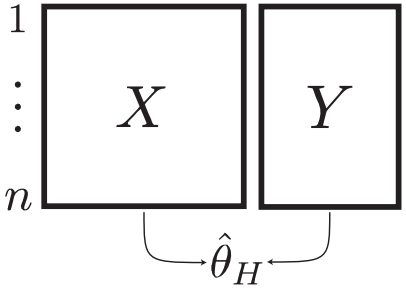
The mixed subjects design decreases costs of precise estimates and maintains validity

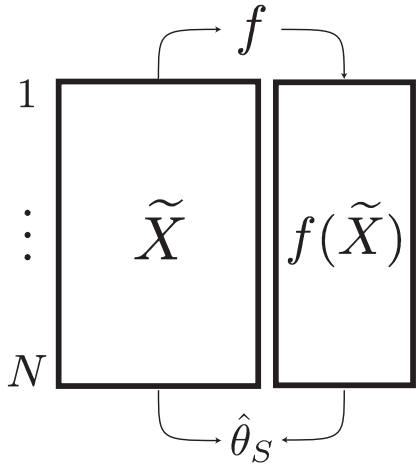
Human Subjects



Estimate $\hat{\theta}_H$ with classic inference, e.g. by using OLS to regress observed outcome Y on X.

- ✓ valid estimate
- ✗ precise estimate
- ✗ inexpensive data

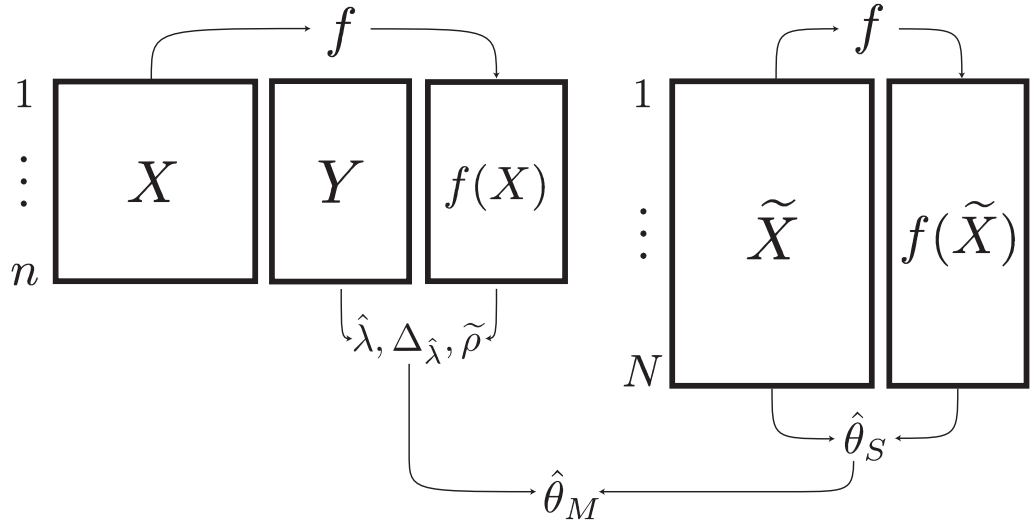
Silicon Subjects



Estimate $\hat{\theta}_S$ with classic inference, e.g. by using OLS to regress outcome $f(\widetilde{X})$ predicted by LLM f on X.

- ✗ valid estimate
- ✓ precise estimate
- ✓ inexpensive data

Mixed Subjects



Estimate $\hat{\theta}_M$ by correcting a possibly inaccurate estimate $\hat{\theta}_S$ from LLM predictions with a rectifier $\Delta_{\hat{\lambda}}$ from data on human subjects. Estimate a tuning parameter $\hat{\lambda}$ for increased statistical precision and the predictive accuracy measure $\widetilde{\rho}$ for conducting power analyses.

- ✓ valid estimate
- ✓ precise estimate
- ✓ inexpensive data