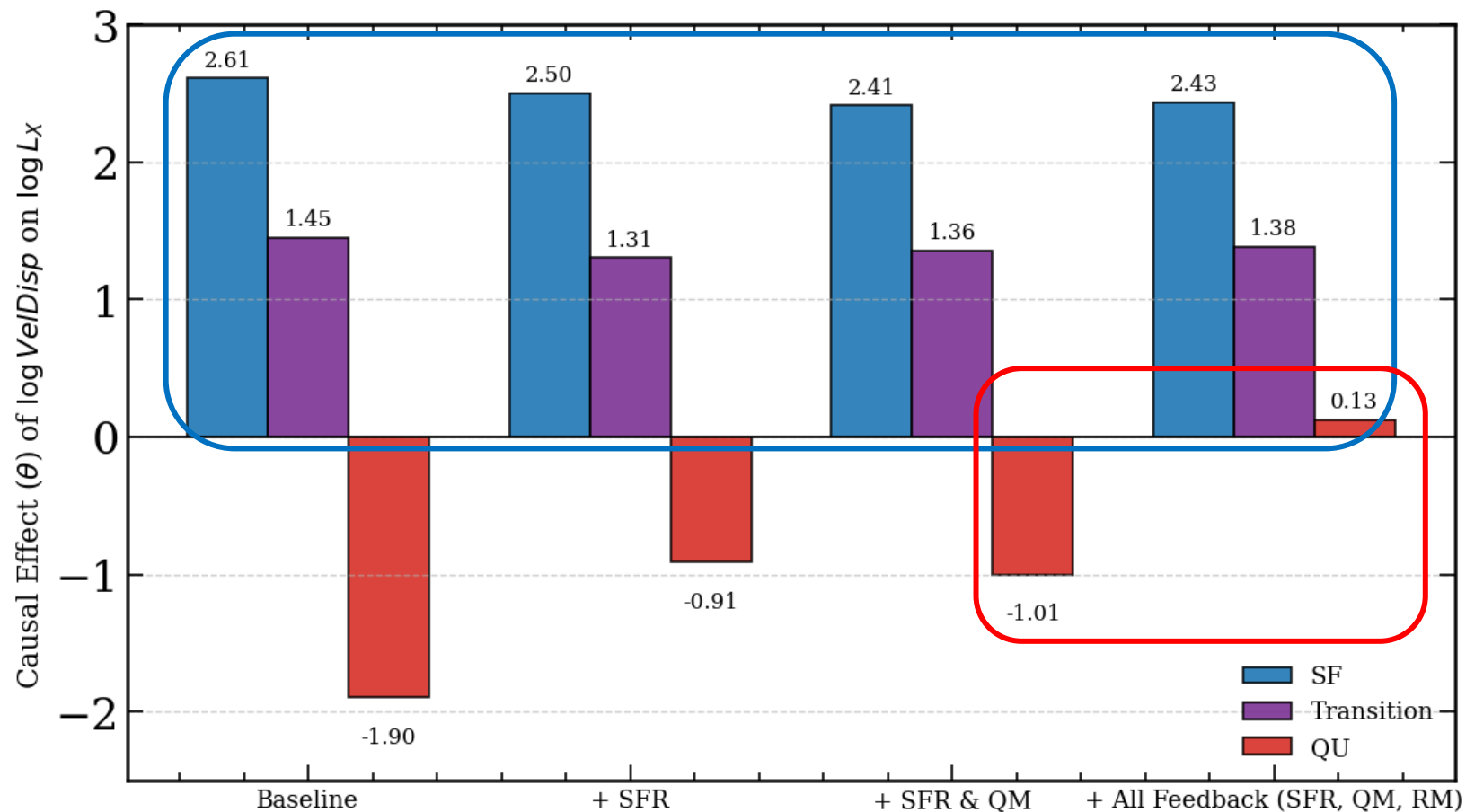
Result : Halo Mass



Direct Effect dominant, consistent with previous study
(Quasi-hydrostatic virialized gas from accretion shock)

Result : Velocity Dispersion

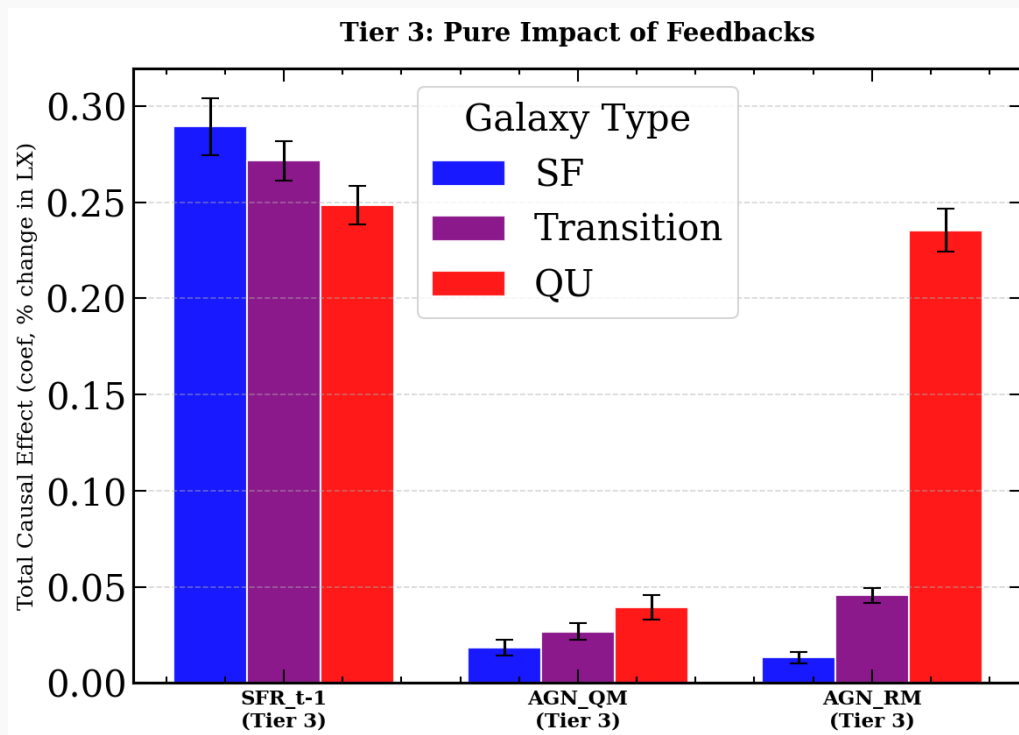
Mediation Analysis: Effect of VelDisp by Controlling Feedback



Star Forming & Transition
Direct path dominant ($\sigma \rightarrow L_X \uparrow$)

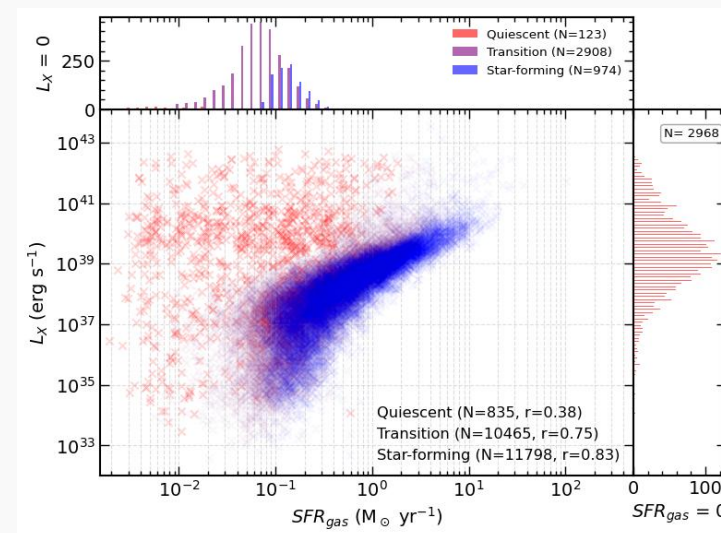
Quiescent
Indirect, RM mediated path dominant
($\sigma \rightarrow AGN_{RM} \rightarrow L_X \downarrow$)

Result : Feedback



SFR still plays a unneglectable role

Radio mode of AGN is also significant in Quiescent galaxy



Caveat : Population-level biases in θ

$$T = g(X) + \mu$$

Confounder (X) $\rightarrow$ Treatment (T)



$$\hat{g}(X) = a \times g(X)$$

Residue of Treatment ($\tilde{T}$)



$$\begin{aligned}\tilde{T} &= T - \hat{g}(X) = g(X) + \mu - a \times g(X) \\ &= (1 - a)g(X) + \mu\end{aligned}$$

$$\tilde{T} = \mu \text{ if } a = 0$$

$$\begin{aligned}Y &= \theta T + h(X) + \varepsilon \\ &= \theta g(X) + h(X) + \theta\mu + \varepsilon\end{aligned}$$

Confounder (X) $\rightarrow$ Outcome (Y)



$$\hat{f}(X) = b \times f(X)$$

Residue of Outcome ($\tilde{Y}$)

$$\begin{aligned}\tilde{Y} &= Y - \hat{f}(X) = f(X) + \theta\mu + \varepsilon - b \times f(X) \\ &= (1 - b)f(X) + \theta\mu + \varepsilon\end{aligned}$$

$$\tilde{Y} = \theta\mu + \varepsilon \text{ if } b = 0$$

$0 < a, b < 1$: Shrinkage coefficient

Caveat : Population-level biases in θ

$$\tilde{T} = \mu \text{ if } a = 0$$

$$\tilde{Y} = \theta\mu + \varepsilon \text{ if } b = 0$$

θ = unbiased, genuine value



$$\theta_{biased} = \frac{\theta Var(\mu) + \theta(1-a)(1-b)Var(g(X)) + (1-a)(1-b)Cov(g(X), h(X))}{Var(\mu) + (1-a)^2 Var(g(X))}$$

Ghost Causality



ML biased causal effect : could be precisely biased

$0 < a, b < 1$: Shrinkage coefficient

Caveat : Population-level biases in θ

$$\tilde{T} = \mu \text{ if } a = 0$$

$$\tilde{Y} = \theta\mu + \varepsilon \text{ if } b = 0$$

Open Question : How to restrict a & b close to 1 ? Or, how to get real causal effect from simulation-biased causal effect ?

Simulation (TNG) biased causal effect

$0 < a, b < 1$: Shrinkage coefficient

Questions?

Back up slides

How to estimate bias?

$$E[Y(1) | T = 1] - E[Y(0) | T = 0] \quad \xrightarrow{\text{What we measure (Correlation)}}$$

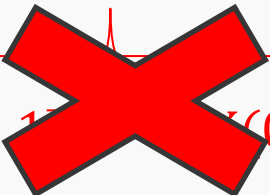
$$= \underbrace{\{E[Y(1) | T = 1] - E[Y(0) | T = 1]\}}_{\text{Average Treatment effect on the treated (ATT) : Causality}} + \underbrace{\{E[Y(0) | T = 1] - E[Y(0) | T = 0]\}}_{\text{Bias}}$$

Average Treatment effect on the treated (ATT) : Causality

Because of **bias**, correlation is not guaranteed to be **causal** relation

How to estimate bias?

$$E[Y(1) | T = 1] - E[Y(0) | T = 0] \quad \text{What we measure (Correlation)}$$

$$= \underbrace{\{E[Y(1) | T = 1] - E[Y(0) | T = 1]\}}_{\text{Average Treatment effect on the treated (ATT) : Causality}} + \underbrace{\{E[Y(0) | T = 1] - E[Y(0) | T = 0]\}}_{\text{Bias}}$$


Average Treatment effect on the treated (ATT) : Causality

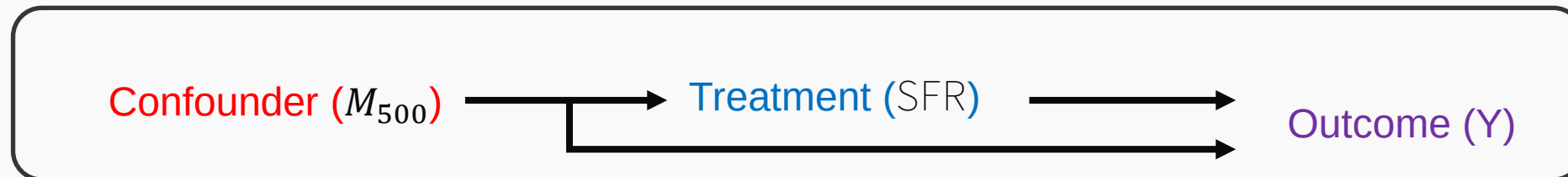
But, making bias = 0 is hard to implement.
Then, how to estimate bias? → Machine Learning !

Problem : Machine Learning is better to the “predict” rather than “what if”

What is Causal Inference?

Science to infer **causal** relationships from correlation and to understand when and why they differ.

- **Treatment (T)** : An intervention applied to influence the outcome of interest.
- **Outcome (Y)** : The outcome or variable we aim to understand (i.e., why it behaves the way it does).
- **Confounder (X)** : A variable that affects both the **treatment (T)** and the **outcome (Y)**, potentially creating a spurious or misleading association between them



Causal Inference with ML : 2 step

1) Identification (식별) : Getting causal estimation from observed data

- Directed Acyclic Graph (DAG)

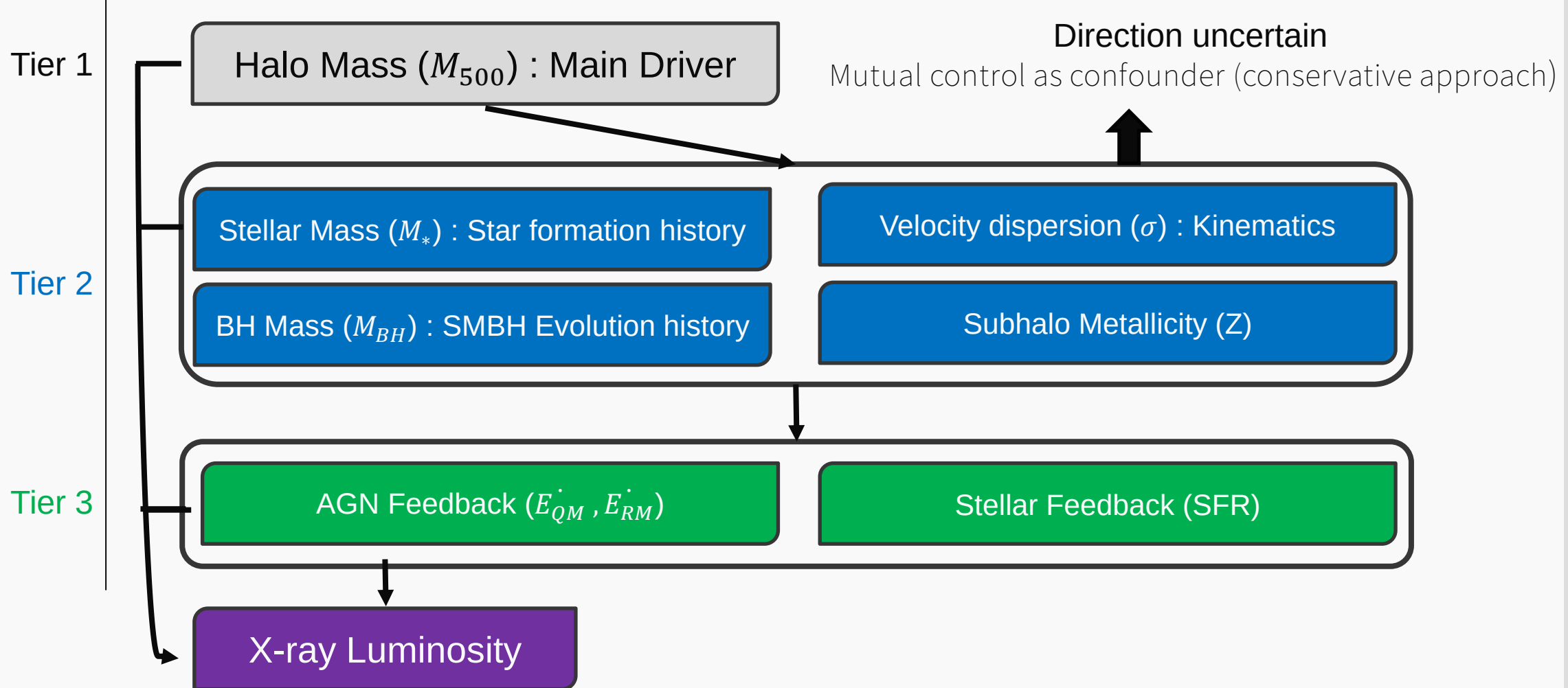
2) Estimation (추정) : Double Machine Learning

- Semi-parametric estimation method with Neyman orthogonality

Neyman orthogonality

Causal estimation is robust to the small error in ML model

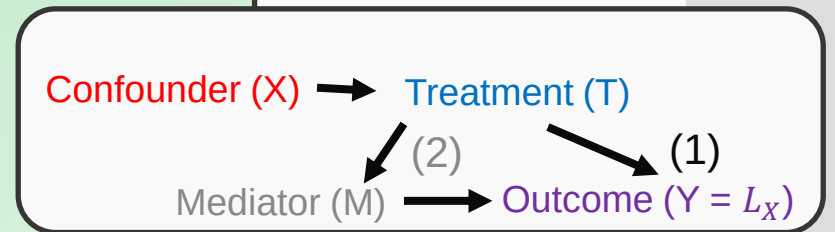
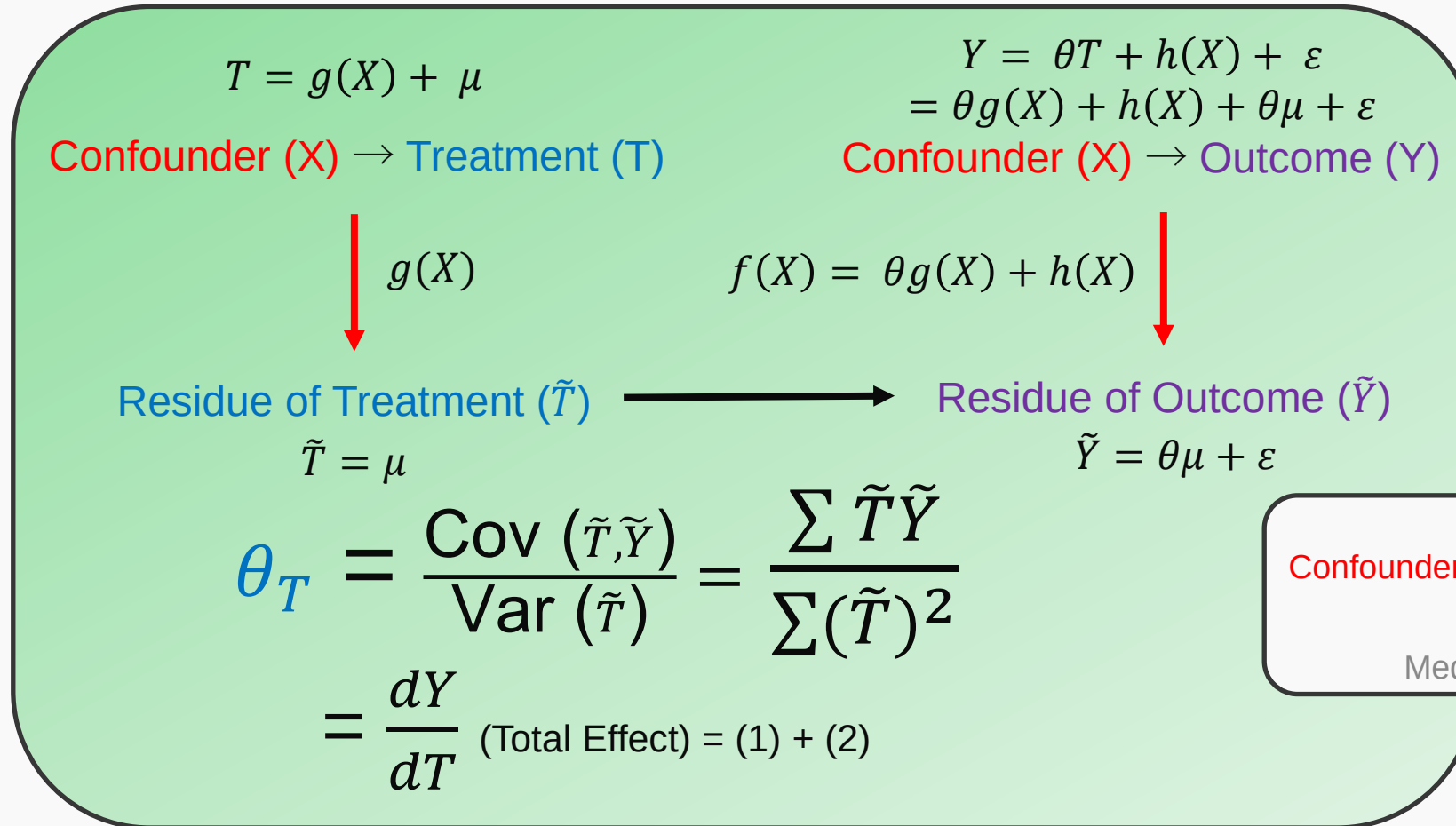
Identification : Directed Acyclic Graph (DAG)



QM & RM

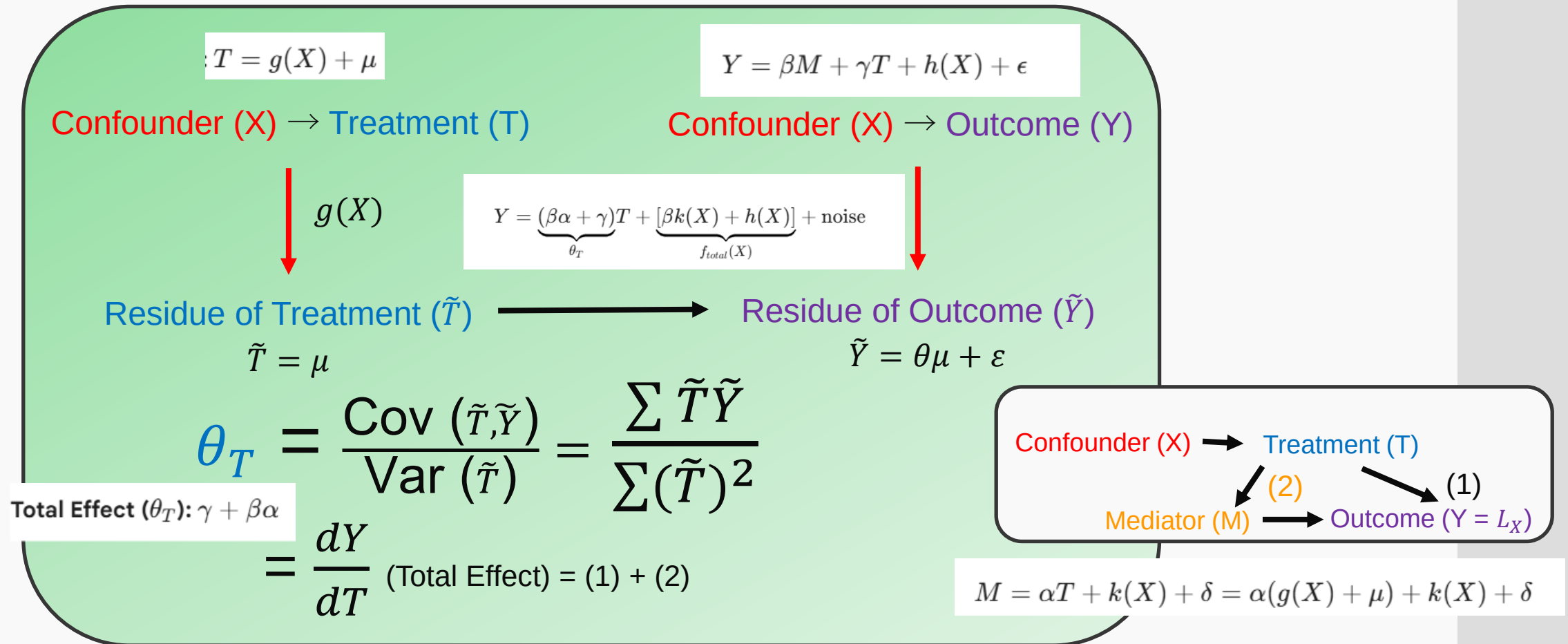
Each refers to the Quasar Mode and Radio Mode. In TNG, there are 2 modes for SMBH

Estimation : Double Machine Learning

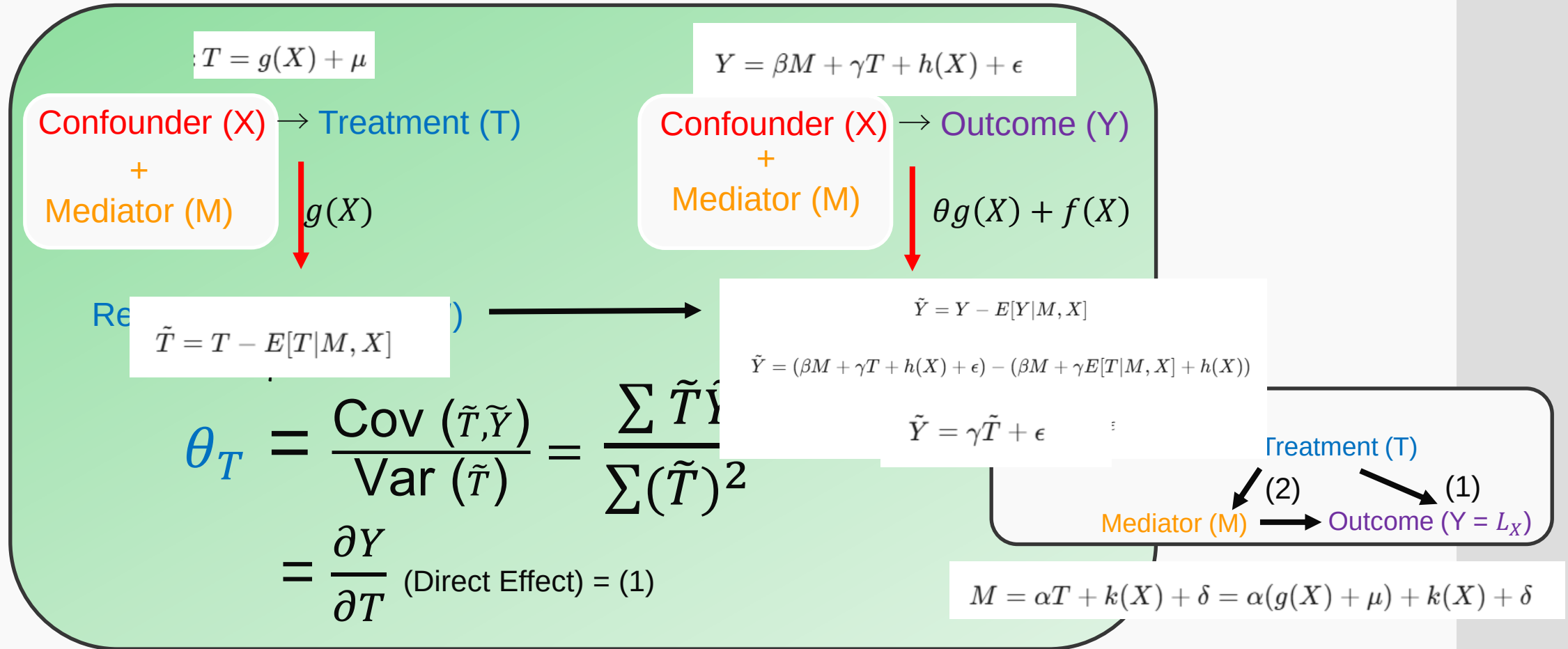


↓ = Random Forest ML model

Estimation : DML with Mediator



Estimation : DML with direct effect



Caveat : Population-level biases in θ

$$T = g(X) + \mu$$

Confounder (X) $\rightarrow$ Treatment (T)



$$\hat{g}(X) = a \times g(X)$$

Residue of Treatment ($\tilde{T}$)



$$\begin{aligned}\tilde{T} &= T - \hat{g}(X) = g(X) + \mu - a \times g(X) \\ &= (1 - a)g(X) + \mu\end{aligned}$$

$$\tilde{T} = \mu \text{ if } a = 0$$

$$\begin{aligned}Y &= \theta T + h(X) + \varepsilon \\ &= \theta g(X) + h(X) + \theta\mu + \varepsilon\end{aligned}$$

Confounder (X) $\rightarrow$ Outcome (Y)



$$\hat{f}(X) = b \times f(X)$$

Residue of Outcome ($\tilde{Y}$)

$$\begin{aligned}\tilde{Y} &= Y - \hat{f}(X) = f(X) + \theta\mu + \varepsilon - b \times f(X) \\ &= (1 - b)f(X) + \theta\mu + \varepsilon\end{aligned}$$

$$\tilde{Y} = \theta\mu + \varepsilon \text{ if } b = 0$$

$0 < a, b < 1$: Shrinkage coefficient

Caveat : Population-level biases in θ

$$\begin{aligned}\tilde{T} &= T - \hat{g}(X) = g(X) + \mu - a \times g(X) \\ &= (1 - a)g(X) + \mu\end{aligned}$$

$$\begin{aligned}\tilde{Y} &= Y - \hat{f}(X) = f(X) + \theta\mu + \varepsilon - b \times f(X) \\ &= (1 - b)f(X) + \theta\mu + \varepsilon\end{aligned}$$

$$\begin{aligned}&\text{Cov}((1 - a)g(X) + \mu, (1 - b)\theta g(X) + (1 - b)h(X) + \theta\mu + \varepsilon) \\ &= \underbrace{\theta\text{Var}(\mu)}_{\text{Signal}} + \underbrace{(1 - a)(1 - b)\theta\text{Var}(g(X))}_{\text{Leakage 1}} + \underbrace{(1 - a)(1 - b)\text{Cov}(g(X), h(X))}_{\text{Leakage 2}}\end{aligned}$$

$$\text{Var}((1 - a)g(X) + \mu) = \text{Var}(\mu) + (1 - a)^2\text{Var}(g(X))$$

Ghost Causality

$$\hat{\theta}_T = \frac{\theta\text{Var}(\mu) + (1 - a)(1 - b)\theta\text{Var}(g(X)) + (1 - a)(1 - b)\text{Cov}(g(X), h(X))}{\text{Var}(\mu) + (1 - a)^2\text{Var}(g(X))}$$

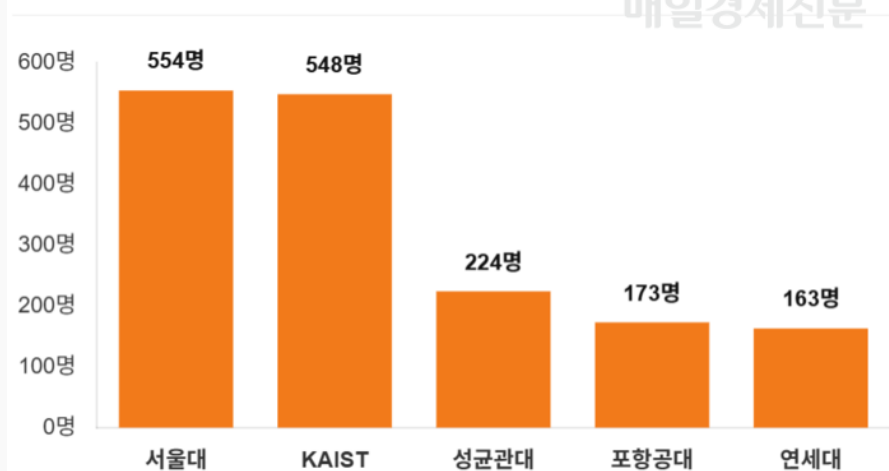
Simulation (TNG) biased causal effect

$0 < a, b < 1$: Shrinkage coefficient

Causal Inference in Social Science



2025년 과학고·영재고 출신 주요 대학 진학 수



단위: 명 / 출처: 대학알리미

If I go to the science high school, I could go KAIST easily...!



Why is it dangerous?

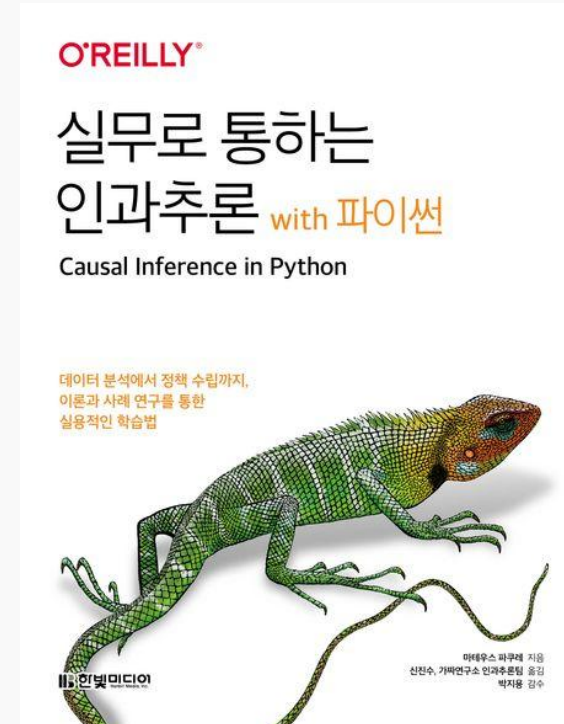
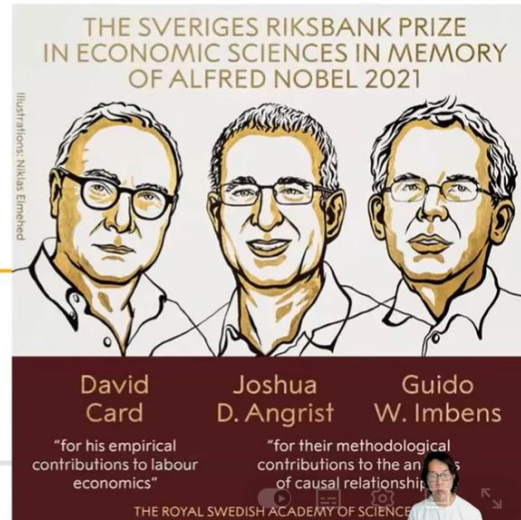
- 1) Selection bias of science high school
- 2) Impossible to implement counterfactuals
- 3) Confounding variable (support of parents)

Causal Inference in Social Science

인과추론 연구 읽기 시리즈 2

2021 노벨경제학상 심층분석
: 그들은 인과추론을
어떻게 혁신하였는가?

0:25 / 53:44
인과추론 연구 읽기 시리즈



What if, Human Population $\rightarrow$ Galaxy Population ?

Caveat : Population-level biases in θ

1. Potential Unmeasured Confounders : hidden variables
2. Limitation of Static DAG (Dynamic Feedback Loops)
3. Additivity Assumption in PLR Model ($Y = \theta T + g(X) + \varepsilon$)
 - a constant treatment effect
4. Implicit priors in TNG