Sequential Underspecified Instrument Selection for Cause-Effect Estimation.

Elisabeth Ailer, Jason Hartford, Niki Kilbertus

HELMHOLTZ MUNICH
Recursion
Mila
TUM

NewScientist

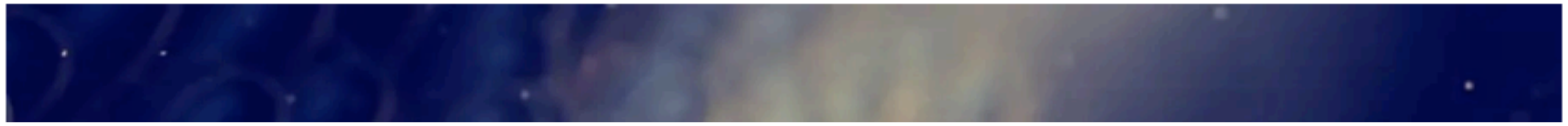
Mind

Could your gut bacteria influence how intelligent you are?

People who are genetically predisposed to have higher levels of Fusicatenibacter bacteria scored better on verbal and mathematical tests, while those with more Oxalobacter scored lower

By Carissa Wong

10 June 2023



NewScientist

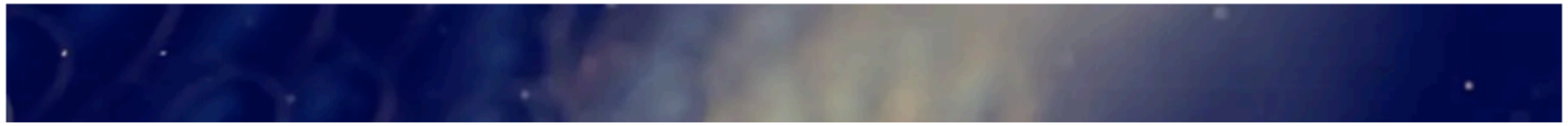
Mind

Could your gut bacteria influence how intelligent you are?

People who are genetically predisposed to have higher levels of Fusicatenibacter bacteria scored better on verbal and mathematical tests, while those with more Oxalobacter scored lower

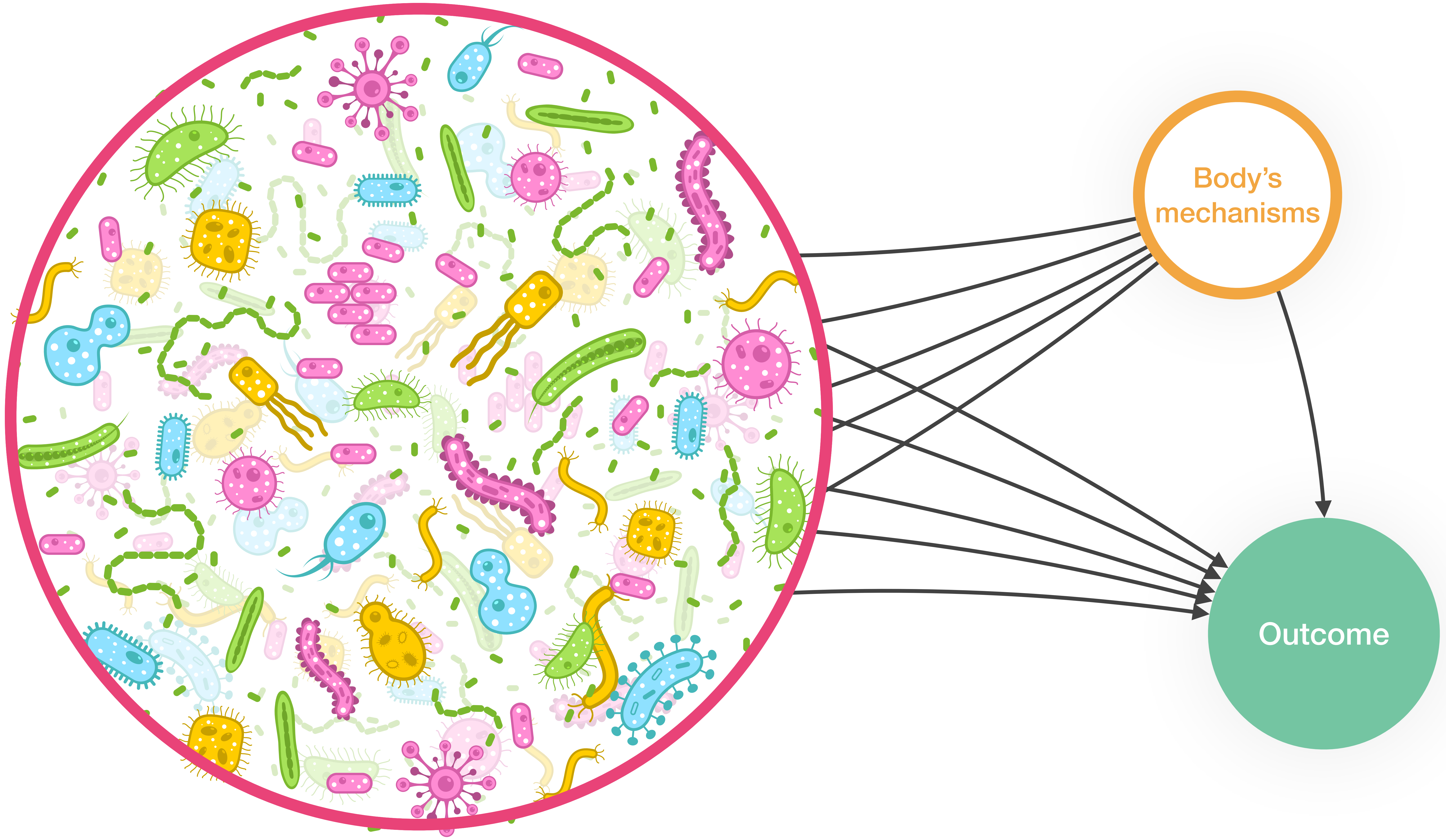
By Carissa Wong

10 June 2023



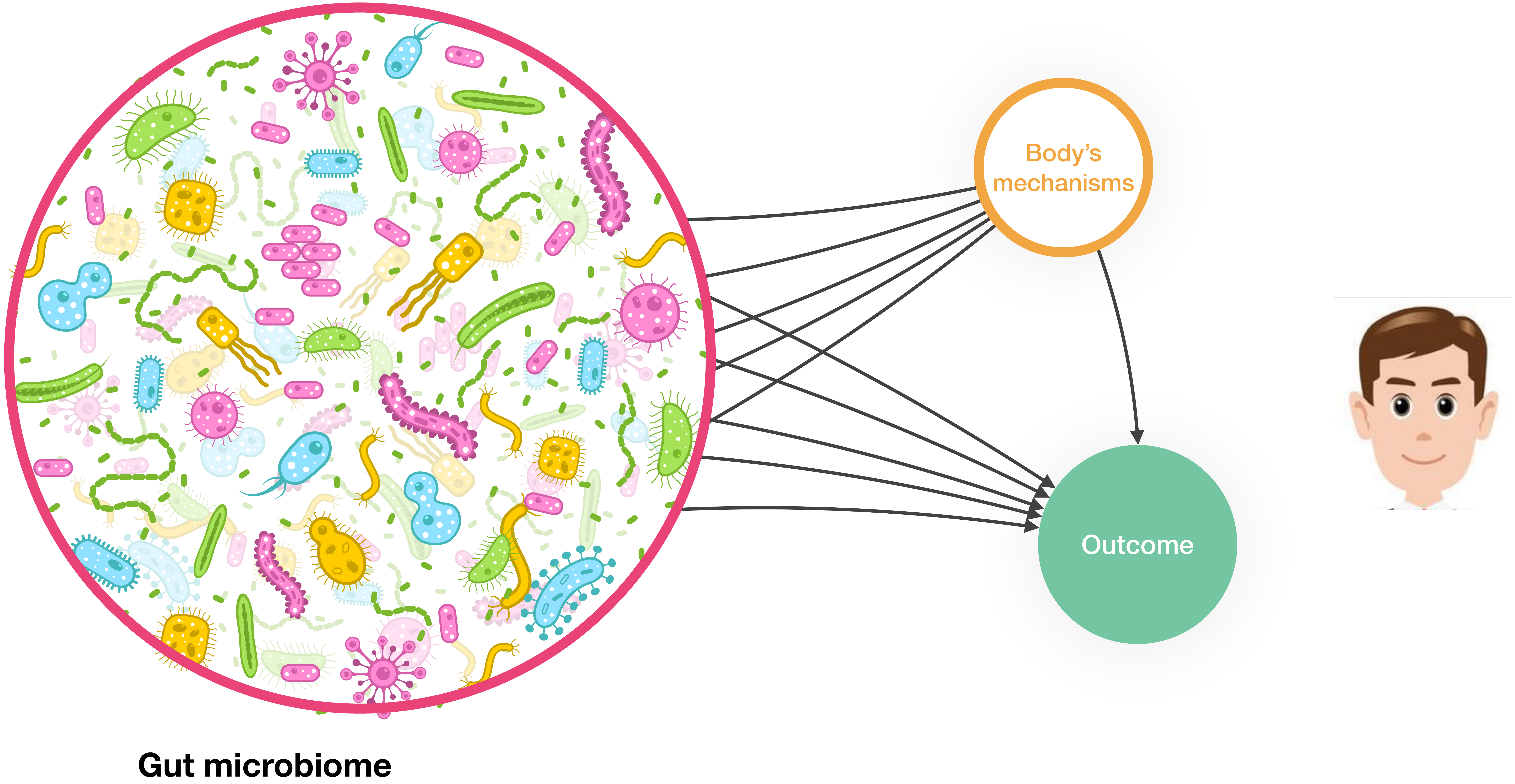
Which bacteria make me smart ???

Which bacteria make me smart ???

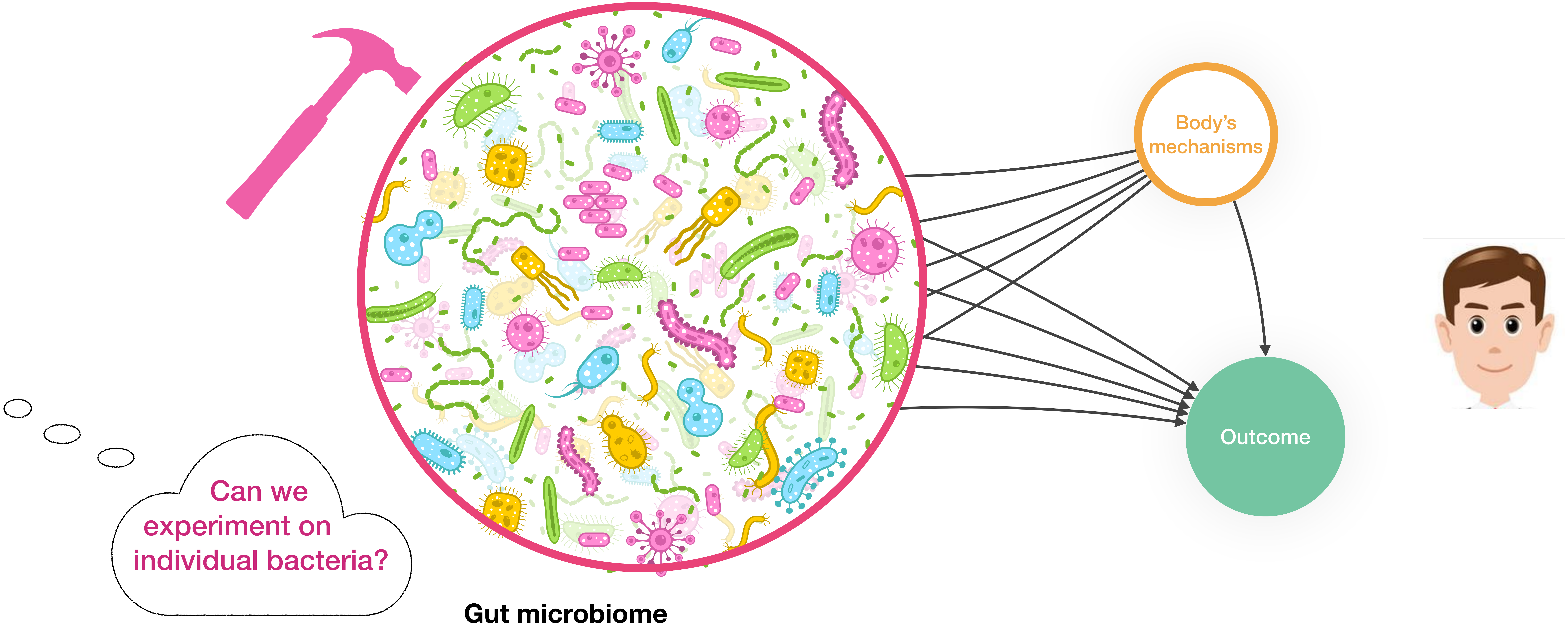


Gut microbiome

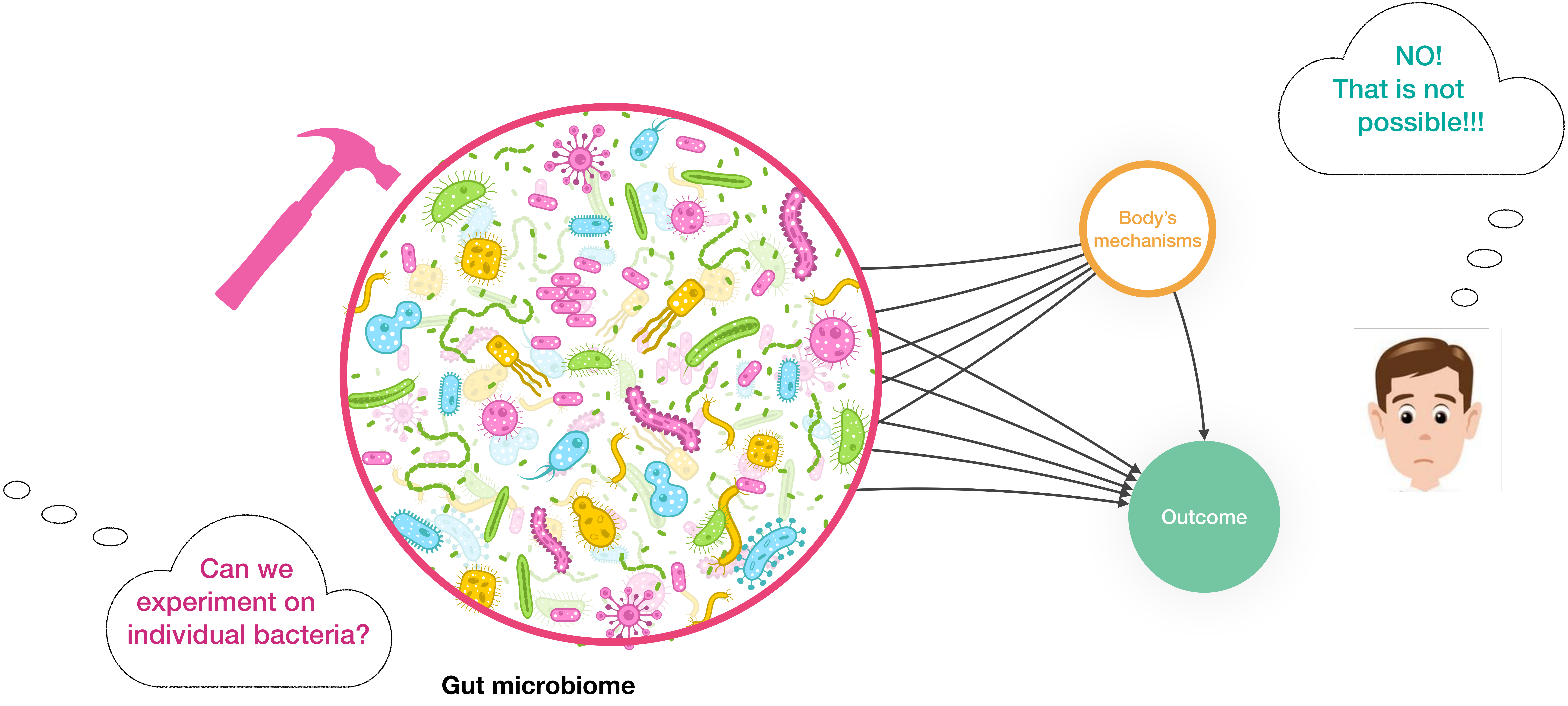
Which bacteria make me smart ???



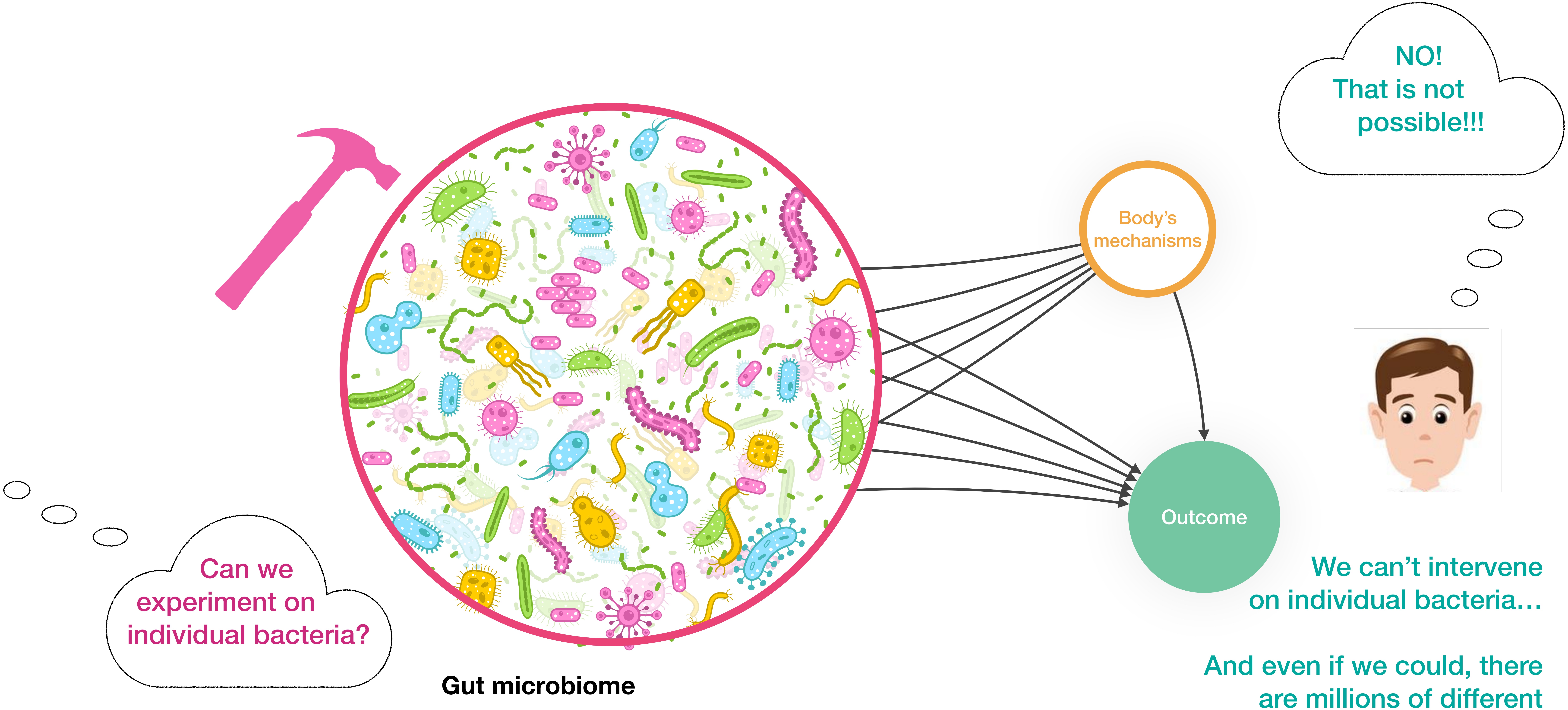
Which bacteria make me smart ???



Which bacteria make me smart ???



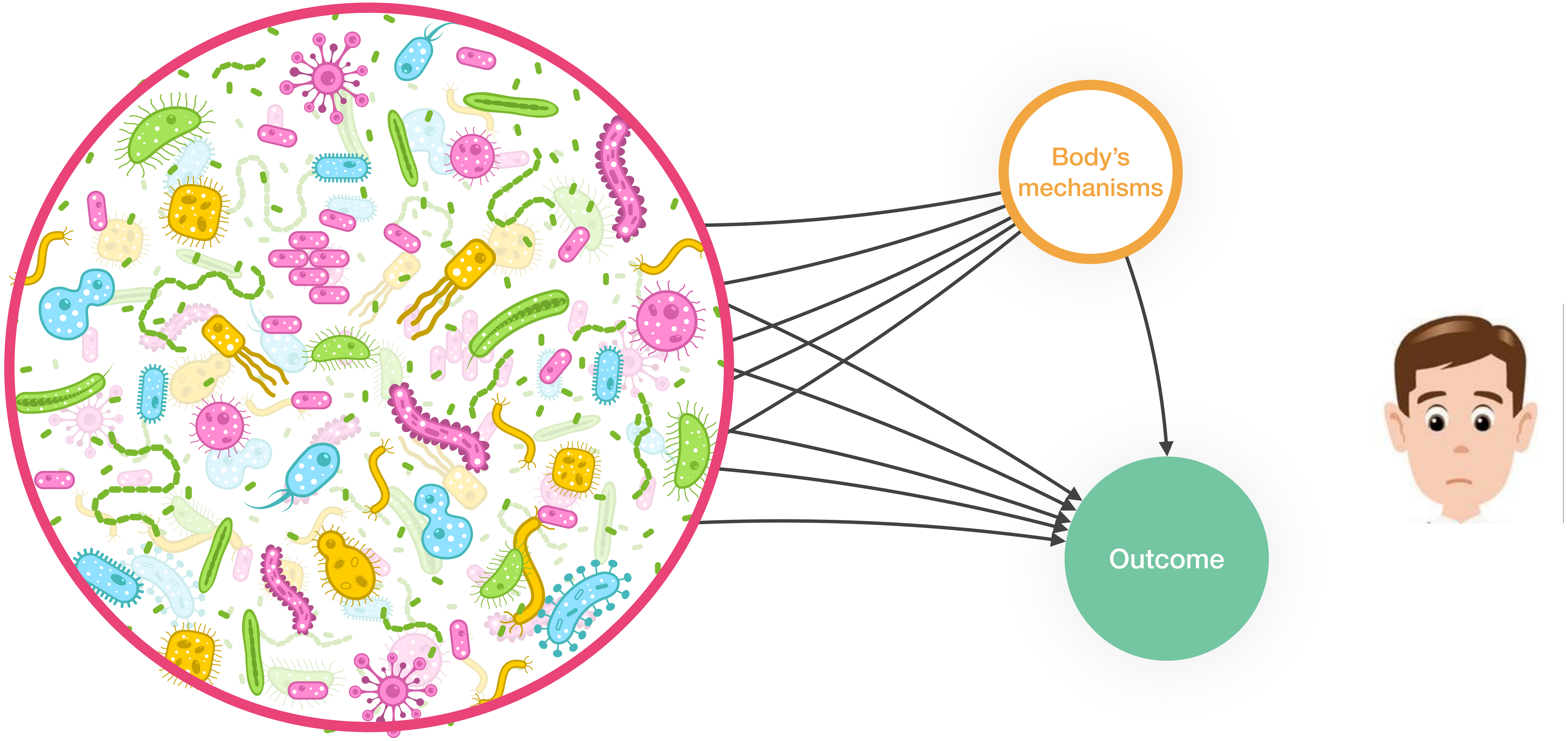
Which bacteria make me smart ???



We can't intervene
on individual bacteria…

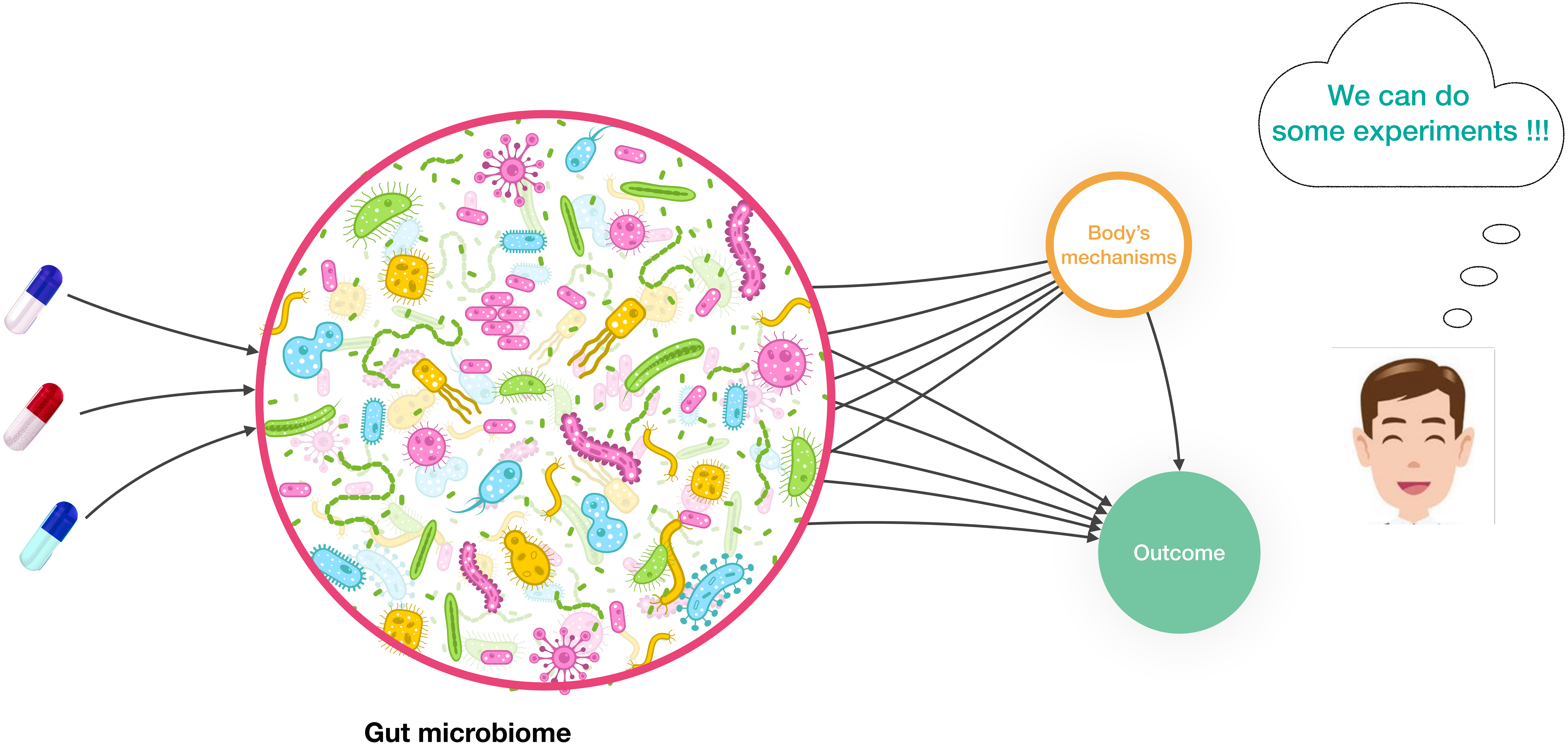
And even if we could, there
are millions of different
types!!!

Which bacteria make me smart ???

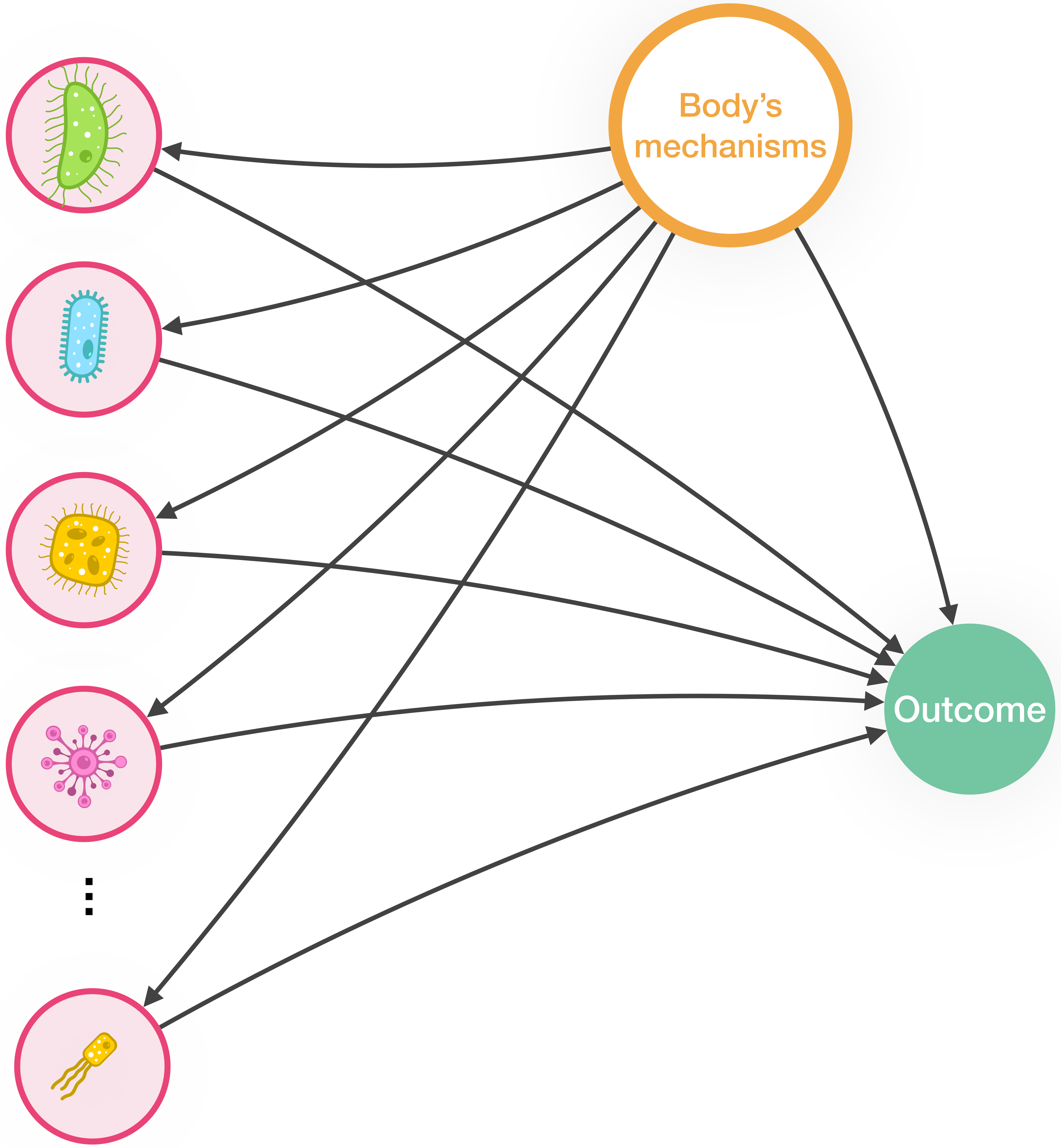


Gut microbiome

Which bacteria make me smart ???

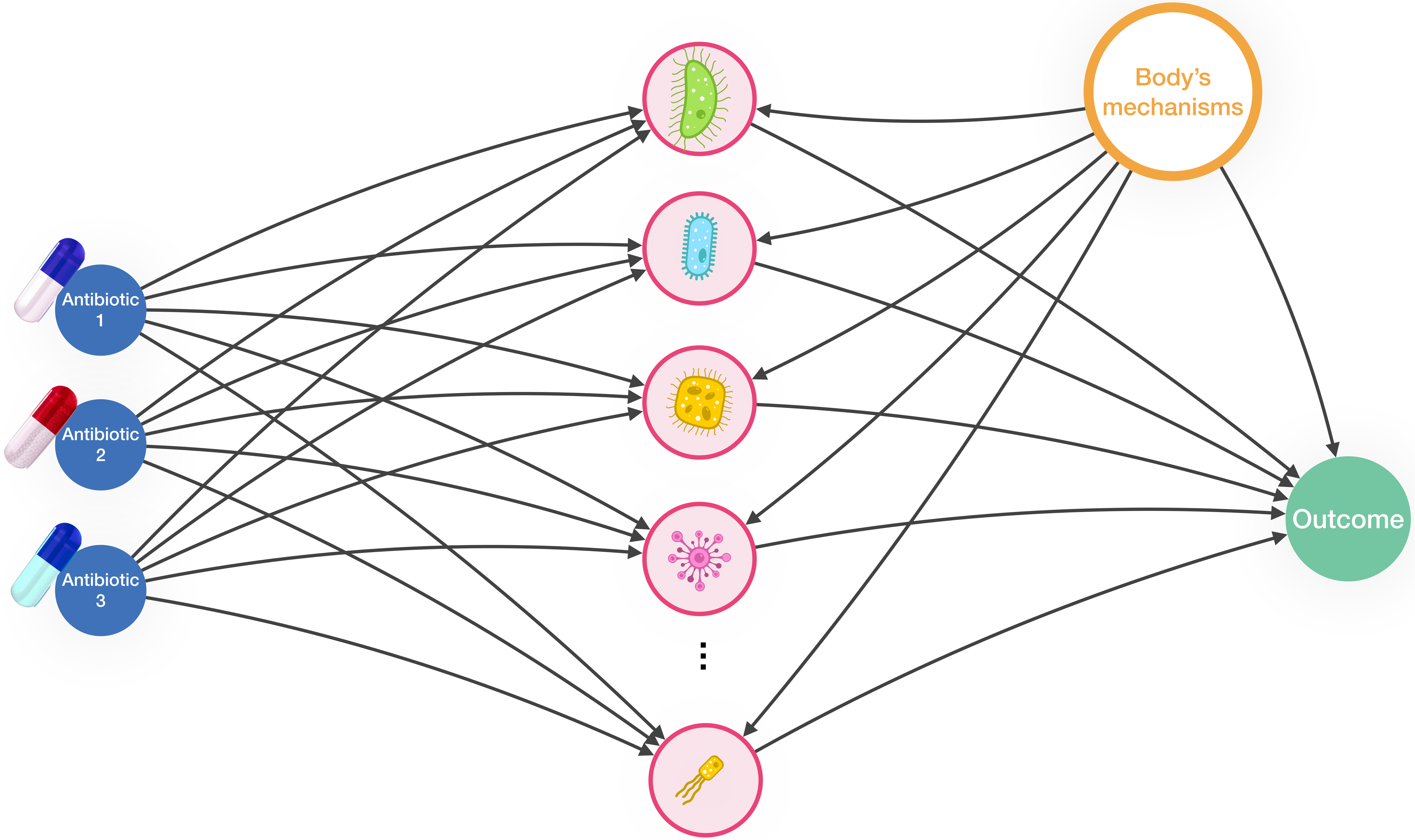


We look at the setting more formally …



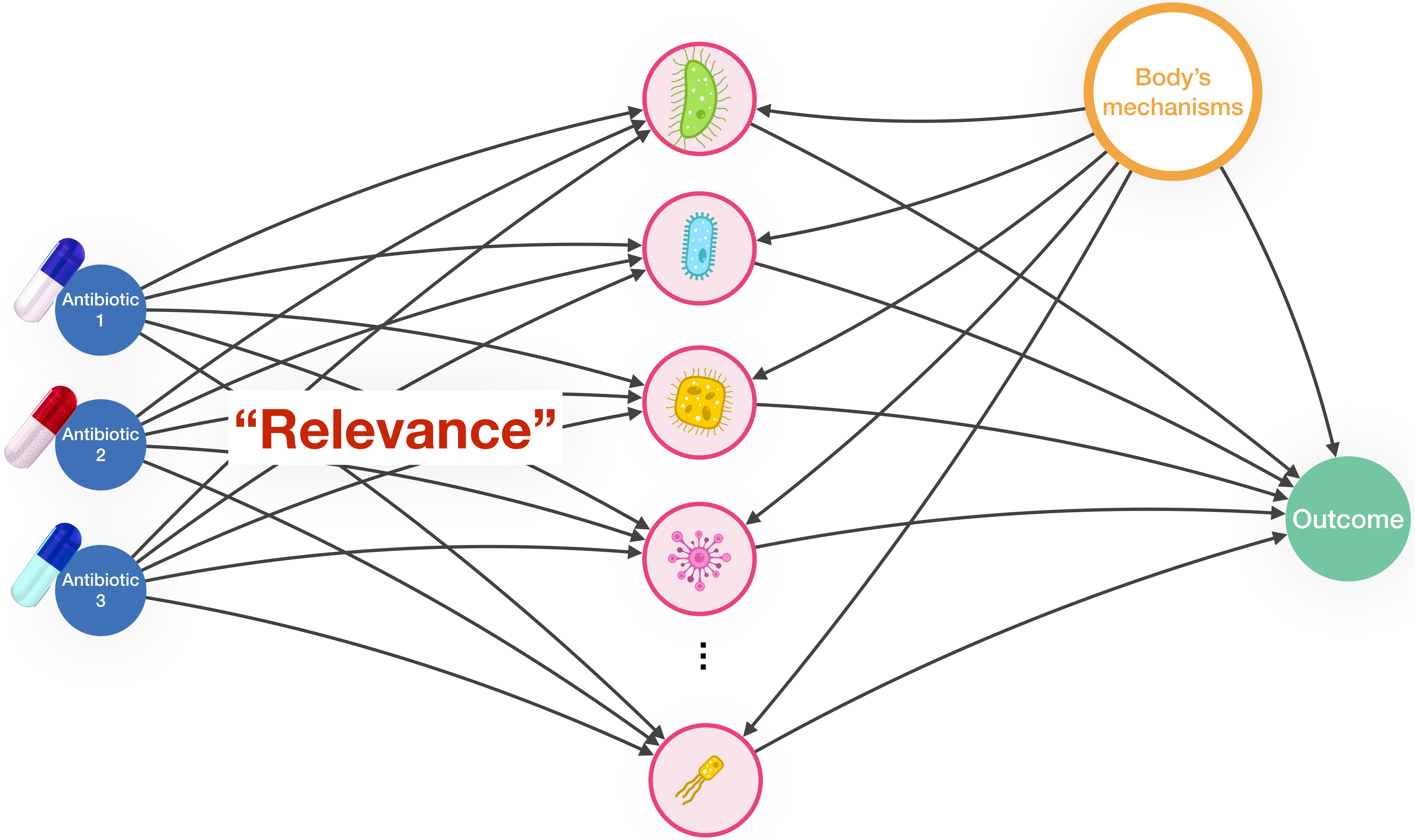
see also: Pfister N. and Peters J. Identifiability of sparse causal effects using instrumental variables. In Proceedings of the 38th Annual Conference on Uncertainty in Artificial Intelligence (UAI), PMLR 180, 1613-1622

We look at the setting more formally ...



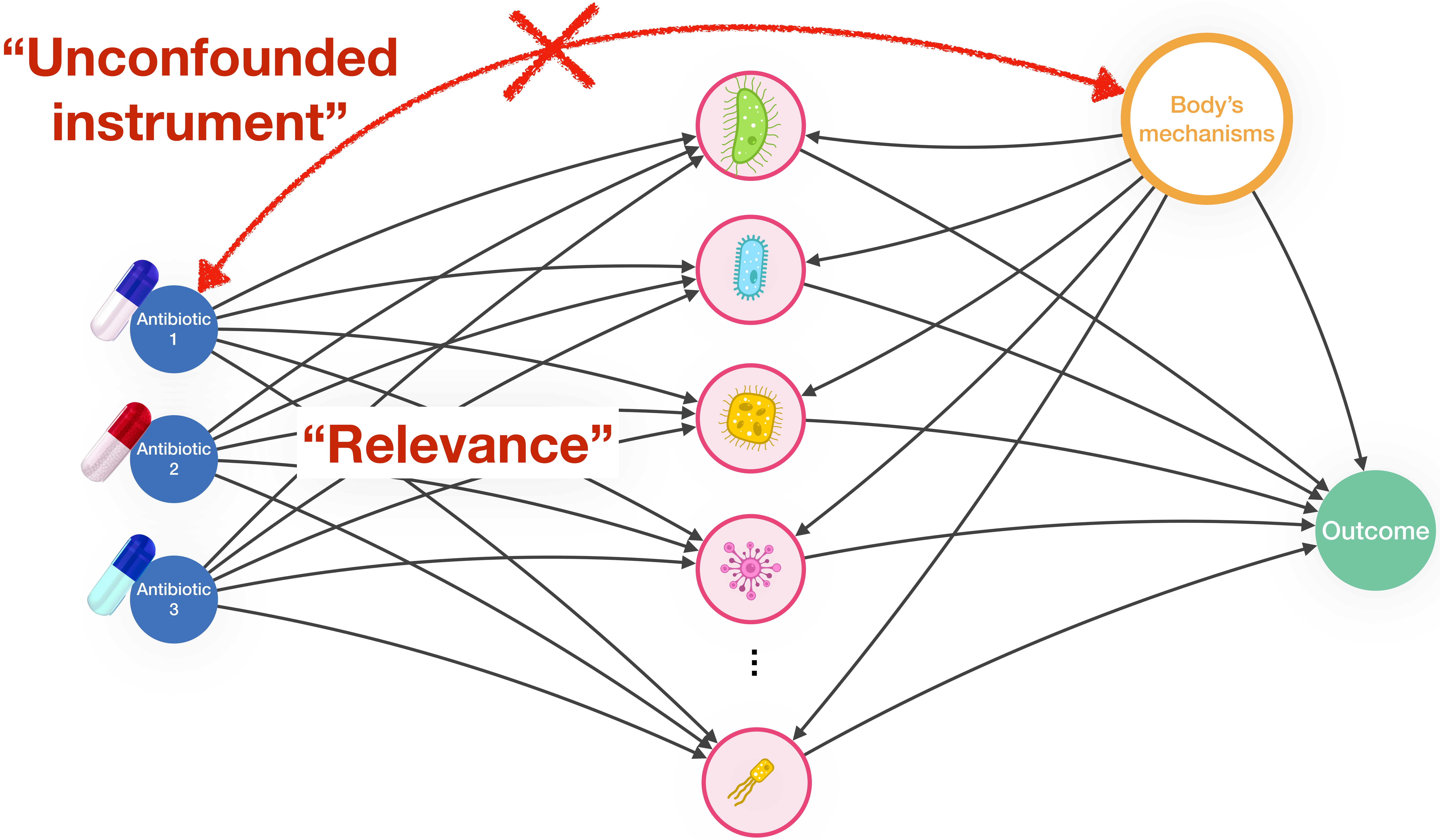
see also: Pfister N. and Peters J. Identifiability of sparse causal effects using instrumental variables. In Proceedings of the 38th Annual Conference on Uncertainty in Artificial Intelligence (UAI), PMLR 180, 1613-1622

We look at the setting more formally …



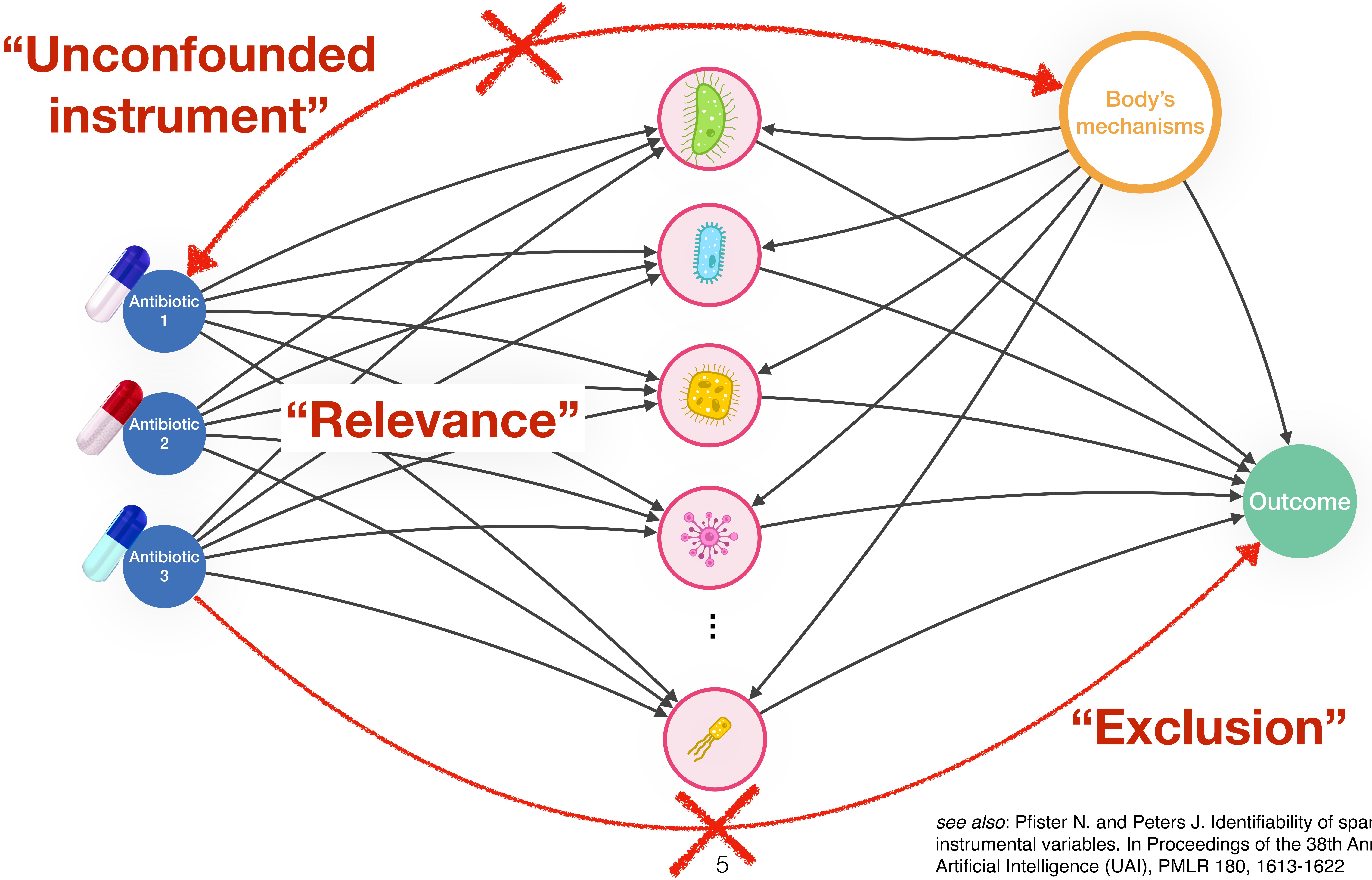
see also: Pfister N. and Peters J. Identifiability of sparse causal effects using instrumental variables. In Proceedings of the 38th Annual Conference on Uncertainty in Artificial Intelligence (UAI), PMLR 180, 1613-1622

We look at the setting more formally ...



see also: Pfister N. and Peters J. Identifiability of sparse causal effects using instrumental variables. In Proceedings of the 38th Annual Conference on Uncertainty in Artificial Intelligence (UAI), PMLR 180, 1613-1622

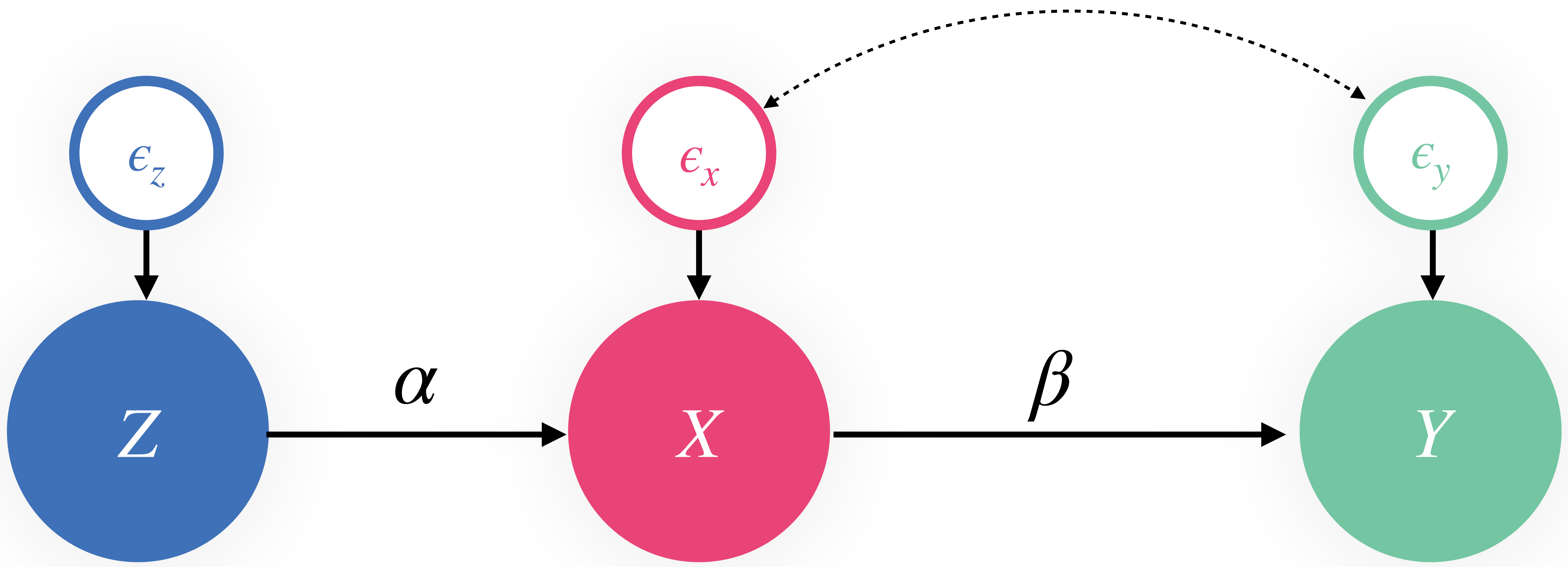
We look at the setting more formally ...



see also: Pfister N. and Peters J. Identifiability of sparse causal effects using instrumental variables. In Proceedings of the 38th Annual Conference on Uncertainty in Artificial Intelligence (UAI), PMLR 180, 1613-1622

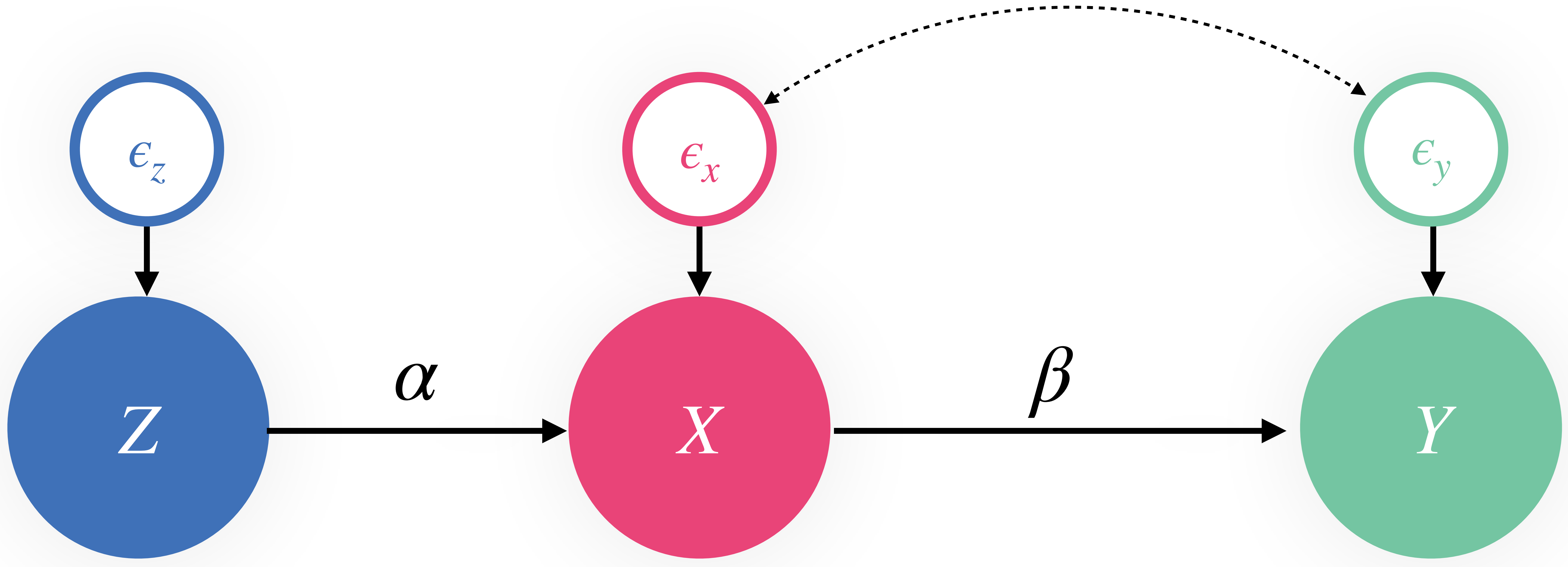
We look at the setting more formally ...

$$X := Z\alpha + \epsilon_x, \quad Y := X\beta + \epsilon_y, \quad X \not\perp \epsilon_y$$



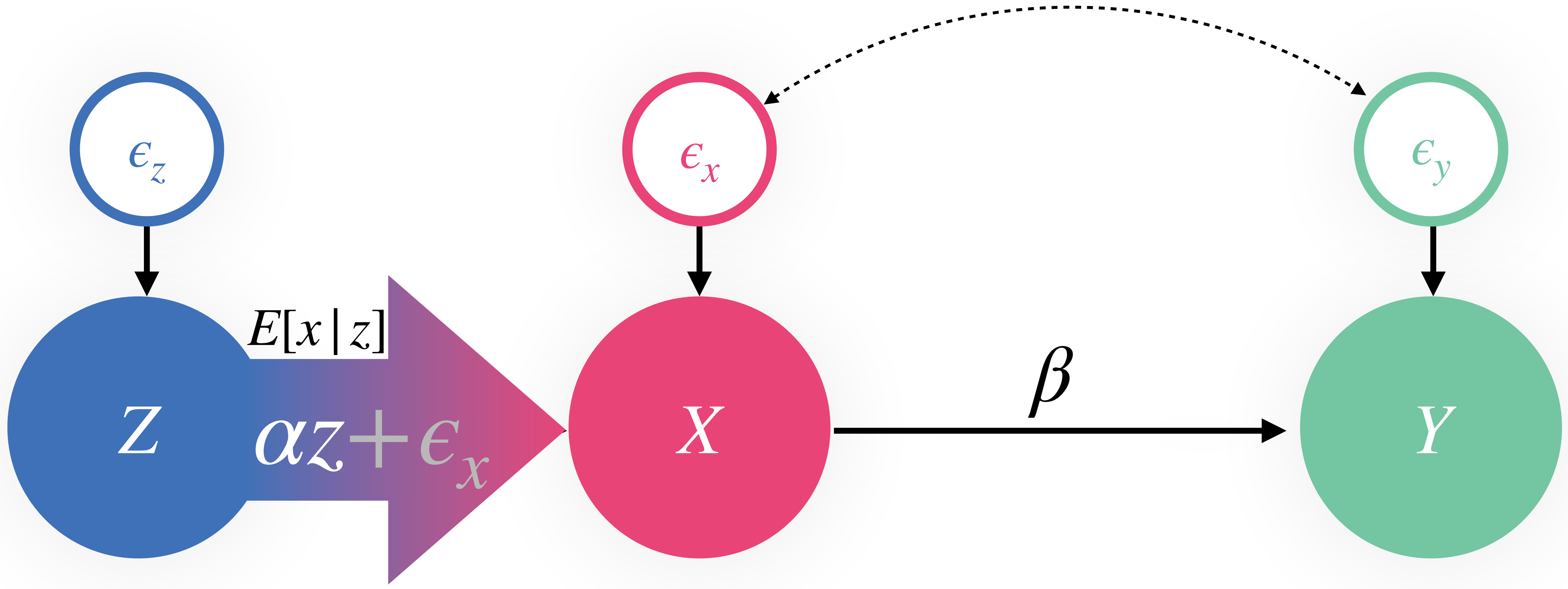
Two Stage Least Squares

$$X := Z\alpha + \epsilon_x, \quad Y := X\beta + \epsilon_y, \quad X \not\perp \epsilon_y$$



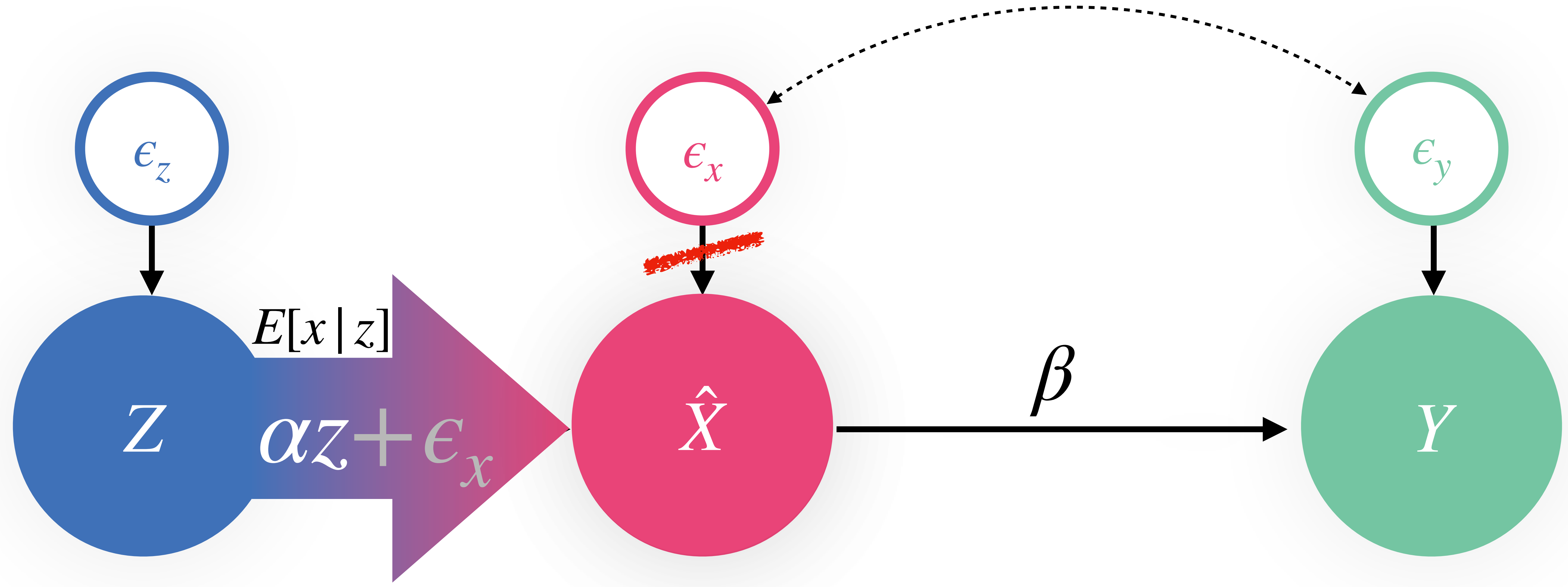
Two Stage Least Squares

$$X := Z\alpha + \epsilon_x, \quad Y := X\beta + \epsilon_y, \quad X \not\perp \epsilon_y$$



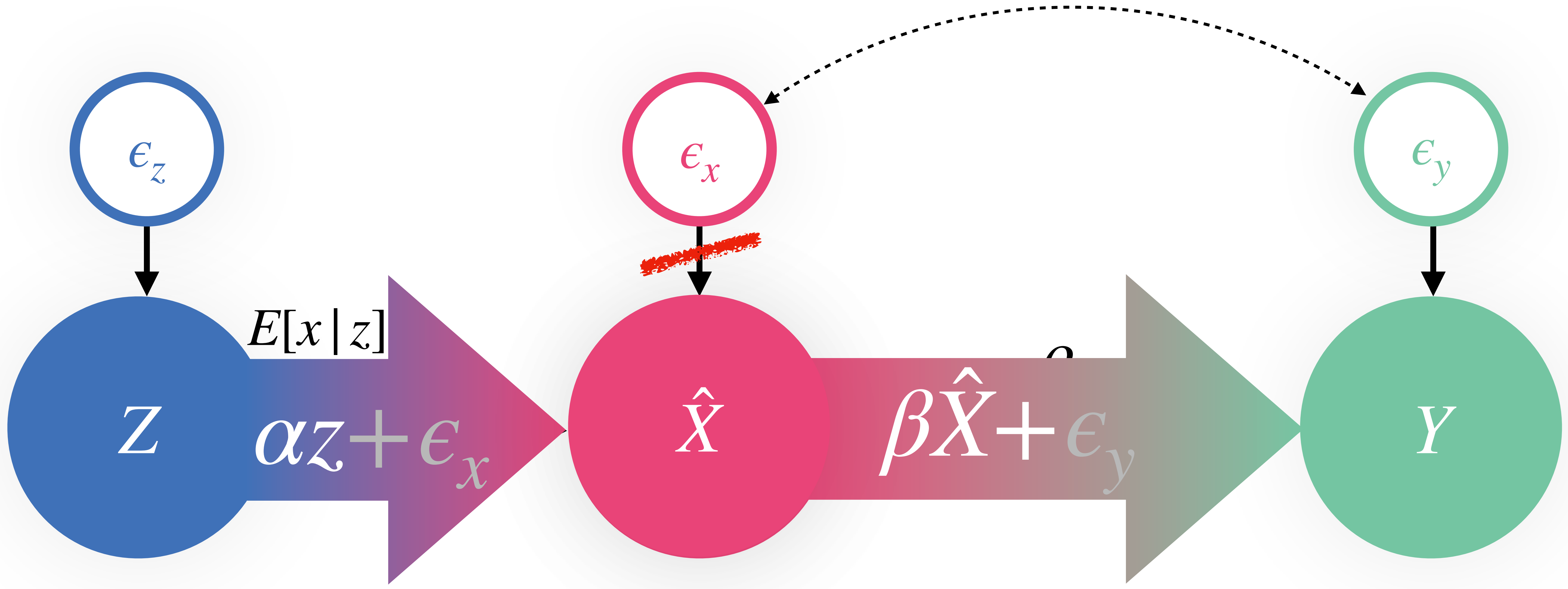
Two Stage Least Squares

$$X := Z\alpha + \epsilon_x, \quad Y := X\beta + \epsilon_y, \quad X \not\perp \epsilon_y$$



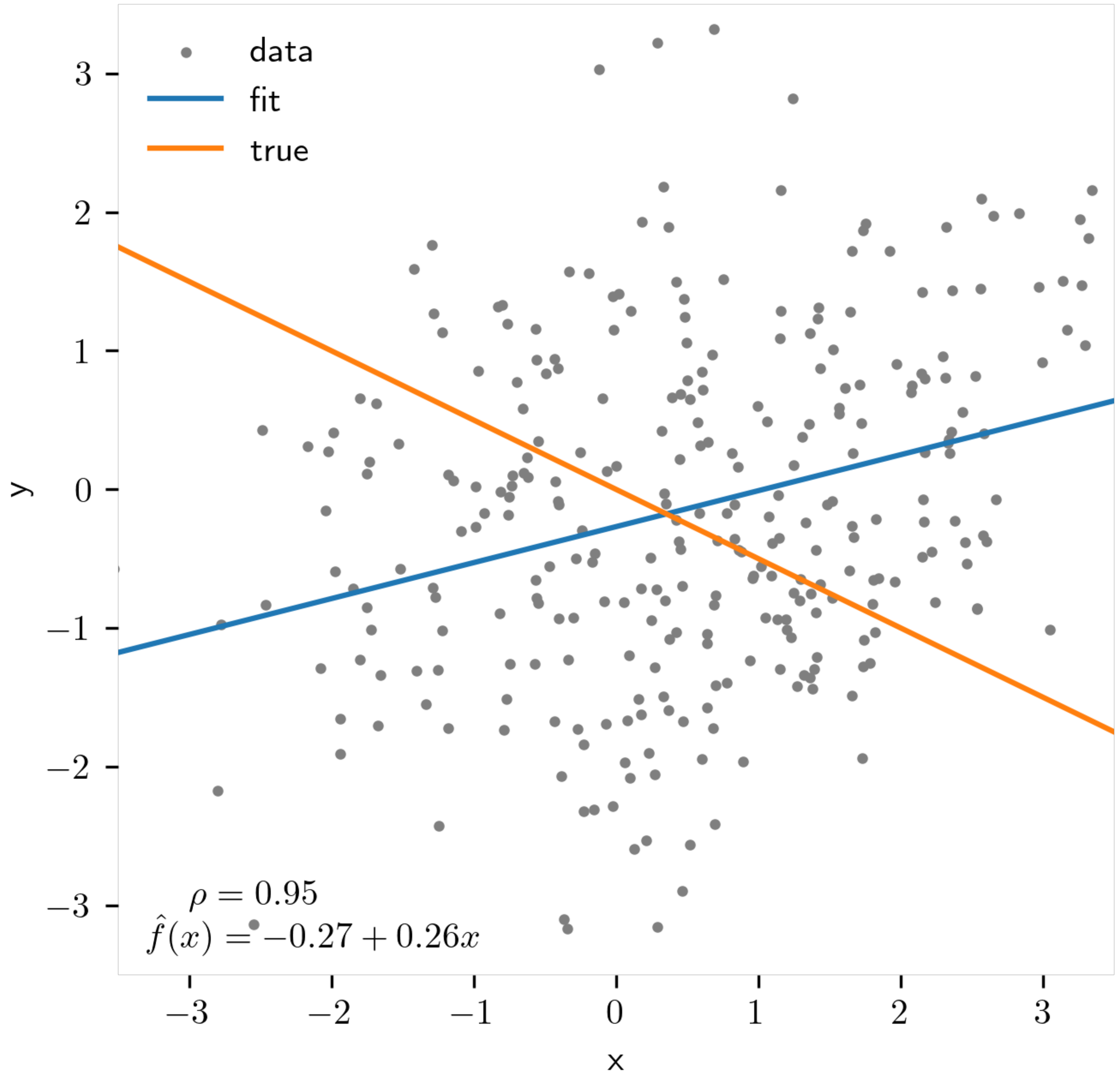
Two Stage Least Squares

$$X := Z\alpha + \epsilon_x, \quad Y := X\beta + \epsilon_y, \quad X \not\perp \epsilon_y$$



$$\hat{\beta}_{2SLS} = (\hat{X}^\top \hat{X})^{-1} \hat{X}^\top y \qquad \hat{X} = E[X|Z]$$

Two Stage Least Squares

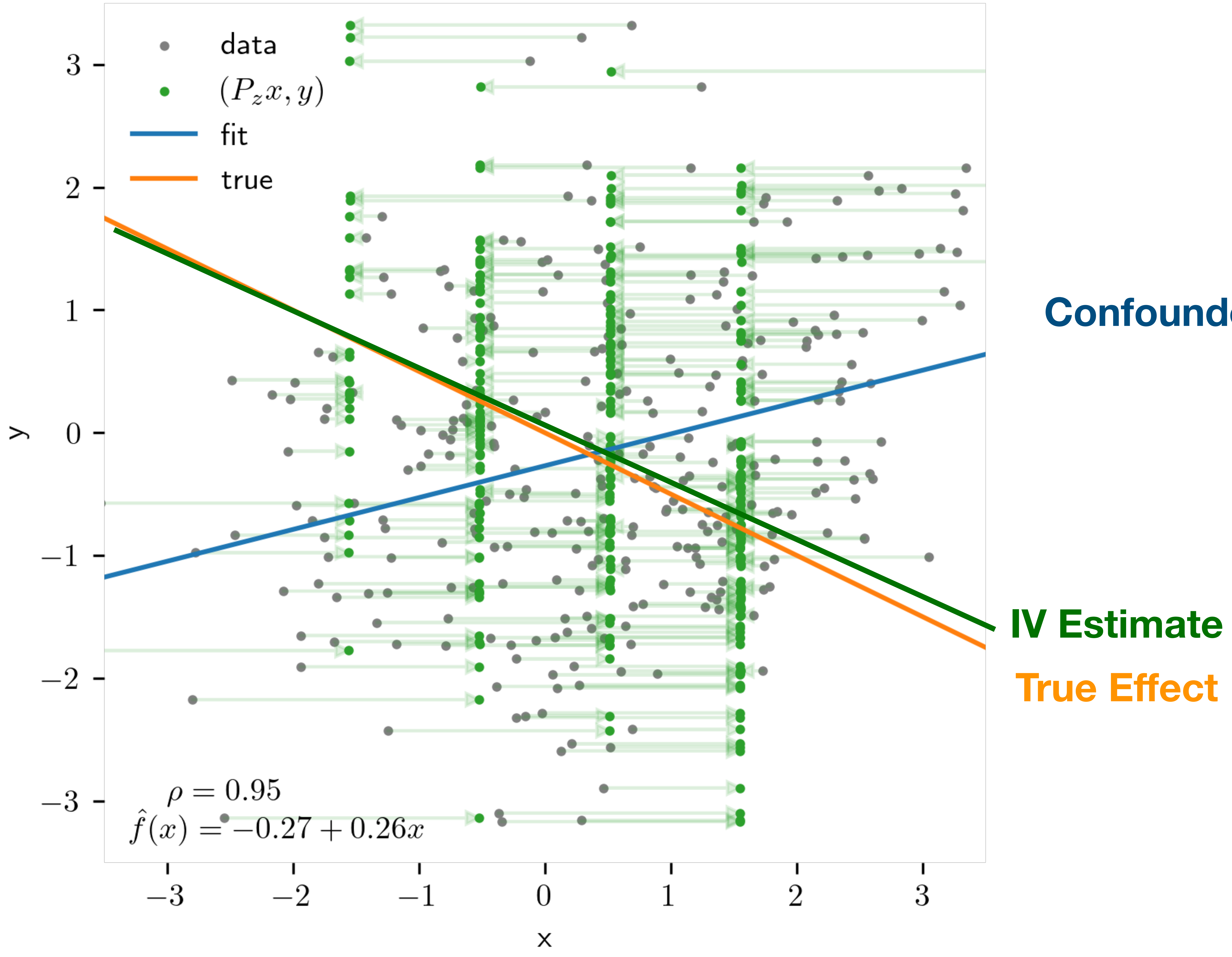


Confounded Effect

True Effect

$$\hat{\beta}_{2SLS} = (\hat{X}^\top \hat{X})^{-1} \hat{X}^\top y$$

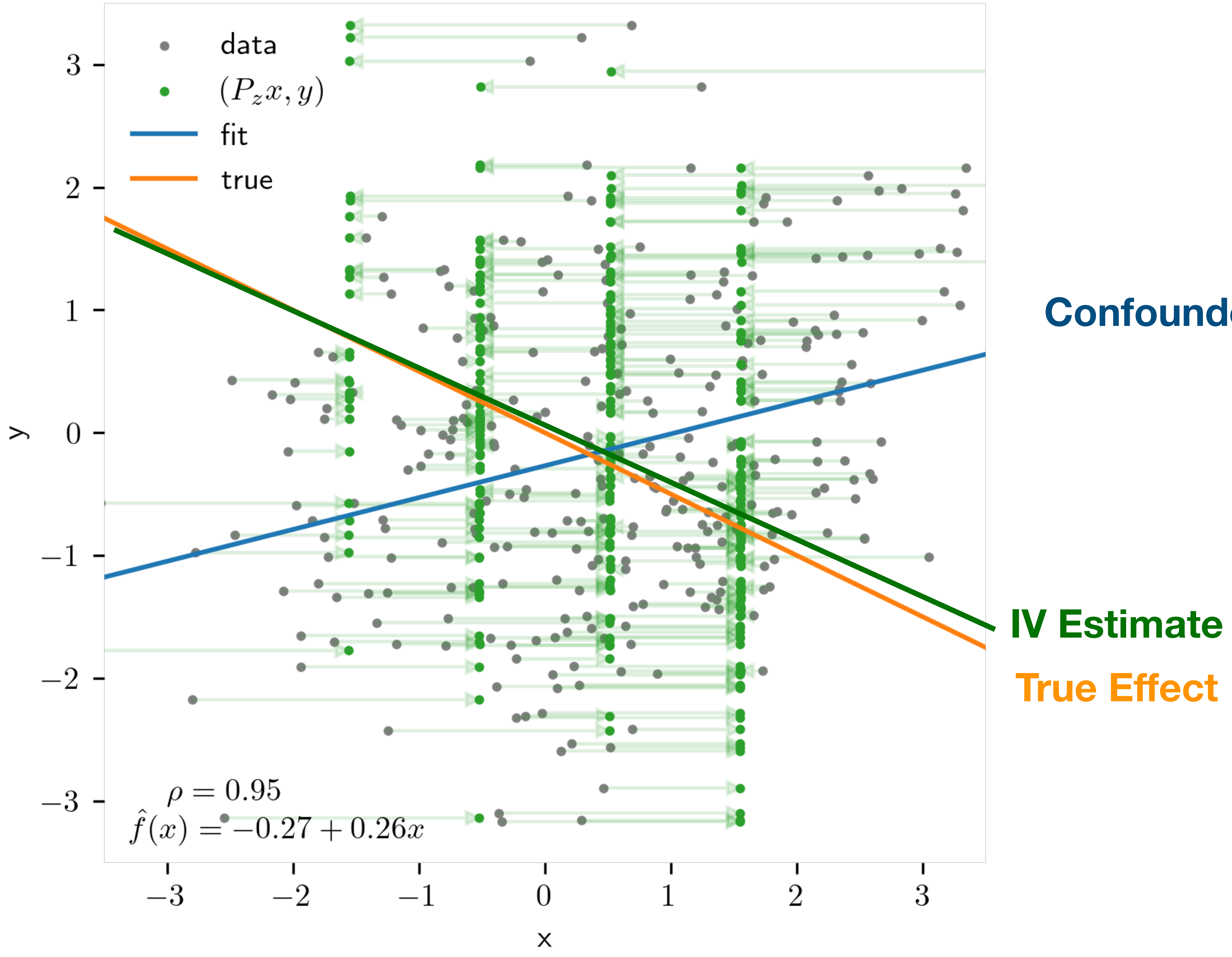
Two Stage Least Squares



$$\hat{\beta}_{2SLS} = (\hat{X}^\top \hat{X})^{-1} \hat{X}^\top y$$

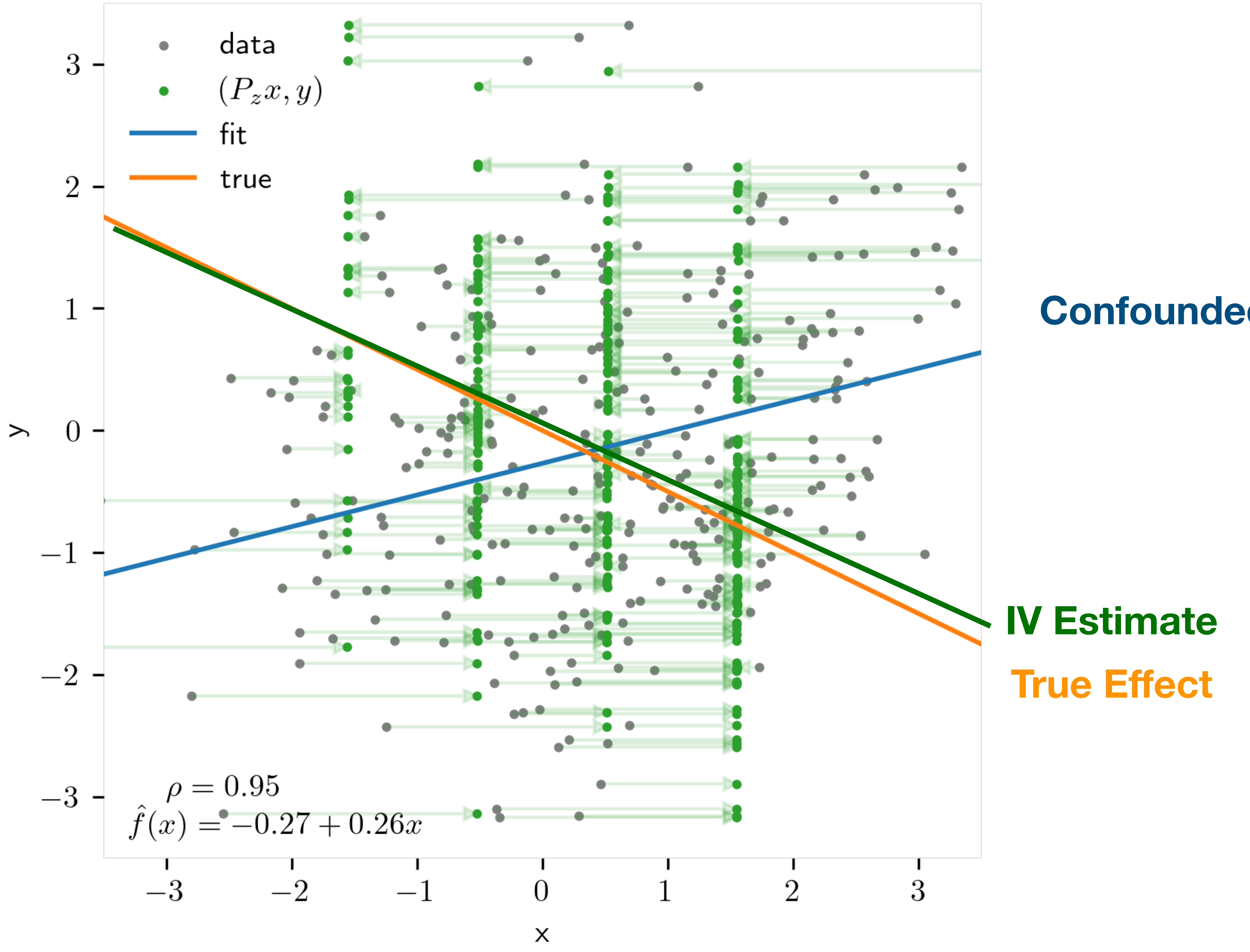
Two Stage Least Squares

… and we are done ???



$$\hat{\beta}_{2SLS} = (\hat{X}^\top \hat{X})^{-1} \hat{X}^\top y$$

Actually we forgot something …

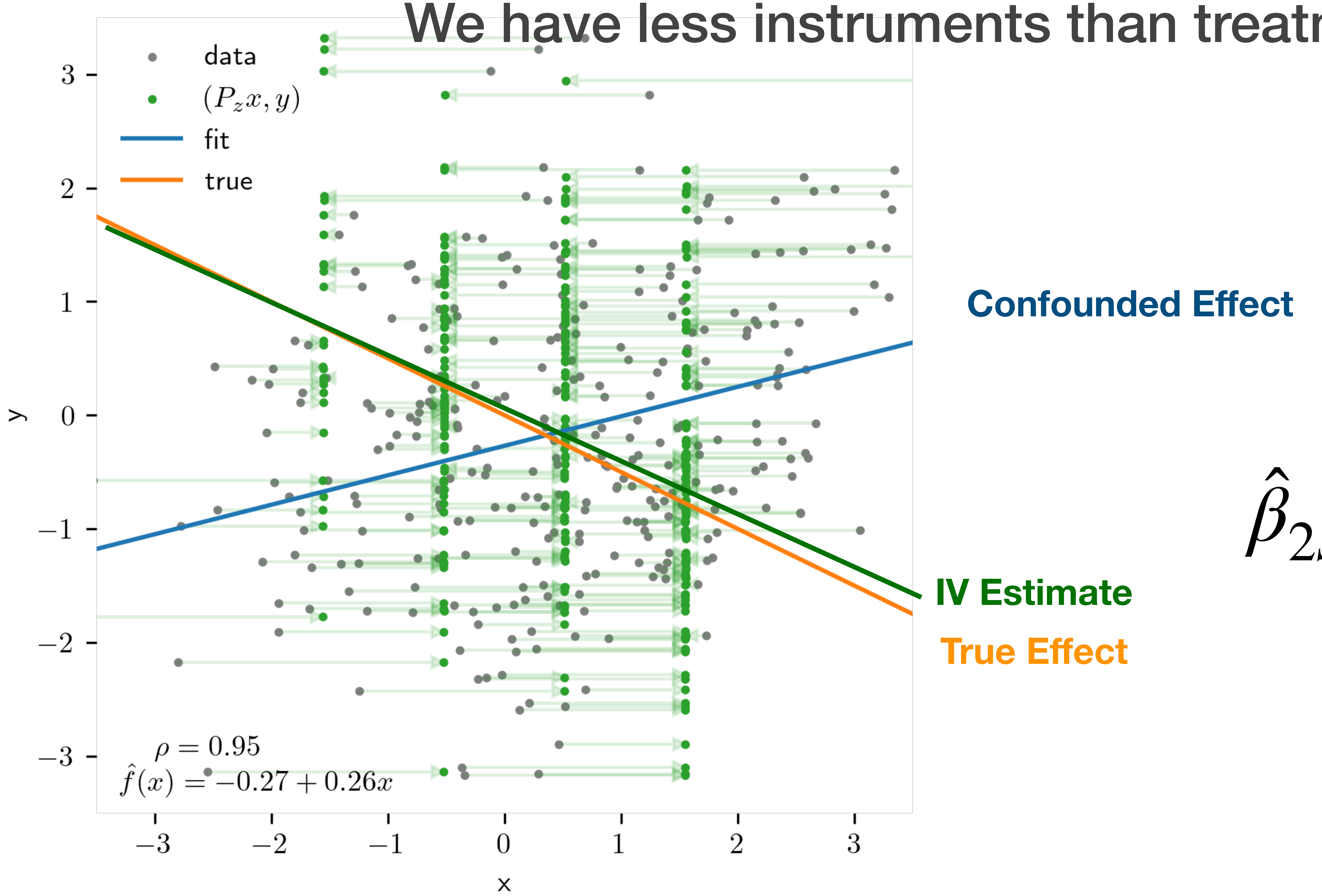


$$\hat{\beta}_{2SLS} = (\hat{X}^\top \hat{X})^{-1} \hat{X}^\top y$$

Actually we forgot something …

Unless we have million antibiotics,

We have less instruments than treatments $d_z < d_x$



This matrix is not invertible if $d_z < d_x$

$$\hat{\beta}_{2SLS} = (\hat{X}^\top \hat{X})^{-1} \hat{X}^\top y$$

A natural machine learning approach :

We use the pseudoinverse…

$$\hat{\beta}_{2SLS} = (\hat{X}^\top \hat{X})^{-1} \hat{X}^\top y$$

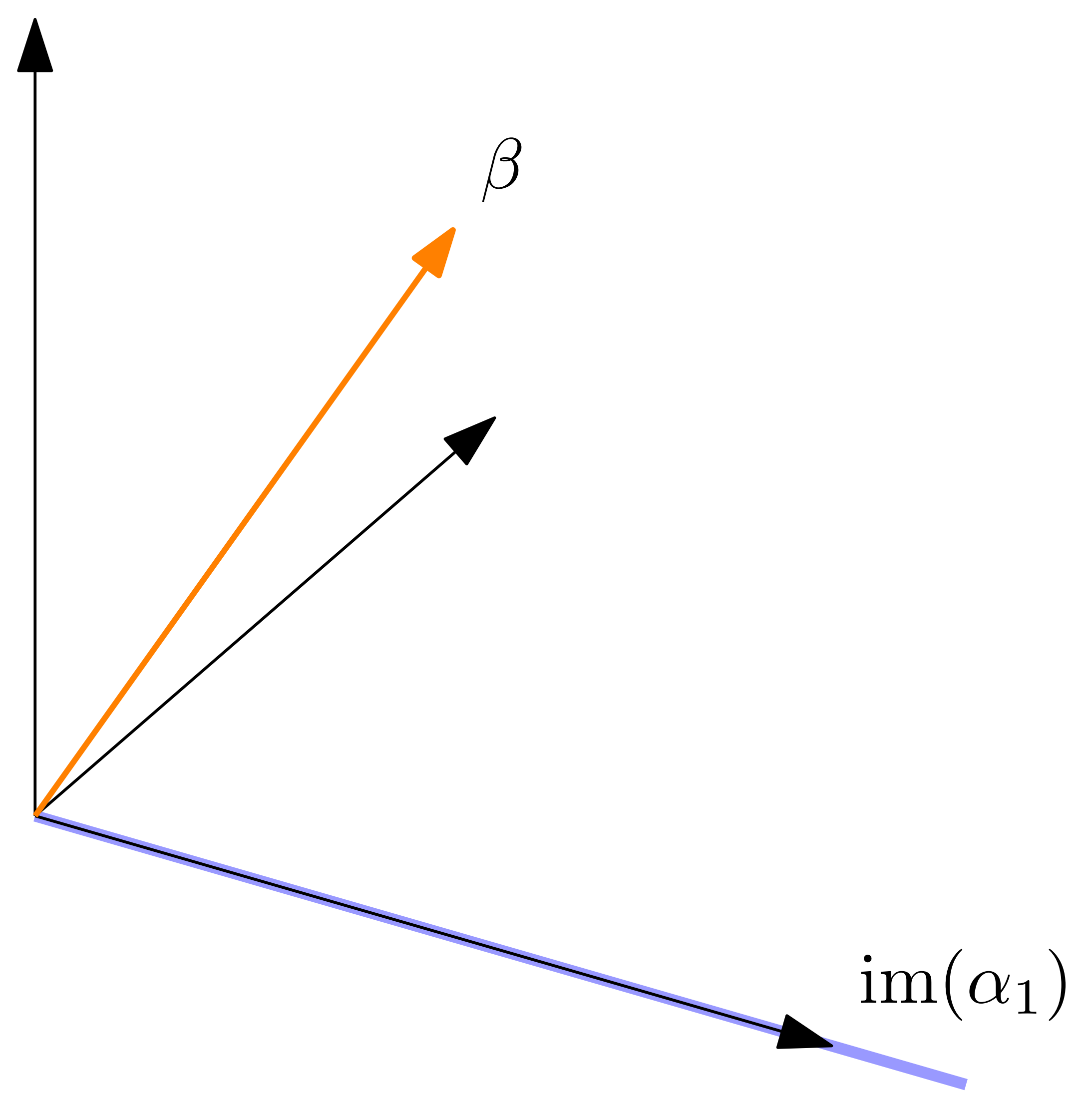
A natural machine learning approach :

We use the pseudoinverse…

$$\widehat{P_\alpha \beta}_{\cancel{2SLS}} = (\hat{X}^\top \hat{X})^{+} \hat{X}^\top y$$

A natural machine learning approach :

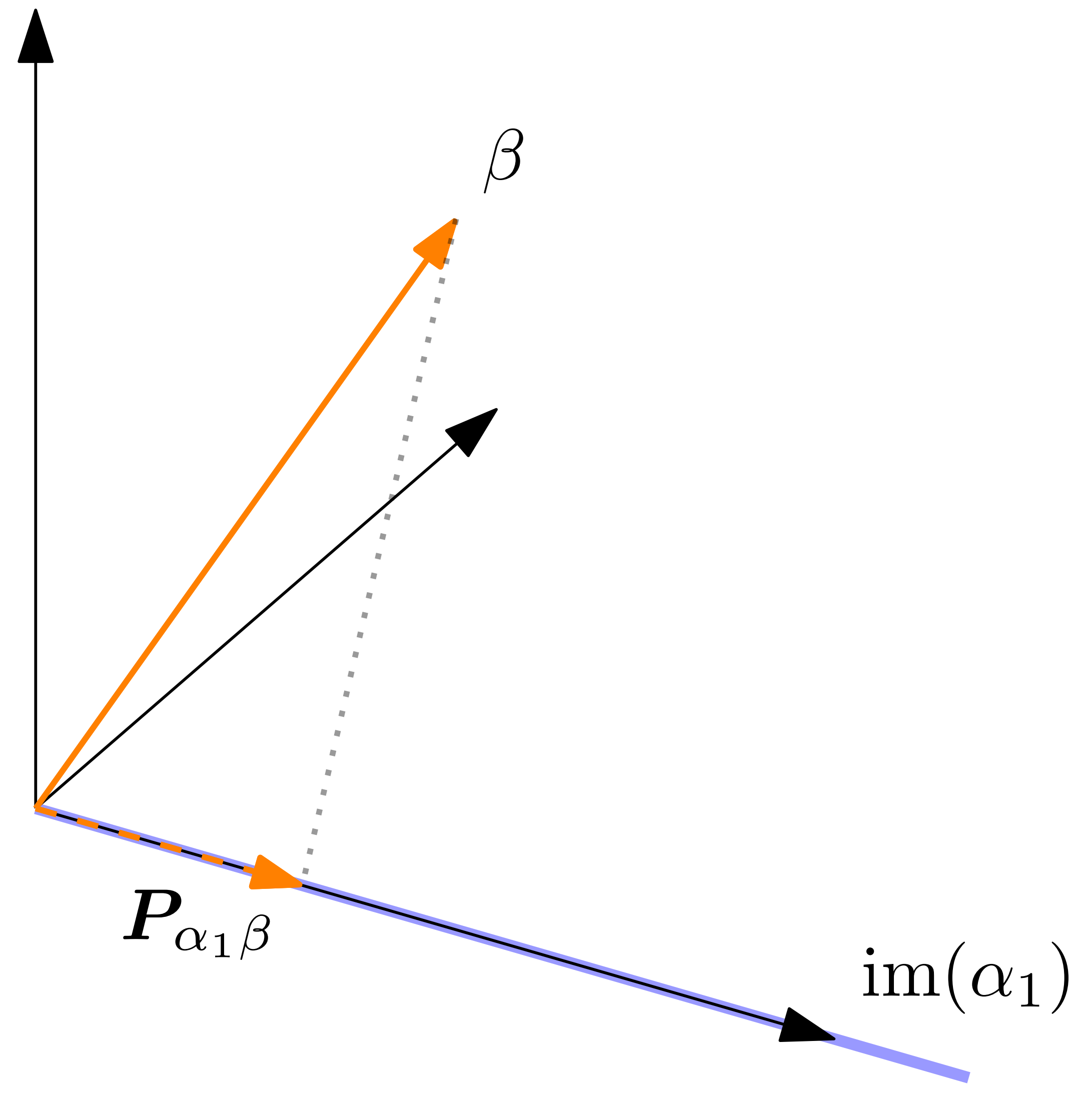
We use the pseudoinverse…



$$\widehat{P_\alpha \beta}_{\cancel{2SLS}} = \boxed{(\hat{X}^\top \hat{X})^{+}} \hat{X}^\top y$$

A natural machine learning approach :

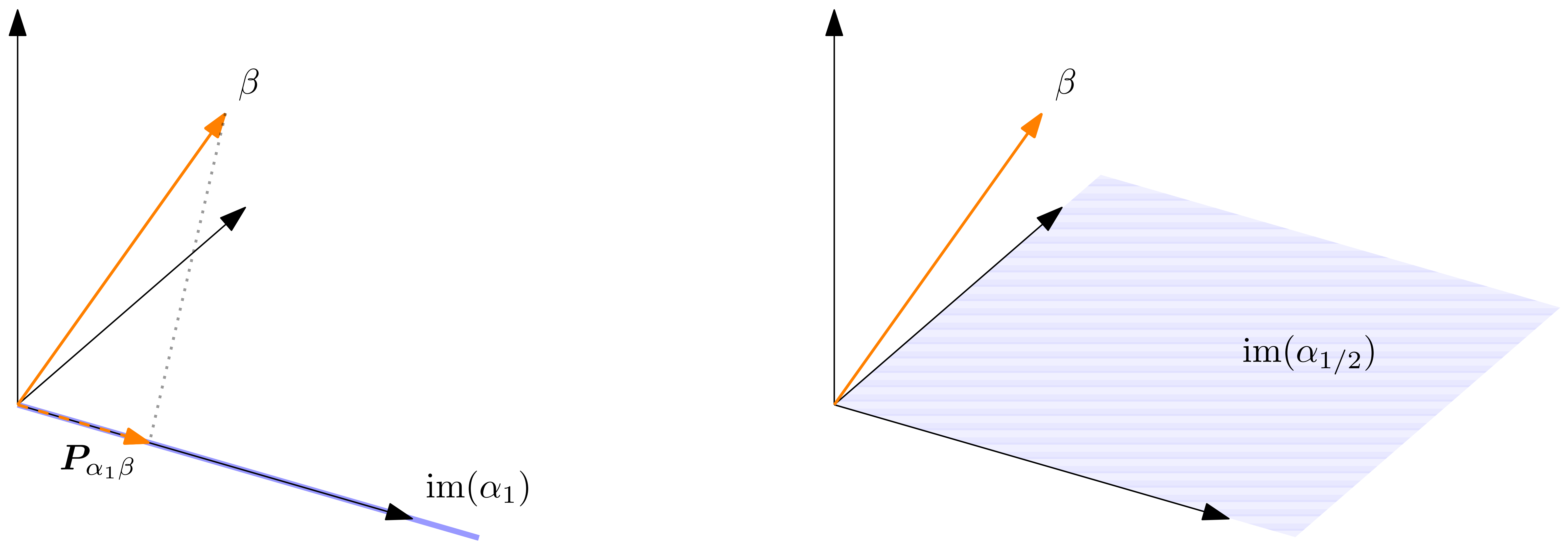
We use the pseudoinverse…



$$\widehat{P_\alpha \beta}_{\sout{2SLS}} = (\hat{X}^\top \hat{X})^{+} \hat{X}^\top y$$

A natural machine learning approach :

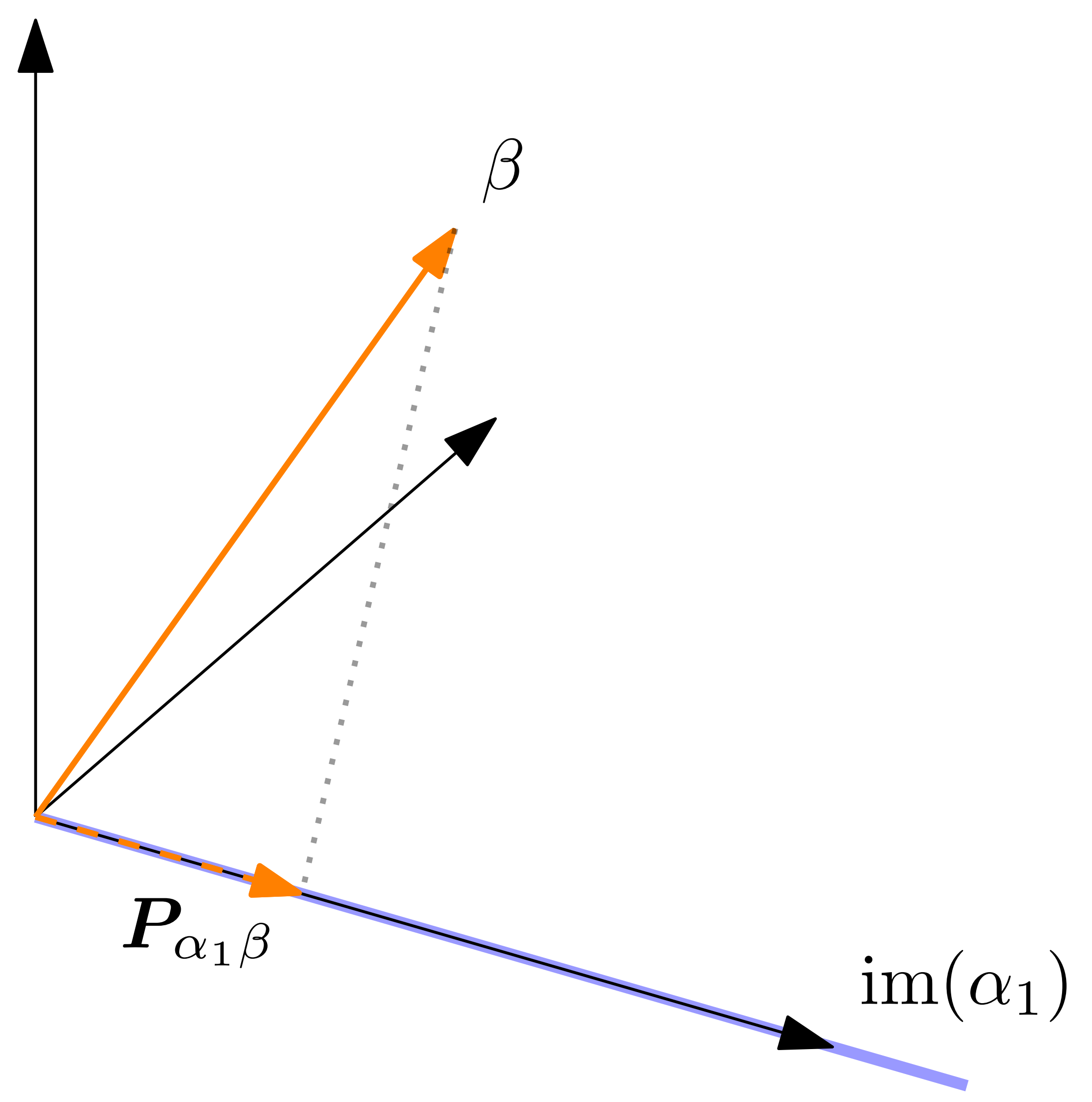
We use the pseudoinverse…

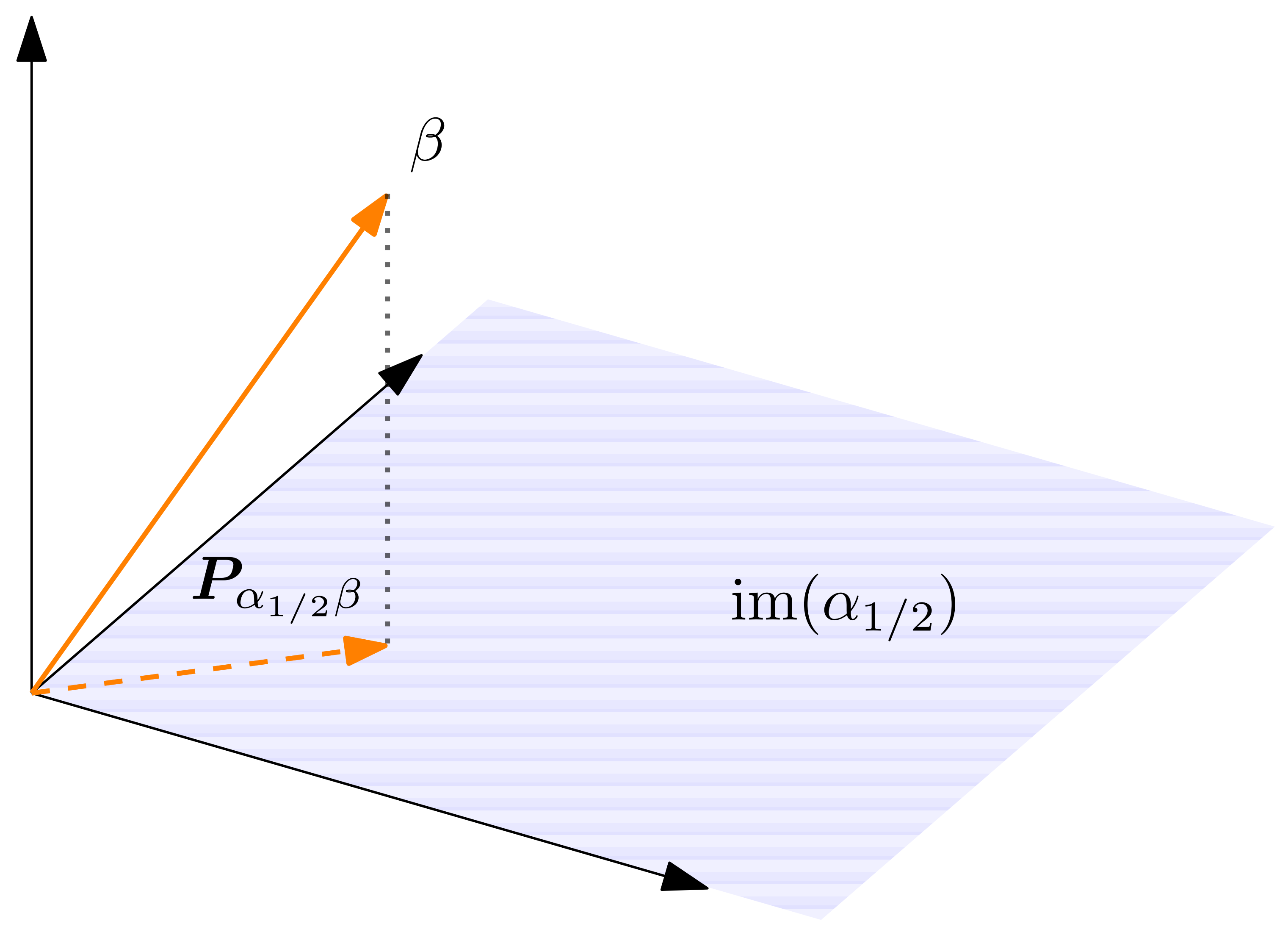


$$\widehat{P_\alpha\beta}_{\cancel{2SLS}} = (\hat{X}^\top \hat{X})^{+} \hat{X}^\top y$$

A natural machine learning approach :

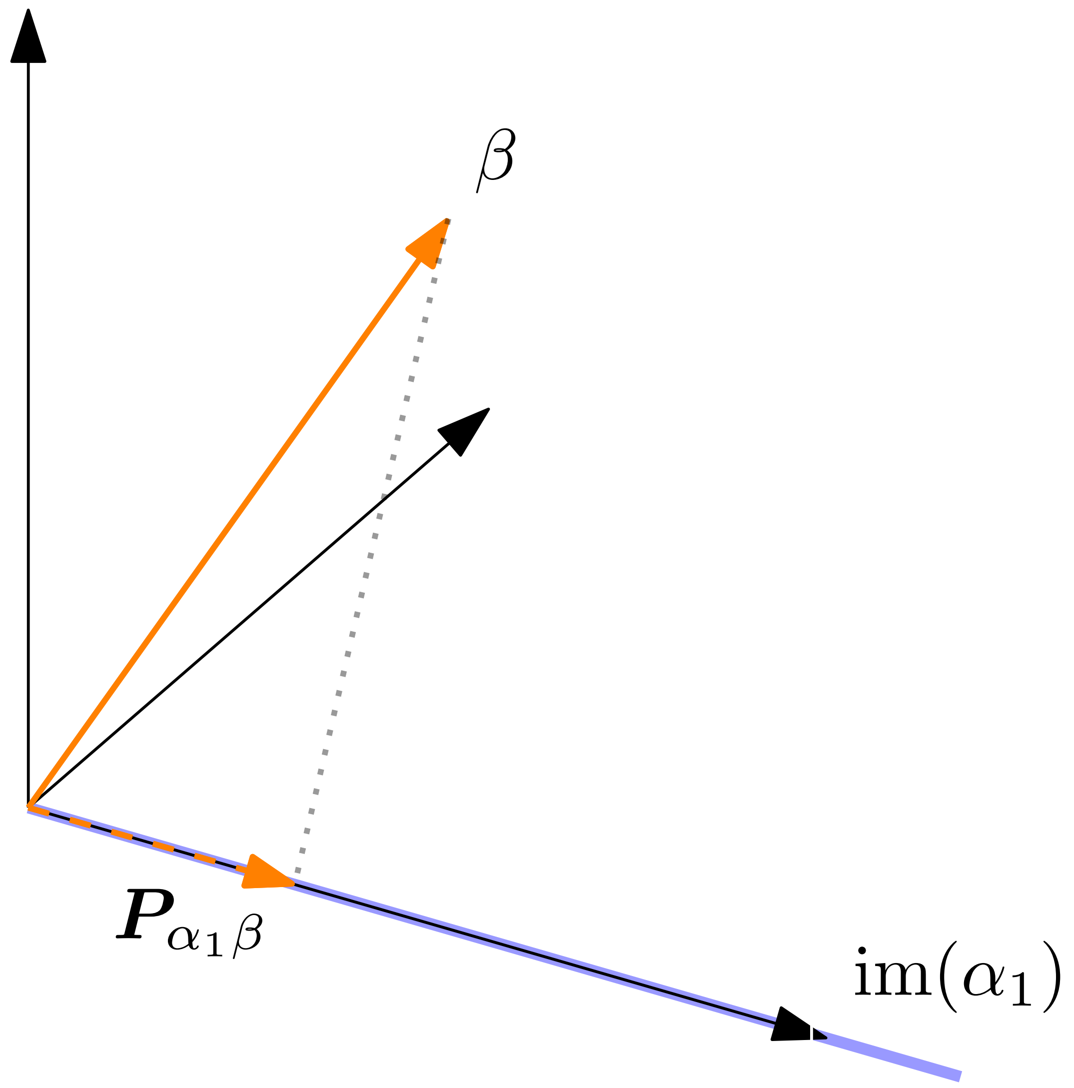
We use the pseudoinverse…

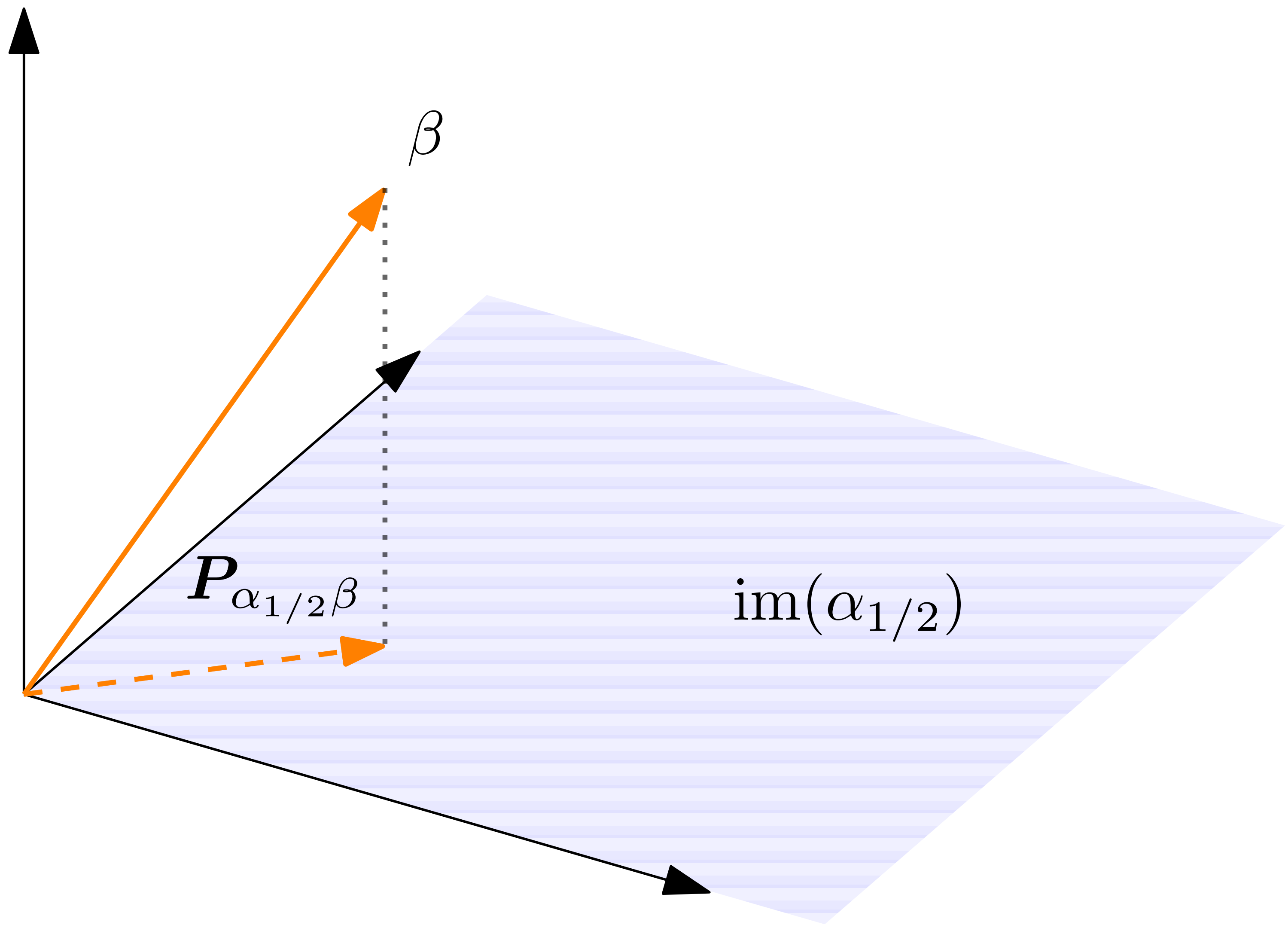


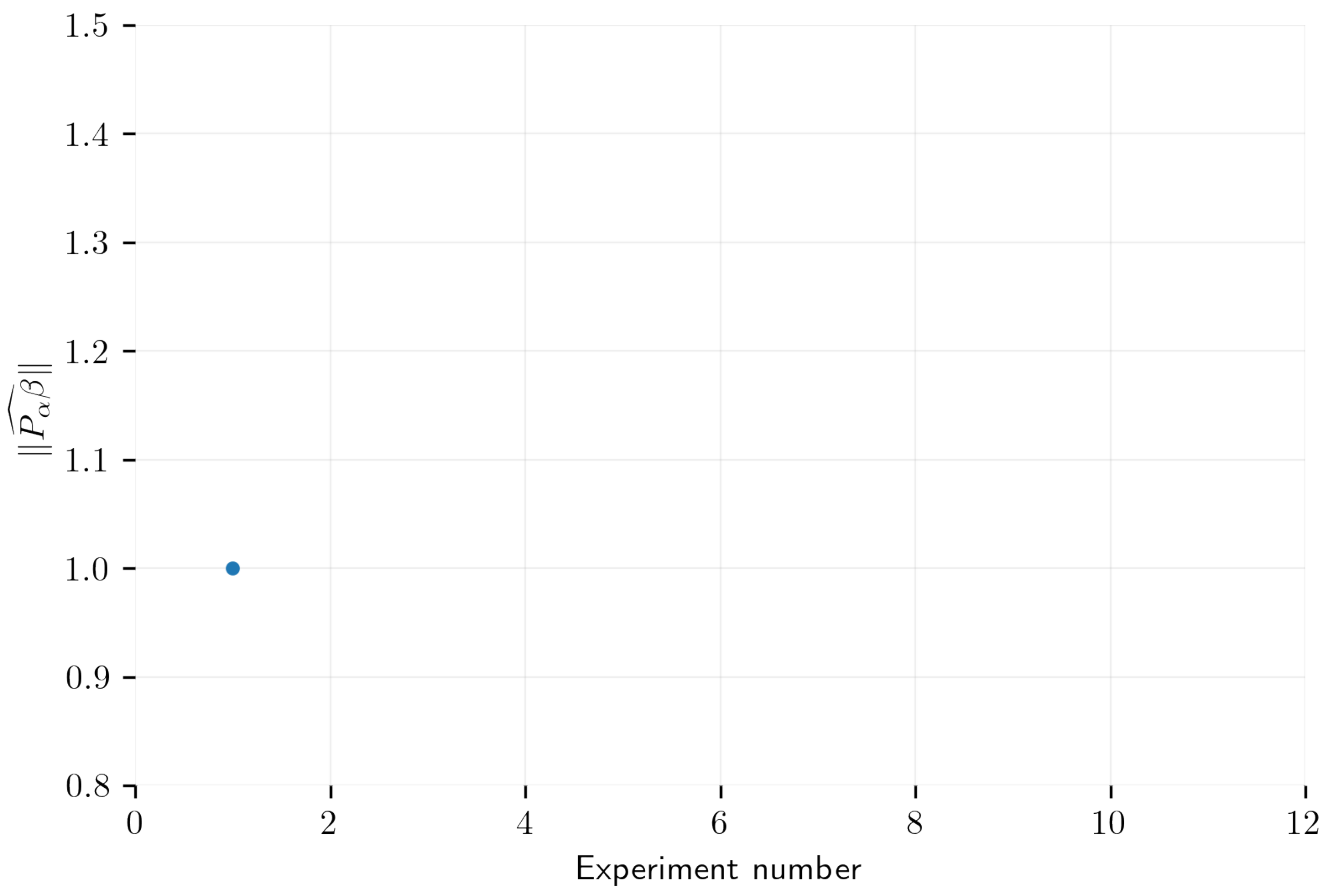


$$\widehat{P_\alpha \beta}_{\cancel{2SLS}} = (\hat{X}^\top \hat{X})^{+} \hat{X}^\top y$$

We can add experiments until $\|P_{\alpha_{1,2,\ldots}}\beta\| = \|\beta\|$

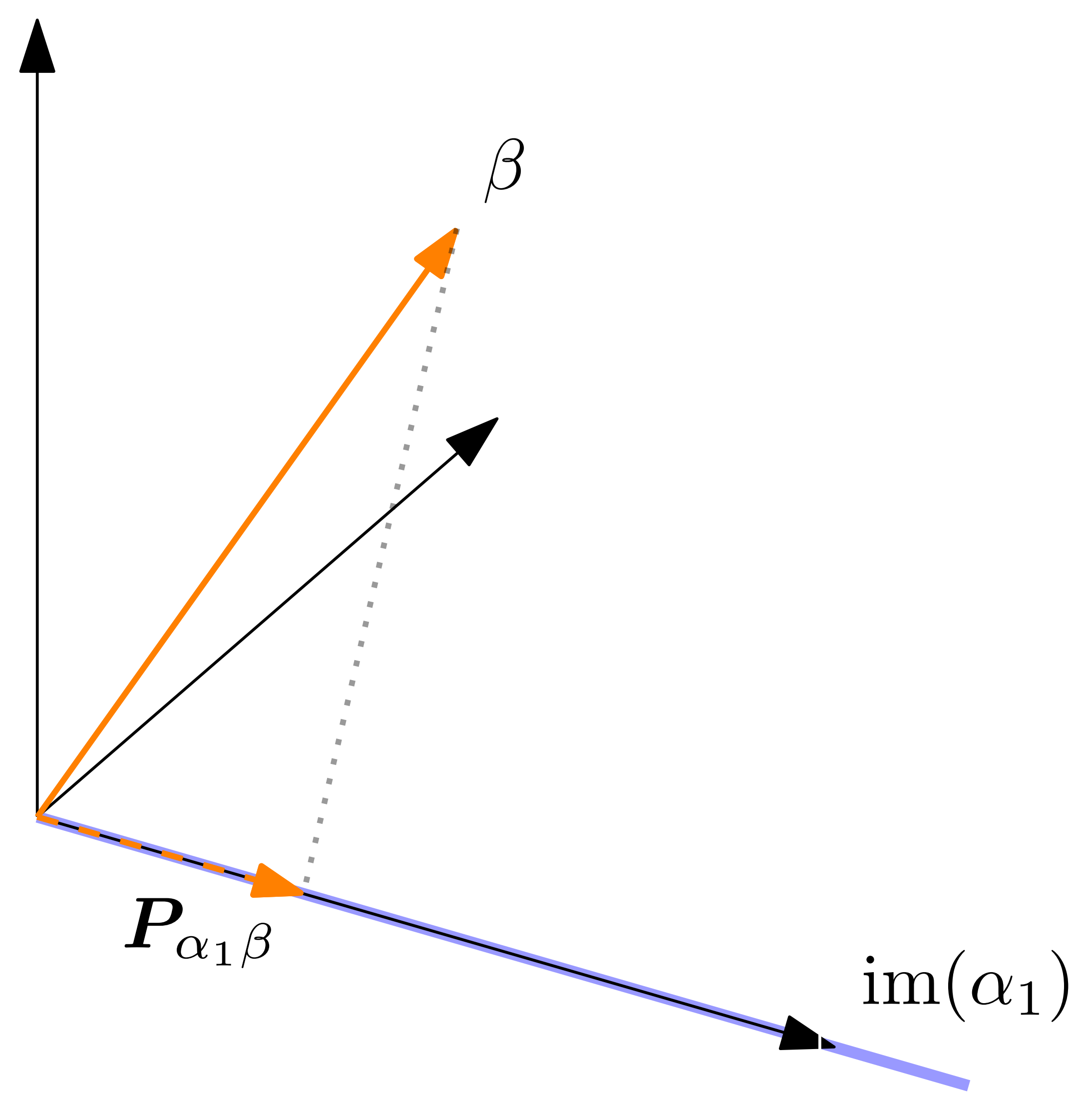


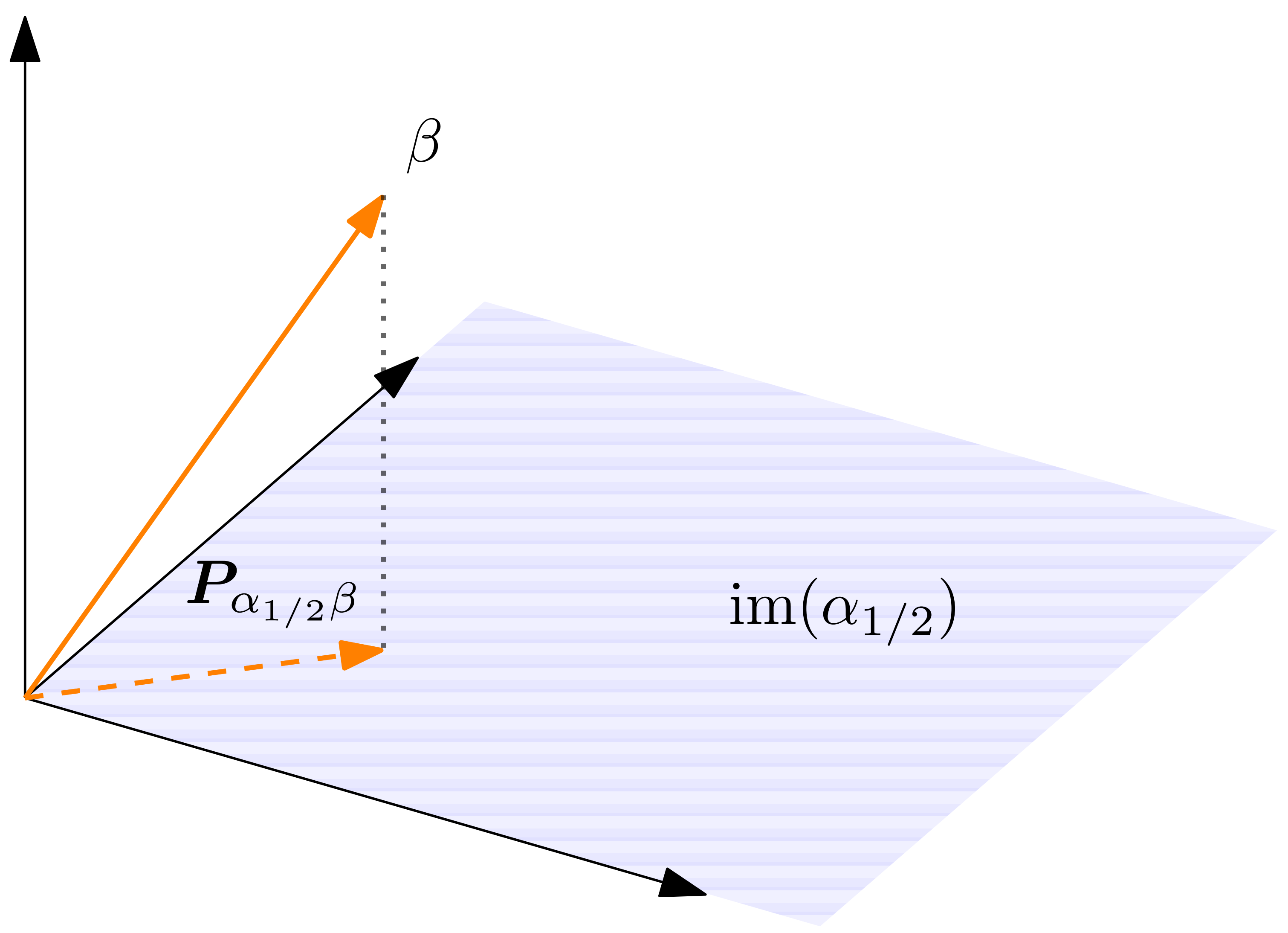


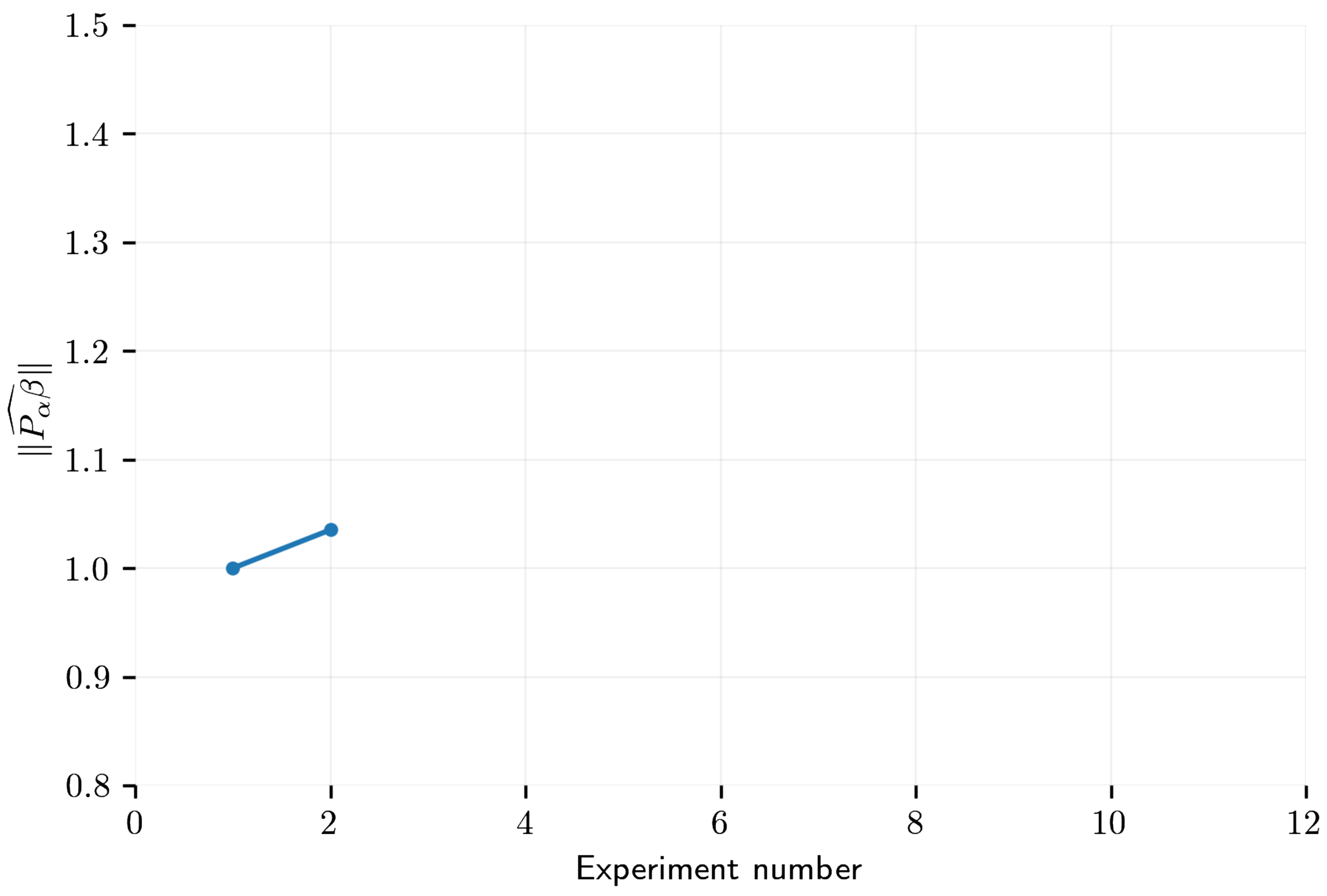


$$\widehat{P_\alpha\beta}_{\text{2SLS}} = (\hat{X}^\top\hat{X})^{+}\hat{X}^\top y$$

We can add experiments until $\|P_{\alpha_{1,2,\ldots}}\beta\| = \|\beta\|$

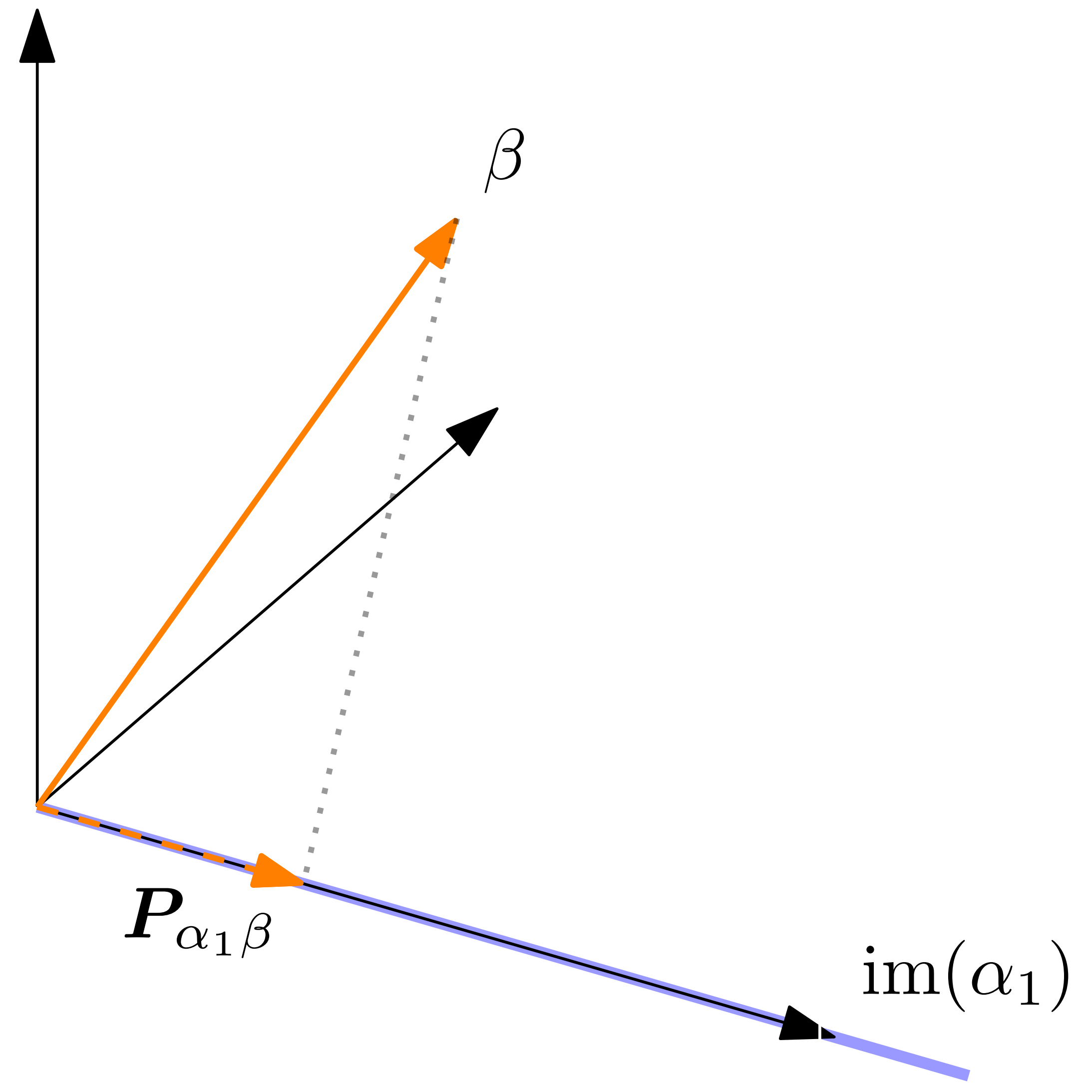


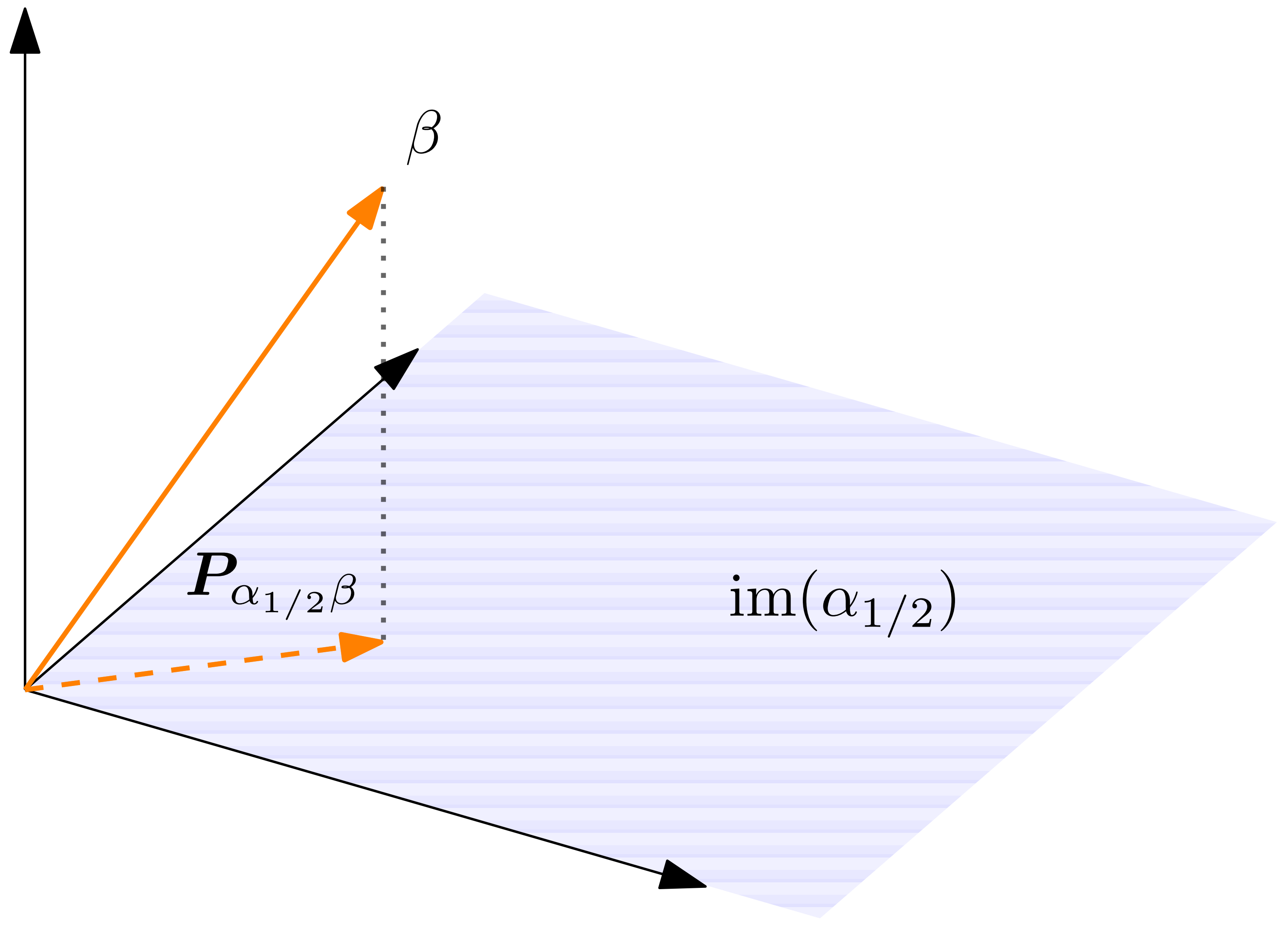


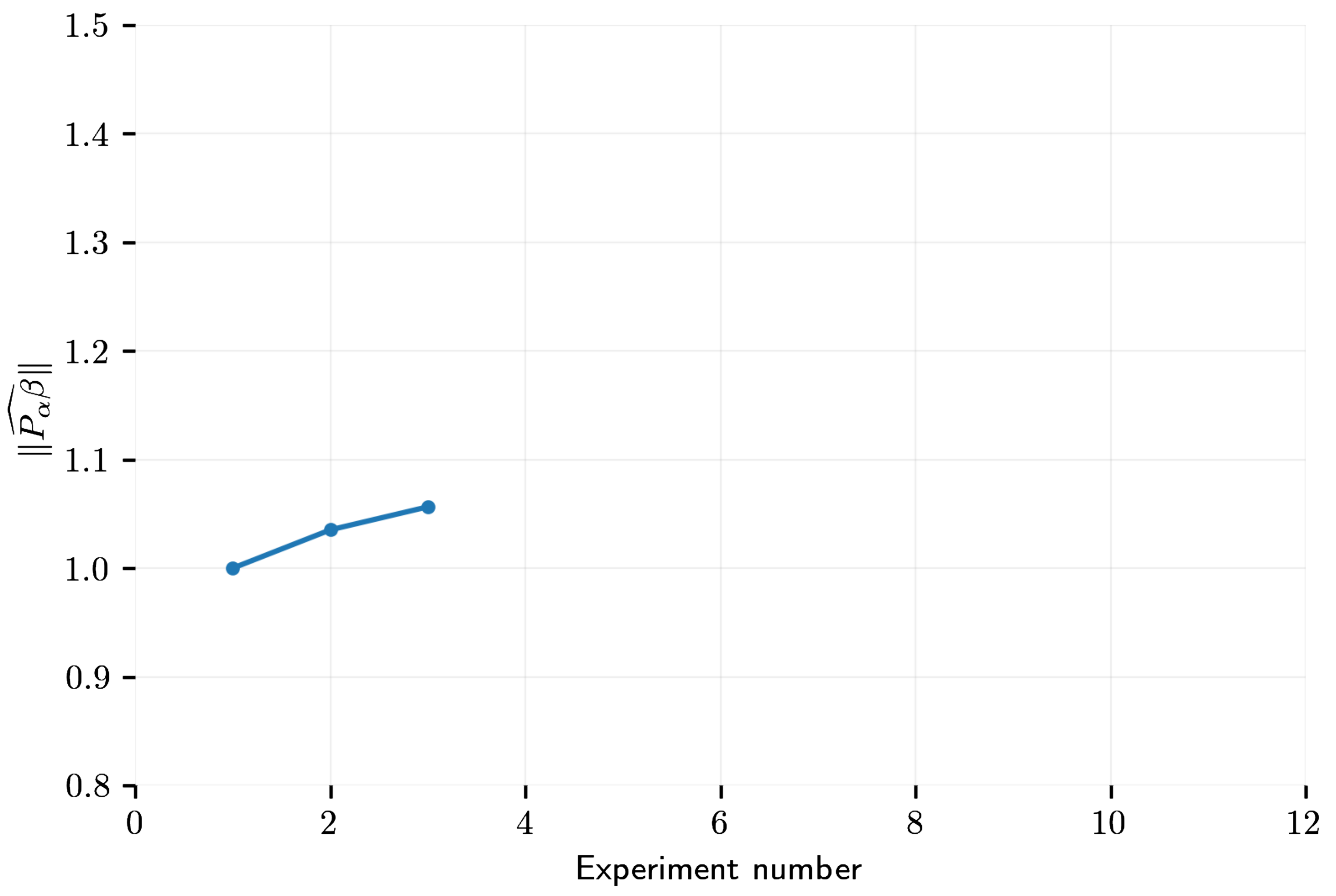


$$\widehat{P_\alpha\beta}_{\text{2SLS}} = (\hat{X}^\top\hat{X})^{+}\hat{X}^\top y$$

We can add experiments until $\|P_{\alpha_{1,2,\ldots}}\beta\| = \|\beta\|$

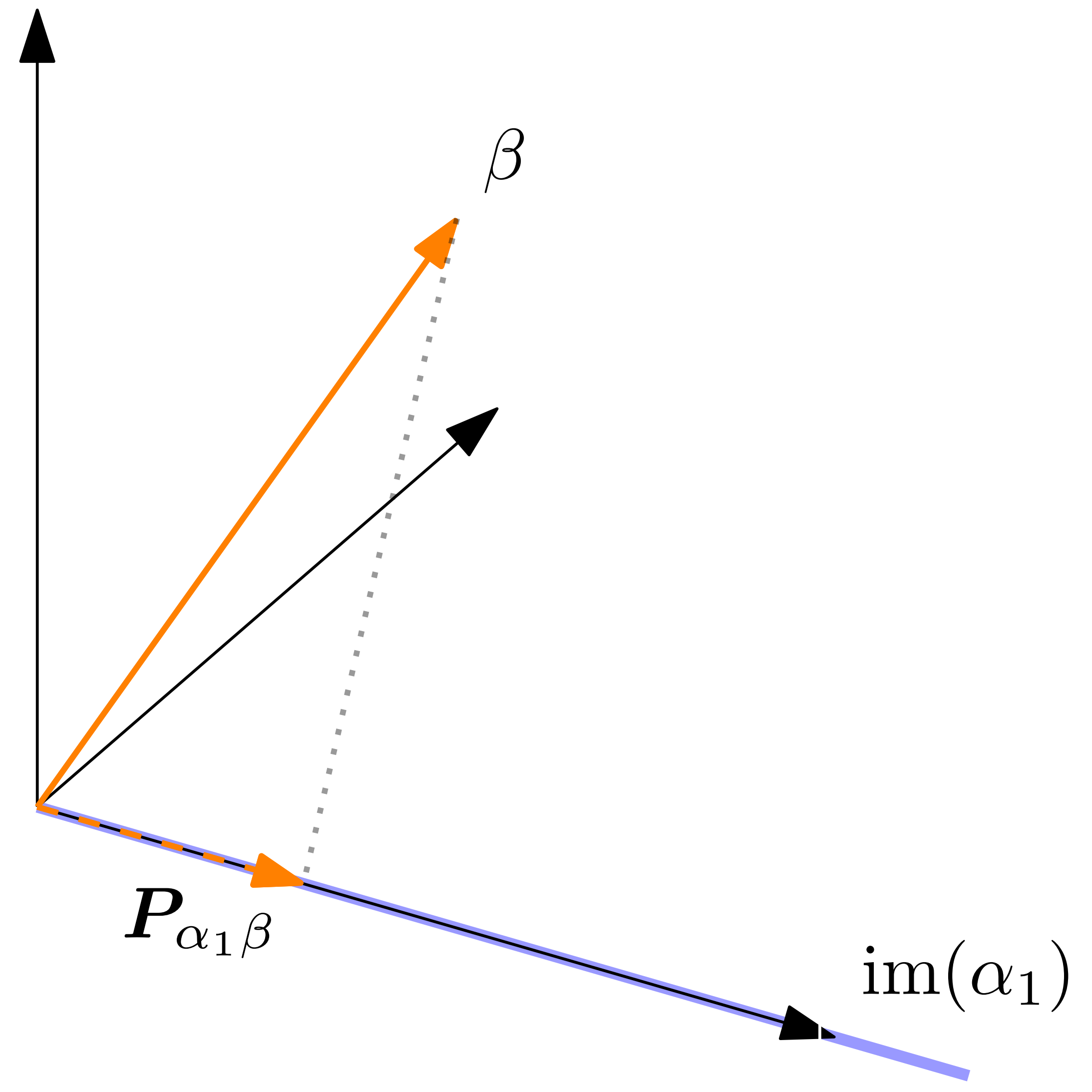


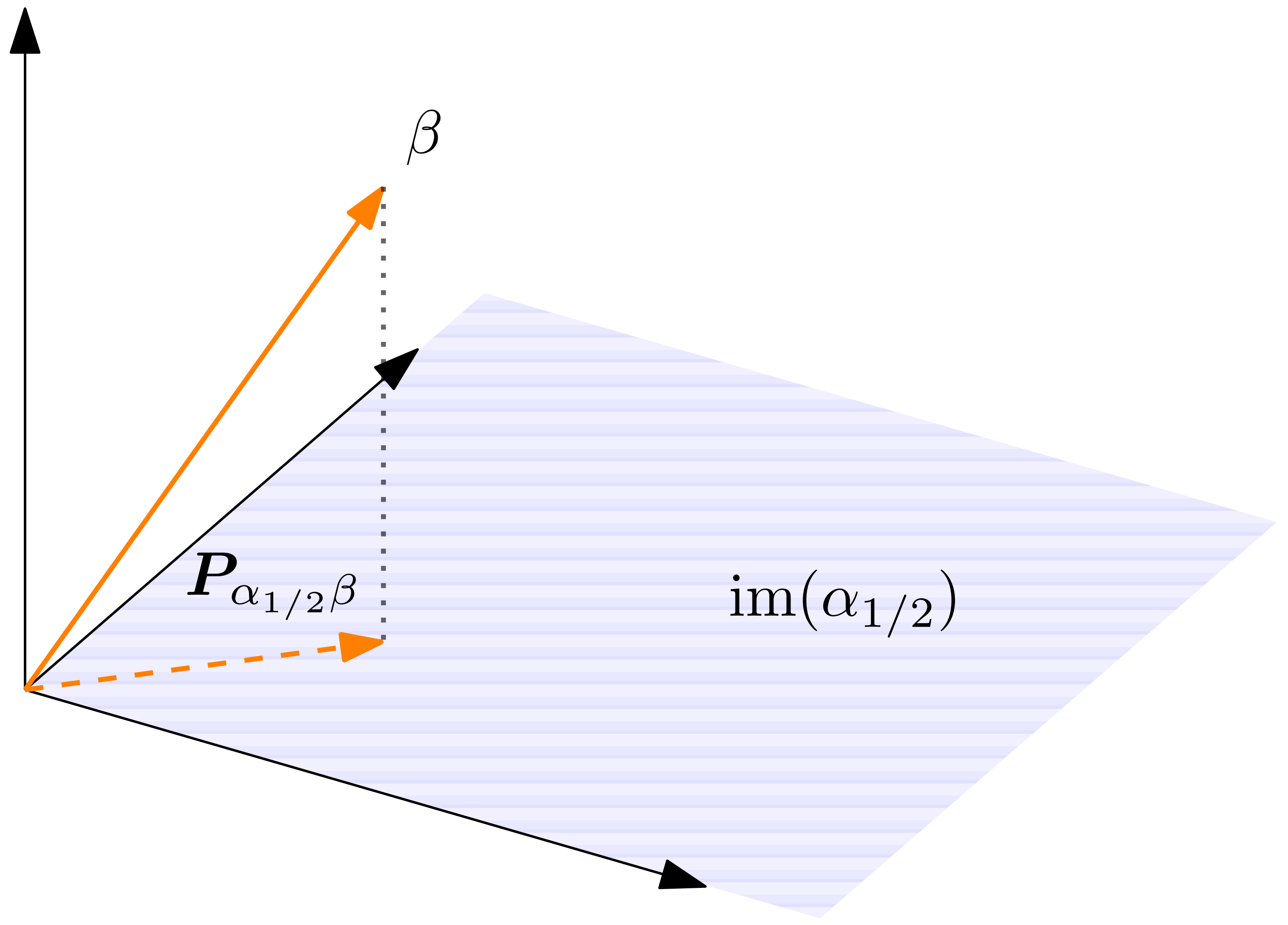


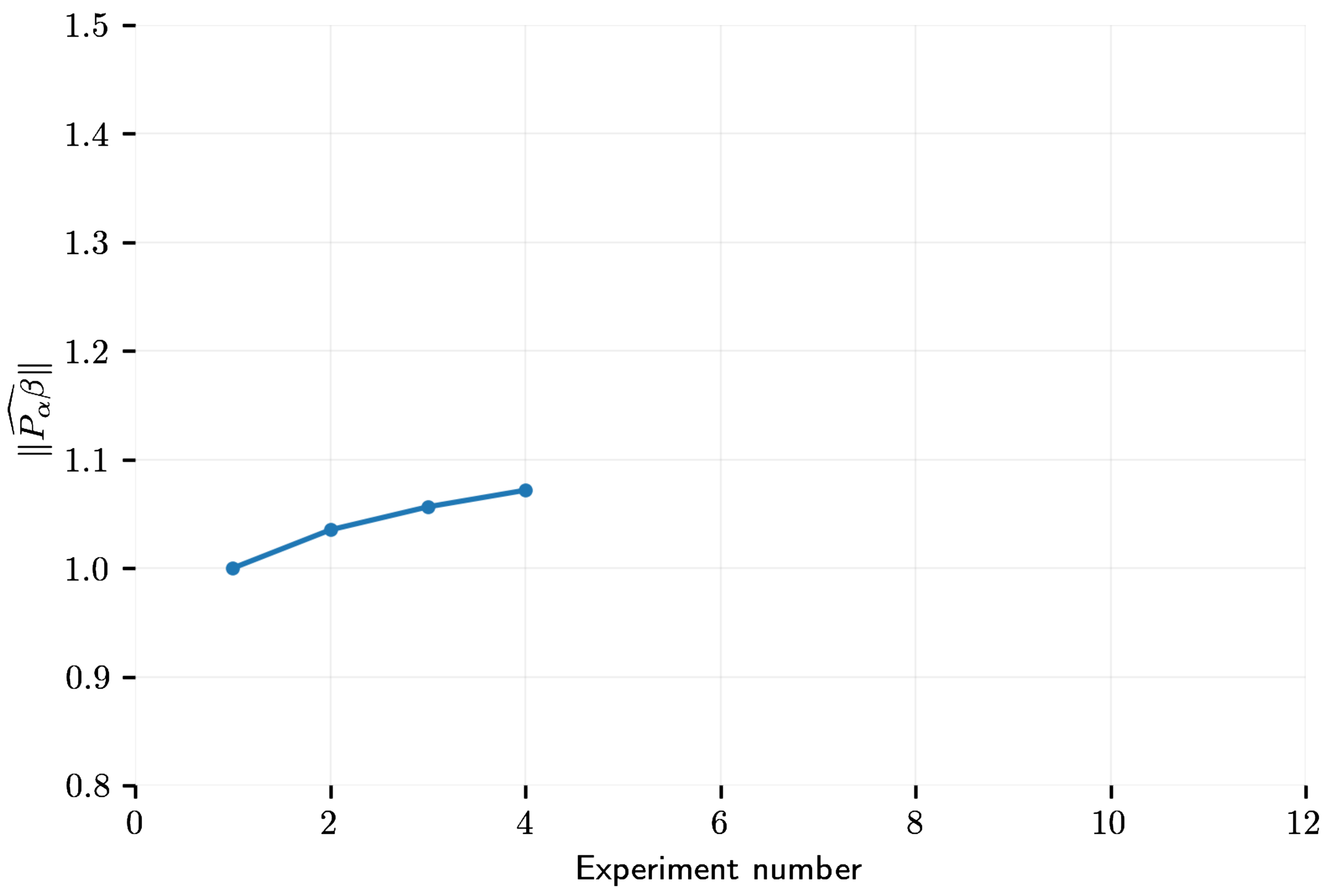


$$\widehat{P_\alpha\beta}_{\text{2SLS}} = (\hat{X}^\top \hat{X})^{+} \hat{X}^\top y$$

We can add experiments until $\|P_{\alpha_{1,2,\ldots}}\beta\| = \|\beta\|$

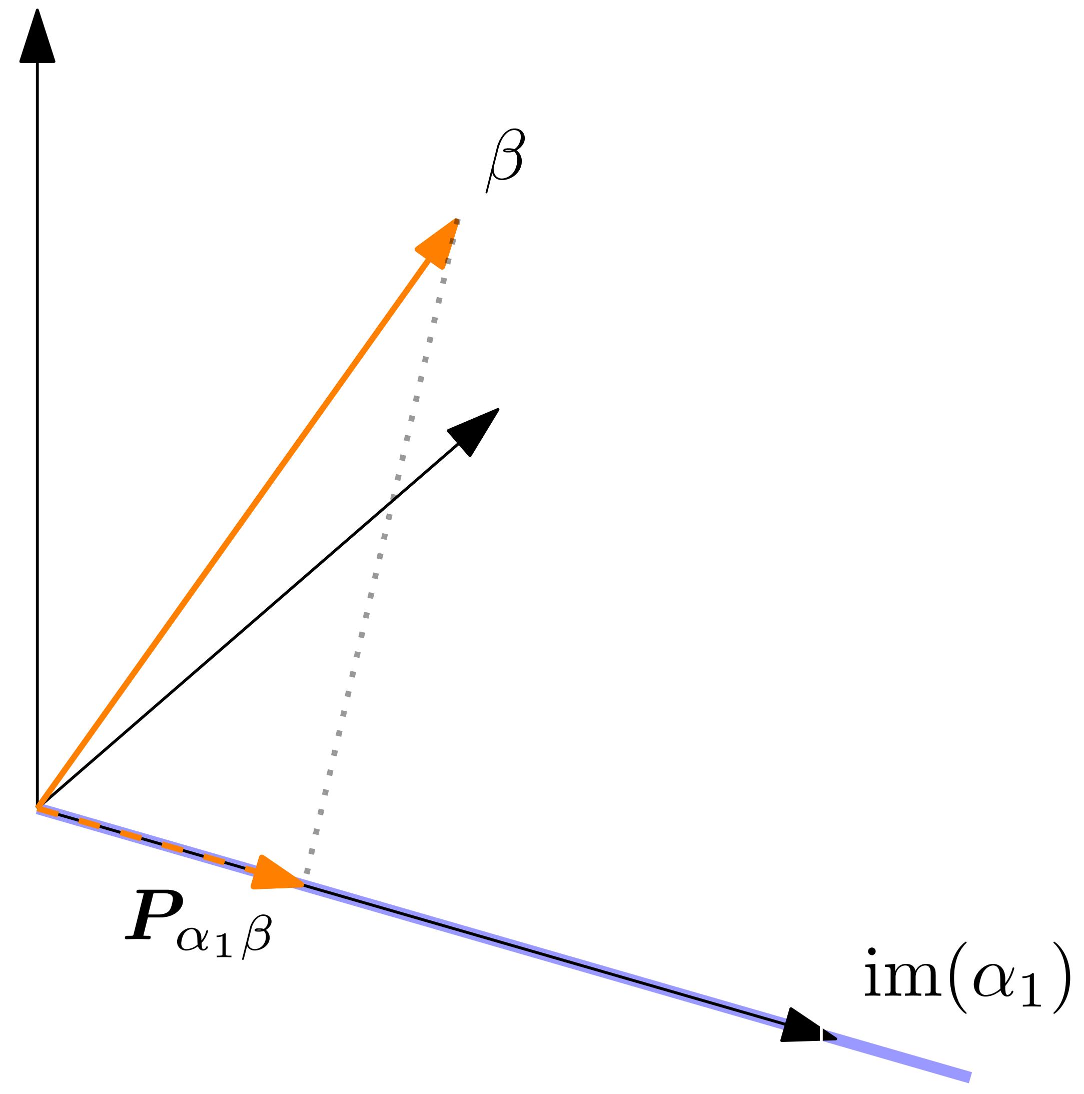


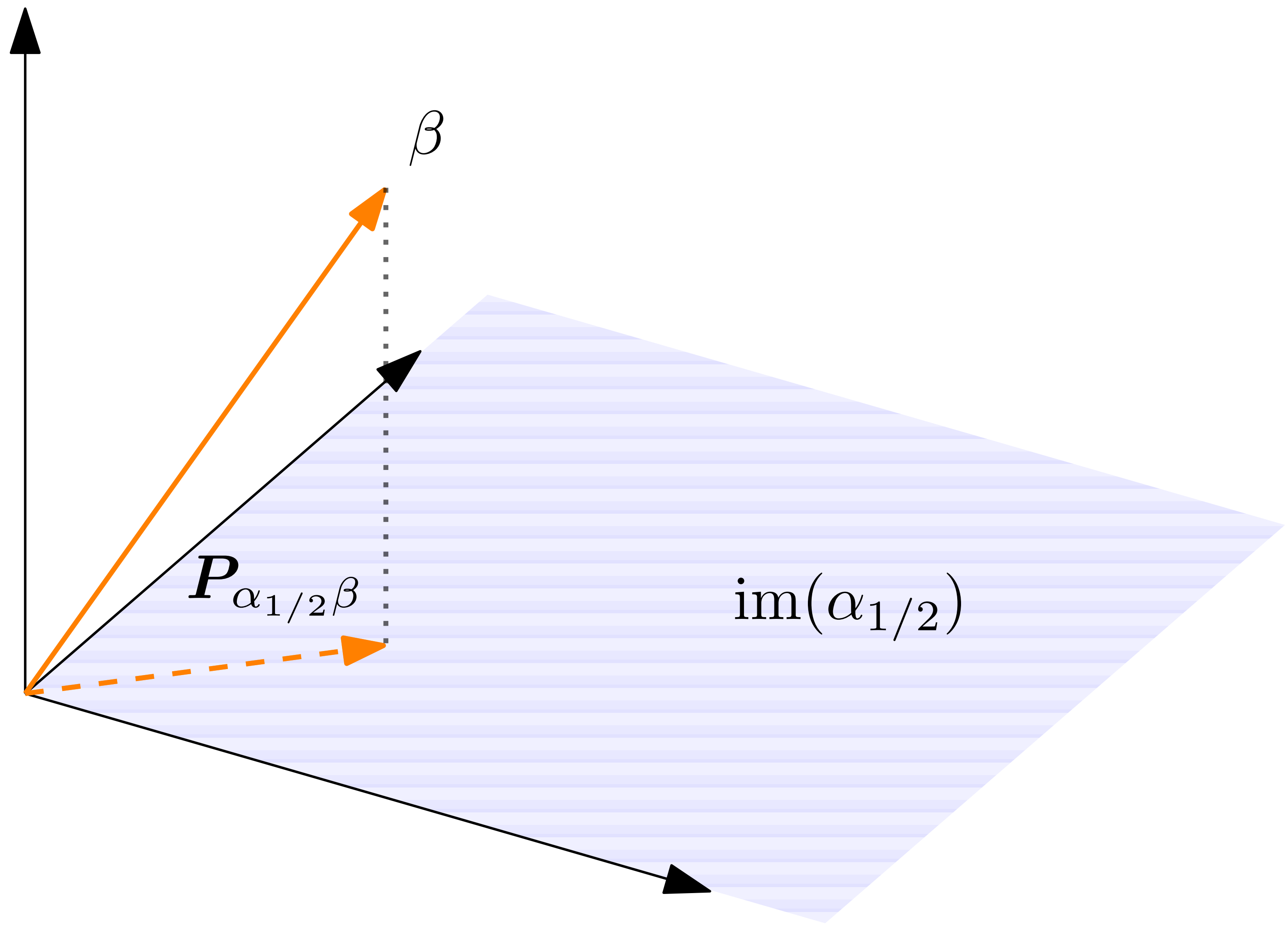


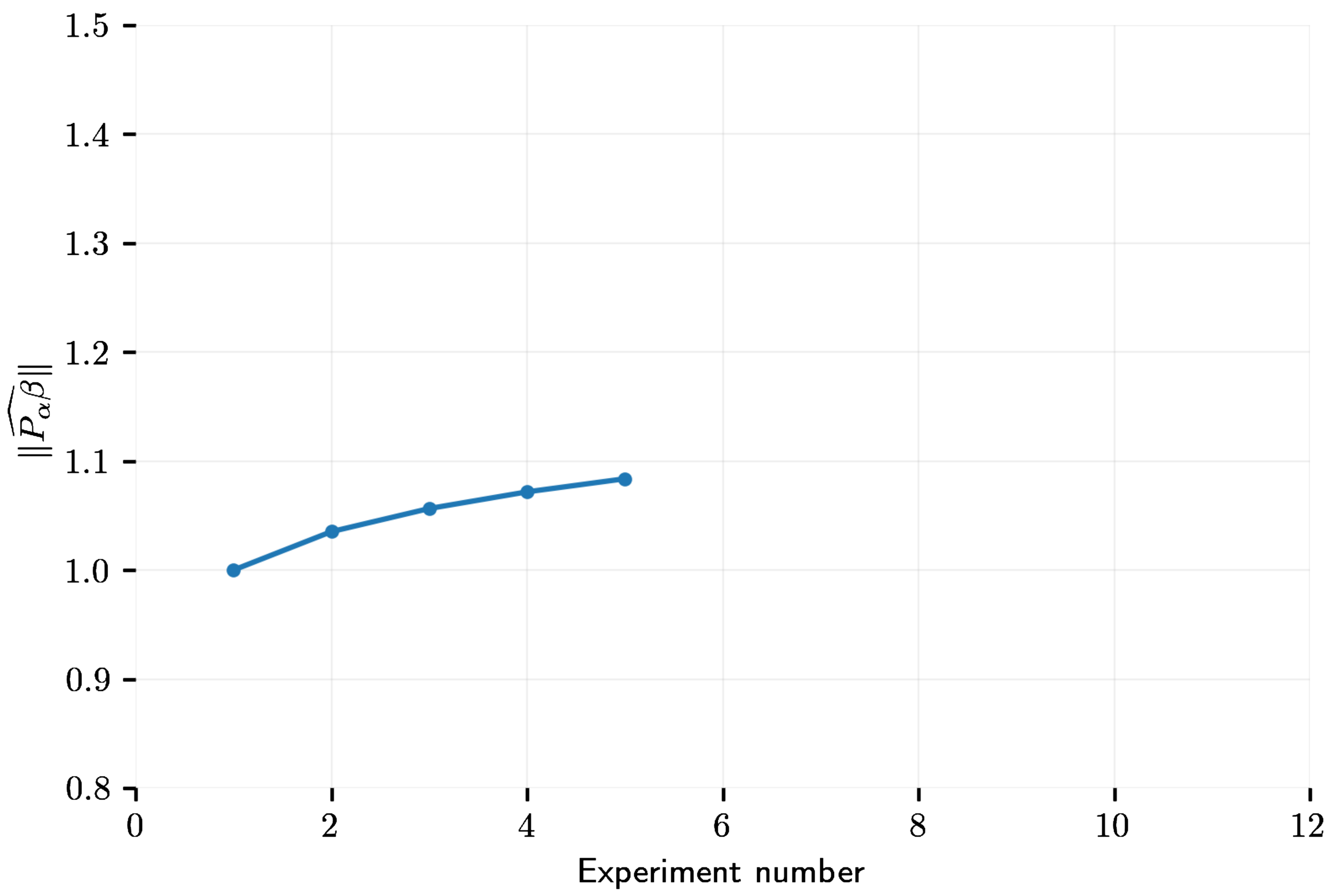


$$\widehat{P_\alpha\beta}_{\text{2SLS}} = (\hat{X}^\top \hat{X})^{+} \hat{X}^\top y$$

We can add experiments until $\|P_{\alpha_{1,2,\dots}}\beta\| = \|\beta\|$

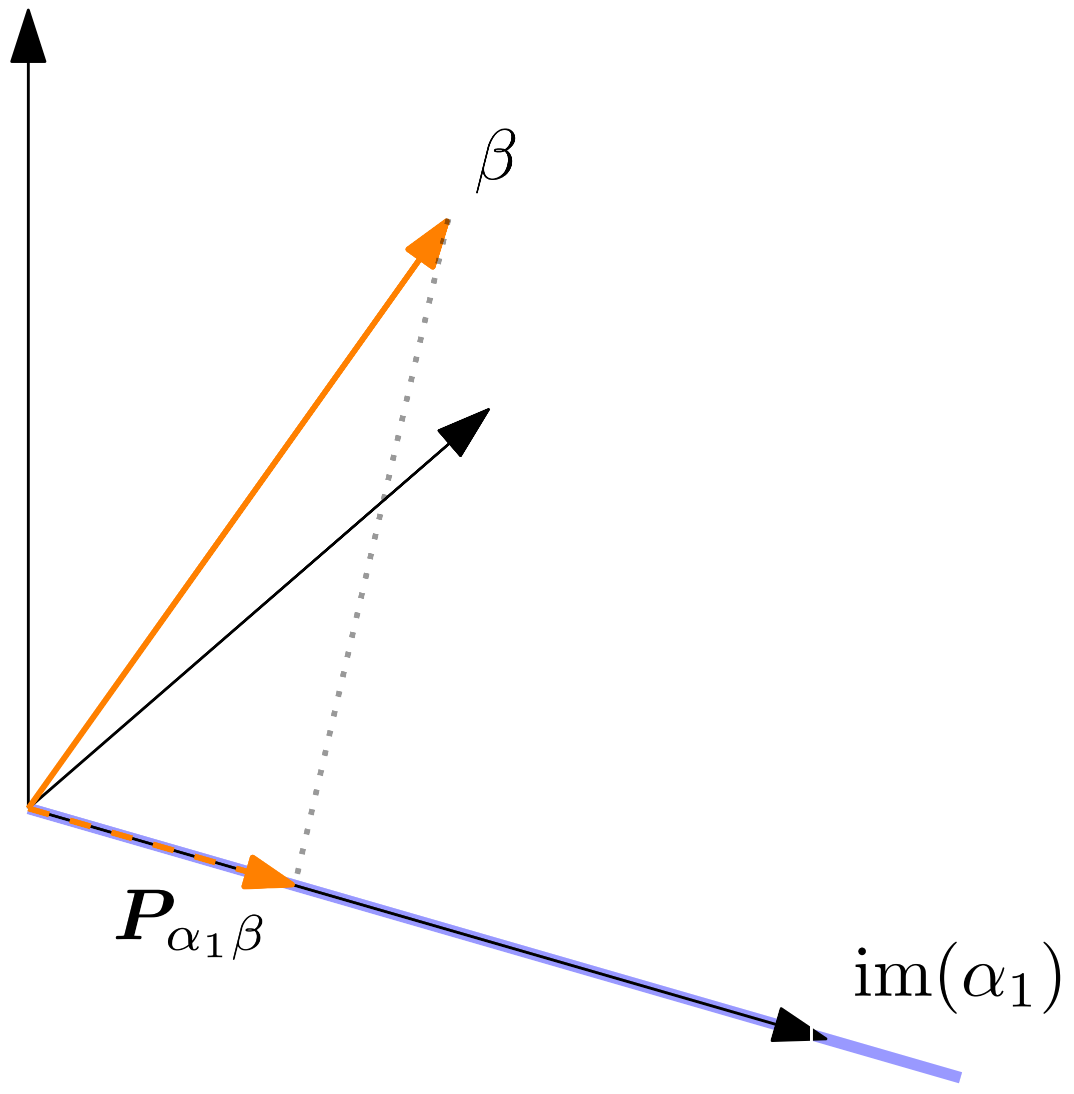


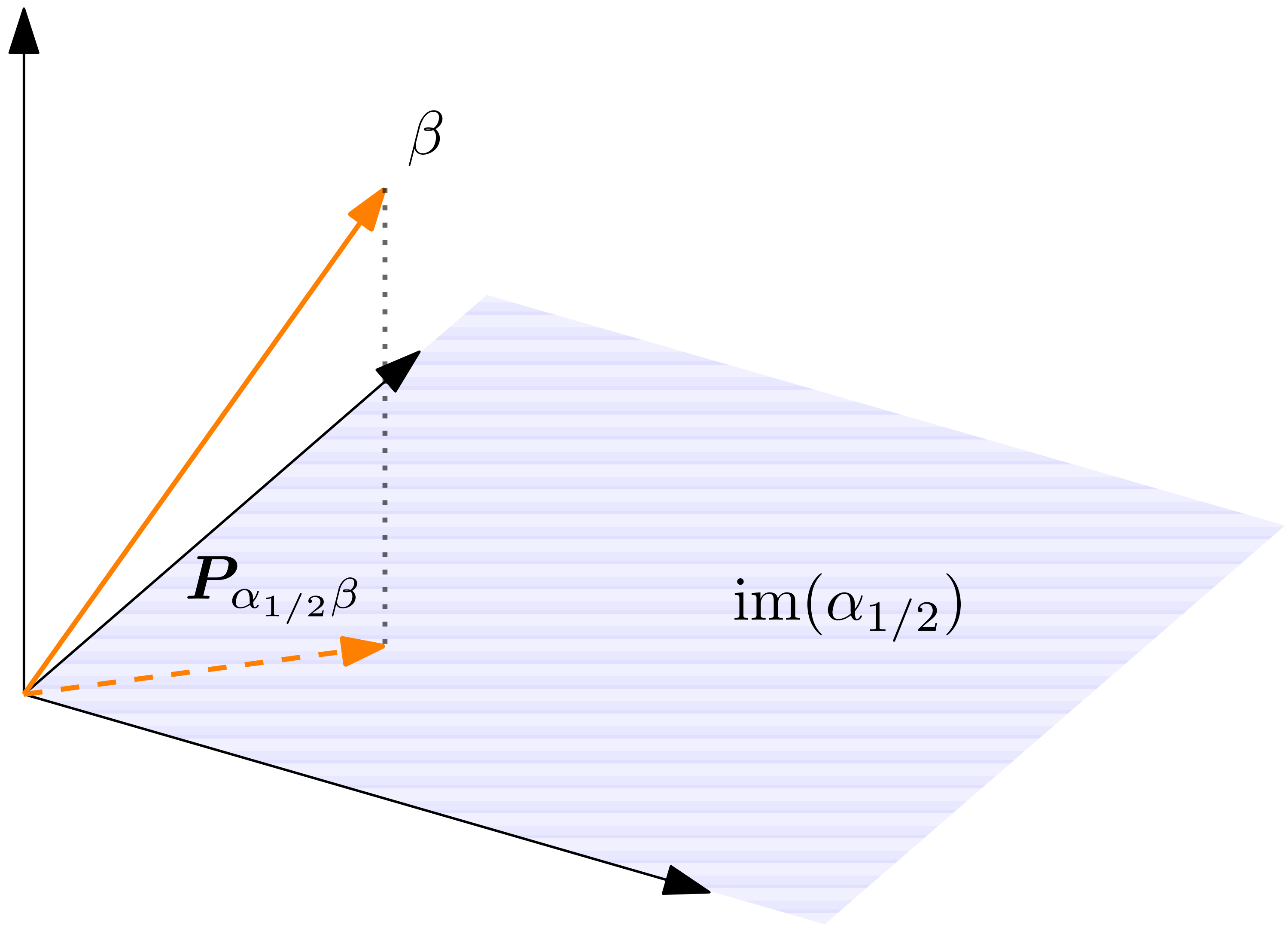


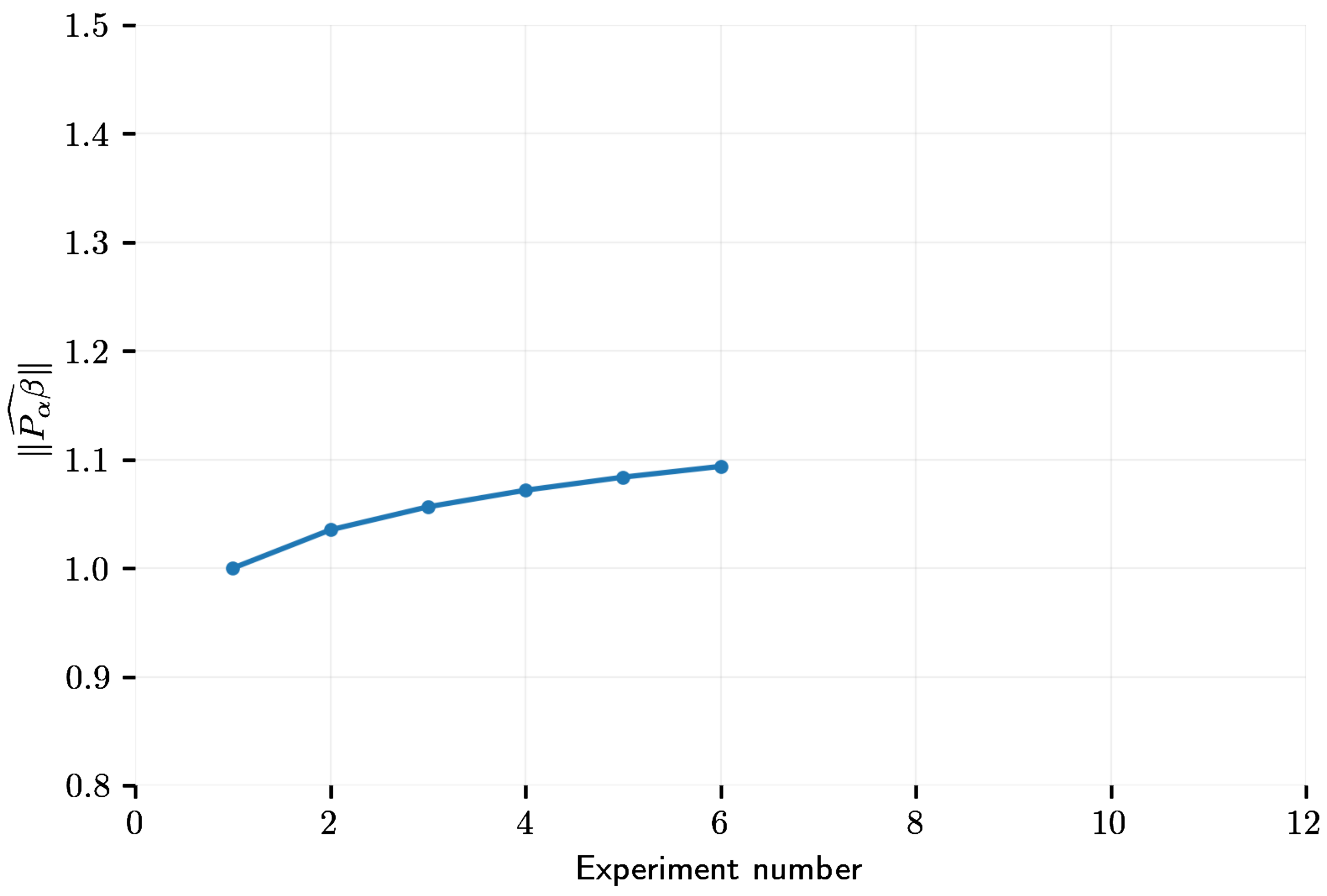


$$\widehat{P_\alpha\beta}_{\cancel{2SLS}} = (\hat{X}^\top \hat{X})^{+} \hat{X}^\top y$$

We can add experiments until $\|P_{\alpha_{1,2,\ldots}}\beta\| = \|\beta\|$

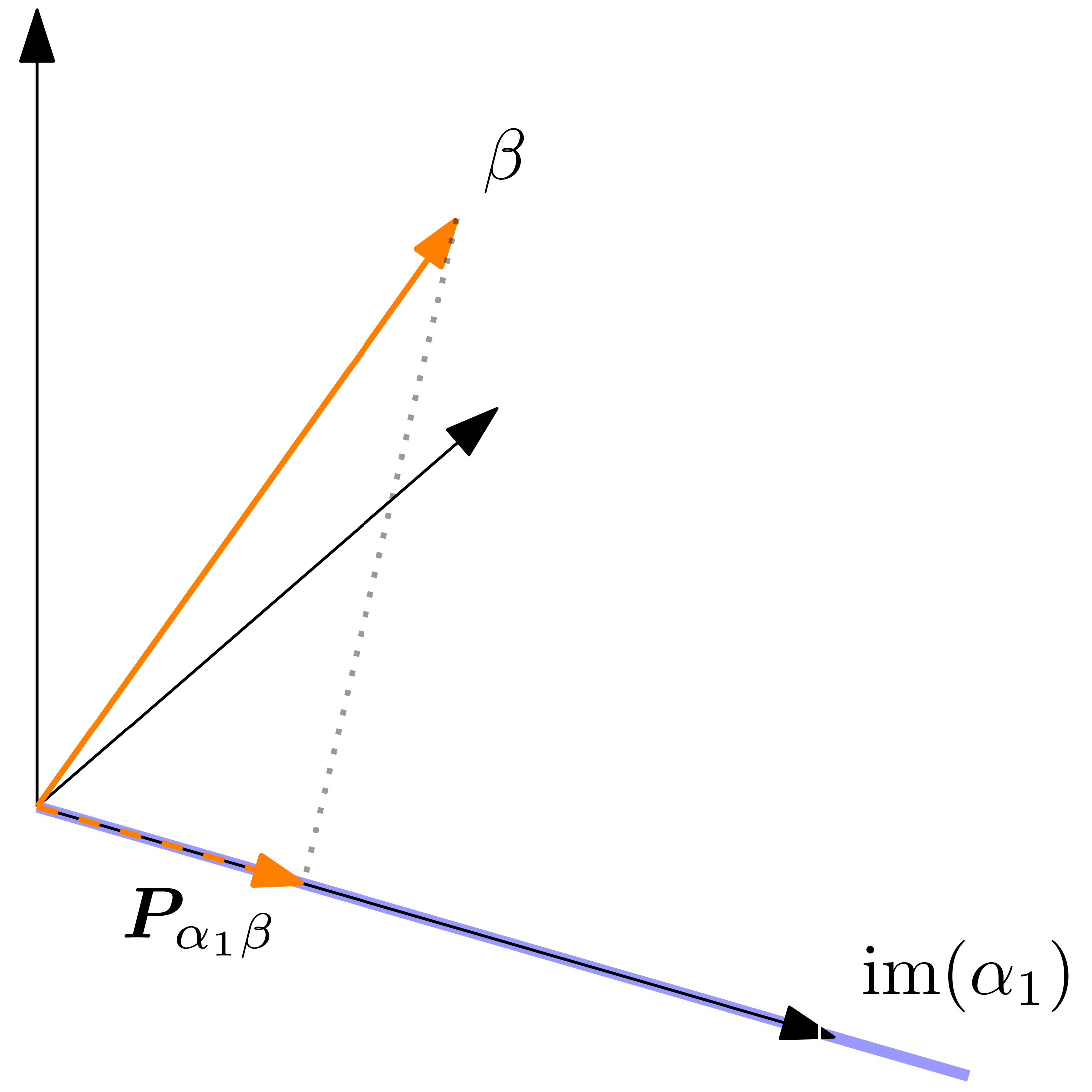


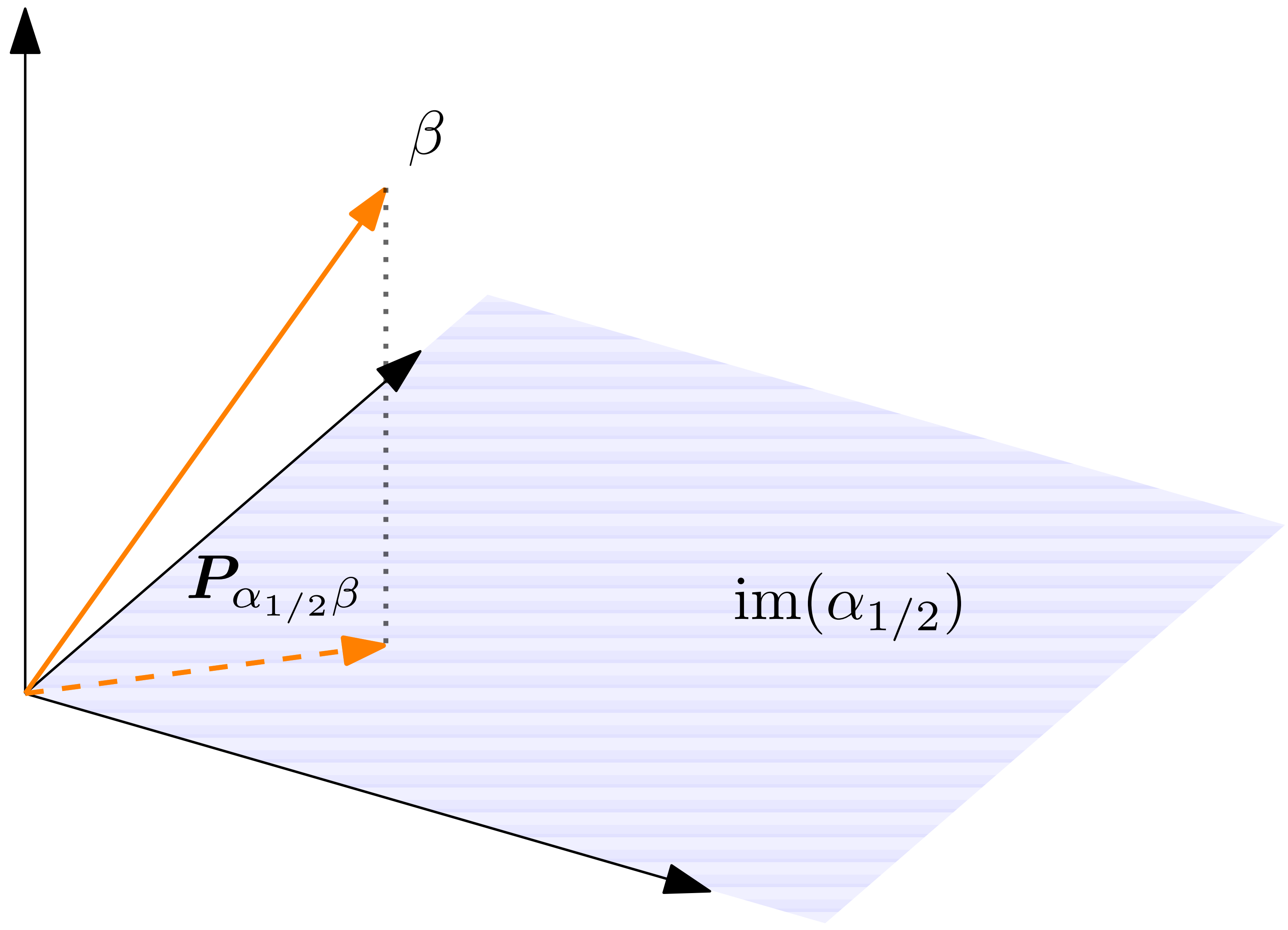


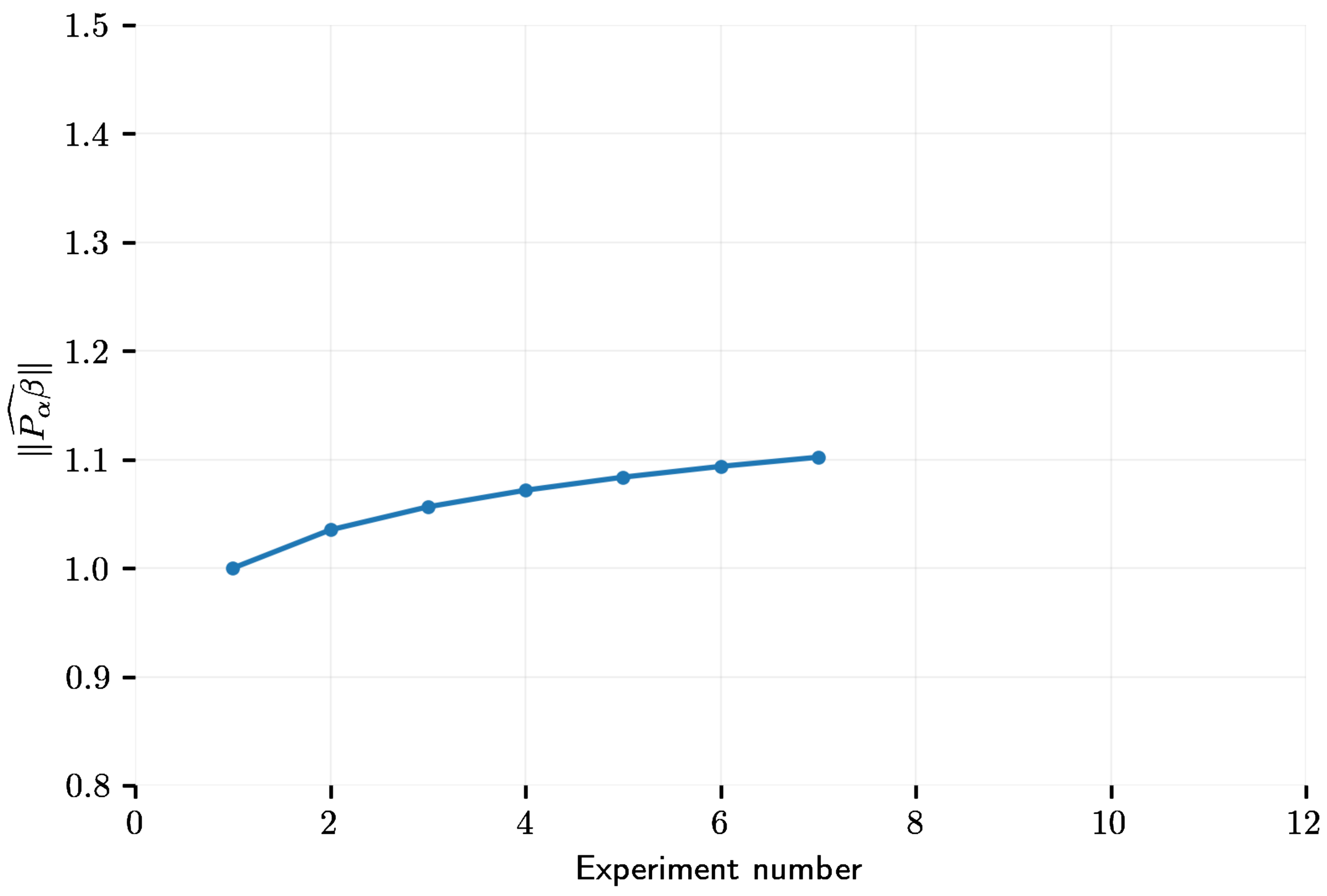


$$\widehat{P_\alpha\beta}_{\text{2SLS}} = (\hat{X}^\top \hat{X})^{+} \hat{X}^\top y$$

We can add experiments until $\|P_{\alpha_{1,2,\ldots}}\beta\| = \|\beta\|$

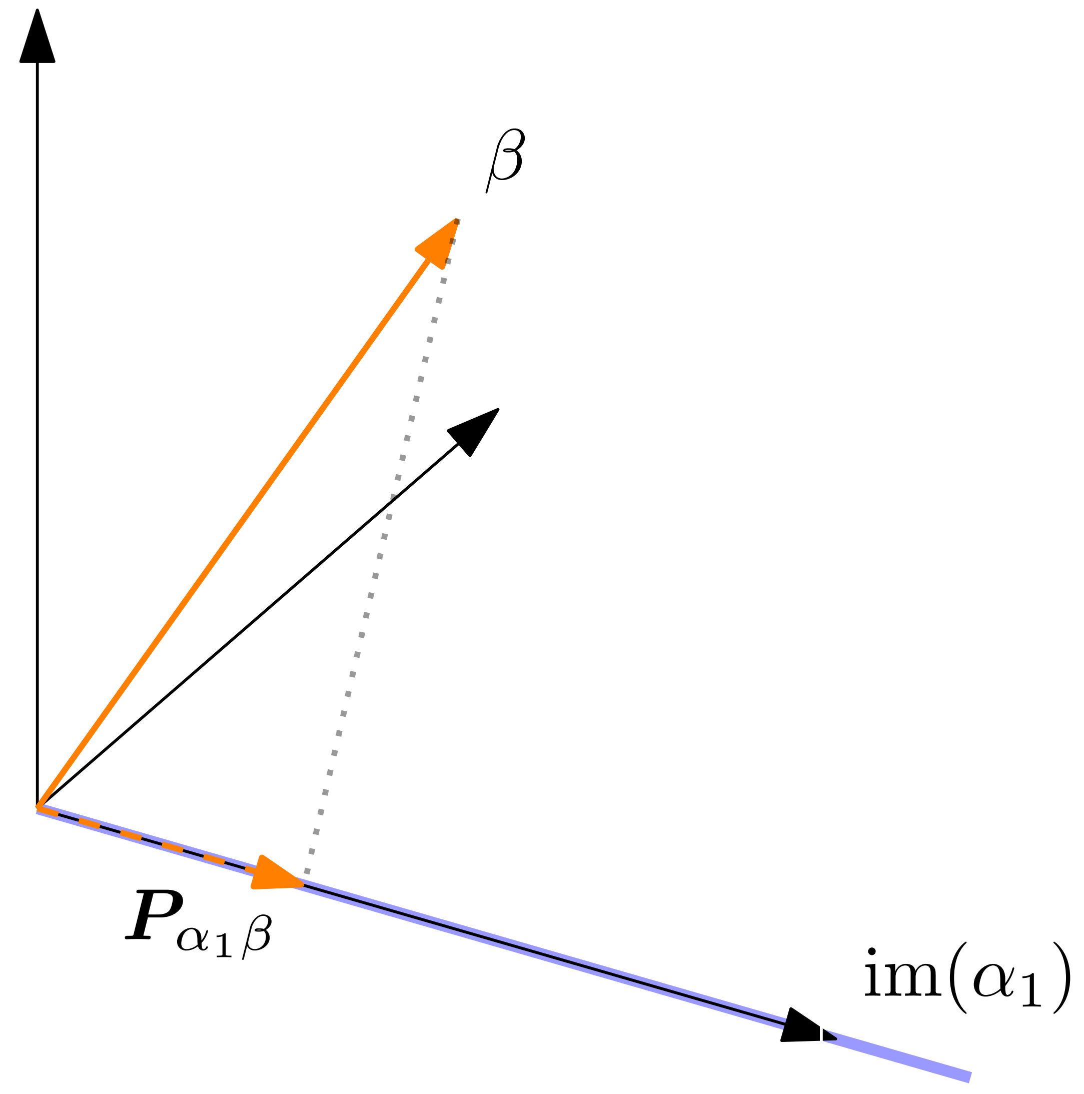


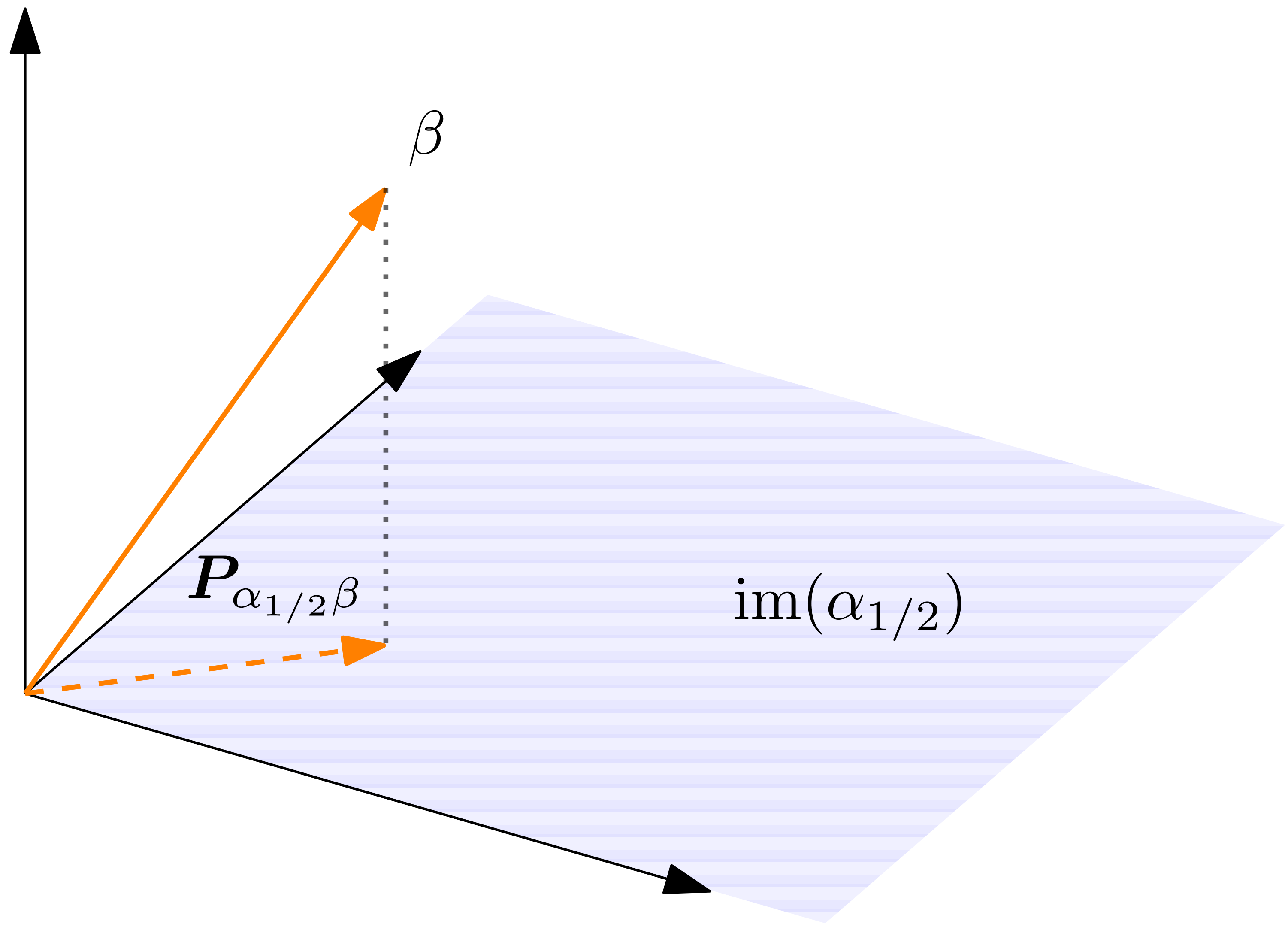


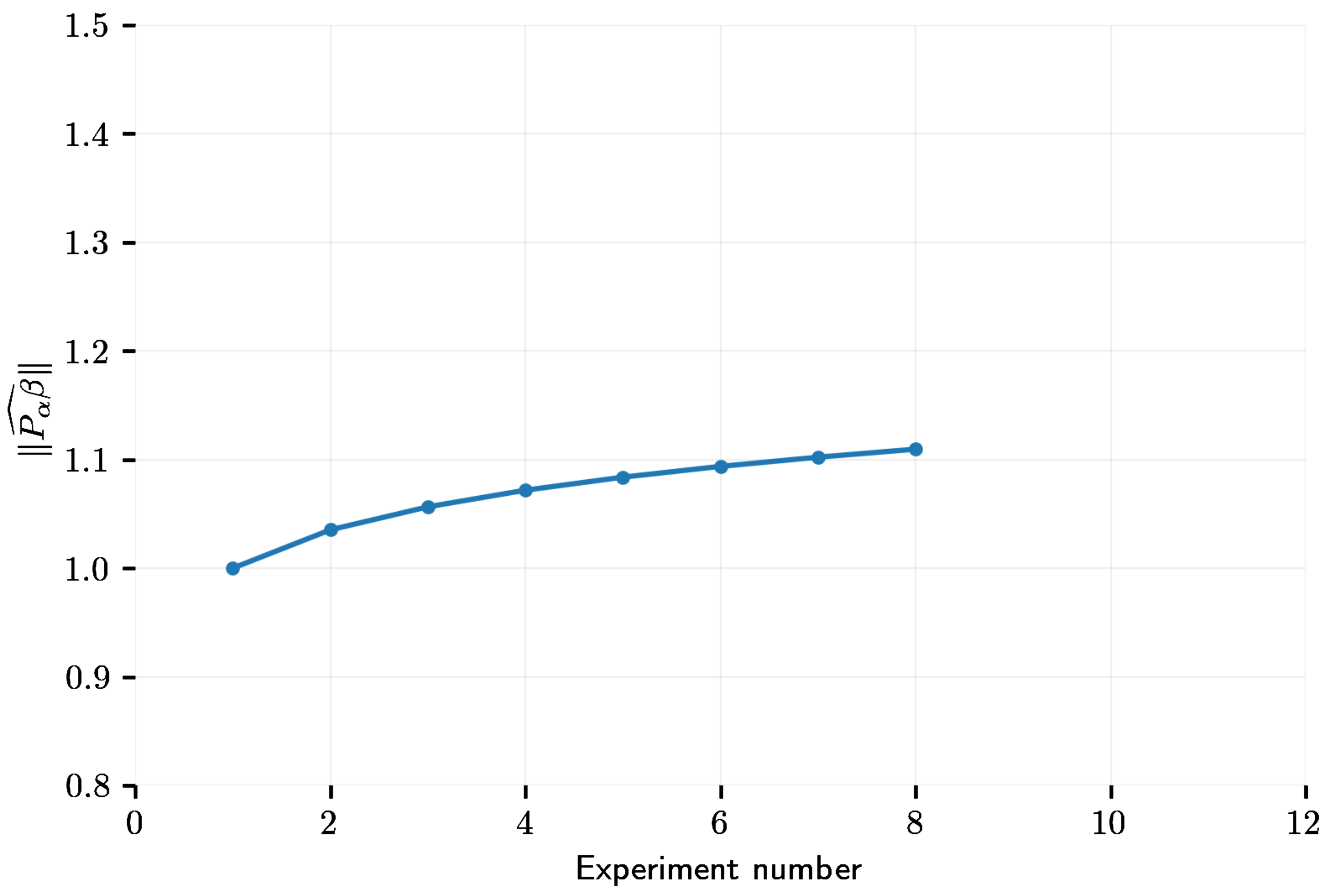


$$\widehat{P_\alpha\beta}_{\cancel{2SLS}} = (\hat{X}^\top\hat{X})^{+}\hat{X}^\top y$$

We can add experiments until $\|P_{\alpha_{1,2,\ldots}}\beta\| = \|\beta\|$

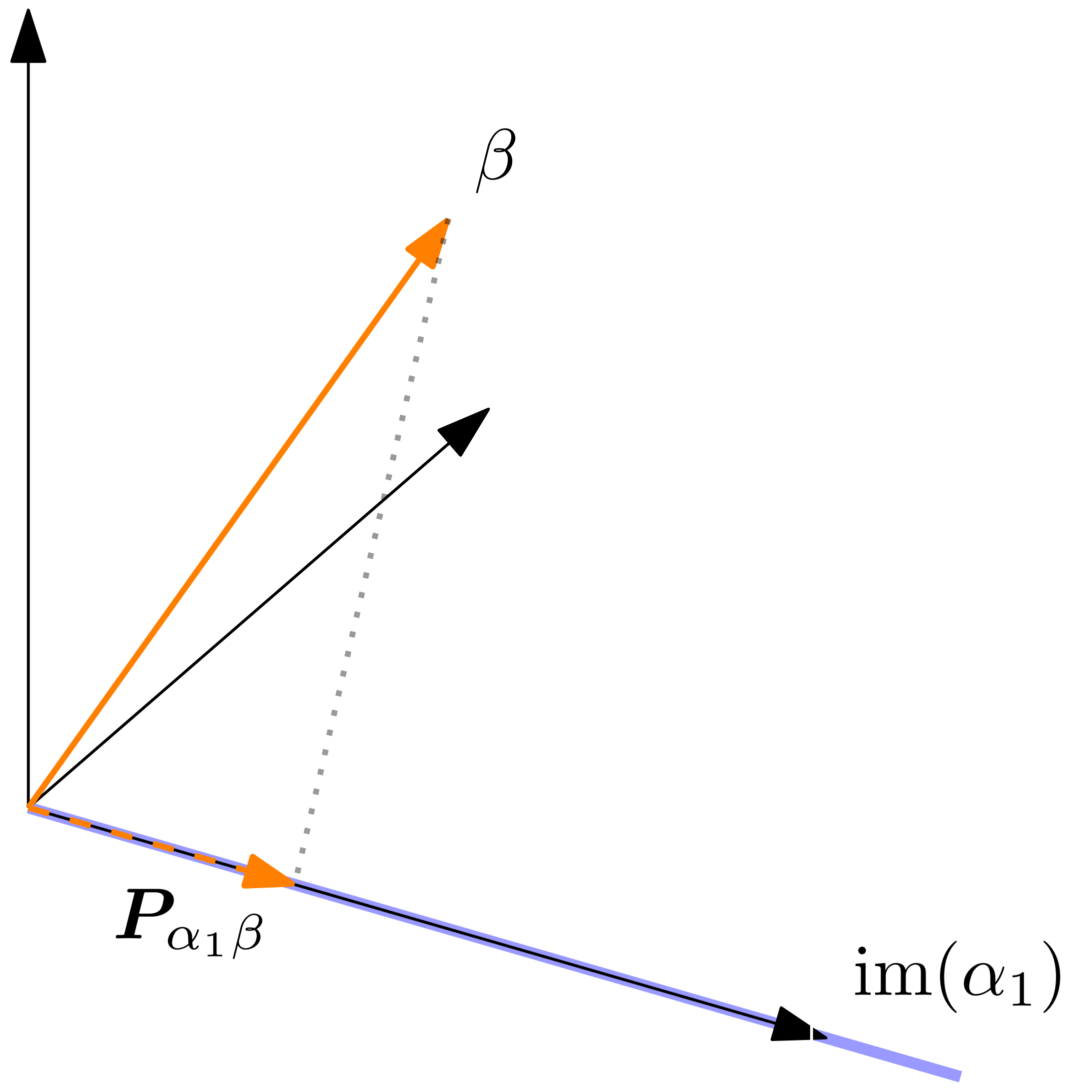


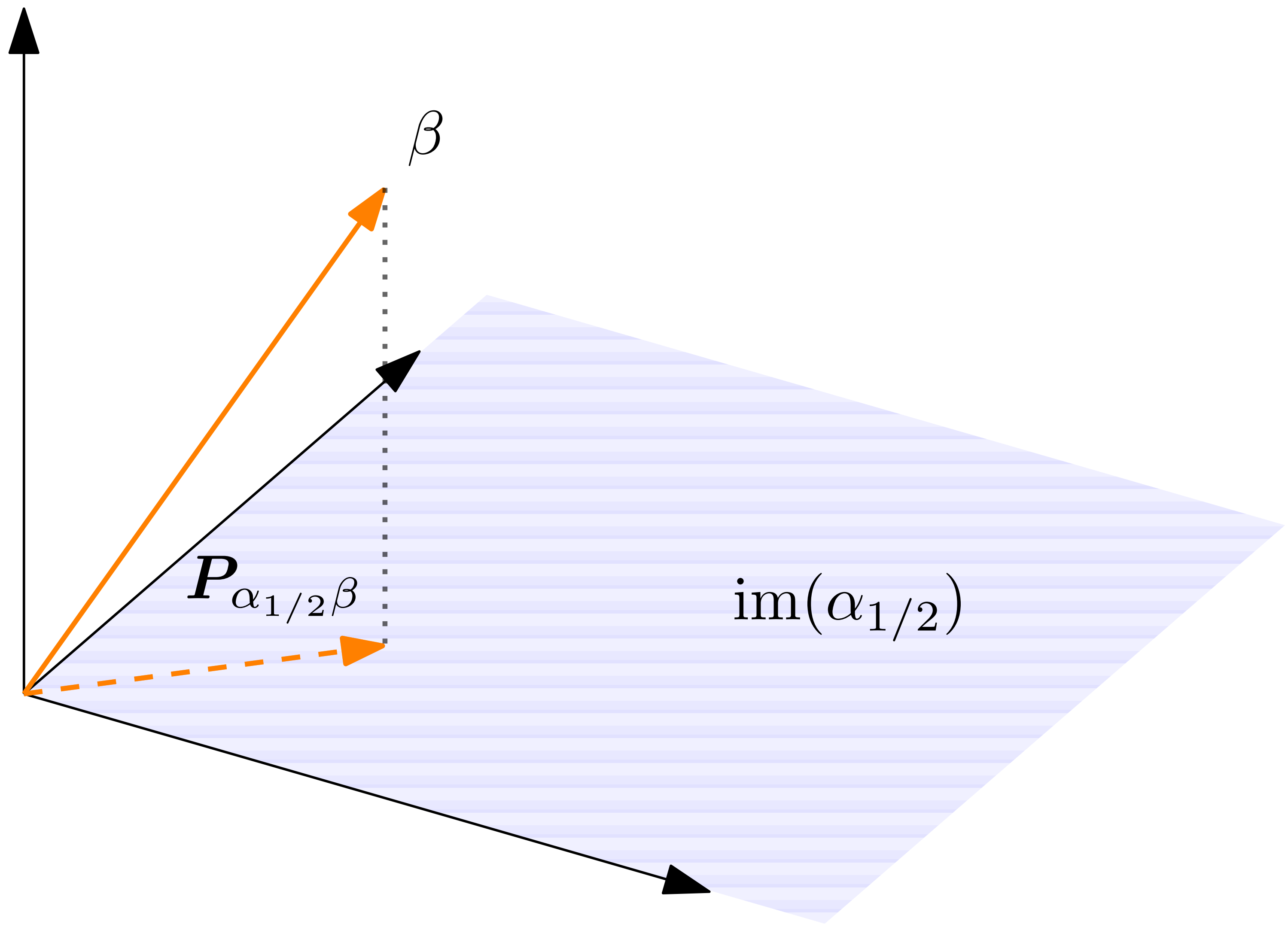


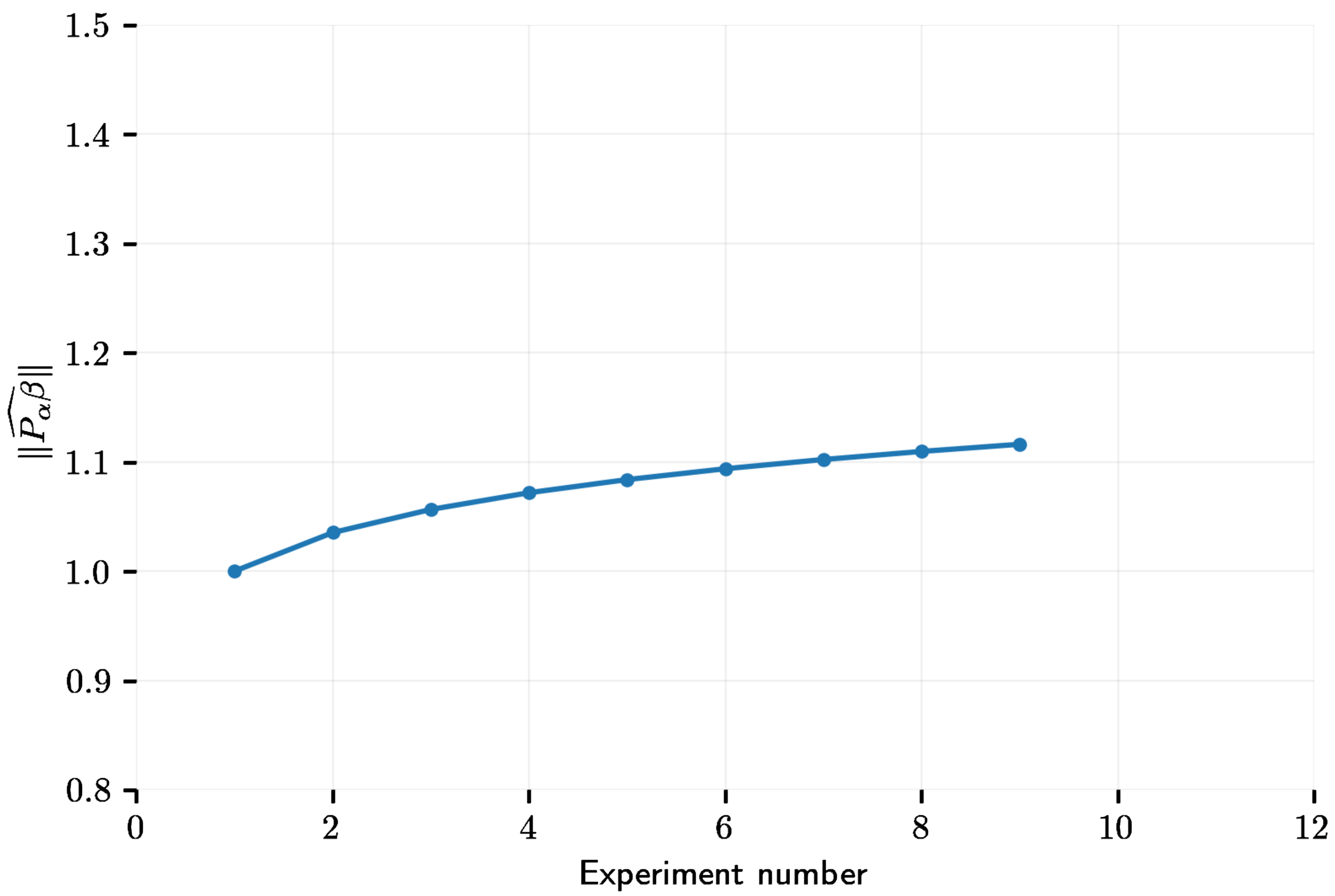


$$\widehat{P_\alpha\beta}_{\cancel{2SLS}} = (\hat{X}^\top \hat{X})^{+} \hat{X}^\top y$$

We can add experiments until $\|P_{\alpha_{1,2,\ldots}}\beta\| = \|\beta\|$

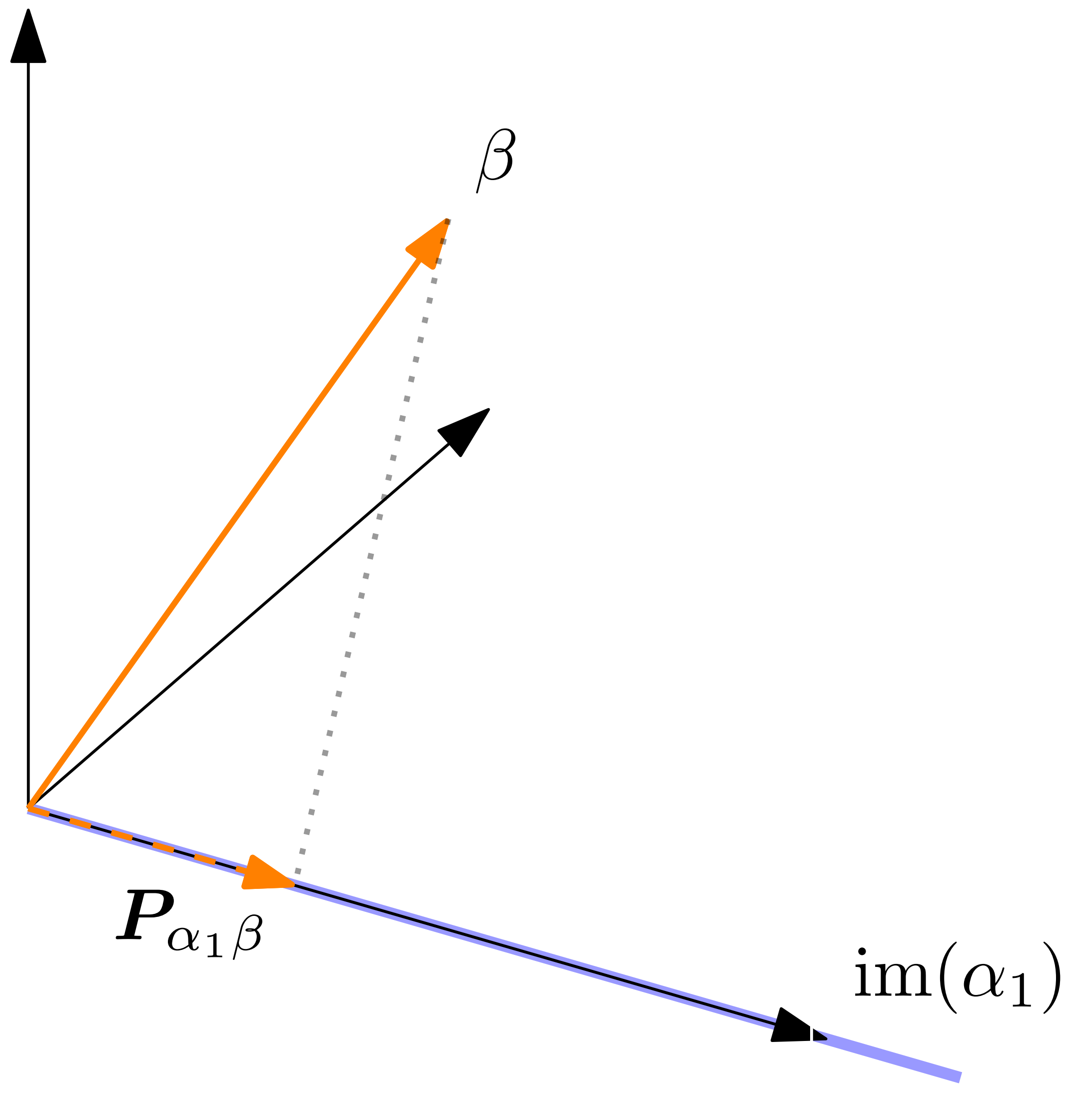


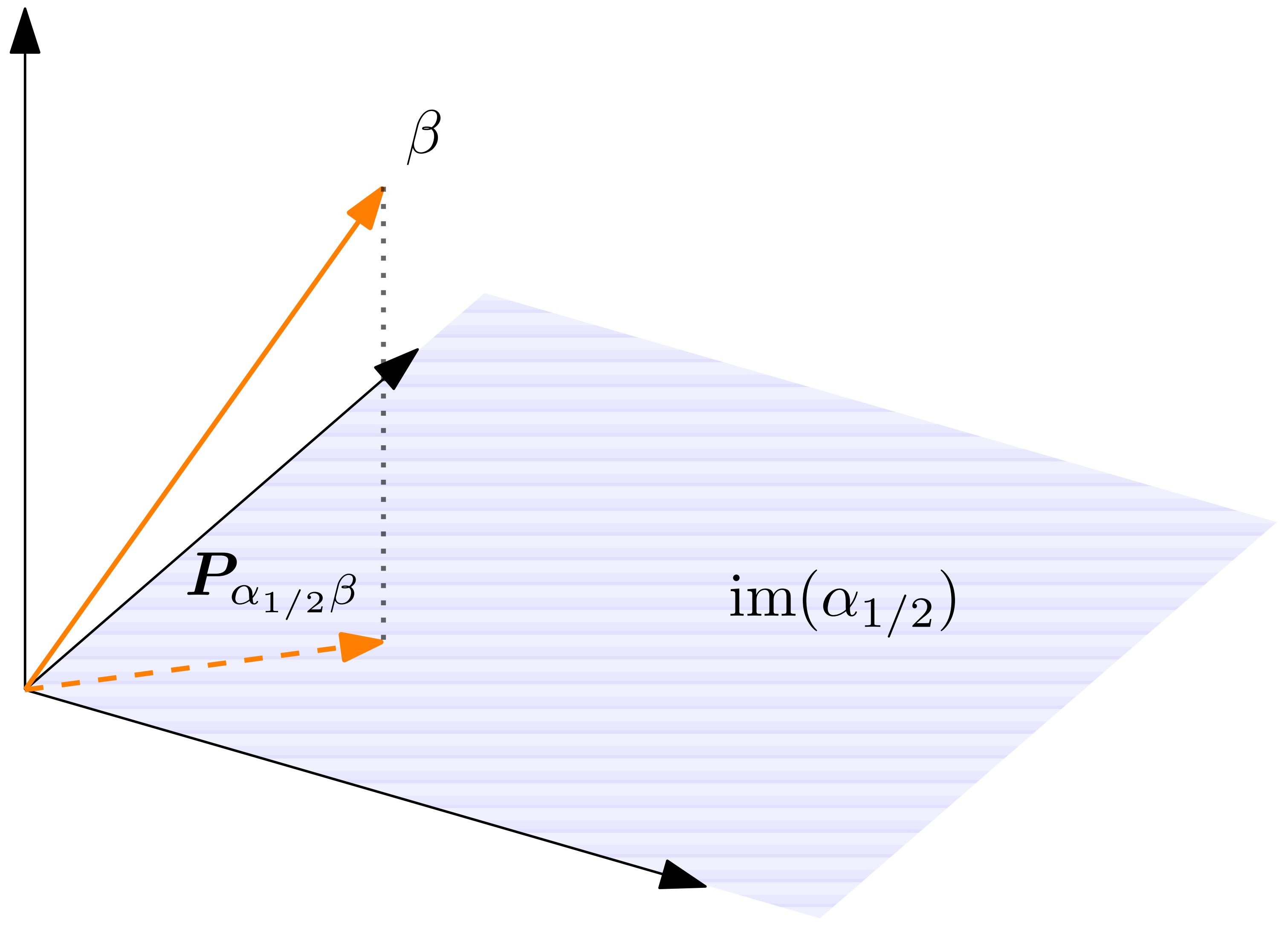


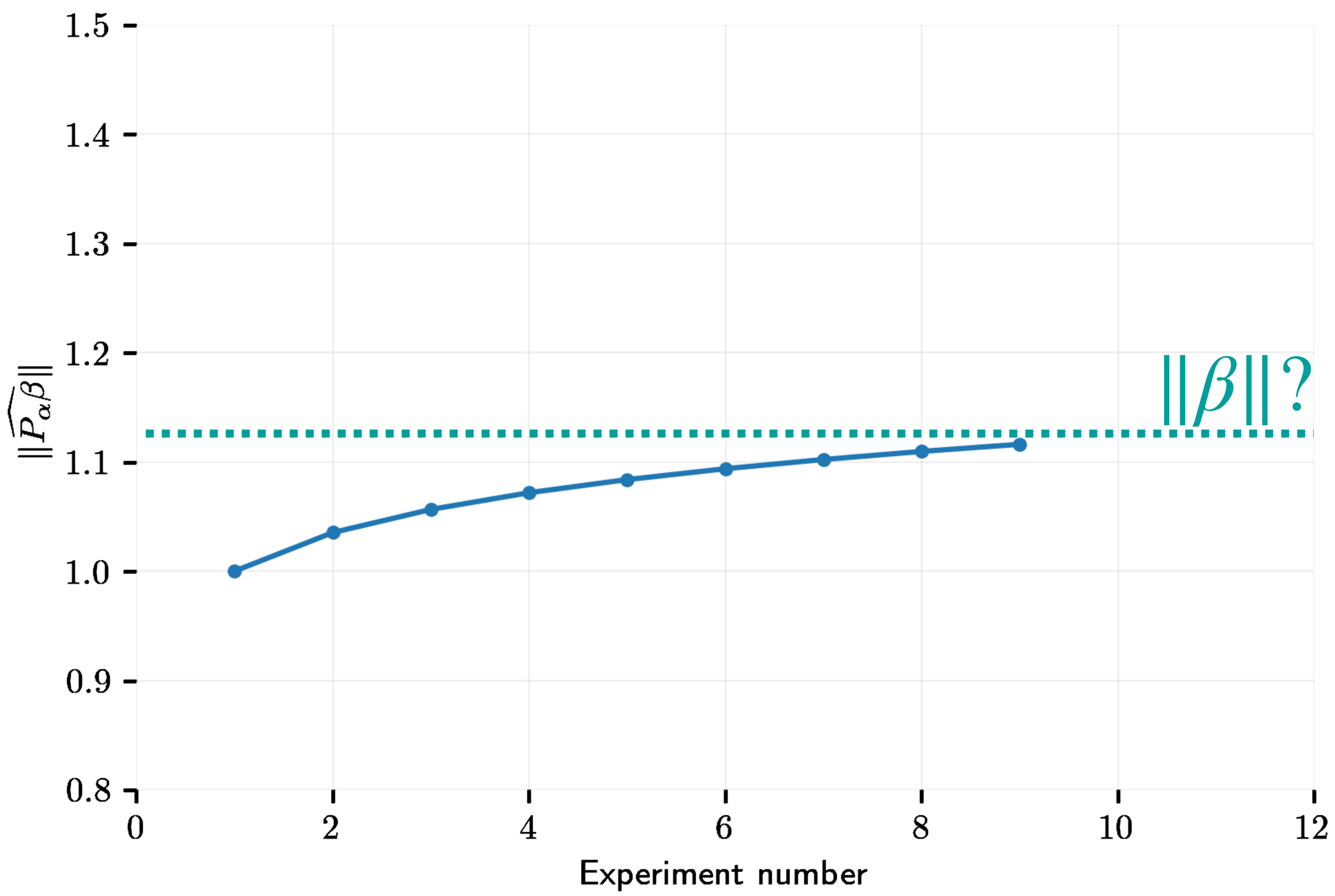


$$\widehat{P_\alpha\beta}_{\text{2SLS}} = (\hat{X}^\top\hat{X})^{+}\hat{X}^\top y$$

We can add experiments until $\|P_{\alpha_{1,2,\ldots}}\beta\| = \|\beta\|$

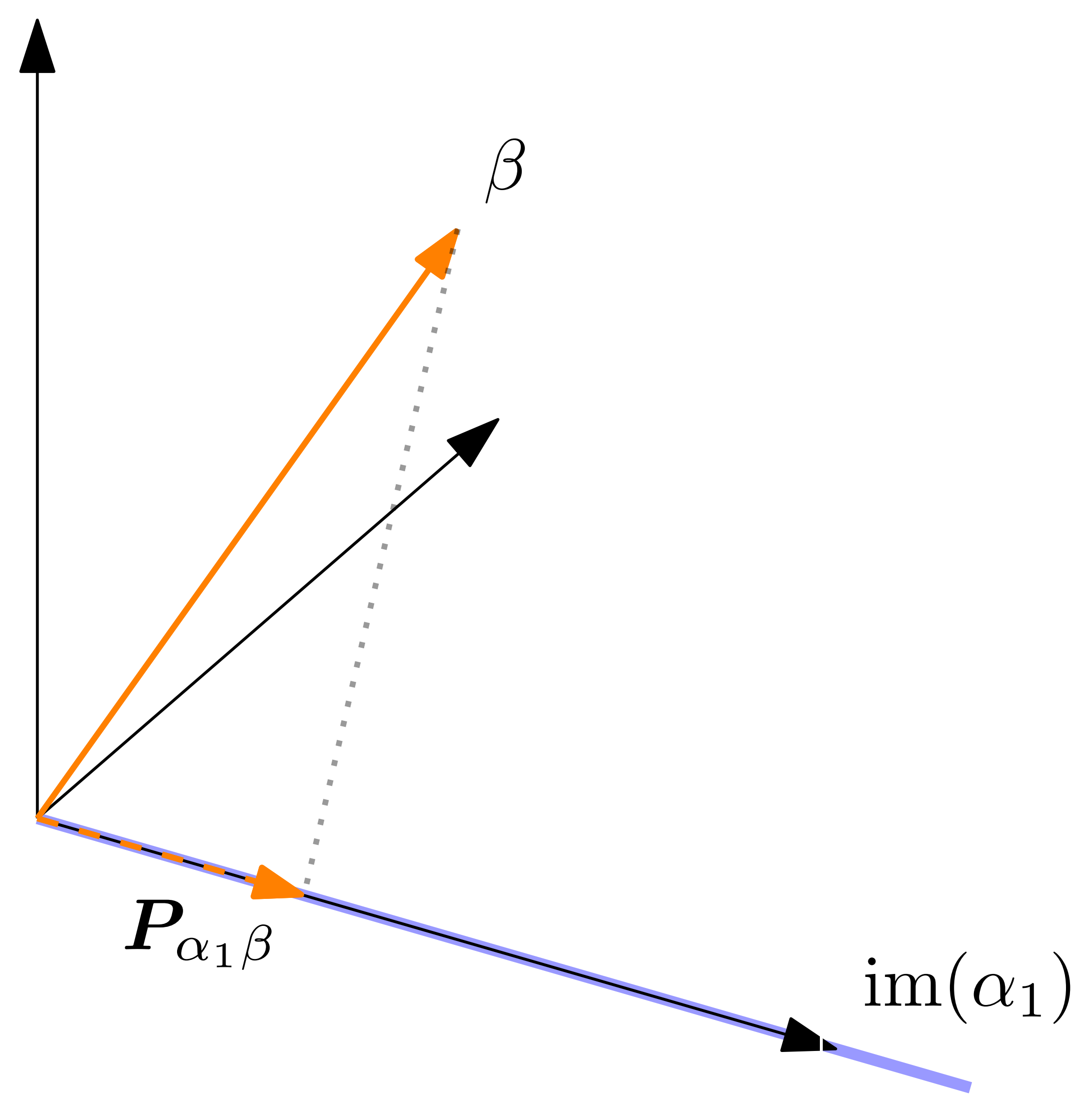


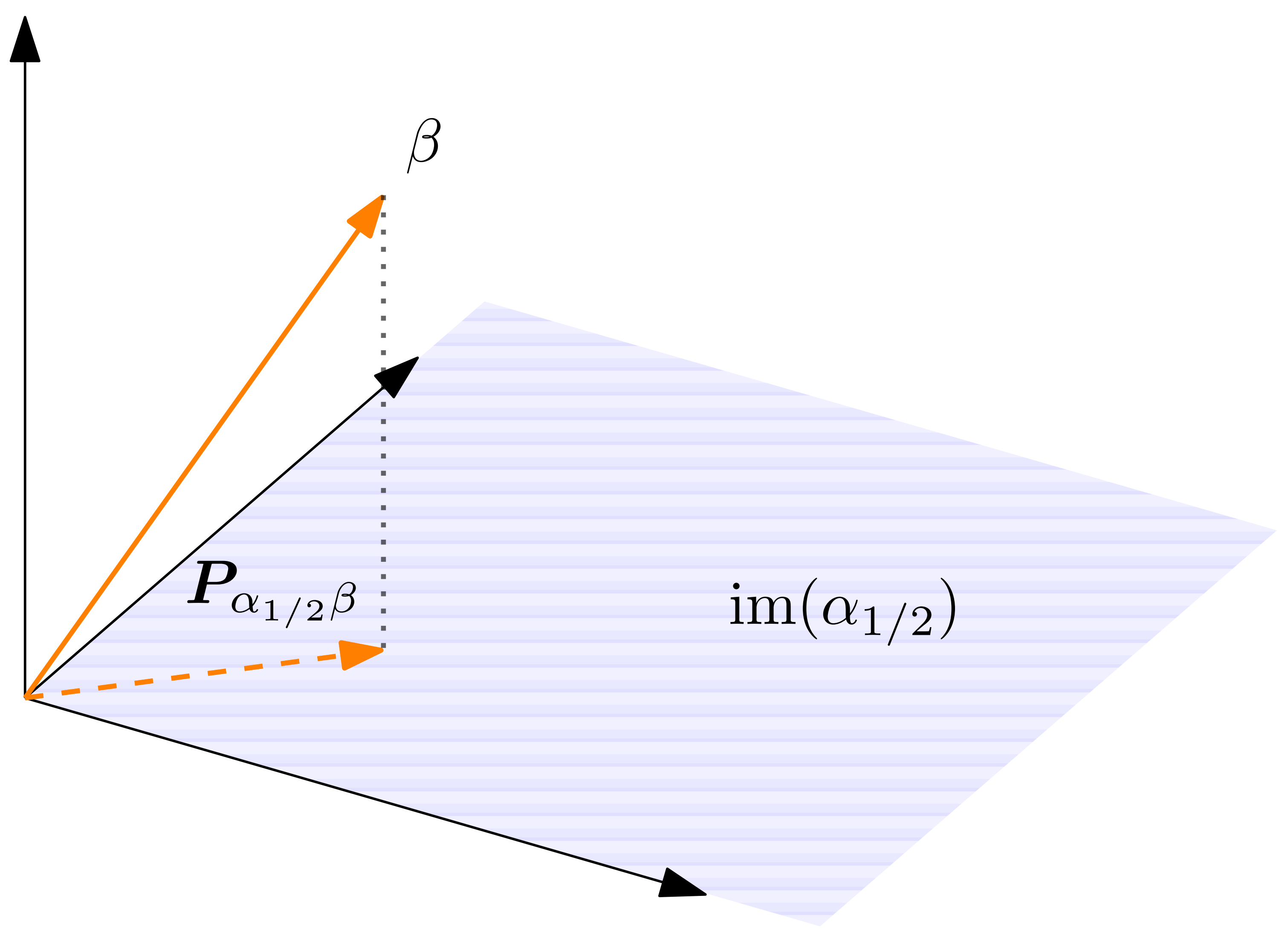


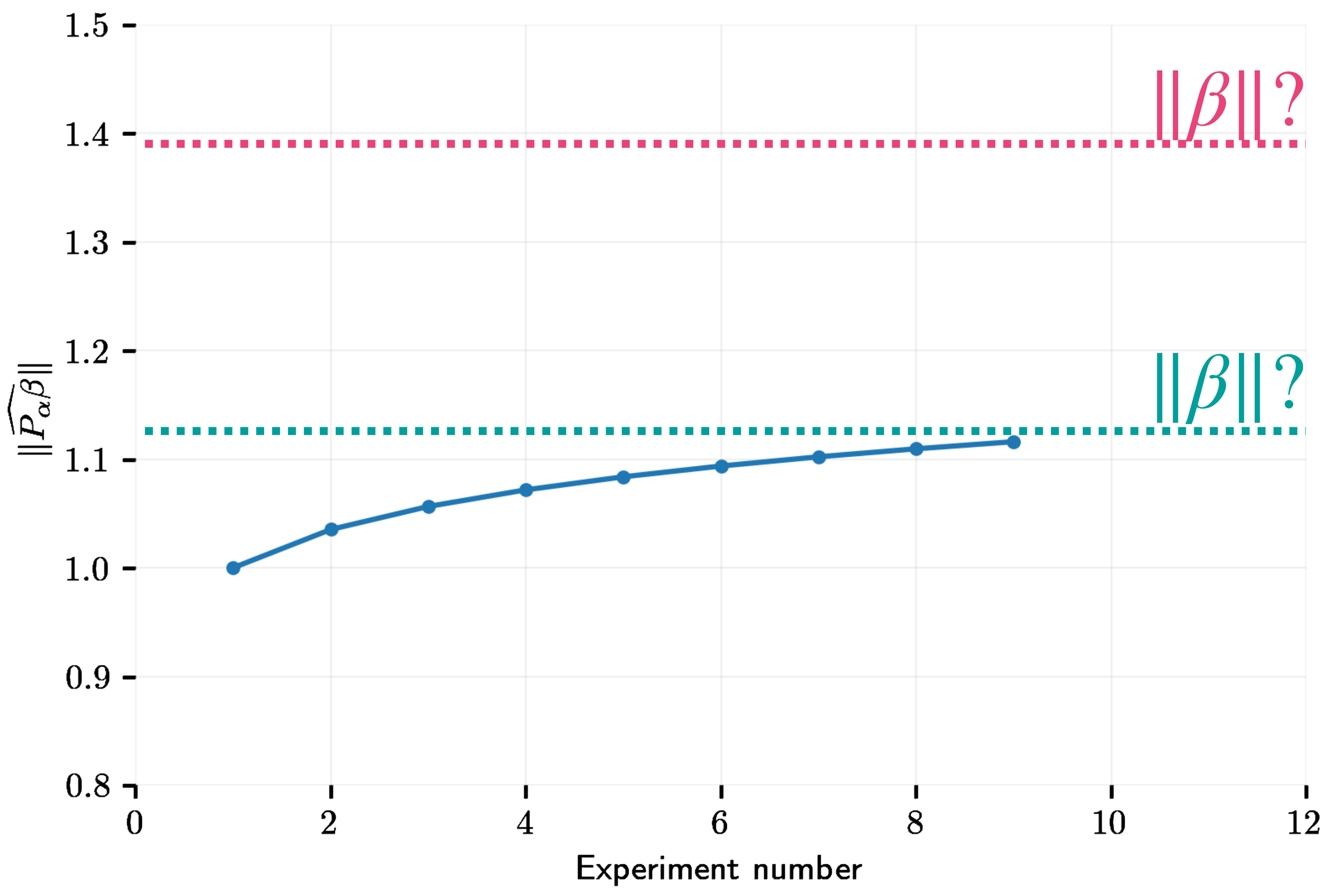


$$\widehat{P_\alpha\beta}_{\text{~~2SLS~~}} = (\hat{X}^\top \hat{X})^{+} \hat{X}^\top y$$

We can add experiments until $\|P_{\alpha_{1,2,\ldots}}\beta\| = \|\beta\|$

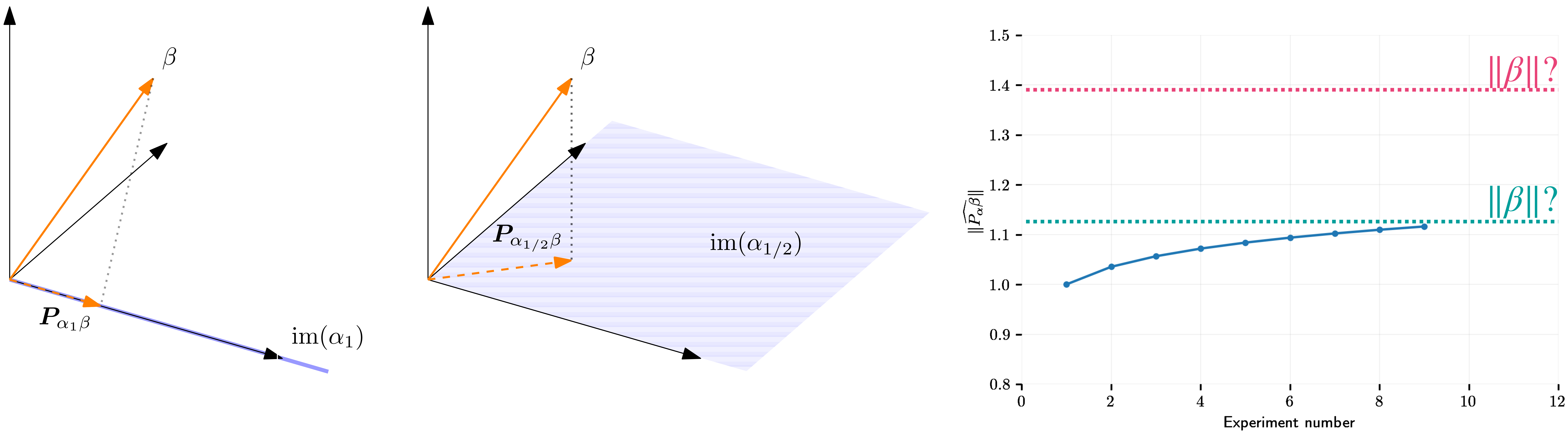






$$\widehat{P_\alpha\beta}_{\cancel{2SLS}} = (\hat{X}^\top \hat{X})^{+} \hat{X}^\top y$$

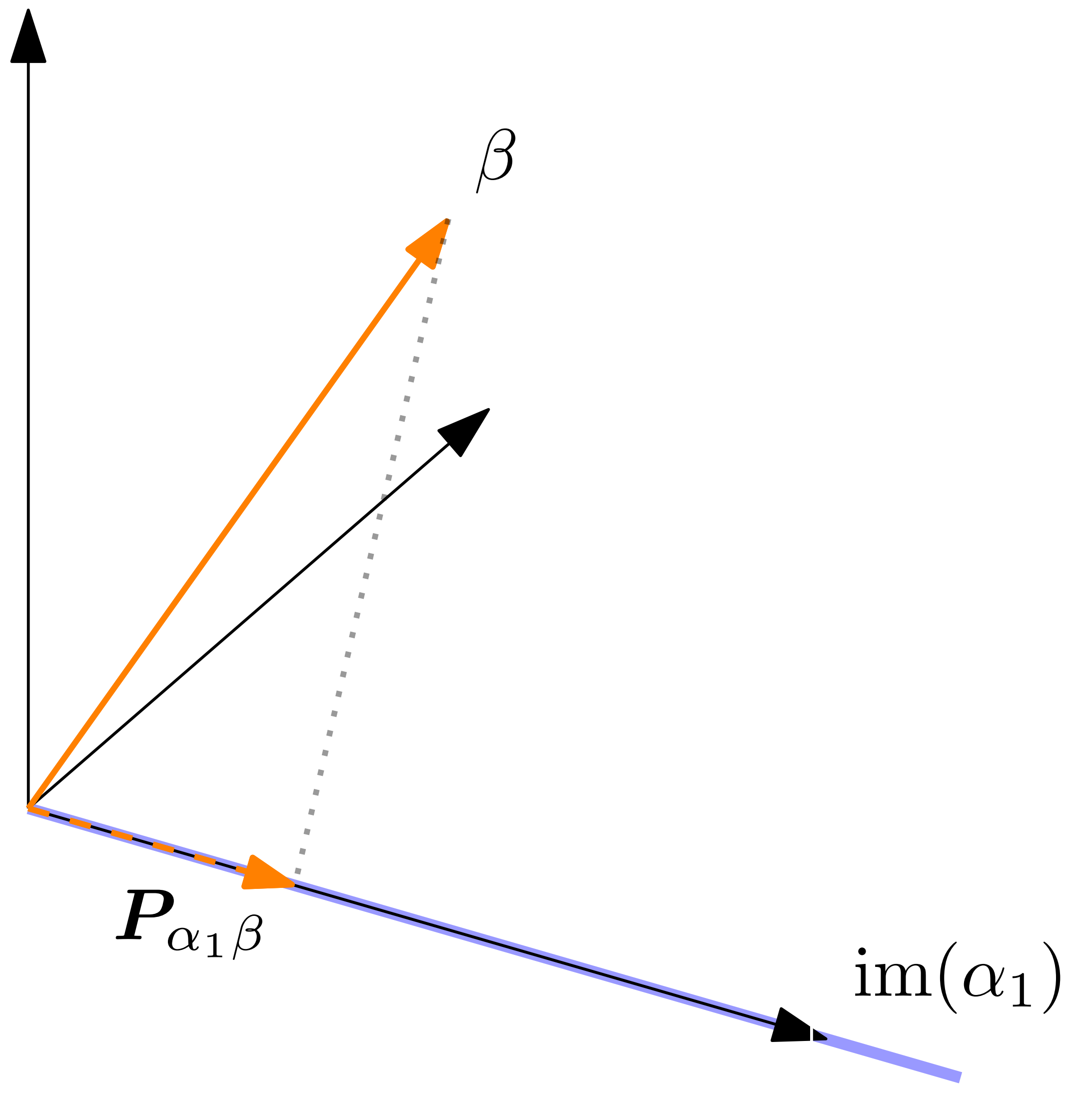
We can add experiments until $\|P_{\alpha_{1,2,\ldots}}\beta\| = \|\beta\|$

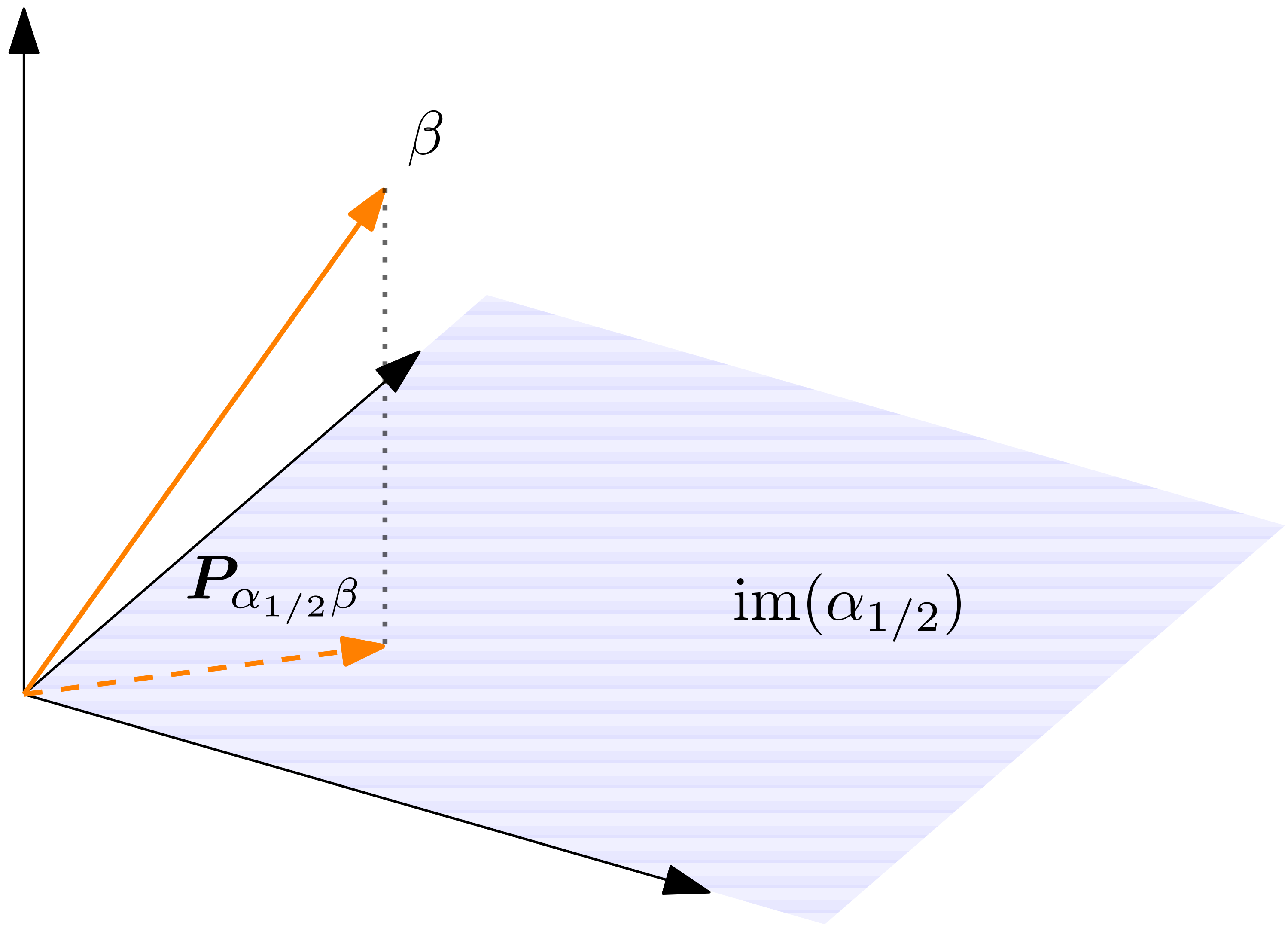


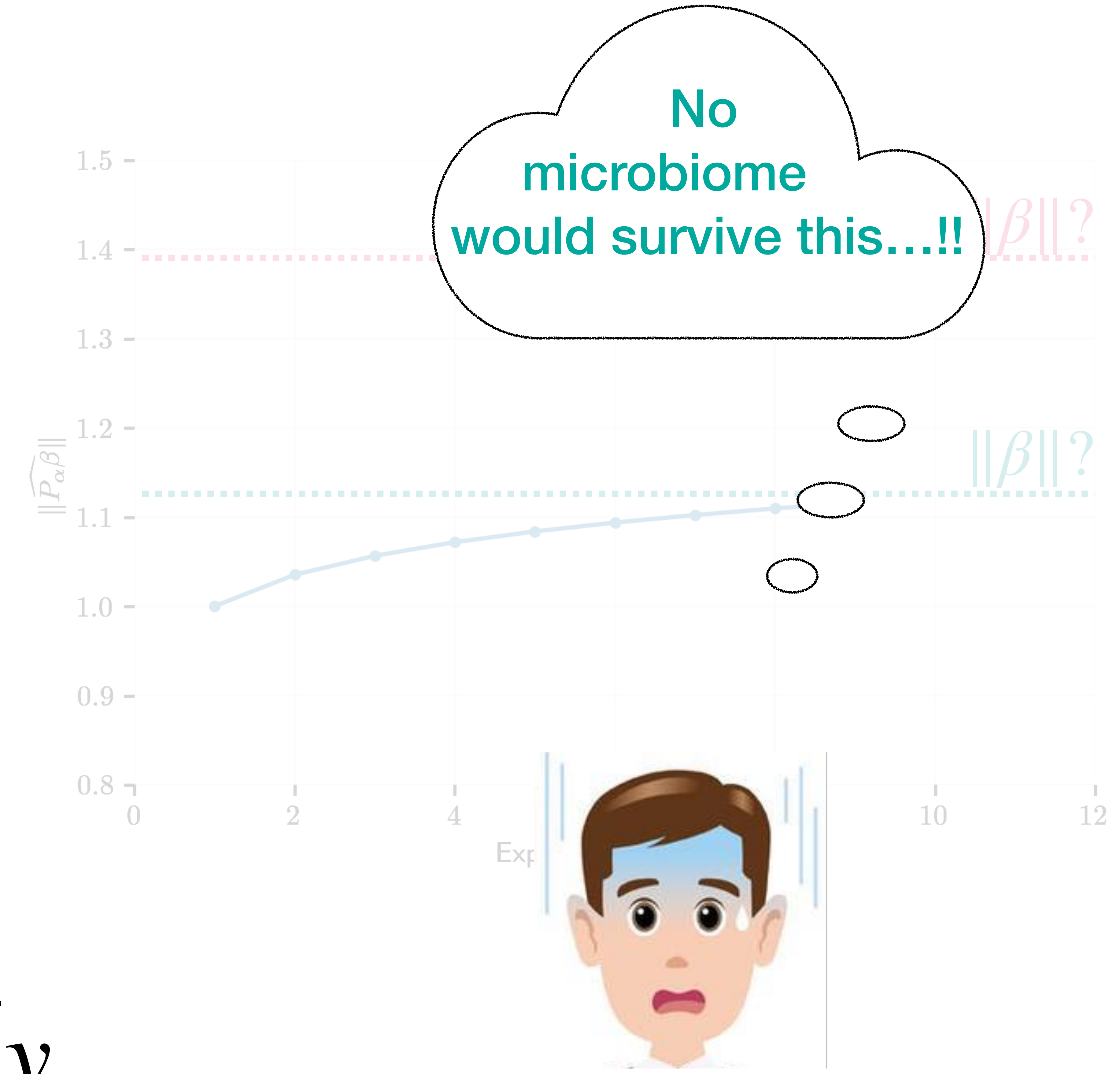
$$\widehat{P_\alpha\beta}_{\cancel{2SLS}} = \boxed{(\hat{X}^\top\hat{X})^{+}}\hat{X}^\top y$$

Janzing, D. and Schölkopf, B. Detecting non-causal artifacts in multivariate linear regression models. In International Conference on Machine Learning, pp. 2245–2253. PMLR, 2018. 5, 7, 12

We can add experiments until $\|P_{\alpha_{1,2,\ldots}}\beta\| = \|\beta\|$



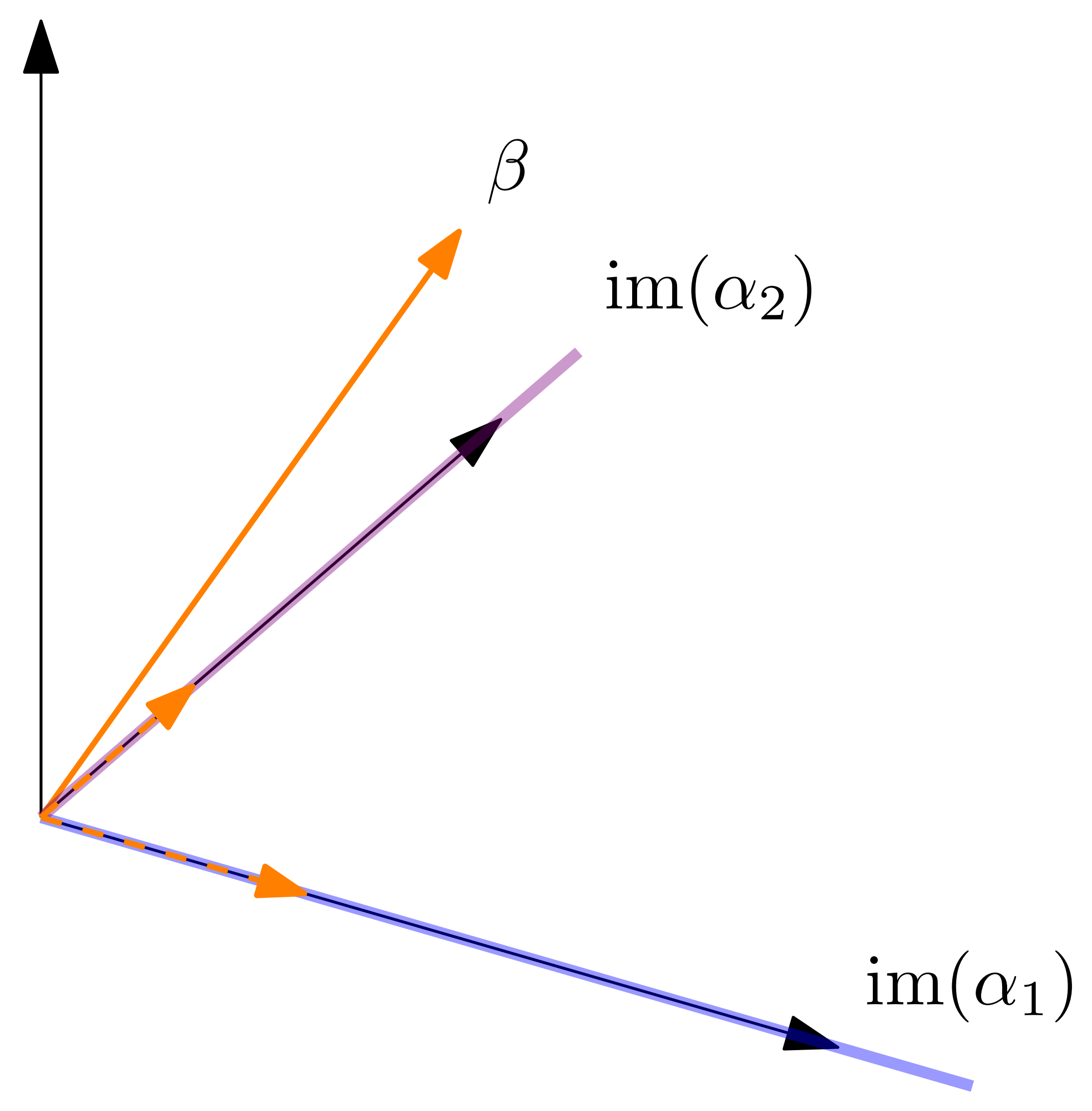




$$\widehat{P_\alpha \beta}_{\text{2SLS}} = (\hat{X}^\top \hat{X})^{+} \hat{X}^\top y$$

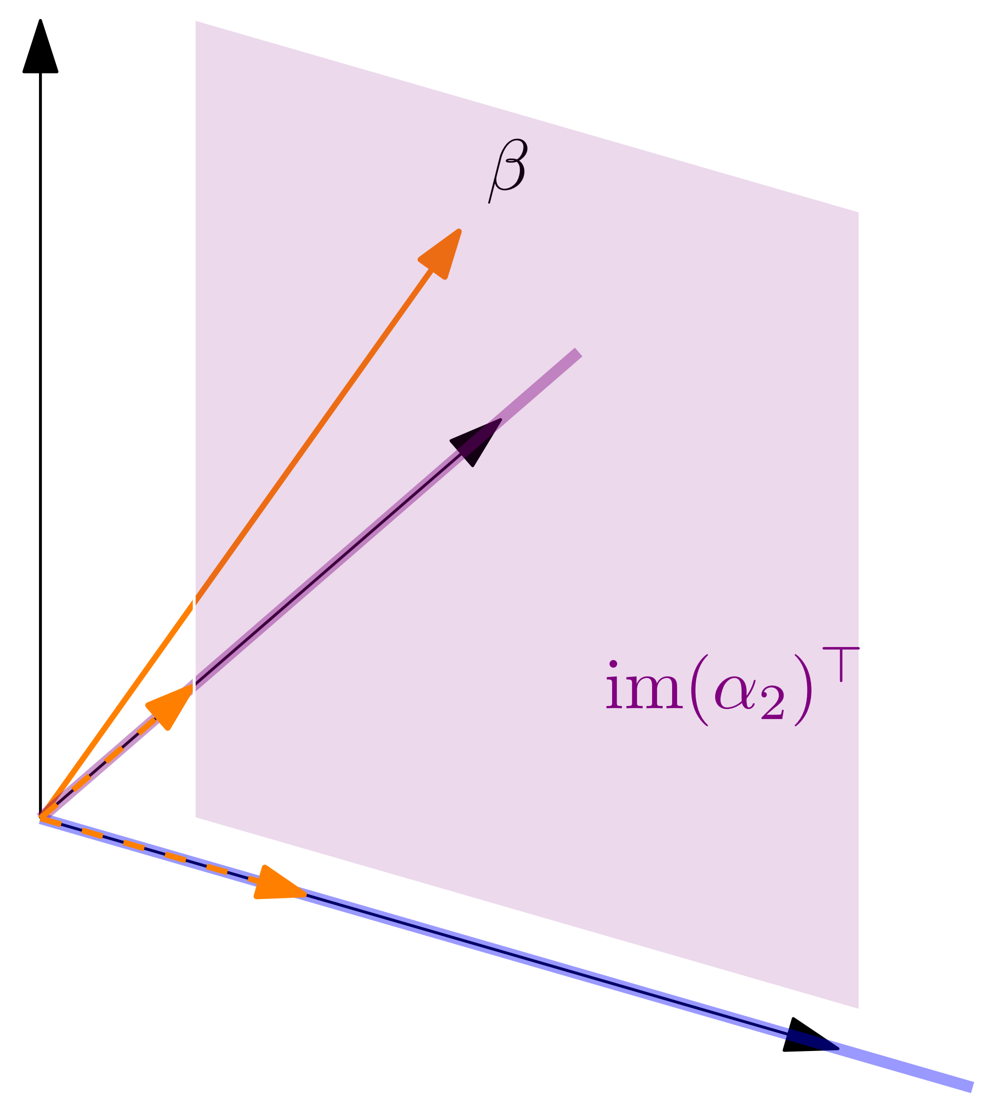
Janzing, D. and Schölkopf, B. Detecting non-causal artifacts in multivariate linear regression models. In International Conference on Machine Learning, pp. 2245–2253. PMLR, 2018. 5, 7, 12

We need to combine the estimators …



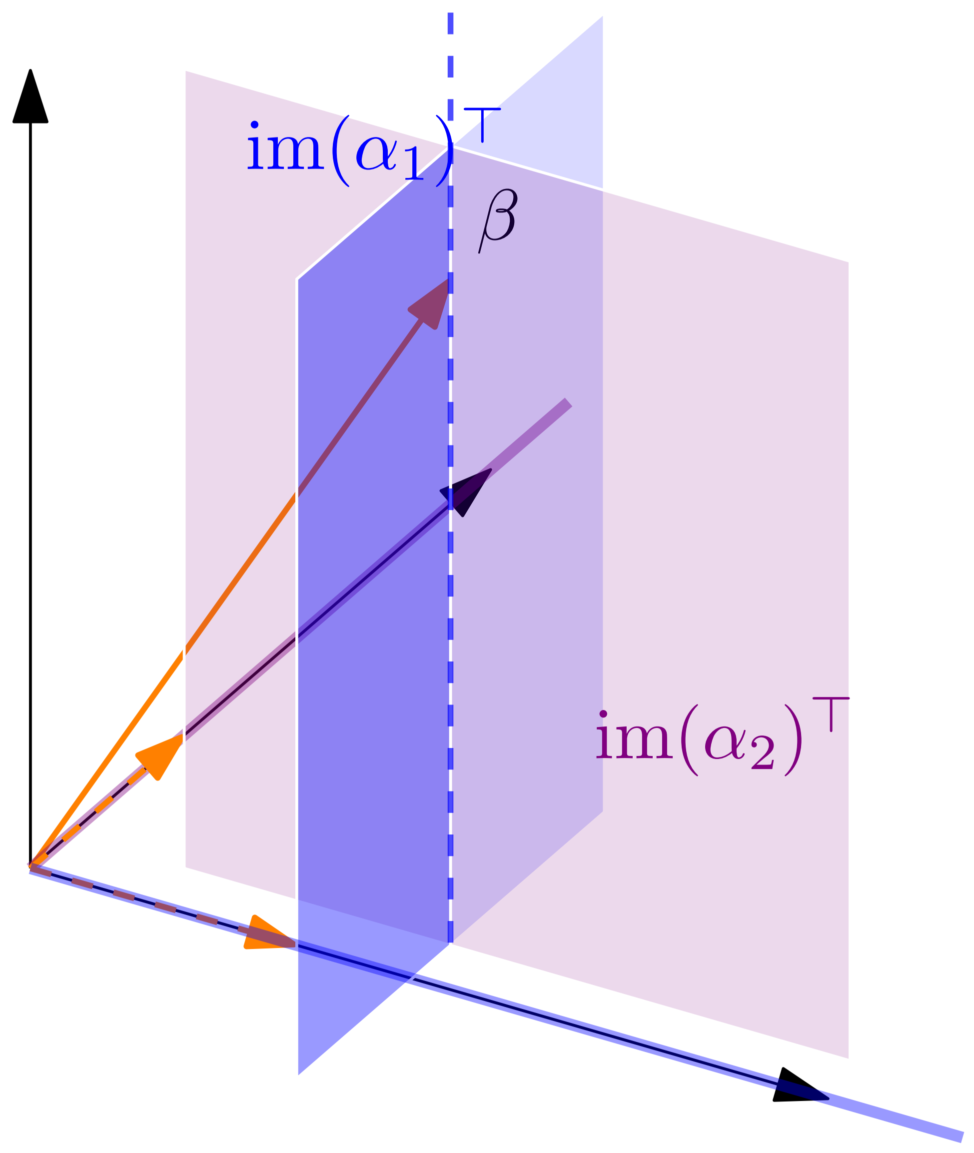
… use the previous experiments as constraints …

$$\widehat{P_{\alpha_2}\beta} = V_{\alpha_2}V_{\alpha_2}^\top \gamma$$



… use the previous experiments as constraints …

$$\widehat{P_{\alpha_2}\beta} = V_{\alpha_2} V_{\alpha_2}^\top \gamma$$

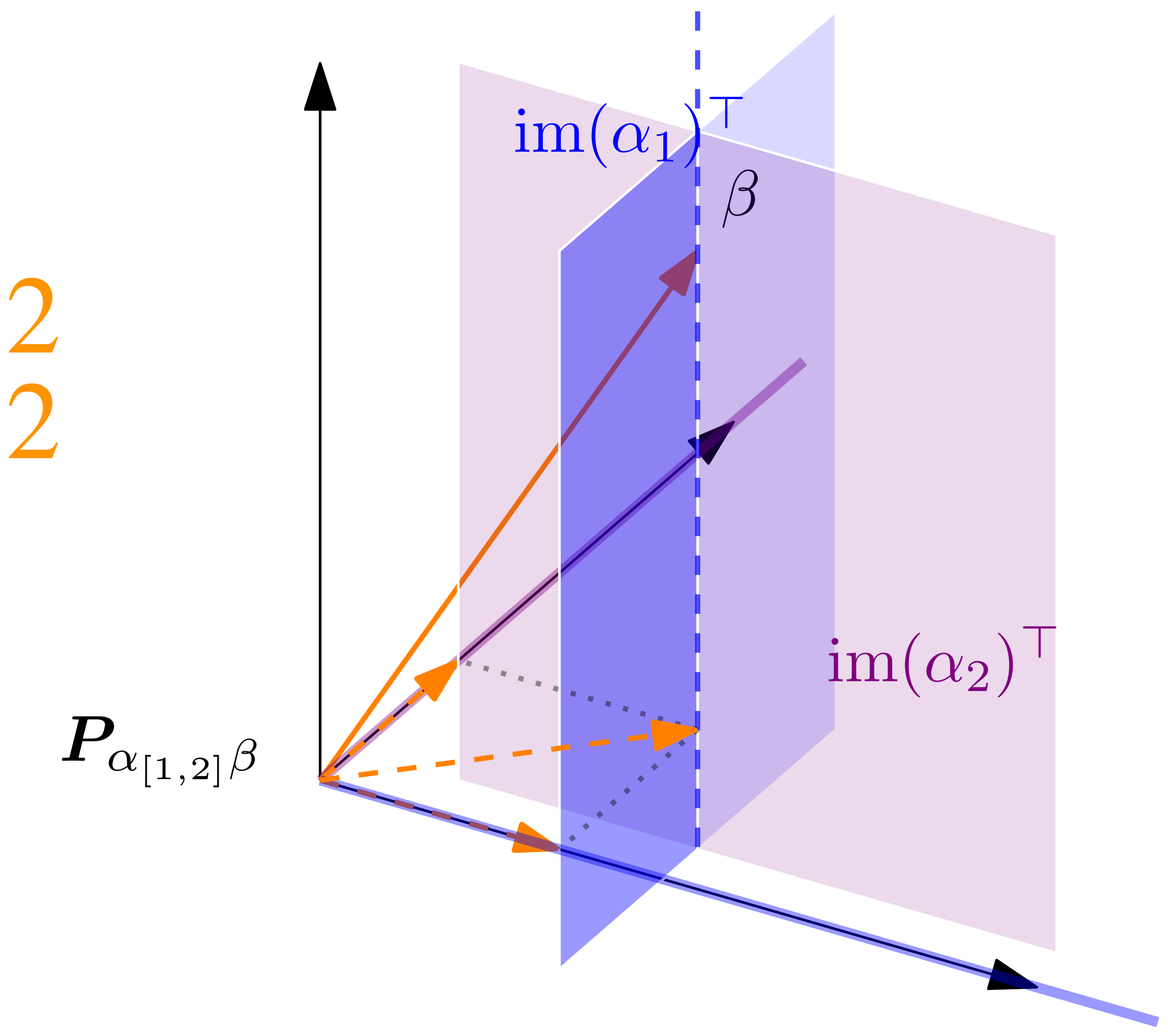


$$\widehat{P_{\alpha_1}\beta} = V_{\alpha_1} V_{\alpha_1}^\top \gamma$$

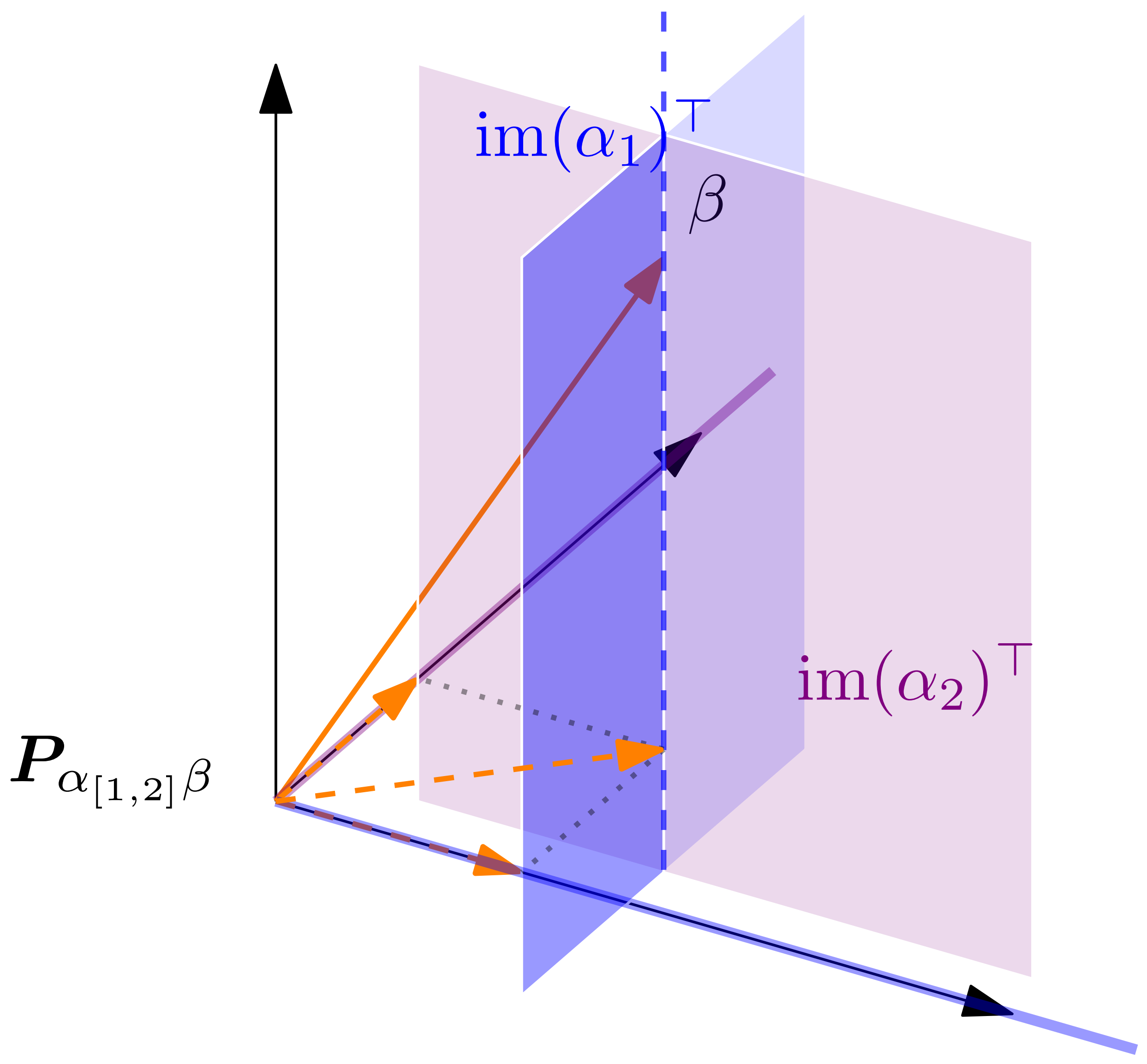
… and minimise the norm of β

$$\widehat{P_{\alpha_2}\beta} = V_{\alpha_2} V_{\alpha_2}^\top \gamma$$

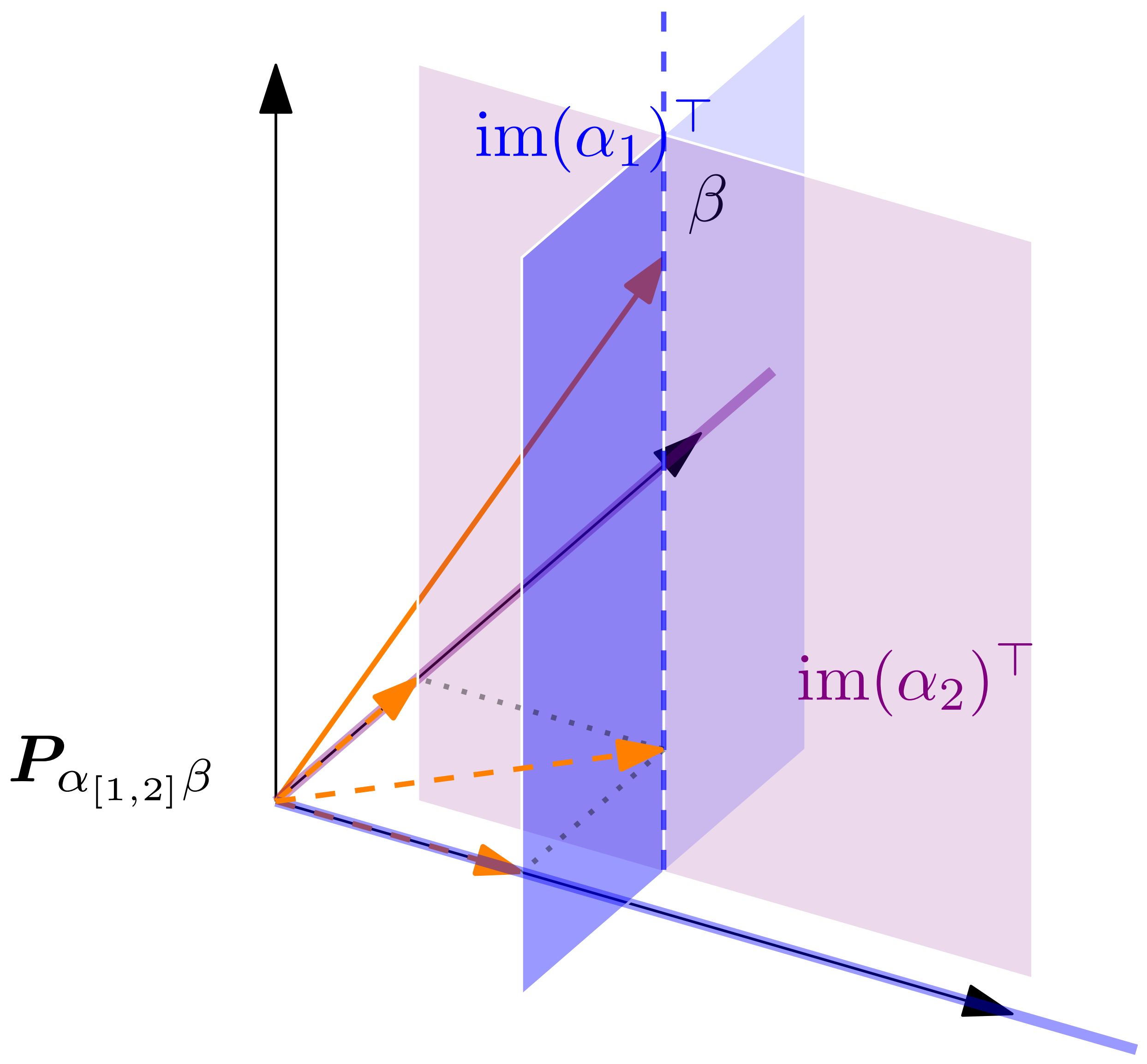
$$\min_{\gamma \in \mathbb{R}^{d_x}} \|\gamma\|_2^2$$

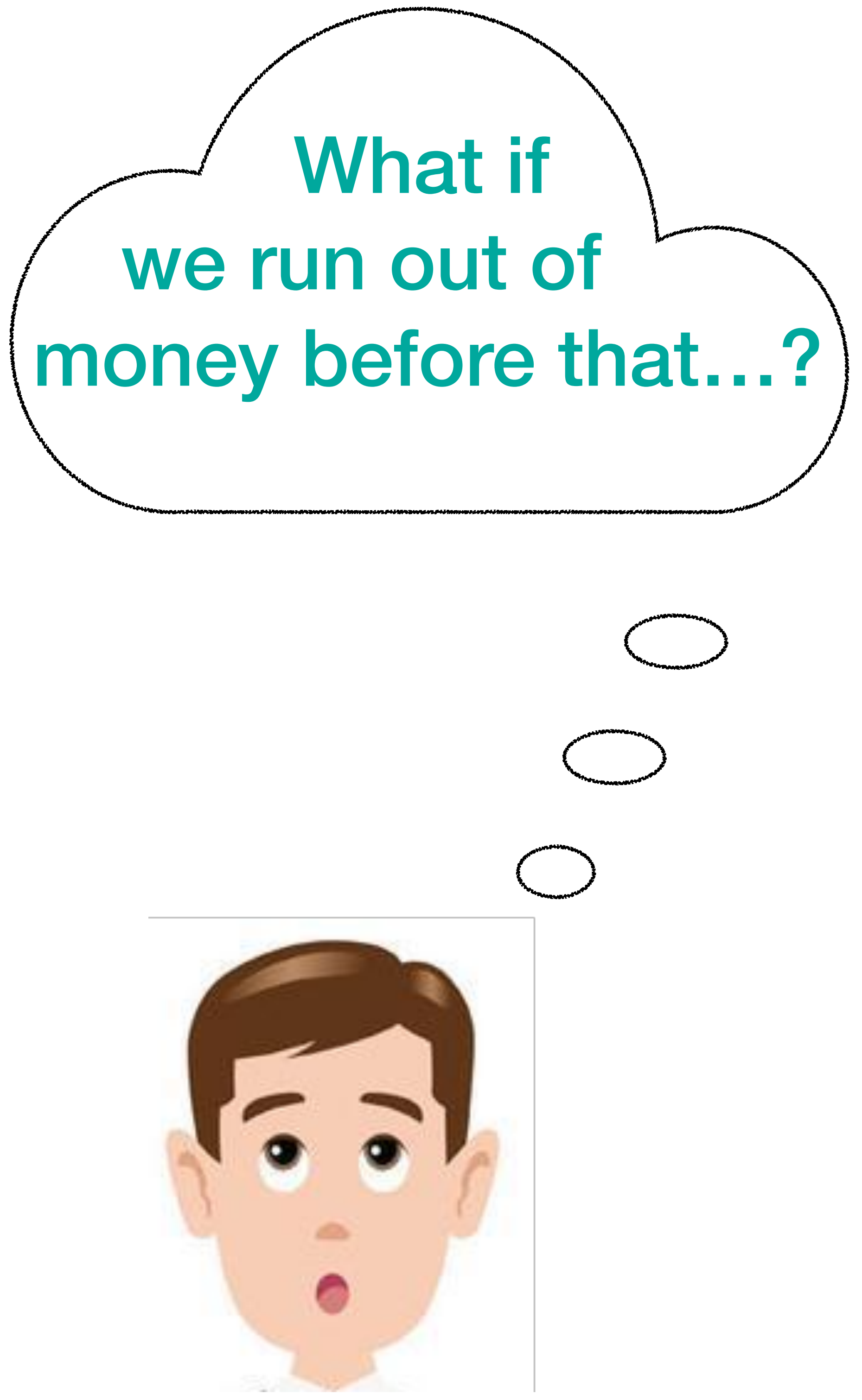


$$\widehat{P_{\alpha_1}\beta} = V_{\alpha_1} V_{\alpha_1}^\top \gamma$$

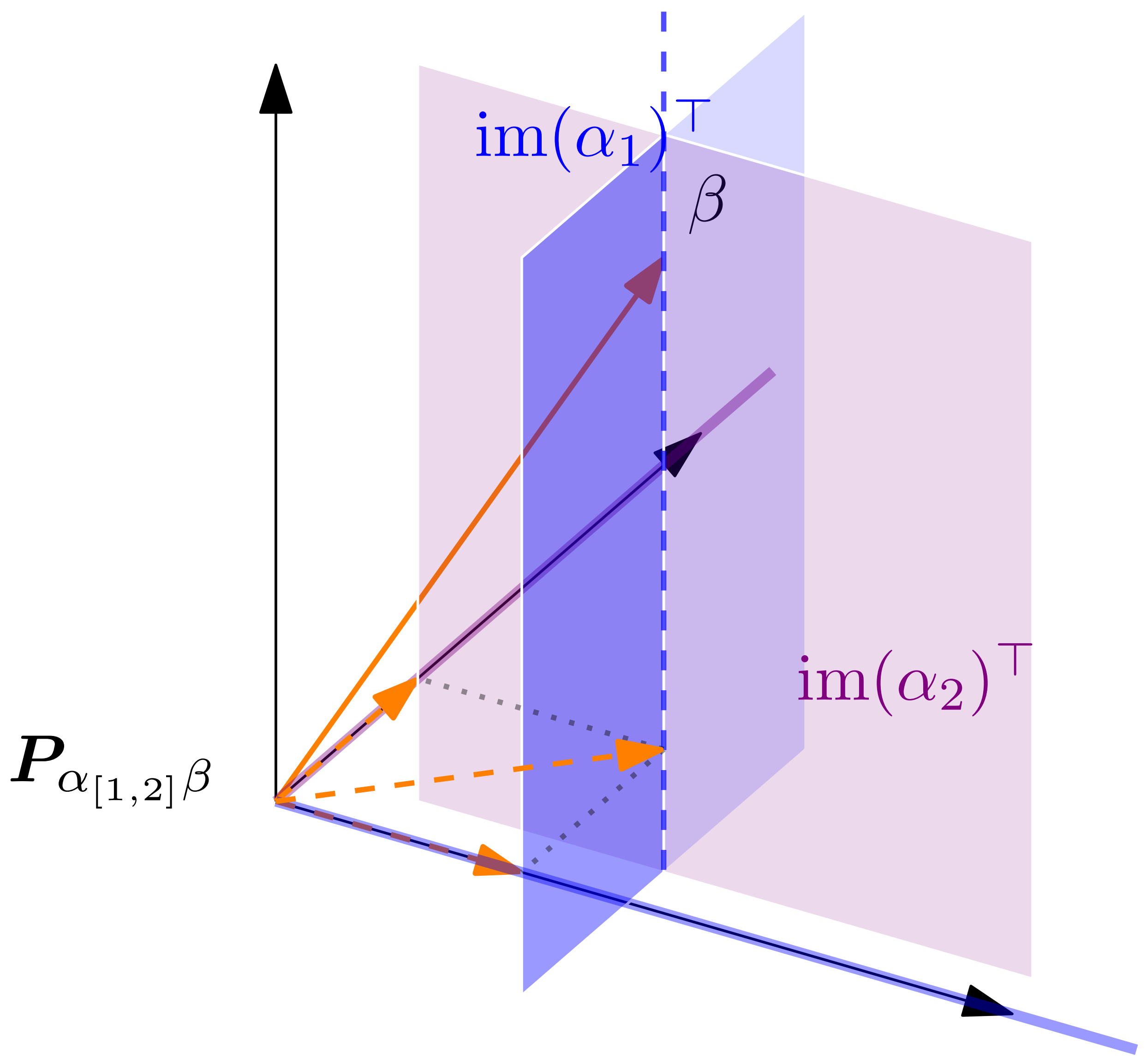


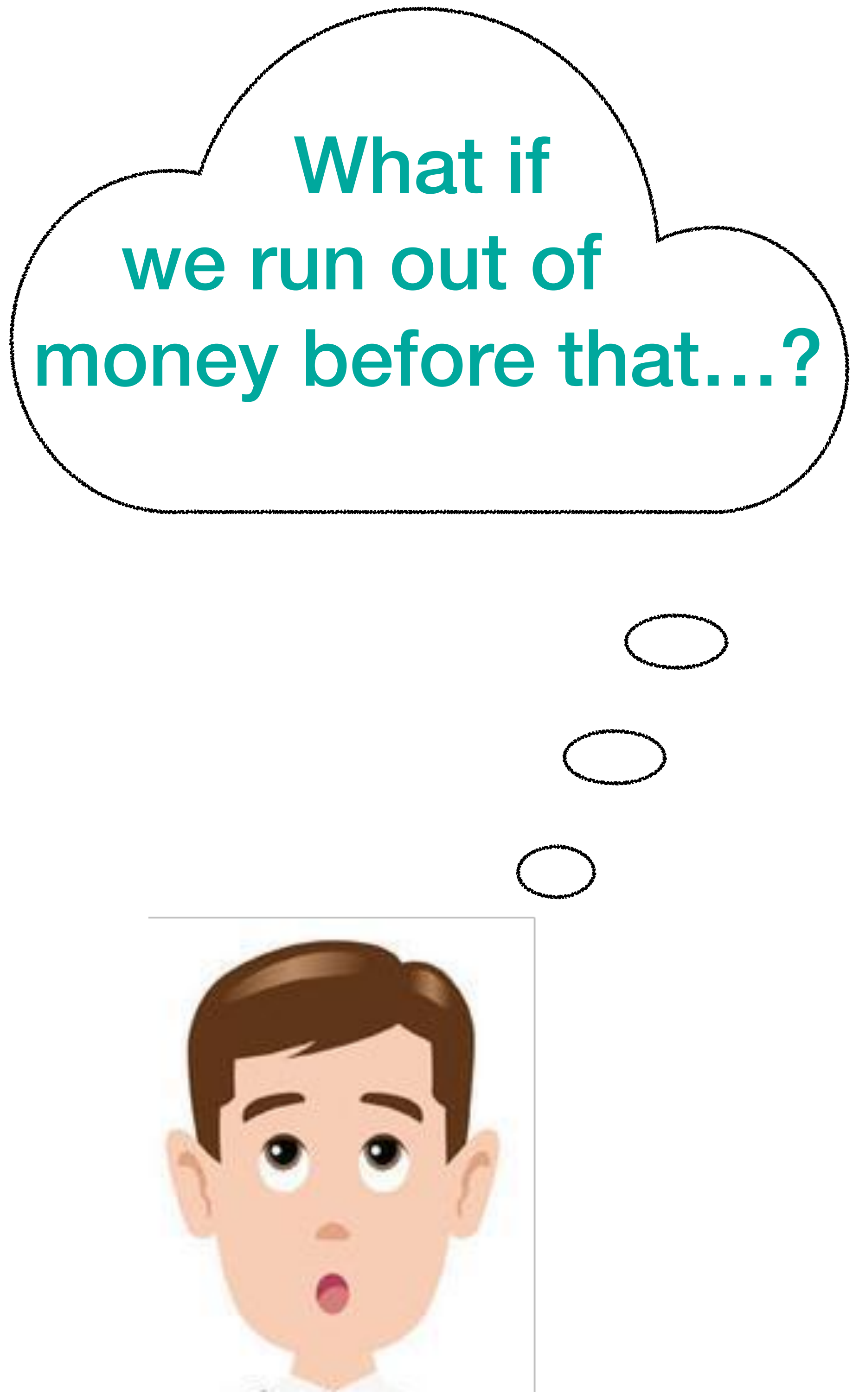
$\mathrm{im}(\alpha_1)^\top$
β
$\mathrm{im}(\alpha_2)^\top$
$\boldsymbol{P}_{\alpha_{[1,2]}\beta}$


im$(\alpha_1)^\top$
β
im$(\alpha_2)^\top$
$\boldsymbol{P}_{\alpha_{[1,2]}\beta}$

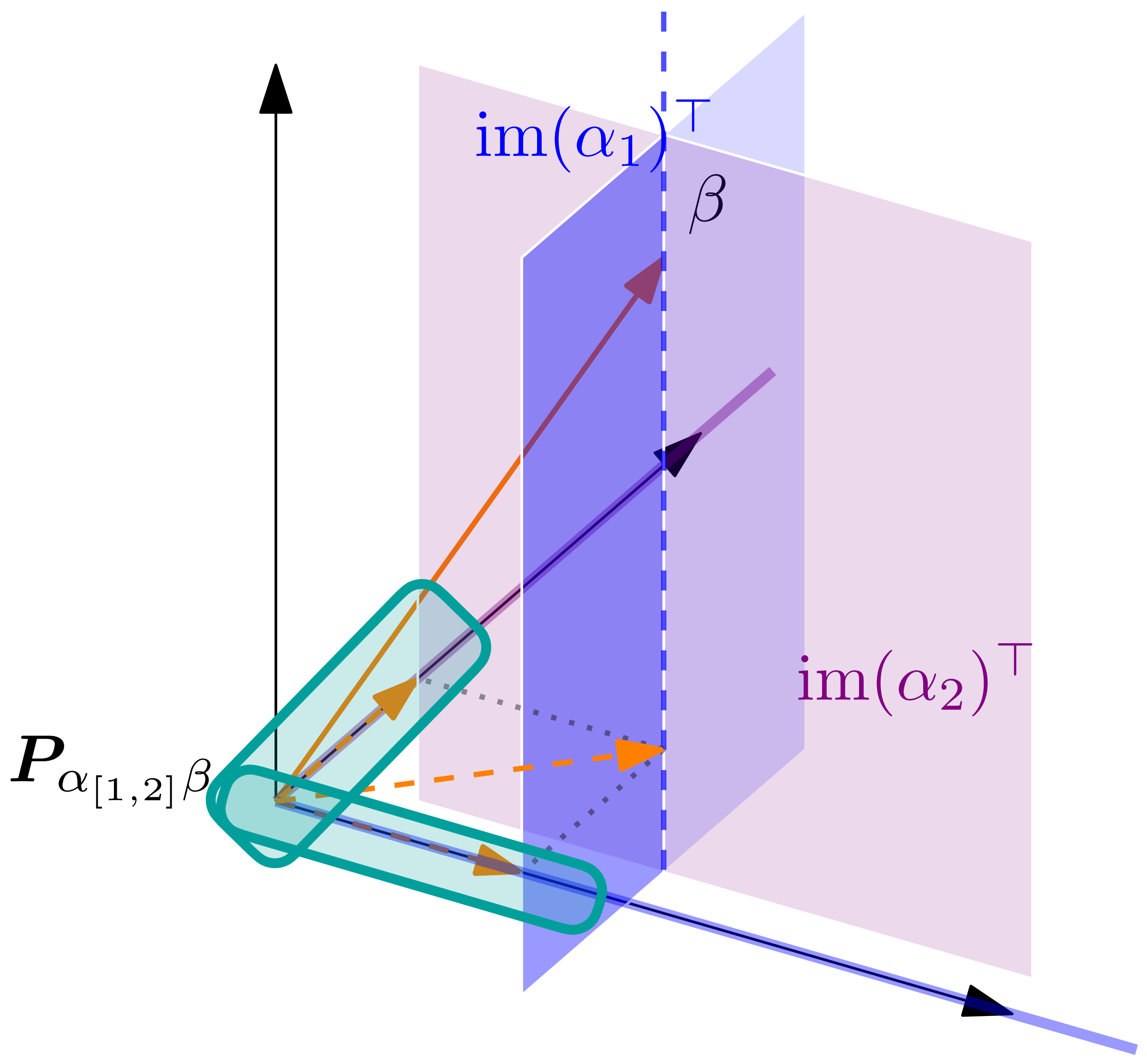

What if
we run out of
money before that...?

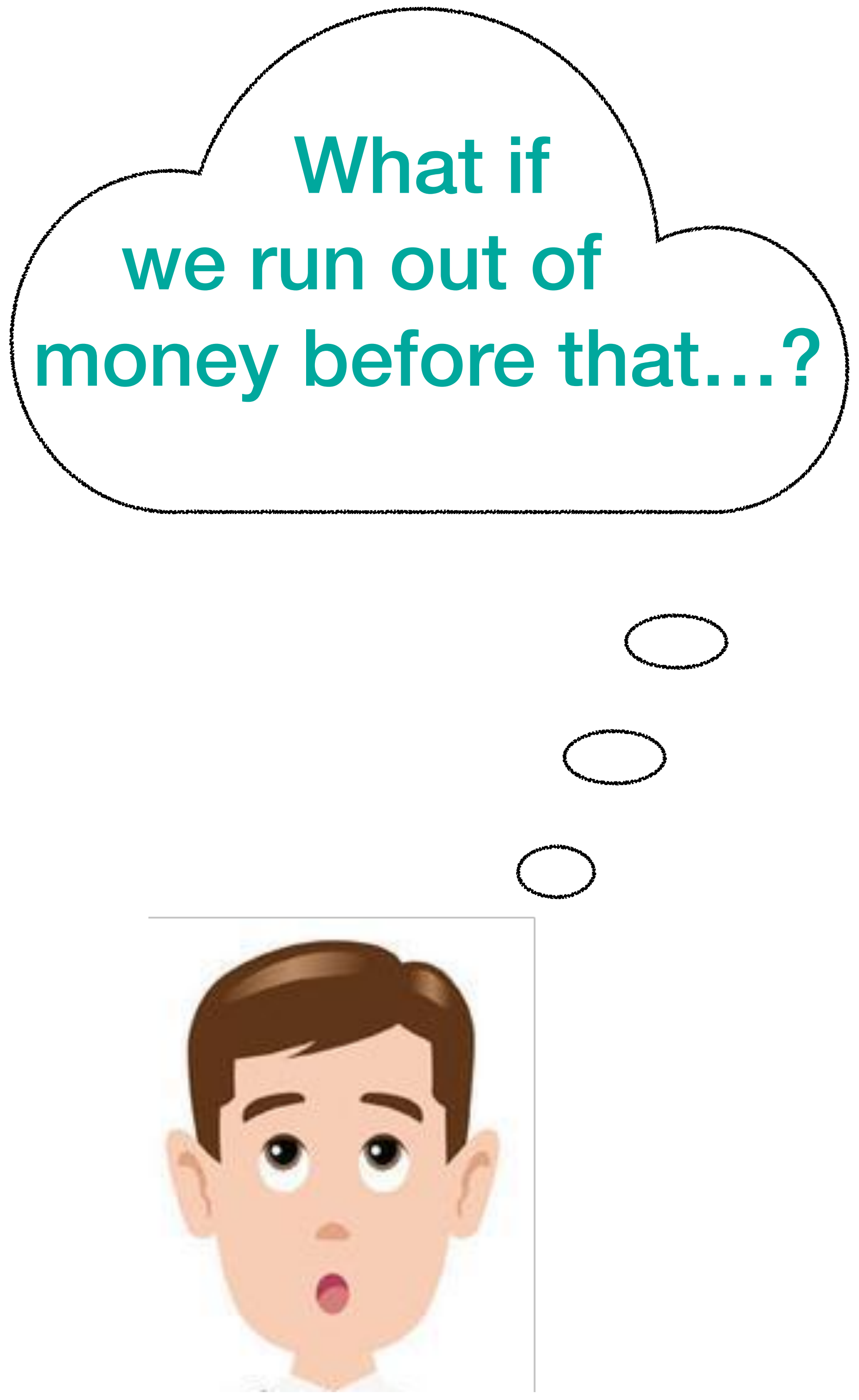
We still know the effect for some bacteria …





We still know the effect for some bacteria …



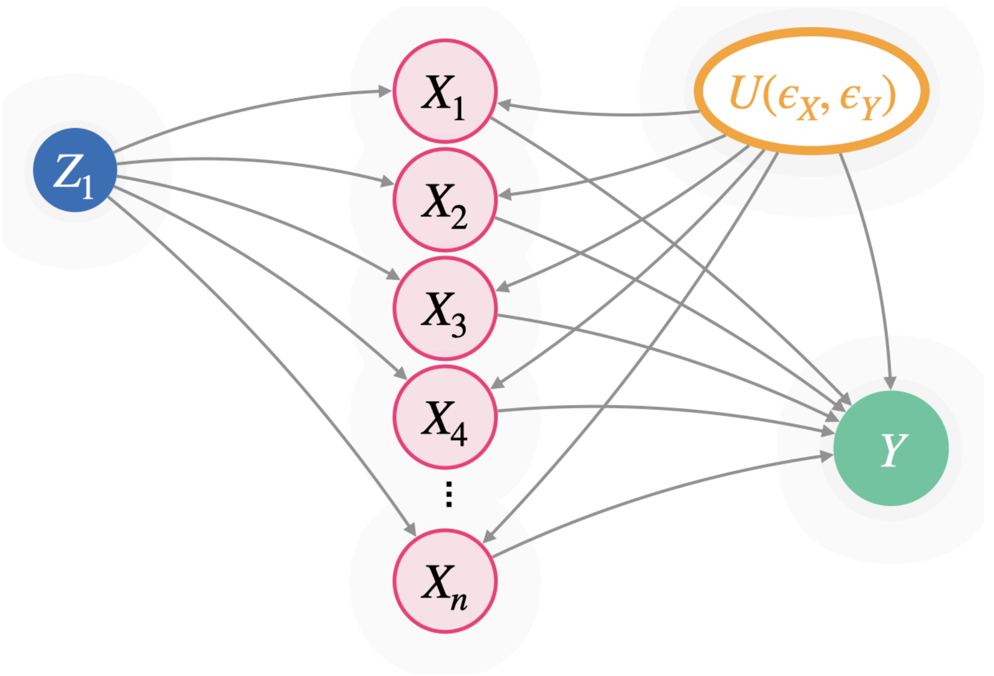


If $P_\alpha e_i = e_i$ then β_i is identified with e_i as the standard basis.

Putting it all together … :

Putting it all together … :

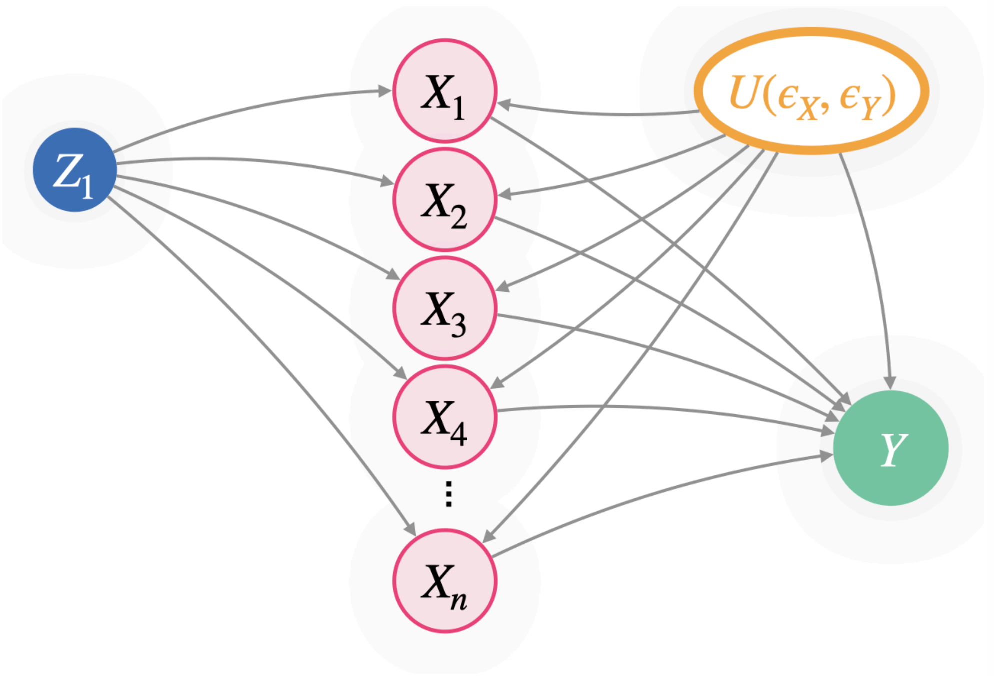
1



Single estimator for the **underspecified** case for one round of experiments.

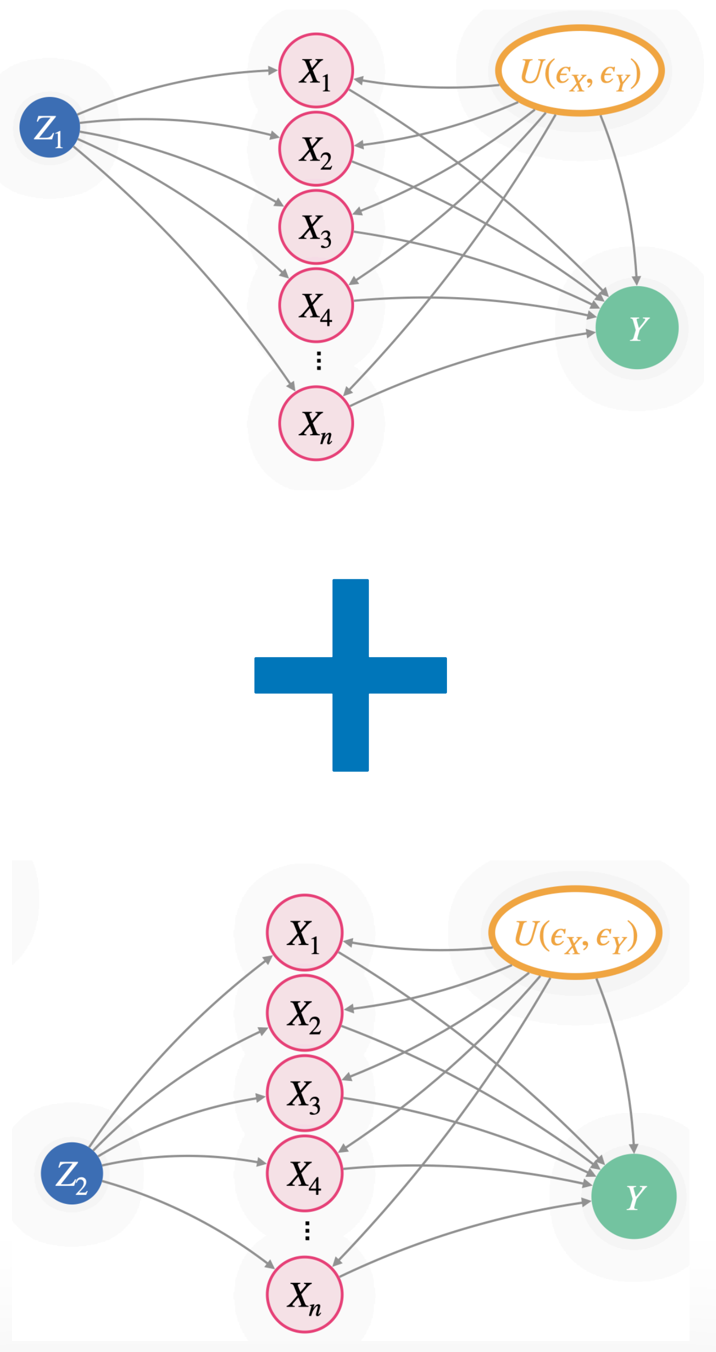
Putting it all together … :

1



Single estimator for the **underspecified** case for one round of experiments.

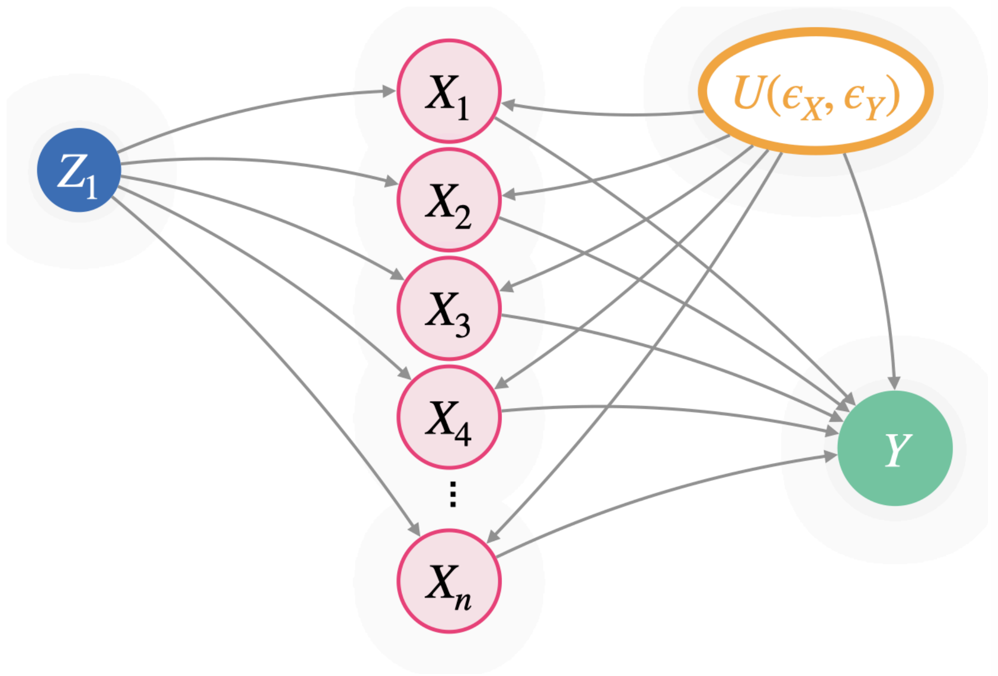
2



Combined estimator for the **multiple** rounds of experiments.

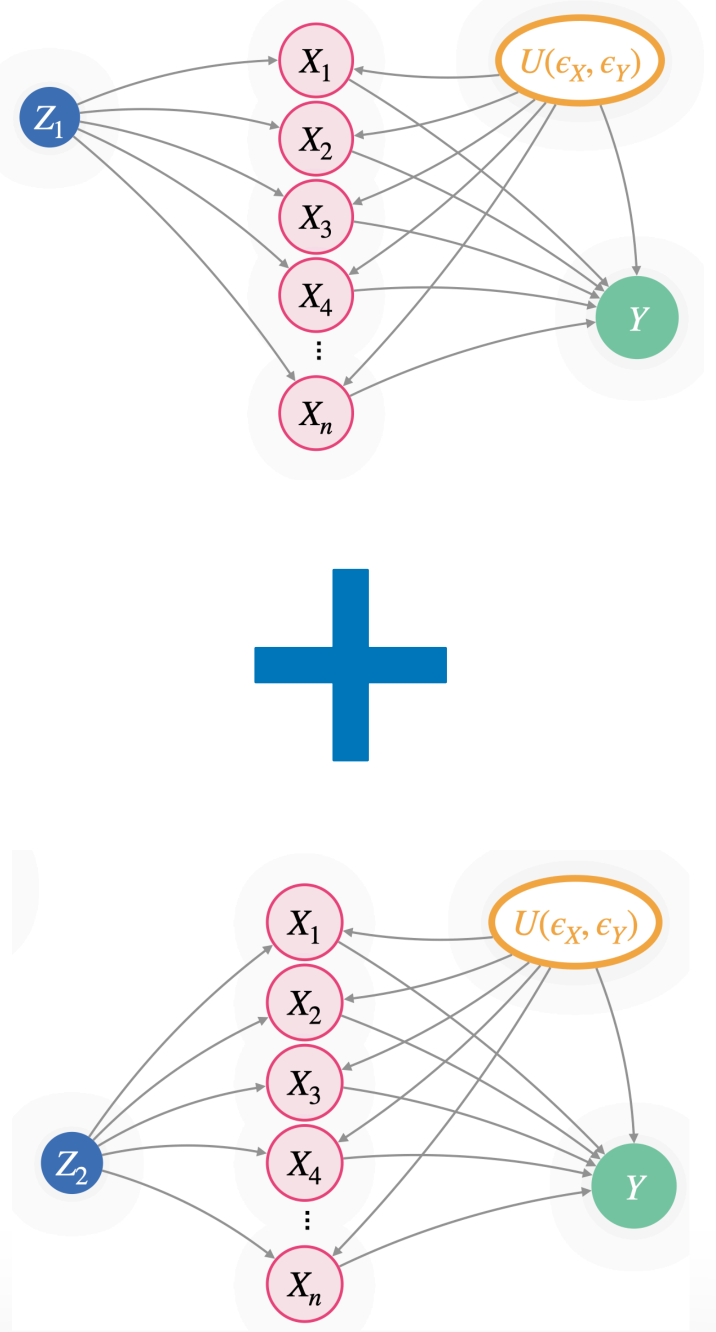
Putting it all together … :

1



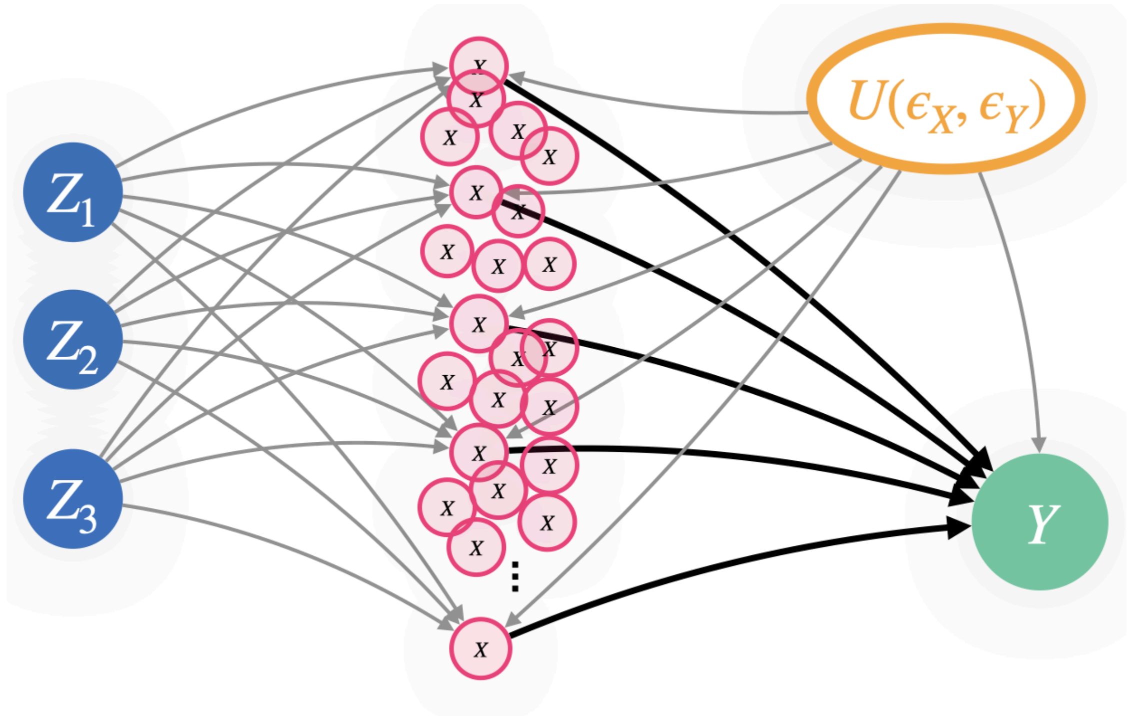
Single estimator for the **underspecified** case for one round of experiments.

2



Combined estimator for the **multiple** rounds of experiments.

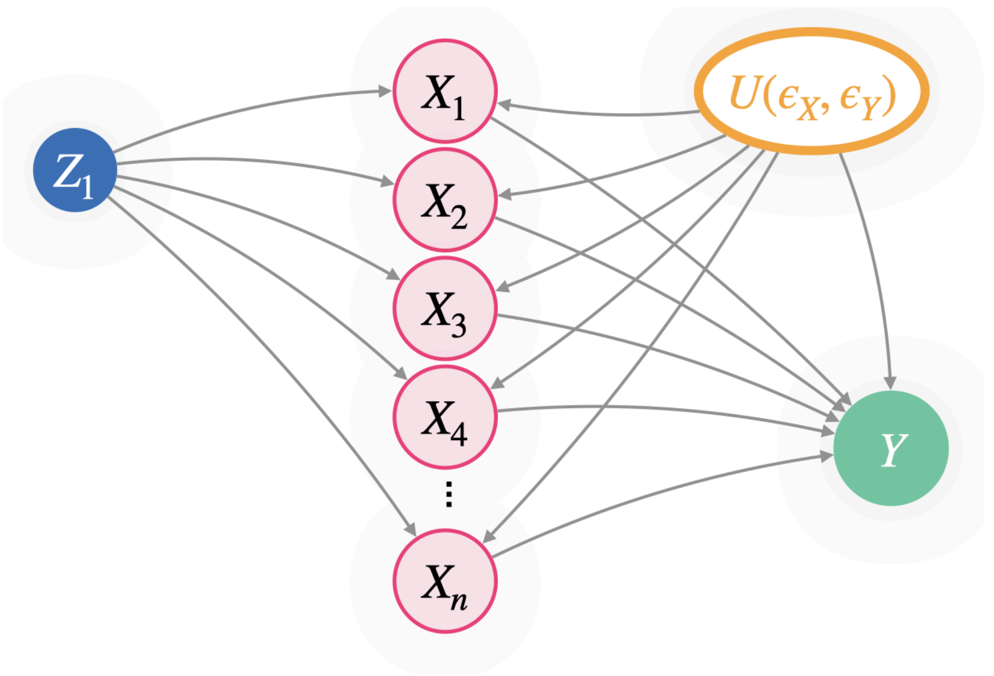
3



Algorithm for **sequential selection** and a **stopping criterion** for effect identification.

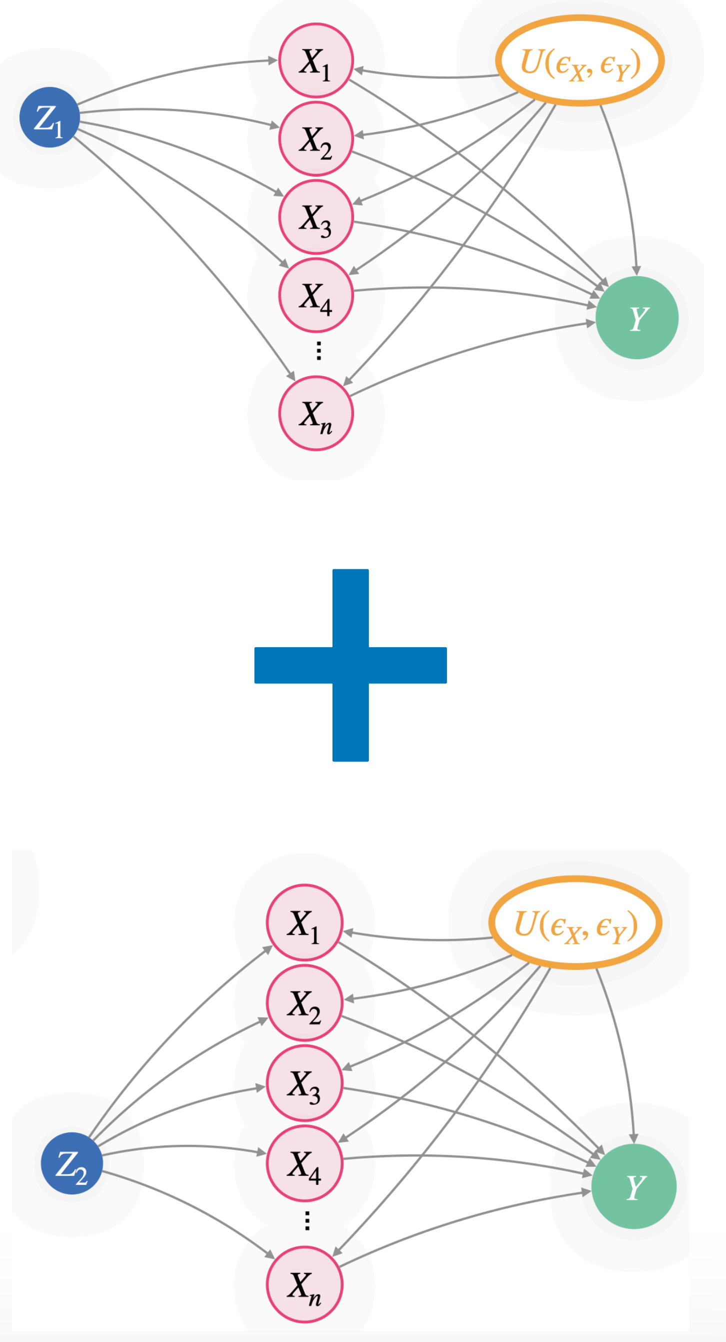
Putting it all together ... :

1



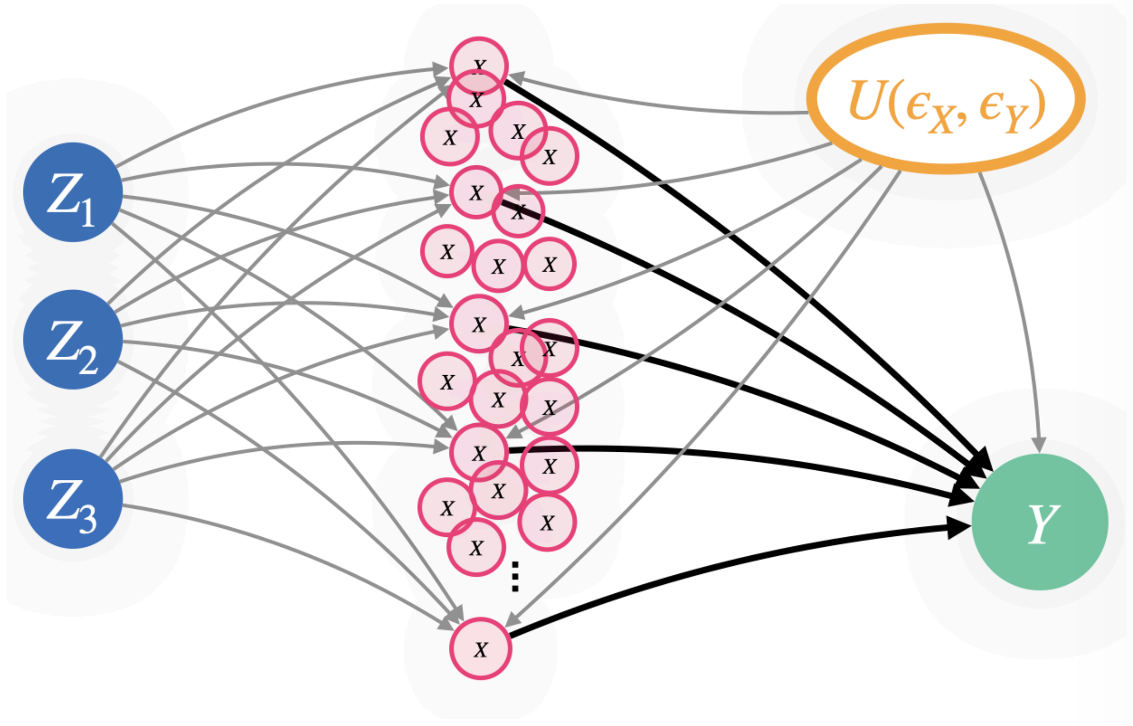
Single estimator for the **underspecified** case for one round of experiments.

2



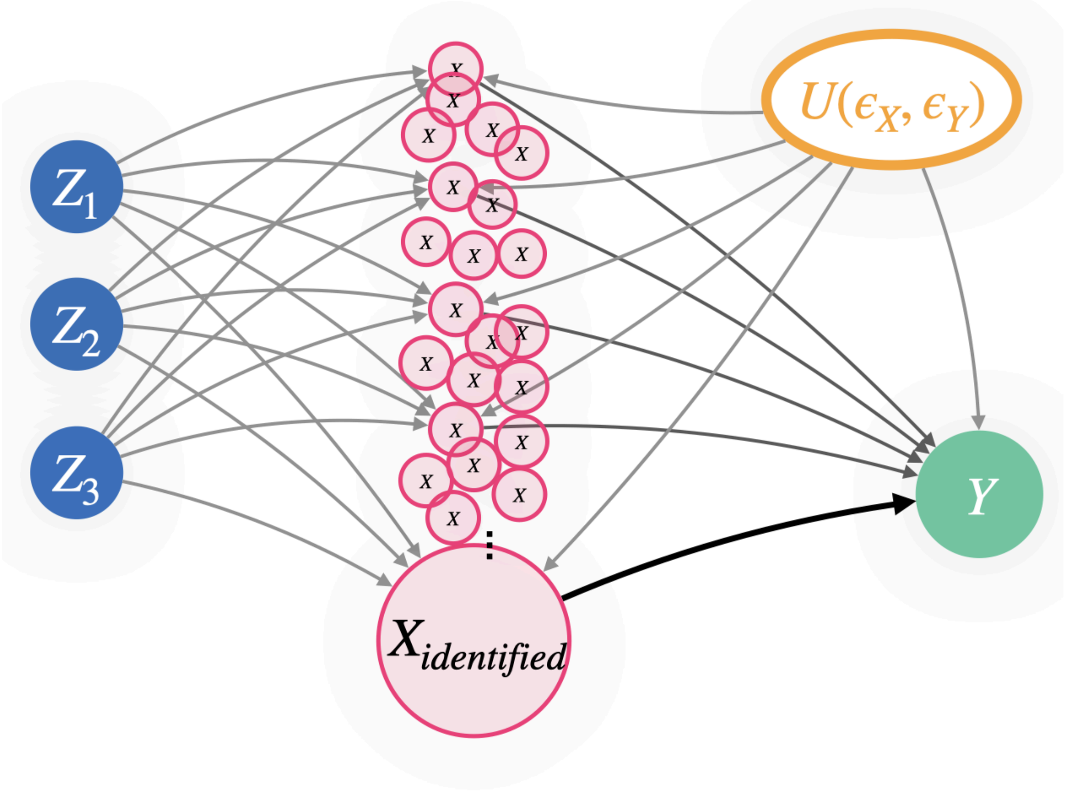
Combined estimator for the **multiple** rounds of experiments.

3



Algorithm for **sequential selection** and a **stopping criterion** for effect identification.

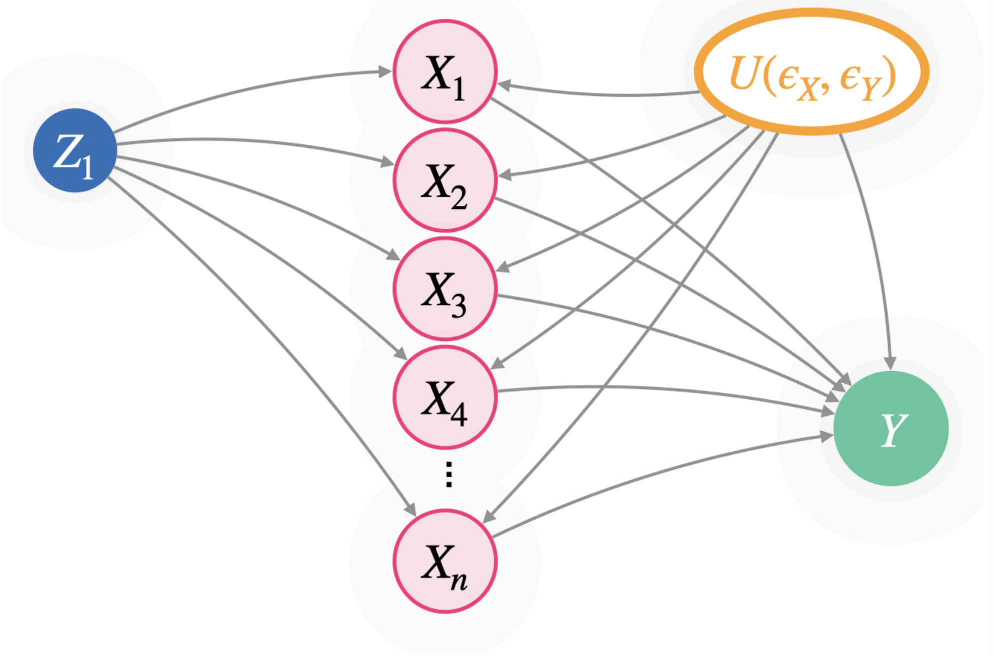
4



Method for checking individual **component-identification.**

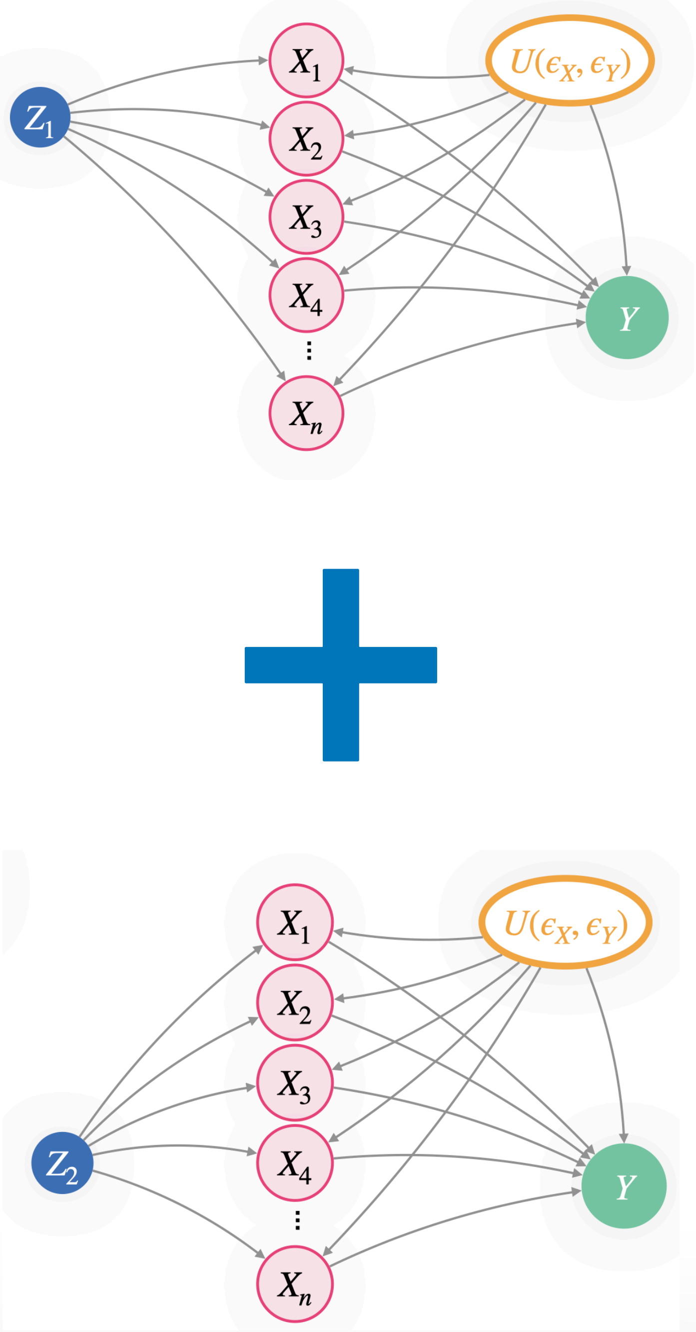
Putting it all together … :

1



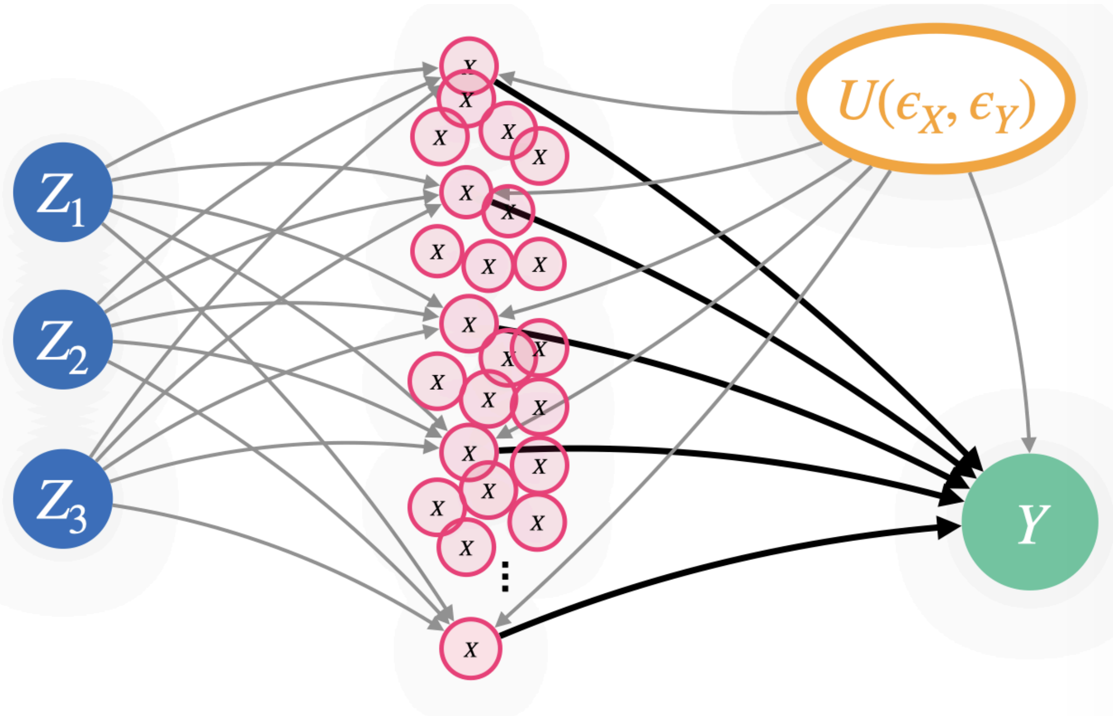
Single estimator for the **underspecified** case for one round of experiments.

2



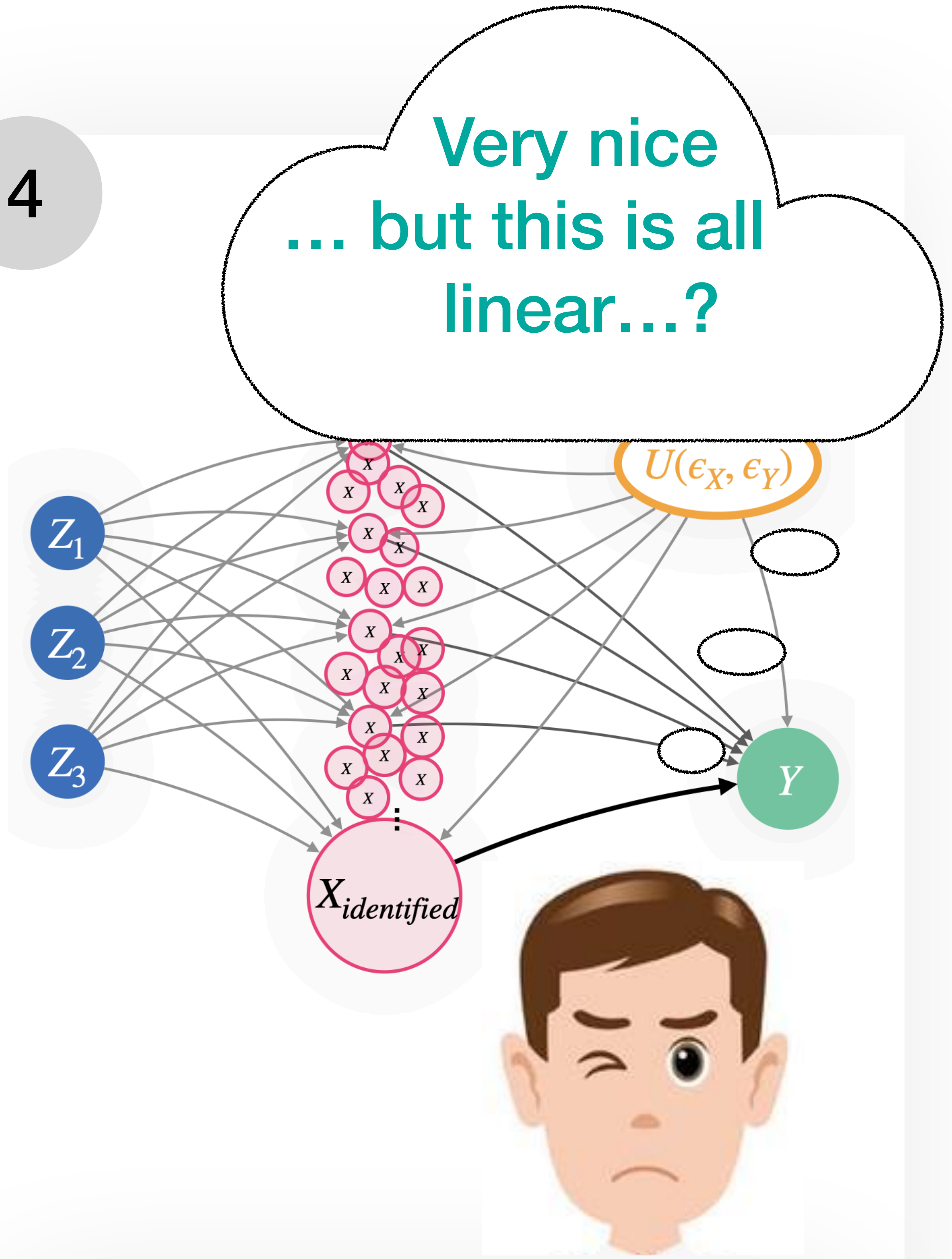
Combined estimator for the **multiple** rounds of experiments.

3



Algorithm for **sequential selection** and a **stopping criterion** for effect identification.

4



Method for checking individual **component-identification.**

Discussion

1. **Extension to nonlinear functional relationships:** Given the motivational dataset, the linearity assumption is restrictive.

    1. How can we interpret a nonlinear instrumented subspace?

    2. How can we transfer the confounding strength to nonlinear settings? Do we need a different stopping criterion?

2. **Similarity metric**, i.e. proposal of instruments: how does this transfer to real world data?

Sequential Underspecified Instrument Selection for Cause-Effect Estimation.

Elisabeth Ailer, Jason Hartford, Niki Kilbertus

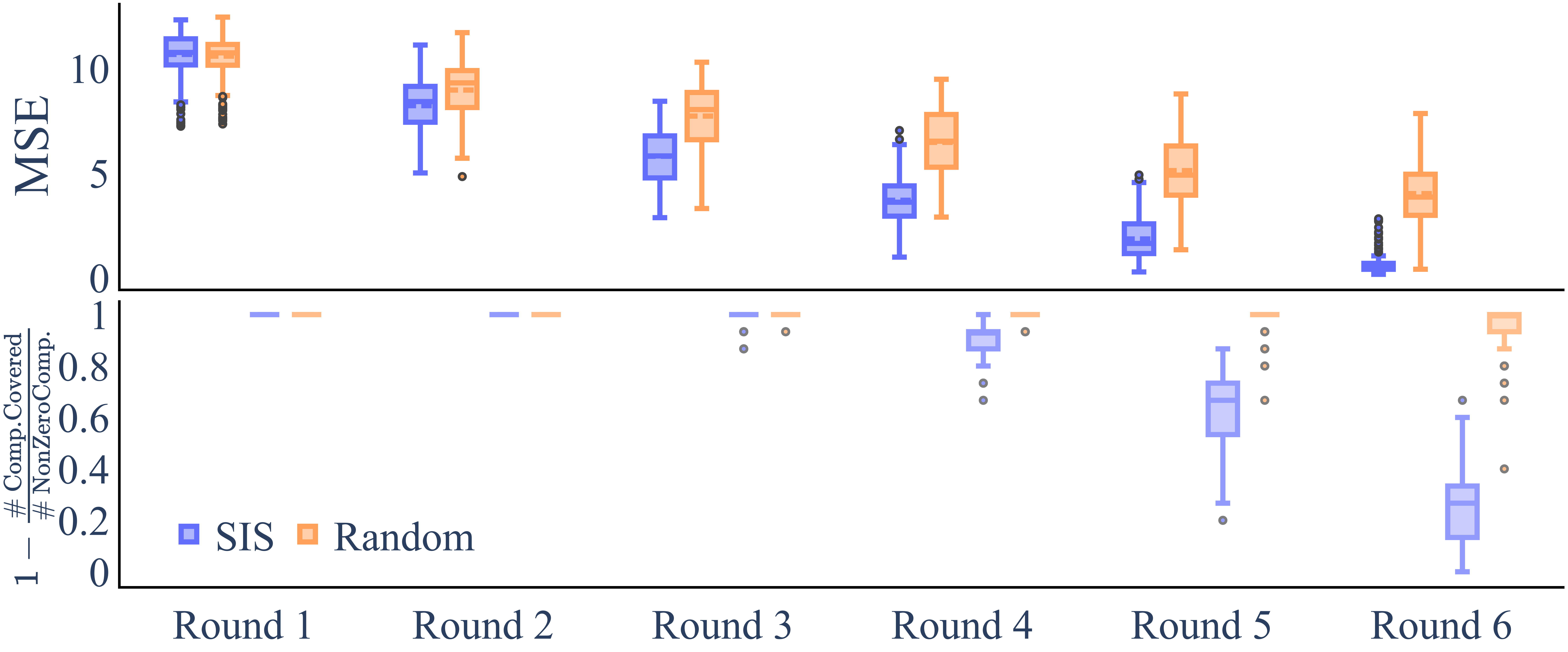
HELMHOLTZ MUNICH
Recursion
Mila
TUM

Appendix

Elisabeth Ailer, Jason Hartford, Niki Kilbertus



Results Sequential Underspecified Instrument Selection



Results Sequential Underspecified Instrument Selection